Université Bordeaux Montaigne

École doctorale Montaigne Humanités (ED 480)

Thèse de doctorat en sciences de l'information et de la communication

À la recherche d'une maîtrise des données : une étude des discours scientifiques et professionnels sur la *data literacy*

Présentée et soutenue publiquement le 16 juillet 2024 par

Ugo Verdi

Sous la direction d'Olivier Le Deuff, Professeur des universités, Université Bordeaux Montaigne

Membres du jury :

Présidente du jury

Lise Verlaet, Professeure des universités, Université Montpellier 3

Rapporteuses

Sarah Labelle, Professeure des universités, Université Montpellier 3

Anne Lehmans, Professeure des universités, Université de Bordeaux

Examinateurs

Gérald Kembellec, Maître de conférences, Conservatoire national des arts et métiers

Vincent Liquète, Professeur des universités, Université de Bordeaux

À la recherche d'une maîtrise des données : une étude des discours scientifiques et professionnels sur la *data literacy*

Résumé : Depuis deux décennies, l'évolution et la démocratisation des logiciels et des technologies numériques, à l'instar des applications des *smartphones*, ont généré de nouveaux usages liés aux données. Ces dernières, considérées comme un « nouvel or noir » nécessaire aux prises de décisions, ont été investies d'enjeux centraux au sein de discours qui véhiculent de nombreux mythes essentialisant. Leur maîtrise est de plus en plus sollicitée par tout un ensemble d'acteurs et la *data literacy*, également nommée « culture des données » ou « littératie des données », est en ce sens évoquée comme une, si ce n'est « la » solution miracle qui permettrait une acculturation efficace des individus. L'objectif de cette thèse est de présenter un état de l'art des caractéristiques de la *data literacy* issues de représentations véhiculées par les discours scientifiques et professionnels. Elle doit ainsi permettre d'une part de répondre à un ensemble de questionnements, à savoir quelles sont les caractéristiques de la *data literacy* et ont-elles évolué dans le temps ? quelles sont les visions portées sur les données et quel lien est tissé avec l'information ? quels sont les enchevêtrements conceptuels et littéraciques ? depuis quand pouvons-nous véritablement parler de *data literacy* et a-t-il existé des proto *data literacies* aux objectifs similaires ? que recouvre une « acculturation » aux données et quelles sont ses modalités de déploiement ? quels sont les acteurs qui se sont exprimés sur cette thématique ? dans quels contextes et selon quelle temporalité ? pouvons-nous parler d'une vision purement française de la *data literacy* ? D'autre part, de déconstruire les mythes qui entourent la *data literacy* et *in extenso* les usages des données. Pour ce faire, trois grand corpus seront exploités : (1) plusieurs centaines de publications scientifiques et professionnelles, (2) 44544 tweets contenant le terme « data literacy » produits entre septembre 2021 et septembre 2022 et (3) 32 entretiens semi-directifs réalisés auprès d'acteurs français.

Descripteurs : Data literacy, culture des données, analyse de discours, représentation, littératie, donnée, acculturation.

Université Bordeaux Montaigne
École doctorale Montaigne Humanités (ED 480)
Thèse de doctorat en Sciences de l'information et de la communication
CC BY-NC-ND 2024 Ugo Verdi

Ce document est mis à disposition selon les termes de la licence Creative Commons « Attribution Pas d'Utilisation Commerciale 4.0 France (CC BY-NC 4.0 FR) ». Pour plus d'informations, visitez https://creativecommons.org/licenses/by-nc-nd/4.0/.

In search of data mastery : an analysis of scientific and professional discourses on data literacy

Abstract: For about two decades, the evolution and spreading of digital softwares and technologies, like smartphone applications, have generated new data practices. Data, considered as a "new black gold" crucial for decision-making, have been invested with serious challenges within discourses conveying numerous essentializing myths. Data mastery is increasingly in demand by a vast array of actors and *data literacy* is in this sense mentioned as one, if not "the" solution which would allow effective training policies. The aim of this essay is to present the characteristics of *data literacy* resulting from representations conveyed by scientific and professional discourses. It must therefore, on the one hand, answer a set of questions, namely what are the characteristics of *data literacy* and have they evolved over time? how data are represented and what link is made with information? what are the existing entanglements with concepts and literacies ? since when can we truly speak of *data literacy* and have there been proto *data literacies* with similar aims? what does "acculturation" to data covers and what are its deployment modalities ? who are the actors who have expressed themselves on *data literacy* ? in what contexts and period of time? can we talk about a French vision of *data literacy*? On the other hand, it must deconstruct the myths surrounding *data literacy* and *in extenso* practices of data. To do so, three large corpora will be examined: (1) several hundred scientific and professional publications, (2) 44,544 tweets containing the term "data literacy" produced between September 2021 and September 2022 and (3) 32 semi-structured interviews carried out with French actors.

Keywords: Data literacy, discourse analysis, representation, literacy, data, acculturation.

This document is made available according to the Creative Commons « Attribution NonCommercial 4.0 France (CC BY-NC 4.0 FR) » license. For additional information, visit https://creativecommons.org/licenses/by-nc-nd/4.0/.

Table des matières

Avant-propos **xv**
Remerciements xv
Notes à l'attention du lecteur xvi

Introduction générale **1**
Objectifs de recherche 4
Problématique et hypothèses 6
Soubassement théorique 7
Les études sur la littératie 7
L'analyse qualitative des discours 8
Plan de la thèse 11

I. Définir la *data literacy* 13

1. Les caractéristiques de la *data literacy* **15**
1.1. Introduction 15
1.1.1. Un premier état de l'art 19
1.1.2. Un paysage définitionnel en perpétuelle évolution 21
1.2. La nécessité pédagogique 25
1.2.1. Le *data literate* 25
1.2.2. Enseigner la *data literacy* 31
1.3. L'écosystème des acteurs 44
1.3.1. Les initiatives entrepreneuriales 45
1.3.2. Les associations et coopératives 46
1.3.3. Les organismes d'enseignement 48
1.3.4. Enquête sur Twitter / X 50
1.4. Conclusion du chapitre 63

2. Des influences contrastées **67**
2.1. Introduction 67

2.2. Les influences attestées 68
2.2.1. La littératie 68
2.2.2. Les données 82
2.3. La datalogie, l'influence oubliée 99
2.3.1. Une opposition aux premiers modèles de formation en informatique . . . 100
2.3.2. Les premières définitions 100
2.3.3. Des tentatives de traductions avortées 102
2.3.4. Les concepts clé de la datalogie 102
2.3.5. Les visées de la datalogie 107
2.3.6. Les liens avec les autres disciplines 108
2.3.7. L'aspect pédagogique 110
2.4. Conclusion du chapitre 111

II. Des visions plurielles 113

3. Les enchevêtrements conceptuels et littéraciques 115
3.1. Introduction 115
3.2. Les enchevêtrements conceptuels 116
3.2.1. La science des données 116
3.2.2. L'esprit critique vis-à-vis des données 121
3.2.3. La gouvernance des données 123
3.2.4. L'éthique des données 127
3.2.5. Vers une convergence des concepts 132
3.3. Les enchevêtrements littéraciques 136
3.3.1. La *data literacy* comme métalittératie 137
3.3.2. La *data literacy* comme littératie équivalente 145
3.3.3. La *data literacy* comme sous-composante de littératie 149
3.3.4. La *data literacy* comme littératie englobante 155
3.4. Conclusion du chapitre 157

4. Les littératies spécialisées des données 159
4.1. Introduction 159
4.2. Les spécialisations de la *data literacy* 162
4.2.1. L'aspect conceptuel 162
4.2.2. L'aspect pédagogique 163
4.2.3. L'aspect journalistique 165
4.2.4. L'aspect des données 165
4.2.5. L'aspect scientifique 166
4.2.6. L'aspect civique et démocratique 167

4.3. Canaliser les externalités négatives du *big data* 168
4.3.1. La datafication 169
4.3.2. La dataveillance 171
4.3.3. La réponse critique 174
4.3.4. La réponse systémique 178
4.3.5. L'activisme des données 179
4.3.6. Contrer les discriminations et révéler les invisibilisations 181
4.4. Conclusion du chapitre 192

III. Des visions françaises ? 195

5. Le contexte français de la *data literacy* 197
5.1. Introduction 197
5.2. La nécessité de la *data literacy* dans les discours 198
5.2.1. Les discours institutionnels européens 198
5.2.2. Les discours institutionnels français 200
5.2.3. Les discours des entreprises françaises 202
5.2.4. Conclusion de la première partie 206
5.3. Les acteurs et initiatives autour de la *data literacy* 207
5.3.1. Les acteurs entrepreneuriaux 207
5.3.2. Les acteurs institutionnels 209
5.3.3. Les acteurs associatifs 211
5.3.4. Les acteurs universitaires et associés 219
5.3.5. Conclusion de la seconde partie 221
5.4. Les conceptualisations universitaires et pédagogiques de la *data literacy* 222
5.4.1. Les définitions 225
5.4.2. Les connaissances 226
5.4.3. Les compétences 227
5.4.4. L'aspect culturel 229
5.4.5. Mettre en place la *data literacy* 231
5.4.6. La problématique des traductions 233
5.4.7. Conclusion de la troisième partie 245
5.5. Conclusion du chapitre 247

6. Un point de vue français de la *data literacy* ? 249
6.1. Introduction 249
6.2. Première partie : terrain et méthodologie 249
6.2.1. L'entretien semi-directif 250
6.2.2. L'échantillon d'analyse 250

6.2.3. Les étapes suivies 252
6.2.4. Les modalités de l'entretien 254
6.3. Deuxième partie : synthèse des résultats 255
6.3.1. Les termes employés 256
6.3.2. Les caractéristiques de la *data literacy* 262
6.3.3. Une formation en *data literacy* 271
6.4. Conclusion du chapitre 294

Conclusion générale **297**
Réponse aux hypothèses de recherche 298
Qu'est-ce qu'une littératie des données ? 301
Les perspectives de recherche 302

Bibliographie **307**

Index **352**

Annexes **359**
Annexe 1 : les définitions de la *data literacy* 359
Tableau des définitions de la *data literacy* 359
Bibliographie de l'annexe 1 363
Annexe 2 : les formules R 366
Importer un fichier CSV 366
Ouvrir les libraries 366
Récupérer les termes 367
Description 367
Récupérer les hashtags 368
Grams 368
Bigrammes séparés 369
Nombre de retweets 369
Termes employés par KirKDBorne 370
hashtags employés par KirKDborne 370
Termes employés par Williamstyles23 371
Hashtags employés par Williamstyles23 371
Lister les bots 372
Annexe 3 : les tweets de William D Styles 372
Annexe 4 : l'agencement des données 376
Bibliographie de l'annexe 4 378
Annexe 5 : La data-philosophie 379
Bibliographie de l'annexe 5 381

Annexe 6 : les *critical data literacies* 382
Tableau des *critical data literaries* 382
Bibliographie de l'annexe 6 385
Annexe 7 : Les réglementations et organismes sur les données 386
Annexe 8 : les feuilles de route des ministères français 389
Annexe 9 : les définitions de la *data culture* 393
Tableau des définitions de la *data culture* 393
Bibliographie de l'annexe 9 397
Annexe 10 : le guide d'entretien 398
Annexe 11 : les entretiens avec le public « associatif » 399
Associatif 1 (04/03/22) 399
Associatif 2 (21/03/22) 409
Associatif 6 (07/12/22) 419
Annexe 12 : les entretiens avec le public « entrepreneur » 430
Entrepreneur 1 (19/10/21) 430
Entrepreneur 2 (09/02/22) 446
Entrepreneur 3 (10/02/22) 459
Entrepreneur 7 (24/06/22) 468
Entrepreneur 9 (05/10/22) 474
Annexe 13 : les entretiens avec le public « institutionnel » 488
Institutionnel 1 (20/12/21) 488
Institutionnel 2 (24/03/22) 496
Institutionnel 3 (06/05/22) 505
Institutionnel 7 (10/02/23) 513
Annexe 14 : les entretiens avec le public « universitaire » 527
Universitaire 1 (28/10/21) 527
Universitaire 2 (06/01/22) 534
Universitaire 3 (01/04/22) 541
Universitaire 4 (20/04/22) 552
Universitaire 6 (01/07/22) 559
Universitaire 7 (11/07/22) 569
Universitaire 8 (13/07/22) 579
Universitaire 9 (25/11/22) 594

Liste des Figures

1. Imaginaires du discours - extrait de Soulages (2015, p. 106) 9

1.1. Requêtes mondiales obtenues à partir du terme « data literacy » dans *Google Trends* sur une période s'étalant de 2004 à 2023 - export réalisé le 23 octobre 2023. 16
1.2. Requêtes obtenues à partir du terme « data literacy » par pays dans *Google Trends* sur une période s'étalant de 2004 à 2023 - export réalité le 23 octobre 2023 16
1.3. Pourcentage d'évolution du nombre de citations par source entre 2020 et 2023 . . . 21
1.4. Nombre de définitions originales de la *data literacy* par années - extrait de (Verdi, 2023b) 24
1.5. Cadre de compétences en *data literacy* - extrait de (Calzada-Prado & Marzal, 2013) . 36
1.6. Compétences du domaine 1 - extrait de la version 2.0 du DigComp 41
1.7. Nombre de tweets publiés par jour (du 06 septembre 2021 au 05 septembre 2022) . 51
1.8. Nombre de tweets par catégorie 51
1.9. Nombre de tweets originaux par comptes uniques 54
1.10. Profil X de Kirk Borne le 16 octobre 2023 54
1.11. Comptes ayant le plus publié de retweets 55
1.12. Tweet le plus retweeté sur la data literacy 56

2.1. Evocations de l' *illiteracy* - extraits du *New England of Education* du 18 janvier 1883, Volume XVII, n°3, pp. 39-40 72
2.2. Niveaux de lecture - extrait de (Hillerich, 1976) 78
2.3. Niveaux de littératie - extrait de (Pill & Harding, 2013) 79
2.4. Pyramide d'Ackoff reconstituée à partir d' (Ackoff, 1989) 84
2.5. Procédure d'enquête selon une vision relationnelle des données - extrait de (Leonelli, 2019) 87
2.6. Granularité des données numériques - créée à partir de (Naur, 1974, pp. 76-111) . . . 103

3.1. Enchevêtrements de la *data literacy* 138
3.2. Interactions des littératies avec la data literacy - extrait de (Bhargava et al., 2015) . . . 139

4.1. Synthèse des littératies (critiques) des données - extraite et traduite de Sander (2023) 177

5.1. Tweet du Ministre de l'Économie, des Finances et de la Souveraineté industrielle et numérique du 12 septembre 2022 . . . 202
5.2. Requêtes françaises obtenues à partir des termes « culture des données », « culture de la donnée », « culture data », « datalphabétisation » et « data lettrisme » dans Google Trends sur une période s'étalant de 2004 à 2024 - export réalisé le 24 avril 2024 203
5.3. Logo de la data literacy conference - extrait du tweet du 12/09/2017 du compte @DataLitConf . . . 213
5.4. Ecosystème simplifié des acteurs français ayant travaillé sur la *data literacy* sur une période comprise entre 2015-2024 - constituée à partir des sources publiques précitées et cartographiée avec le logiciel Cosma . . . 223
5.5. Déployer une culture de la donnée de l'organisation : publics & phases - extrait de (Labelle et al., 2017, p. 8) . . . 232
5.6. Data literacy et data culture . . . 245

6.1. Occurrence des termes employés par les acteurs . . . 256
6.2. Occurrence des termes selon la typologie des acteurs . . . 257
6.3. Termes cités selon les typologies d'acteurs . . . 262
6.4. Nature(s) de la data literacy . . . 264
6.5. Liste des données citées selon les typologies de publics . . . 278
6.6. Liste des grandes catégories de projets . . . 292
6.7. Dénominations et traductions de la *data literacy* en France et à l'international . . . 299

Liste des Tables

1.1. Les sources les plus citées sur la *data literacy* (exports combinés du 20 octobre 2020 et du 06 octobre 2023) . . . 20
1.2. Les principales compétences citées dans les définitions de la *data literacy* . . . 24
1.3. Les typologies de compétences « donnée » - extrait de Ridsdale et al. (2015) . . . 37
1.4. Ensemble des initiatives sur la *data literacy* . . . 45
1.5. Pics de tweets publiés . . . 52
1.6. Exemples de bots présents dans le corpus de tweets . . . 53
1.7. Les comptes les plus retweetés . . . 56
1.8. Les vingt termes les plus employés dans les tweets . . . 56
1.9. Les hashtags employés au sein des tweets . . . 57
1.10. Correspondance en pourcentage entre les termes et les hashtags employés . . . 58
1.11. Les vingt bigrammes comportant le plus d'occurrences avec le terme « dataliteracy » 59
1.12. Les termes les plus utilisés par @KirkDBorne . . . 60
1.13. Les hashtags les plus employés par Kirk D Borne . . . 61
1.14. Les cinq tweets les plus retweetés de Kirk D Borne . . . 62

2.1. La classification des représentations de données - résumée à partir de (Naur, 1974, pp. 23-24) . . . 105
2.2. Définitions des concepts liés à la résolution de problèmes - extrait de (Naur, 1992, p. 3) 107
2.3. Comparaison entre le langage naturel et le langage de programmation - extrait de (Naur, 1992, pp. 26-35) . . . 109

3.1. Principes généraux en éthique des données - extrait d'Harfoush (2021), traduction de (Verdi, 2023a) . . . 128

4.1. Les dates d'apparition des définitions des littératies spécialisées . . . 160
4.2. Dangers de la *dataveillance* - compilation et traduction réalisées à partir de Clarke (1988) . . . 173

5.1. Les compétences en données du PIX - tirées de Pix.fr, partie « compétences » . . . 210
5.2. Valeurs et missions de la charte des infolabs . . . 217
5.3. Définitions des *data literacies* françaises . . . 225

5.4. Compétences associées aux *data literacies* françaises 227

6.1. Les personnes interrogées entre 2021 et 2023 253
6.2. Concepts mobilisés selon les typologies d'acteurs 257
6.3. Facettes de l'esprit critique vis-à-vis des données 270
6.4. Les types d'enjeux autour des données 275
6.5. Les types de données cités 277
6.6. Les réglementations citées 278
6.7. Les projets représentatifs évoqués par les acteurs 291
6.9. Termes les plus employés par @Williamstyles23 372
6.10. Les hashtags les plus employés par @Williamstyles23 373
6.11. Les tweets les plus retweetés de @Williamstyles23 373
6.12. Les dispositifs d'un agencement de données - extraits et traduits de Kitchin & Lauriault (2014) 376
6.13. Les champs d'investigation de la data-philosophie - compilés à partir de Bressler (2020) et Bressler (2023) 380
6.15. Compilation des feuilles de route des ministères français 390

Avant-propos

Remerciements

Si le travail du doctorant est évidemment central pour aboutir à la validation de sa thèse et au développement de son expertise scientifique, il reste toutefois limité, voire impossible, sans les nombreuses personnes qui l'accompagnent dans sa réflexion. Ces quatre années de thèse (2020 - 2024) ont été l'occasion de rencontres dont il est central de rendre ici un hommage :

Olivier Le Deuff qui a accepté de m'encadrer, qui a toujours répondu présent et qui a su dispenser ses conseils, m'apportant une vision fine de ce que recouvre le métier de chercheur de par son attachement à la rigueur scientifique. C'est grâce à lui que j'ai pu aujourd'hui confirmer ma vocation scientifique.

Arthur Perret qui m'a accompagné tout au long de mon parcours, tant d'un point de vue technique que moral : sans lui, cette thèse n'aurait jamais pu atteindre ce degré d'aboutissement.

Gérald Kembellec, Sarah Labelle, Anne Lehmans, Vincent Liquète et Lise Verlaet pour avoir accepté de faire partie de mon jury de soutenance et pour leur lecture attentive de ma thèse.

Ghislaine Chartron et l'ensemble de l'équipe de l'INTD du CNAM pour ces deux années formatrices (2021 - 2023) en tant qu'ATER qui ont contribué à forger mon expertise, ma pédagogie et ma volonté de poursuivre mes recherches au sein de l'Enseignement Supérieur et de la Recherche (par ordre alphabétique) : Xavier Aimé, Carole Briend, Evelyne Broudoux, Catherine De Laitre, Gérald Kembellec, Loïc Lebigre, Adriana Lopez Uroz, Anne-Solenne Maroulle, Josselin Noirel, Nadia Raïs, Claire Scopsi et Colline Urcun.

Thomas Grignon pour nos longs échanges, ses conseils répétés et son soutien maintes fois renouvelé.

Guy Melançon et Nathalie Pinède pour leur confiance pour le poste d'ingénieur d'études sur la thématique de la gouvernance des données au sein du laboratoire LaBRI de l'Université de Bordeaux qui a nourri ma thèse et continue d'alimenter mes réflexions sur les données.

Antoine Blanchard et Anne Pletinckx pour leur accompagnement et leur ouverture qui ont grandement facilité mon intégration dans l'écosystème ACT et qui m'ont permis de comprendre de manière plus fine ce que recouvre la mise en place d'une gouvernance des données.

Sarah Medjek et l'ensemble de l'équipe d'Open Data France pour la chaleur de nos échanges et pour m'avoir permis de toucher au plus près le terrain de ma thèse lors de ma participation au projet « Culture-D ».

L'ensemble des personnes qui ont accepté de répondre à mon enquête sur la *data literacy* et qui m'ont permis de développer et d'affiner mes connaissances. Nos échanges encore aujourd'hui sont une source infinie de découvertes et cette thèse leur est directement dédiée.

Orélie Desfriches-Doria qui m'a initié et accompagné dans mes premières recherches scientifiques et dont l'investissement à mon égard n'a jamais été oublié.

Mes proches qui me demandent encore quand j'arrêterai de « faire des études ».

Notes à l'attention du lecteur

Sauf indication contraire, l'ensemble des traductions sont de notre fait et sont pour la plupart inédites.

Introduction générale

> « Pour l'heure, la mise en place d'un *monde des données* est souvent appréciée, avec enthousiasme ou frayeur, comme une nouvelle puissance susceptible, pour les uns, de réinventer les marchés et l'organisation, de rendre la démocratie plus transparente, de faciliter les interactions avec les choses et l'environnement ou d'élaborer des connaissances prédictives, alors que d'autres s'alarment des usages commerciaux des fichiers, des menaces sur la vie privée, de la dictature de l'hyper-visibilité ou de la colonisation du monde vécu par un esprit de calcul et d'évaluation. Les données sont moins interrogées dans leur matérialité, pour ce qu'elles sont et pour ce qu'elles font, que pour mettre en scène les espoirs et les craintes qu'elles suscitent. Or, pendant que s'affole le débat sur les risques et les opportunités des *big data*, d'autres s'affairent à régler les algorithmes destinés à faire "parler" les données. Ils entretiennent un rapport beaucoup plus "réaliste" à la façon dont la multiplicité des traces numériques n'engage aucune signification *a priori* et s'échinent à construire de nouveaux artefacts computationnels destinés à mettre en signification les *mondes de données*. Ils composent avec l'hétérogénéité grandissante des traces, testent des modèles, mobilisent de nouveaux savoirs statistiques et conçoivent de nouvelles formes de visualisation. Car des données, il faut faire sens et forme.»
>
> Dominique Cardon, *Regarder les données* , 2012

Par ces mots, Dominique Cardon listait un ensemble de problématiques qui entravaient et entravent toujours aujourd'hui les réflexions autour des usages des données. Ces dernières sont au cœur de nombreux discours depuis plusieurs décennies et leur étude, originellement située dans le champ de la recherche scientifique, s'est étendue à un nombre conséquent d'acteurs. En effet, la généralisation et l'appropriation des évolutions techniques, en particulier l'informatique et plus globalement « le » numérique, ont modifié à un niveau systémique les manières de traiter les données dont la production exponentielle suit une courbe toujours ascendante et dont l'estimation pour 2030 est de l'ordre d'un yettaoctet (Huawei, 2021). On aurait ainsi affaire à une « révolution des données » [1] qui rebattrait les cartes de tout un pan des pratiques professionnelles.

[1] Nous l'avons précisé dans (Verdi, 2023b), le terme de « révolution des données », en anglais *Data revolution*, apparaît en 2013 et 2014 dans deux rapports : *The Report of the High-Level Panel of Eminent Persons on the Post-2015 Develop-*

> « Assurément, l'intelligence des données va transformer notre rapport à l'information aussi sûrement que l'internet des objets va bouleverser notre rapport au quotidien (d'ailleurs, l'un n'ira pas sans l'autre) » (Guillaud, 2012, p. 19).

Les données ont ainsi été investies dans les discours d'enjeux centraux, scindés en deux grands axes découlant d'une vision pharmaconique au sens de Stiegler (2008), à savoir d'un côté les données perçues comme un poison dans les cas de la *datafication* et de la *dataveillance*[2], d'un autre côté les données définies comme catalyseurs de « l'économie de la donnée » [3] et vectrices d'émancipation des individus. Elles ont été ainsi considérées comme un « enjeu politique fort » (Theviot, 2023), une monnaie (Eggers, 2013) et plus précisément une « monnaie de pouvoir » (*currency of power*) (D'Ignazio, 2017) pour l'ensemble des organisations, publiques comme privées, qui cherchaient et cherchent encore à tirer parti de leur exploitation afin de structurer leurs activités et planifier leurs décisions.

Au-delà de cette dichotomie instituée par une vision pharmaconique, les discours sur les données charrient un certain nombre de mythes au sens de Barthes (1957) : des paroles dépolitisées qui abolissent la complexité des actes humains et leur donnent la « simplicité des essences ». Ces mythes les plus tenaces prennent la forme, dans le cas d'une vision techno-utopique [4], de techno-discours : des discours ni strictement techniques, ni « autonomes » qui empêchent toute remise en question du phénomène technique (Janicaud, 1985). Ceux-ci ont pour tendance d'enfermer leurs récepteurs dans un impensé technique qui installent les technologies dans la posture de l'évidence (Robert, 2020) où « la » donnée est à la fois essentielle et essentialisée, érigée en emblème d'une « nouvelle utopie technologique » (Chartron & Broudoux, 2015). Elle serait une « essence agissante » (Spinoza, 1677), un

ment Agenda (HLPP 2015, 2013) et *A world that counts* (Data Revolution Group, 2014). Elle y est décrite comme une démultiplication des données du fait de l'évolution des nouvelles technologies, et plus globalement des évolutions numériques, et est perçue comme le moyen permettant d'améliorer les politiques publiques (en particulier du point de vue écologique) et de développer l'économie sur la base de sources de données venant alimenter des logiques statistiques. La critique apportée par MacFeely (2020) souligne le fait qu'il ne s'agit nullement d'une révolution homogène mono-causale mais d'un ensemble de révolutions conjointes (numérique, informationnelle, culturelle et sociale).

[2] La *datafication* est « la transformation de l'action sociale en données quantifiées en ligne permettant le traçage en temps réel et l'analyse prédictive » (Mayer-Schönberger & Cukier, 2013) tandis que la *dataveillance* est « la pratique disciplinaire et de contrôle consistant à surveiller, agréger et trier les données » (Raley, 2013). Nous y reviendrons en détail au chapitre 4.

[3] Nous citons à titre d'exemple les propos d'Axelle Lemaire tenus en 2015 : « L'objectif du gouvernement est de hisser la France sur le podium des pays les plus avancés numériquement. Nous proposons de tirer parti au maximum de l'économie de la donnée », a expliqué la ministre, invitant les parlementaires à « construire ensemble la République, à l'heure du numérique, la République numérique ». Rapporté dans Cuny (2015).

[4] Pour Sfez (2002), des marqueurs de l'utopie que sont (1) l'isolement (pour servir de modèle à la société future, elle doit se déployer sur un territoire coupé du reste du monde), (2) la pleine maîtrise du narrateur sur son récit (il a un regard total ou totalitaire de maitre du monde), (3) la technique qui résout tous les problèmes et (4) change l'ordre du monde, (5) le changement qui nous fait parvenir à un « Eden », deux s'appliquent au domaine des techniques de l'information. D'une part, le marqueur de la toute puissance de la technique, qui règle tous les problèmes sociaux et économiques, directement, sans médiation. D'autre part, le marqueur de l'hygiène qui, appliqué à la vie sociale, signifie transparence et sécurité totale sans aucun risque.

être vivant non organique (Moreau, 2022), pouvant répondre à l'ensemble des demandes de la société notamment grâce à la puissance des calculs automatisés qu'elle alimenterait par l'intermédiaire des algorithmes et de l'intelligence artificielle :

> « Les idolâtres de la machine présentent en général le degré de perfection d'une machine comme proportionnel au degré d'automatisme » (Simondon, 2012, p. 11).

Par le truchement des mouvements de l'*open Data* et du *big data*, traduits et incarnés différemment selon les pays, a été apportée une certaine visibilité sur les problématiques liées aux données telles que l'opacité, la non-accessibilité et les mésusages liés au non-respect de la vie privée (Calzada-Prado & Marzal, 2013; Deahl, 2014), comme cela a été le cas en France dans les années 2010 (Chignard & Marchandise, 2012). Plus qu'une mise en visibilité, ces mouvements ont porté le besoin d'une appropriation des données, renforcée depuis par les affaires Snowden, *Cambridge Analyticia* ou encore la pandémie de Covid-19.

Si une vision techniciste, principalement incarnée par la science des données (*data science*) et la figure du *data scientist*, a prédominé jusqu'au début du XXIe siècle, notamment en raison du fait que les données étaient l'affaire de spécialistes et / ou réservées à un public professionnel spécifique, la simplification et la démocratisation des logiciels et des technologies numériques, comme dans le cas des applications des *smartphones*, ont généralisé à l'ensemble de la société les questionnements de cet écosystème autrefois siloté ; la thématique des données est devenue une « donnée » du quotidien des individus.

Ces derniers, consciemment ou non, produisent des données qui alimentent des logiques globales d'analyses de tendances qui façonnent à leur tour la génération de nouvelles données et produisent une nouvelle réalité, ce fameux « monde de données » (Guillaud, 2012). En tant qu'« absolue totalité de l'ensemble global des choses existantes » (Kant, 2001, p. 425), ce monde intègre une vision totalisante où les données constituent l'essence de toute chose [5], ordonnant [6] les actes telle une « main invisible » dont il s'agirait de déceler l'existence pour espérer échapper à son déterminisme grâce à une stratégie globale d'acculturation portée par la *data literacy*.

Très souvent employé dans les discours, le terme d'acculturation, à l'origine proposé par J.W Powell en 1880 pour parler de la « transformation des modes de vie et de pensée des immigrants au contact de la société américaine » (Graber, Megard Mutezintare, & Gakuba, 2010), a été repris pour désigner « l'ensemble des phénomènes qui résultent d'un contact continu et direct entre des groupes d'individus de cultures différentes et qui entraînent des changements dans les modèles culturels initiaux de l'un ou des deux groupes » (Redfield, Linton, & Herskovits, 1936). L'acculturation

[5] L'individu, et les faits de manière plus globale, ne serait ici plus dissociable de son double numérique constitué de données, la donnée agissant alors comme miroir numérique de l'existence.

[6] Monde vient à l'origine du latin *Mundus*, calqué sur le grec *Kosmos* (*κόσμος*) qui signifie « ordre », « harmonie ».

s'inscrit donc « dans une logique dialectique où viennent se mêler des flux de déculturation (perte de culture) et d'enculturation (appropriation d'une nouvelle culture) » (Vandevelde-Rougale & Fugier, 2016).

Une grande controverse a émergé vis-à-vis de l'emploi de ce terme qui porte en lui une vision coloniale qui impose les caractéristiques d'une culture spécifique et tend à effacer les autres. Néanmoins, du fait que ce terme a été intégré dans une vision à la fois éthologique et psychologique (Vandevelde-Rougale & Fugier, 2016) pour désigner aujourd'hui un ensemble de modalités d'apprentissage qui concourent au partage d'une culture globale sur les données, il reste encore aujourd'hui très utilisé [7].

Or, l'émergence de visions plurielles sur les données freinent les logiques d'apprentissage, voire tendent à les enrayer complètement. Les données ne sont en effet pas directement visibles (*Pour une culture critique des données*, 2015) et leur conceptualisation est difficile à appréhender en-dehors d'un milieu spécialisé ; répondre de manière simple et concise à la question « qu'est-ce qu'une donnée ? » étant une difficulté maintes fois rencontrée dans la littérature. Une difficulté renforcée par ses interactions avec des objets tout aussi difficiles à appréhender que sont les algorithmes et l'intelligence artificielle.

La *data literacy* est ainsi évoquée comme « la » solution miracle qui permettrait d'apporter un éclairage sur les données et de résoudre l'ensemble des problématiques associées auxquelles la *data science* dans son ensemble n'apporte pas suffisamment de solutions (Clara, 2024).

Objectifs de recherche

Notre recherche s'inscrit dans les Sciences de l'Information et de la communication (*SIC*). Nous nous inscrivons en continuité avec l'héritage légué par différents chercheuses et chercheurs sans lesquel(les)s cette thèse n'aurait tout simplement pas existé. Elle a pour but de contribuer aux précédents travaux réalisés sur les littératies et d'apporter un éclairage sur les représentations des données, de saisir ce que recouvrent les logiques d'apprentissage et de transmission en recontextualisant les pratiques et les rôles attribués aux données.

Rappeler et réaffermir le lien entre donnée et information, et plus globalement les logiques informationnelles, est un point important que nous tiendrons à souligner dans cette thèse : non seulement

[7] Cela se voit en particulier dans les discours des entreprises où il est question d' « acculturation à la / aux donnée(s) » ou d'« acculturation data ».

dans une logique d'inscription dans notre discipline de recherche mais également pour affirmer la place centrale des *SIC* dans l'appréhension du phénomène des données.

Déconstruire la « révolution des données » et dépasser la vision essentialiste des données [8], trop souvent formalisées dans « la » donnée qui amène à la production et la généralisation d'applications utilitaires sur un mode unique de type *one size fits all*, est un but recherché afin de cesser d'appréhender leur importance et leurs usages selon un prisme purement pharmaconique ou selon une vision totalisante présentés précédemment.

Nous nous permettons toutefois de préciser que cette thèse n'a pas pour but d'une part de recenser l'ensemble des discours sur les données : ces thématiques seront naturellement abordées mais nous n'oublions pas ici le sujet central de nos recherches qui est la *data literacy*, à savoir cet accompagnement des individus leur permettant de décrypter les logiques d'usages liées aux données, leur offrant la possibilité de développer une attitude réflexive et proactive.

D'autre part, de proposer un modèle data littéracique déclinable permettant d'alimenter des projets de formation : nous nous situons ici en amont de ce travail en réalisant un état de l'art nécessaire afin de poser les contextes, les acteurs, les représentations et les conceptualisations. Il existe en effet de nombreux travaux sur ce sujet : évidemment dans le monde anglo-saxon depuis 2004, date des premières conceptualisations de la *data literacy*, mais également en France et en particulier dans les *SIC* avec des auteurs précurseurs comme Sarah Labelle qui a entamé dès les années 2010 un travail de délimitation de cet objet de recherche, rapidement suivie par un certain nombre de chercheurs et chercheuses comme Anne Lehmans, Vincent Liquète, Camille Capelle, Béa Arruabarrena, Gérald Kembellec, Ghislaine Chartron, Nathalie Pinède, Béatrice Drot-Delange, Olivier Le Deuff et récemment par nous-même. Et au-delà du monde universitaire, c'est toute une littérature foisonnante issue des institutions, des associations et des entreprises qu'il conviendra de répertorier et de synthétiser afin d'offrir aux lecteurs une vision claire de ce que recouvre la *data literacy* et finalement de comprendre ce qui est entendu par une littératie appliquée aux données.

[8] « Contrairement à ce que l'on pourrait croire, ces données ne sont pas des phénomènes à partir desquelles le sujet élabore un savoir sur lui-même, elles forment déjà un savoir fabriqué de toutes pièces par un algorithme. On les croit d'autant plus naturelles qu'elles seront bientôt incorporées : à la question de savoir "qui je suis", il n' y aura plus alors d'autres réponses que celles élaborées dans le langage des machines. Si l'on veut éviter que ce langage devienne celui de toutes nos expériences, il faudra réapprendre à percevoir le sensible sous les *data*. Cela implique sans doute de reprendre ce que les techniques nous ont si généreusement "donné" ». Extrait de Fœssel (2015).

Problématique et hypothèses

La problématique générale de notre recherche est donc la suivante : qu'est-ce qu'une littératie appliquée aux données ?

Vis-à-vis de cette problématique découlent de nombreuses questions. Parmi elles :

- Quelles sont les caractéristiques de la *data literacy* et ont-elles évolué dans le temps ?
- Quelles sont les visions portées sur les données et quel lien est tissé avec l'information ?
- Quels sont les enchevêtrements conceptuels et littéraciques ?
- Depuis quand pouvons-nous véritablement parler de *data literacy* et a-t-il existé des proto *data literacies* aux objectifs similaires ?
- Que recouvre une « acculturation » aux données et quelles sont ses modalités de déploiement ?
- Quels sont les acteurs qui se sont exprimés sur cette thématique, dans quels contextes et selon quelle temporalité ?
- Pouvons-nous parler d'une vision purement française de la *data literacy*?

Pour y répondre, nous formulerons ici plusieurs hypothèses qui accompagneront notre réflexion tout au long de ce travail de thèse.

Première hypothèse : la *data literacy* est un objet frontière.

L'objet frontière est une entité qui améliore la capacité d'une idée, d'une théorie ou d'une pratique à se traduire au-delà de frontières culturellement établies (Fox, 2011). Il « qualifie la manière dont les acteurs établissent et maintiennent une cohérence entre des mondes sociaux en interaction » (Trompette & Vinck, 2009, paragr. 10), qui négocient leurs différences et créent une mise en accord, sans consensus préalable, de leurs points de vue respectifs, facilitée par la « flexibilité interprétative » inhérente à l'objet frontière lui-même (Star, 2010). La *data literacy*, au même titre que la littératie, est en effet l'objet de réappropriation par de nombreux acteurs issus de mondes professionnels distincts, amenant à des évolutions croisées qu'il conviendra de circonscrire.

Deuxième hypothèse : les domaines de rattachement des porteurs des discours influent sur leurs visions de la *data literacy*.

Découlant de la première hypothèse, nous supputons un lien d'influence des domaines de rattachement des acteurs sur les discours. Selon Sfez (2002), il existerait en effet « une collection d'imaginaires particuliers propres à des métiers, professions ou vocations » (pp. 33-34) : ces imaginaires, à la fois individuels et institutionnels, déclinent la *data literacy* en un ensemble de *data literacies* aux visées

parfois contradictoires (économiques, politiques, citoyennes, démocratiques) et aux applications distinctes.

Troisième hypothèse : parler de *data literacies* plutôt que de *data literacy*.

Au même titre qu' Olson (1998) qui parlait de cultures de l'écrit « avec des conséquences et des implications fort différentes les unes des autres » et à la suite des deux premières hypothèses, nous supputons la nécessité de parler de *data literacies* et non de « la » *data literacy* du fait de la dilution d'unicité de ce concept sitôt associé à d'autres littératies, à d'autres thématiques de recherches ou encore à d'autres milieux professionnels.

Soubassement théorique

Notre cadre théorique et méthodologique se scinde en deux grandes parties interreliées : (1) les études sur la littératie et (2) l'analyse qualitative des discours.

Les études sur la littératie

Tout d'abord, nous intégrons notre recherche dans les études sur la littératie (*literacy studies*) définies par Gee (1989) comme « un nouveau champ d'étude, intégrant des approches "psycho" et "socio" au langage issues d'une variété de disciplines [...] le centre de l'attention des études sur la littératie ou de la linguistique appliquée ne devrait pas être le langage, ou la littératie, mais les pratiques sociales [9] ».

Selon Williams (2004), cette étude se scinde en deux traditions. La première est la tradition dite « restreinte » (*narrow*) qui limite la littératie (*literacy*) à la capacité de lire et d'écrire et donc aux compétences individuelles. Les travaux qui en sont issus tendent à séparer nettement les pratiques de lecture et d'écriture et à se centrer autour des connaissances et des compétences nécessaires au lecteur. Une deuxième tradition dite « élargie » (*broad*) étudie le déploiement des pratiques littéraciques dans les sociétés et se concentre sur les significations et les valeurs d'un comportement lettré (*literate behavior*) dans les contextes sociaux.

Ces deux traditions renvoient aux modèles « autonome » et « idéologique » de la littératie développés par Street (2006). Dans le modèle autonome, la littératie est vue comme un ensemble de com-

[9] « *a new field of study, integrating "psycho" and "socio" approaches to language from a variety of disciplines [...] the focus of literacy studies or applied linguistics should not be language, or literacy, but social practices* ».

pétences neutres détachées du contexte social, à l'inverse du modèle idéologique. Ces deux modèles, tout comme les deux traditions précitées, ne sont pas mutuellement exclusifs et des glissements entre les deux peuvent se produire. Nous en parlerons en détail dans le chapitre 2.

L'attention portée aux discours, définis par Gee (1989) comme « des manières d'être au monde ; ils sont des formes de vie qui intègrent des mots, des actes, des valeurs, des croyances, des attitudes et des identités sociales ainsi que des gestes, des regards, des positions du corps et des vêtements [10] » est ici importante et constitue notre second soubassement théorique.

L'analyse qualitative des discours

En tant que « signes de richesse destinés à être évalués, appréciés, et des signes d'autorité, destinés à être crus et obéis » (Bourdieu, 2014, p. 72) et que « supports et langages de la culture » (Jeanneret & Souchier, 2001), les discours construisent la représentation du réel que le locuteur souhaite faire partager par son allocutaire (Seignour, 2011). Selon Maingueneau (2021), ils permettent à la fois de désigner des objets d'analyse et de montrer que l'on adopte un certain point de vue.

En effet, les discours sont des actes de communication par lequel « le locuteur autorise la révélation et l'objectivation partielles de son savoir et de sa volonté sous la forme d'un système de croyances contenu dans l'esprit de l'auditeur » (Kuroda, 2012, paragr. 30). Ils ont pour fonction première l'*enargeia* « que l'on peut traduire par "évidence" ou "visibilité", c'est-à-dire que le discours doit d'abord montrer : l'*enargeia* ou l'*evidentia* doit transformer l'auditeur-lecteur en spectateur. L'*ekphrasis* consiste alors à donner à voir avec "évidence" l'objet par le langage, c'est un discours qui présente en détail et met sous les yeux de façon évidente ce qu'il donne à connaître » (Klock-Fontanille, 2014). Ces démonstrations de savoir concourent à l'avènement de représentations :

> « Lorsque les individus ou les groupes parlent de situations qu'ils vivent, lorsqu'ils s'expriment sur les événements qui les touchent ou qu'ils tiennent pour extérieurs, ils donnent à voir leurs représentations. Résultat d'une activité mentale qui fonde en partie les positions et les conduites, les représentations permettent d'interpréter le monde et renvoient simultanément à différentes sources de significations » (Giust-Desprairies, 2009, introduction, paragr. 1).

Les représentations, en tant que *Weltanschauung*, permettent donc d'interpréter le monde, et dans le cas qui nous intéresse ici, d'interpréter le « monde des données ». La place des acteurs est ici centrale :

[10] « *Discourses are ways of being in the world; they are forms of life which integrate words, acts, values, beliefs, attitudes, and social identities as well as gestures, glances, body positions, and clothes* ».

en tant que « foyers de significations », ils reconfigurent le monde par la médiation des données. En effet, Jeanneret & Souchier (2001) précisaient que les discours, au-delà de les accompagner, configuraient les objets concernés. Les représentations et les imaginaires, intégrés aux discours, doivent donc émerger et être analysés pour déceler les visions des acteurs qui configurent les données et leurs usages.

L'ensemble de ces représentations proviennent des individus et plus précisément de leur imaginaire. Pour Giust-Desprairies (2009), l'imaginaire désigne « l'ensemble des productions d'une fonction mentale appelée imagination, à la fois de l'ordre de la reproduction, par le pouvoir qu'elle a de faire revivre des perceptions déjà éprouvées, et créatrice dans sa capacité à former des images selon des combinaisons inédites » (chapitre 2, paragr. 75). Il est comme une « image de la réalité, mais en tant que cette image interprète la réalité, la fait entrer dans un univers de significations » (Charaudeau, 2014, p. 158). Celui-ci ne s'analyse pas seulement à travers des « contenus pleins » mais aussi « à partir de la façon particulière qu'a une image de faire surgir un monde ou d'empêcher un autre d'advenir » (Giust-Desprairies, 2009, Chapitre 1, paragr. 61). De fait, il existe autant d'imaginaires que d'individus, parfois cristallisés au sein d'une institution sous forme de culture institutionnelle [11].

Soulages (2015) propose à ce sujet une représentation schématique des imaginaires du discours dans laquelle nous nous inscrivons :

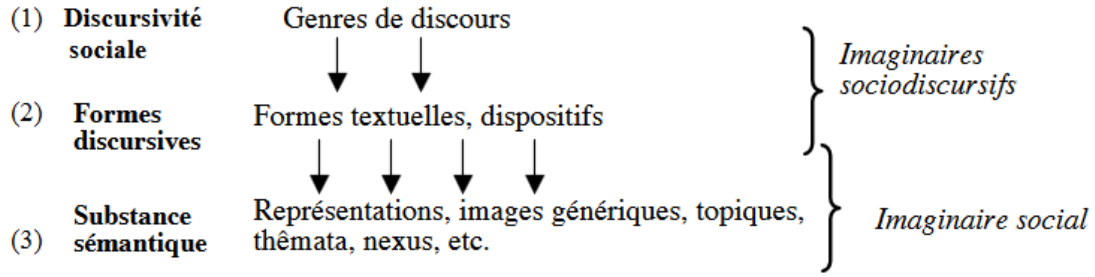

Figure 1 : *La sémiosis sociale ou les imaginaires du discours*

Figure 1.: Imaginaires du discours - extrait de Soulages (2015, p. 106)

Nous nous intéresserons plus particulièrement à la « substance sémantique » définie comme le « contenu du discours, c'est-à-dire ce qui est dicible à l'intérieur d'un type de discours suivant un répertoire sémantique et anthropologique relatif à une société donnée » (p. 107).

Pour Seignour (2011), trois types d'analyses interdépendantes sont applicables aux discours :

[11] Issue de sa synthèse de la littérature, Bertrand (2011) en proposait la définition suivante : « la culture institutionnelle est un processus socio-dynamique et un ensemble de connaissances (perceptions, jugements, intuitions, informations, stratégies, valeurs, etc.) utilisées par des groupes afin de se doter de meilleurs moyens de survivance dans un monde caractérisé par des relations antagonistes entre les personnes et les groupes » (p. 54).

> « 1. Analyse du système d'énonciation, c'est-à-dire de la façon dont l'émetteur et le récepteur s'inscrivent dans l'énoncé ; 2. Analyse du référentiel du discours, c'est-à-dire des principaux champs sémantiques ; 3. Analyse du circuit argumentatif, c'est-à-dire de la nature et de la structure des arguments. » (paragr. 17).

Comme précisé, nous nous attarderons plus particulièrement sur les champs sémantiques qui désignent « l'ensemble des mots utilisés pour caractériser une notion, une activité, une personne, etc. » car « analyser le champ sémantique d'une notion, c'est relever dans un énoncé tous les termes s'y rattachant afin de comprendre la perception que l'énonciateur en a, et/ou qu'il vise à faire partager par l'énonciataire du discours » (Seignour, 2011, paragr. 25).

Dans les études sur la littératie, Gee (1989) notait la présence de deux discours, à savoir le discours primaire et le discours secondaire. Le discours primaire (*primary discourse*) est celui qui se construit au sein du foyer, de la communauté, lors de l'enfance (ce que Gee appelle la socialisation primaire (*primary socialization*)) tandis que le discours secondaire (*secondary discourse*) désigne les discours issus des sphères socio-culturelles situées en-dehors du foyer comme l'école, les lieux de prière, les organisations étatiques, les entreprises, etc. Toujours selon Gee, la littératie est ainsi « la maîtrise ou le contrôle courant d'un discours secondaire [12] » et est ici perçue comme émancipatrice : par la maîtrise des discours secondaires, les individus s'émancipent, s'encapacitent et gagnent en « pouvoir ». C'est ici ce que nous verrons dans les discours liés à la *data literacy* où la maîtrise des données permet cette encapacitation.

Sur la base des études sur la littératie et de l'analyse qualitative des discours, nous effectuerons ainsi une analyse des discours sur la *data literacy* à partir de trois corpus :

- un premier corpus, centré sur les publications scientifiques et professionnelles, réactualisé jusqu'à l'écriture du manuscrit de thèse ;
- un deuxième corpus basé sur 44544 tweets intégrant le terme « data literacy » sur une période comprise entre septembre 2021 et septembre 2022 ;
- un troisième corpus comprenant 32 entretiens semi-directifs réalisés entre 2021 et 2023 auprès d'acteurs français.

[12] « *mastery or fluent control over a secondary Discourse* ».

Plan de la thèse

Nous déclinerons notre réflexion en trois parties divisées en six chapitres.

La première partie portera sur les caractéristiques générales de la *data literacy* et sur les influences diverses qui ont accompagné sa conceptualisation.

Le premier chapitre fait office d'introduction générale au concept de *data literacy* où nous rappelons ses origines, les différentes étapes de son développement et l'ensemble des publications qui l'ont premièrement conceptualisées. Il s'agit ici de rappeler que le paysage définitionnel, de provenances plurielles, associé à ce concept est en évolution constante. Par la suite, l'ensemble des acteurs internationaux seront présentés pour offrir un point de vue général sur l'effervescence des initiatives en *data literacy*. Enfin, le chapitre se terminera sur une exploration de tweets recueillis de 2021 à 2022. Celle-ci nous a permis de prendre en compte le rôle joué par ce réseau social en termes de stratégie communicationnelle et de déterminer le positionnement de certains acteurs vis-à-vis de la *data literacy*.

Le second chapitre proposera un récapitulatif des influences passées ayant amené à la formation du concept de *data literacy*. Il posera les concepts de « littératie » et de « donnée » et présentera l'évolution de leurs définitions, visions et périmètres. Il décrira par ailleurs la datalogie (*datalogi*) en tant qu'influence oubliée.

La seconde partie traitera des visions plurielles de la *data literacy*.

Le troisième chapitre listera tout d'abord les enchevêtrements de la *data literacy* avec un ensemble de concepts liés aux données que sont la science des données (*data science*), la gouvernance des données, l'éthique des données et l'esprit critique vis-à-vis des données. Nous décrirons ensuite les interactions littéraciques. Il existe en effet de nombreuses littératies qui ont toutes été étudiées au cours des précédentes décennies comme l'*information literacy*, la *digital literacy* ou encore la *statistical literacy* qui ont influencé certaines visions de la *data literacy*. Le but de cette analyse est de déterminer les types d'enchevêtrements proposés par les auteurs présents dans la littérature, de souligner les influences des domaines d'études respectifs et desquels découlent des didactiques spécifiques.

Le quatrième chapitre explorera les évolutions de la *data literacy* présentes sous la forme de « littératies spécialisées ». La première partie sera centrée sur les aspects couverts par les littératies spécialisées des données comme l'aspect scientifique, l'aspect démocratique ou encore l'aspect pédagogique, tandis que la seconde partie présentera les littératies spécialisées des données issues du rejet de la *datafication* et de la *dataveillance*.

La troisième partie explorera la potentialité d'une *data literacy* française.

Pour ce faire, le cinquième chapitre se basera un triple état de l'art. Un premier état de l'art sur les discours des institutions et des entreprises françaises traitant de la nécessité de maîtriser les données. Un second état de l'art des acteurs français ayant travaillé sur la *data literacy* et de leurs initiatives dans ce domaine. Un dernier état de l'art sur les transpositions et les (re)conceptualisations universitaire et pédagogique françaises de la *data literacy*.

Et enfin le sixième chapitre s'articulera autour des résultats tirés de notre enquête auprès de 32 acteurs représentatifs de l'écosystème français ayant travaillé sur les formations aux données. Celle-ci a eu pour but d'approfondir des points peu abordés ou explicités dans la littérature scientifique et professionnelle et de déterminer l'évolution des pensées des acteurs au regard de leurs anciennes publications (ces dernières étant pour certaines antérieures de plusieurs années à notre enquête).

Partie I.

Définir la *data literacy*

Chapitre 1.

Les caractéristiques de la *data literacy*

> « *L'Humanité est à un tournant de son histoire. La masse des données acquises est formidable. Il faut de nouveaux instruments pour les simplifier, les condenser ou jamais l'intelligence ne saura ni surmonter les difficultés qui l'accablent, ni réaliser les progrès qu'elle entrevoit et auxquels elle aspire.* »
>
> Paul Otlet, *Traité de documentation*, 1934

1.1. Introduction

La *data literacy* est un concept d'origine nord-américaine. Parfois nommée *data wise* (Verbiest et al., 2014), *datacy* (Radermacher, 2017) et/ou confondue avec le concept de *data acumen*[1] (Bargagliotti et al., 2020; Keller & Shipp, 2021; Keller, Shipp, Schroeder, & Korkmaz, 2020; National Academies of Sciences, Engineering, and Medicine, 2018), son usage s'inscrit dans les discours promouvant la valeur des données et la nécessité de leur exploitation pour le développement économique (Commission Européenne, 2020b, p. 16).

[1] Décrit par Sallie Ann Keller et Stephanie Shipp comme la capacité de poser de bons jugements sur les usages des données pour trouver des solutions adaptées aux problèmes du monde réel, ce terme est présenté comme une littératie « moderne » qui se distancie d'une précédente vision de la *data literacy* jugée trop proche des statistiques et englobée au sein de la science des données (*data science*), elle-même perçue comme l'apprentissage à partir de données dans un contexte de problèmes (Keller & Shipp, 2021). Ce « flair des données » doit permettre (1) de connaître les données et les outils pour les traiter, (2) de développer des solutions (stratégiques, anticipatives, tactiques) aux problèmes et (3) d'appliquer les analyses nécessaires pour les soutenir. De plus, il englobe à la fois les compétences de base en *data literacy* mais également les compétences quantitatives et informatiques affiliées à la science des données. Il s'agit donc d'un continuum intégrant des niveaux variés de maîtrise des données. L'aspect éthique est également une prérogative de cette nouvelle littératie, nécessitant un recul critique pour déterminer les biais potentiels dans les données.

Avec l'émergence de nouvelles pratiques et problématiques liées à la démocratisation des objets connectés, la *data literacy* a acquis une place de plus en plus centrale dans les discours. En nous basant simplement sur une première recherche, nous constatons que l'intérêt pour cette thématique a connu une croissance significative à partir de 2022 (figure 1.3) :

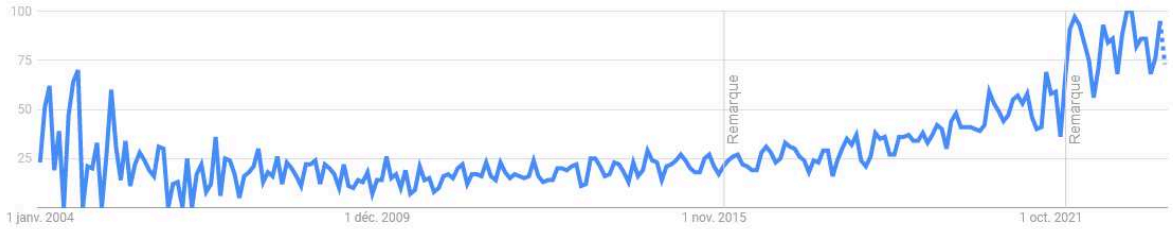

Figure 1.1.: Requêtes mondiales obtenues à partir du terme « data literacy » dans *Google Trends* sur une période s'étalant de 2004 à 2023 - export réalisé le 23 octobre 2023.

Si la corrélation entre de simples requêtes et les évènements importants, telle la pandémie de Covid-19, ne peut être que supposée, nous notons toutefois l'apparition de deux pics : l'un en 2004 lorsque la *data literacy* a été définie officiellement dans le monde universitaire, l'autre en 2022 peu après la fin de la pandémie.

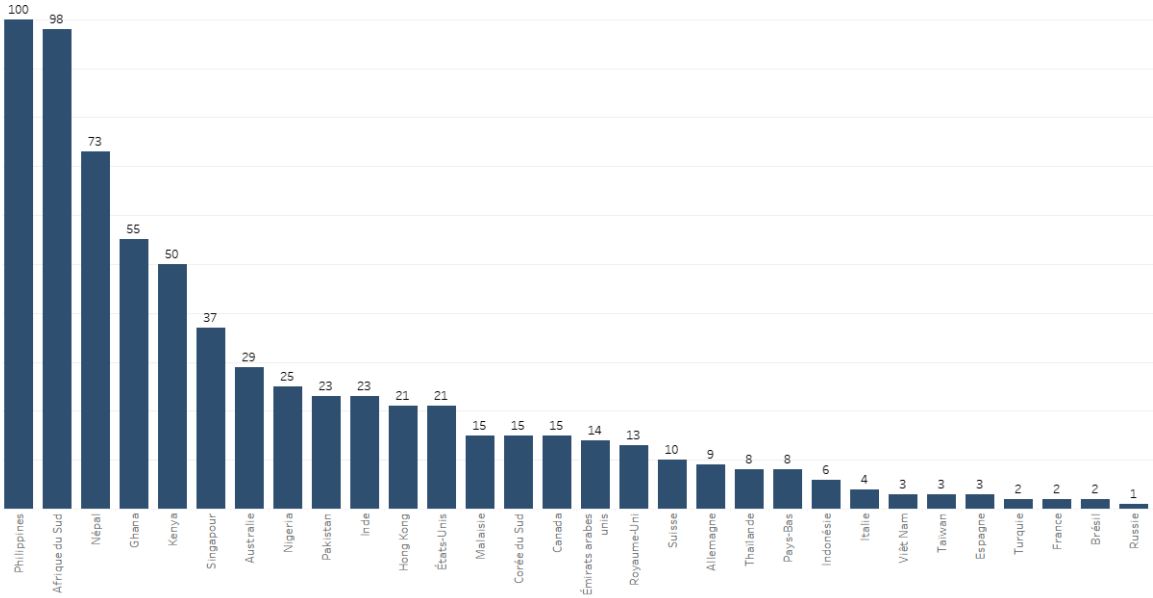

Figure 1.2.: Requêtes obtenues à partir du terme « data literacy » par pays dans *Google Trends* sur une période s'étalant de 2004 à 2023 - export réalité le 23 octobre 2023

Quand nous nous attardons sur les requêtes réalisées pays par pays (figure 1.2), nous remarquons un fort intérêt de la part de pays aux PIB faibles et intermédiaires, rejoignant ici l'étude précédemment

effectuée par Kuonen (2021). La *data literacy* est un enjeu mondial dont chaque acteur souhaite s'emparer.

En nous plaçant à un niveau plus granulaire, nous avons remarqué que nombre d'études ont sollicité le développement et l'acquisition de la *data literacy* au cours des dernières années. L'étude de l'entreprise *Qlik* de 2022 la citait comme « la compétence la plus demandée d'ici 2030 » (Qlik, 2022b), s'intégrant ainsi dans la lignée d'un certain nombre d'articles entrepreneuriaux requérant son usage [2]. L'entrepreneur Brent Dykes parlait même en 2017 d'un fossé à combler, le *data literacy gap* (Dykes, 2017), apparaissant lorsqu'une technologie n'est plus en mesure de compenser les compétences et connaissances déficitaires d'une personne en matière d'usages de données.

Cette volonté d'instituer la *data literacy* comme pivot central des décisions intervient en corollaire de la parution de plusieurs études ayant fait état d'une carence profonde des employés en termes de culture et de compétences vis-à-vis des données. Ainsi, le rapport de l'organisme *DataCamp* de 2023 [3] rappellait que si les compétences en *data literacy* se hissaient parmi les principales demandes des entreprises, leur maîtrise ne concernait que 5% d'entre elles (slide 14). Parmi les difficultés rencontrées par les organisations dans l'amélioration des compétences étaient citées un problème de budget suffisant dédié à cette tâche (40%), suivi par des ressources pédagogiques inadéquates (36%), un manque de soutien de la part de la direction (29%) ou encore la non propriété des programmes d'entraînement (29%) (slide 29).

Pour compenser ce manque, les organisations ont orienté leurs demandes, en premier lieu sur le recrutement de *data scientists* (Calvi, 2018), emploi considéré comme le plus « sexy » du XXIe siècle en 2012 (Davenport & Patil, 2012). Ce profil d'expert, bien que toujours recherché (Hughes, 2022), est peu à peu délaissé ces dernières années pour des fonctions plus intermédiaires comme celle de *data analyst* et de *data engineer* [4]. Et plus globalement, c'est tout un pan du marché du travail qui s'est remodelé pour intégrer les données au sein des métiers [5].

Toutefois, l'étude de Forrester Consulting (2021) commandée par Tableau dénonçait un déséquilibre flagrant entre les demandes entrepreneuriales basées sur le recrutement de personnes versées dans les usages des données et les formations universitaires américaines trop peu centrées sur cette thématique amenant à une volonté de recentrage de ces dernières.

A l'inverse, en France, les formations autour des données se sont rapidement développées dès l'émergence de recherche autour du *big data* (Lewandowski, 2017) proposant aujourd'hui une large gamme de cursus universitaires estimée à 1438 selon le rapport de l'observatoire GEN_SCAN (2023).

[2] Voir à ce sujet Korolov (2020), Sabar (2021), Precisely Editor (2022) et Marr (2022).
[3] Voir à ce sujet DataCamp (2023).
[4] Voir à ce sujet Thompson (2016) et APEC (2020).
[5] Dans sa note de recherche, Nieuwbourg (2020) recensait 70 métiers de la donnée.

Toujours selon ce dernier, les offres d'emplois publiées dans le domaine des *data* ont été estimées à 13 345 au 1er janvier 2023 ; les emplois de *data engineer*, *data analyst* et *data scientist* occupant respectivement la seconde, quatrième et cinquième place des postes les plus recherchés.

Au-delà de répondre aux besoins des entreprises en formant les individus à des métiers clairement définis, la *data literacy* est également conçue pour répondre aux besoins des citoyens dans leur vie quotidienne. Ce point, bien que déjà évoqué dans la littérature, est devenu particulièrement prégnant suite à la pandémie de Covid-19 : les usages des données, en particulier leur lecture et leur déchiffrage, ont été au cœur des pratiques quotidiennes des citoyens qui consultaient nombre de plateformes gouvernementales pour se renseigner sur les taux de contamination, taux de contagion, carte de circulation du virus, etc. En France, le tableau de bord « InfoCovidFrance » a été emblématique en ce sens, présentant douze thématiques actualisées quotidiennement et médiatisées sous la forme de datavisualisations, de cartographies ou encore de graphiques animés [6]. Une démultiplication des usages numériques perçus et pris en compte dans de nombreux rapports et études, et ce jusqu'au niveau de l'Union Européenne dès le rapport *DESI* de 2020 [7] :

> « La pandémie de COVID-19 actuelle a montré l'importance que revêtent désormais les ressources numériques pour nos économies et comment les réseaux et la connectivité, les données, l'IA et le calcul haute performance ainsi que les compétences numériques élémentaires et avancées soutiennent nos économies et nos sociétés en permettant la poursuite du travail, le suivi de la propagation du virus et l'accélération des activités de recherche sur les traitements et les vaccins » (p.2).

La pandémie a fait de la *data literacy* l'objet central de revendications de la part des entreprises (Morrow, 2022) mais également du domaine universitaire (Nguyen, 2021). Diego Kuonen, professeur à la faculté d'économie et de management de Genève, lançait justement une alerte sur ce point :

> « La pandémie de COVID-19 illustre clairement que les données constituent une ressource essentielle pour l'élaboration et l'évaluation de stratégies adaptées dans le but de résoudre d'importants problèmes sanitaires et sociaux. À cette fin, il est toutefois essentiel de comprendre ces données et en particulier d'évaluer leur valeur et leur comparabilité dans le contexte approprié. La pandémie de COVID-19 a particulièrement levé le voile sur les nombreuses facettes des problèmes liés au manque de littéracie ou de culture des données. Que dit vraiment le nombre de cas confirmés? Comment collecter ces données de manière pertinente? Est-ce que les données des différentes régions et pays sont totalement comparables? Comment interpréter correctement ces données?

[6] https://www.santepubliquefrance.fr/les-actualites/2021/infocovidfrance-le-nouveau-tableau-de-bord-de-sante-publique-france-pour-suivre-l-epidemie-de-covid-19

[7] Commission Européenne (2020b), *op. cit.*

> Quelle confiance pouvons-nous accorder aux données? Que faisons-nous de ces données, comment pouvons-nous les partager et qui y a accès? Répondre à ces questions montre à quel point l'élaboration, l'accès et l'utilisation des données sont complexe [...] Par conséquent, cette pandémie de COVID-19 doit être perçue comme une occasion de renforcer la culture des données afin de garantir à l'avenir un traitement des données responsable, critique, éthique et durable. Donner aux citoyens les compétences nécessaires au traitement des données est non seulement essentiel pour leur santé et leur qualité de vie, mais aussi pour leur participation informée et responsable à notre société démocratique.» (Kuonen, 2022).

Les compétences en *data literacy* des citoyens n'étant pas optimales, notamment en Europe [8], une nécessité de formation des publics s'est faite de plus en plus pressante dans un contexte de *datafication* sur lequel nous reviendrons au chapitre 4. La question de son application présente toujours à ce jour de grandes difficultés : quelles méthodes employer ? Quels acteurs solliciter ? A quelles connaissances et compétences former ? Dans quel contexte ? Dans quels buts ?

Pour apporter des éléments de réponse, nous débuterons par une présentation des caractéristiques générales de la *data literacy* et de l'écosystème qui l'accompagne.

1.1.1. Un premier état de l'art

Nous avions réalisé en octobre 2020 un premier état de l'art et listé les dix sources les plus citées sur la *data literacy*; l'objectif était de dessiner un premier paysage définitionnel de la thématique (Verdi & Le Deuff, 2020). De plus, nous avions pu constater que les sources les plus citées provenaient de différents domaines : les sciences des bibliothèques et de la documentation, *l'information literacy*, les sciences de l'éducation ainsi que les sciences dites « dures » (statistiques, mathématiques, etc.). Les contenus se scindaient entre des définitions de la *data literacy* et/ou la présentation de divers projets impliquant son élaboration.

Nous avons reproduit cette expérience trois ans plus tard et avons précisé les évolutions du nombre de citations des différentes sources (tableau 1.1):

[8] Voir à ce sujet Eurostat (2023).

Tableau 1.1.: Les sources les plus citées sur la *data literacy* (exports combinés du 20 octobre 2020 et du 06 octobre 2023)

Titre de la source	Auteur(s)	Année	Citations (2020)	Citations (2023)	Rang (2020)	Rang (2023)
A systemic view of implementing data literacy in educator preparation	Mandinach et Gummer	2013	385	619	1	1
Incorporating data literacy into information literacy programs: Core competencies and contents	Cazalda Marzal et Prado	2013	204	413	2	2
Information literacy, statistical literacy, data literacy	Schield	2004	196	328	3	3
Building a Conceptual Framework for Data Literacy	Gummer et Mandinach	2015	138	282	4	4
Data literacy for researchers and data librarians	Koltay	2017	86	210	7	5
Strategies and best practices for data literacy education: Knowledge synthesis report	Risdale et al.	2015	132	204	11	6
Data literacy: in search of a name and identity	Koltay	2015	109	198	5	7
Data literacy is statistical literacy	Gould	2017	44	163	9	8
Data literacy for educators: Making it count in teacher preparation and practice	Mandinach et Gummer	2016	71	153	13	9
Data literacy: What do educators learn and struggle with during a data use intervention?	Kippers et al.	2018	72	135	12	10
Lessons learned from a two-year experience in science data literacy education	Qin et D'Ignazio	2010	91	128	6	11
Research data literacy	Schneider	2013	75	119	8	12
Data carpentry: workshops to increase data literacy for researchers	Teal, Cranston, Lapp, White, Wilson	2015	71	109	10	13

Une certaine constance se reflète ici : les quatre premières sources centrées sur la définition et la mise en application de la *data literacy* restent à ce jour les plus citées. Nous avons évalué le taux d'évolution

de ces sources dans la figure 1.3 pour vérifier si une thématique particulière se démarquait (ex : formation, épistémologie, etc.).

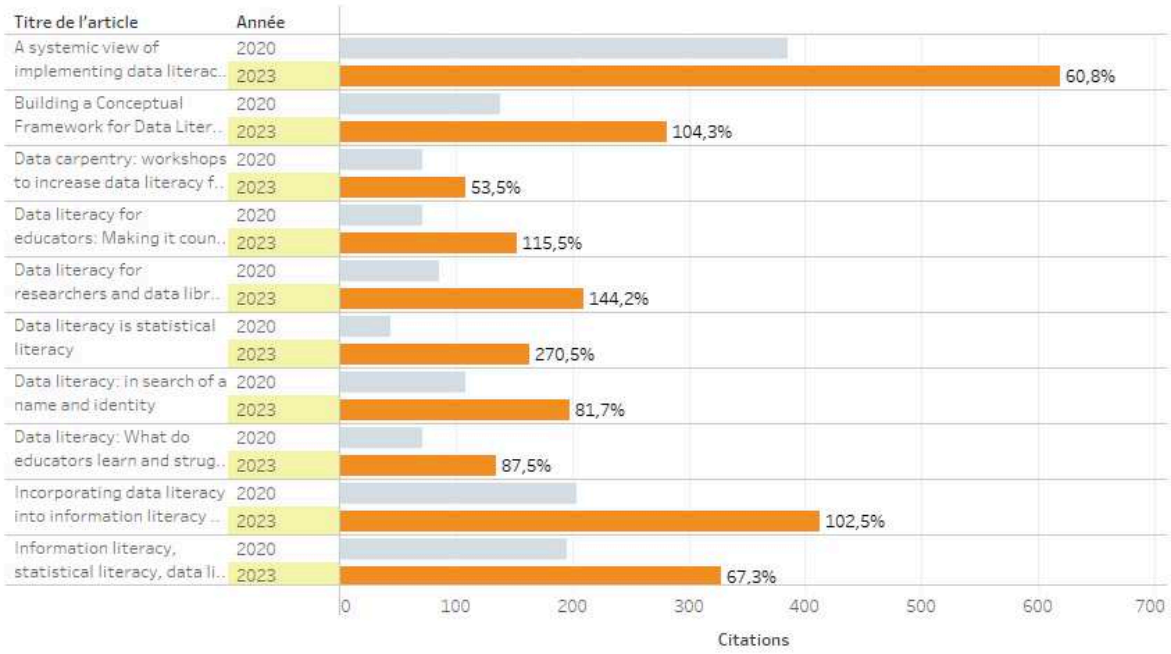

Figure 1.3.: Pourcentage d'évolution du nombre de citations par source entre 2020 et 2023

Ainsi, si toutes les sources ont vu leurs citations croître, une évolution parfois exponentielle pour certaines d'entre elles est particulièrement visible. Nous pouvons prendre le cas de l'article *data literacy is statistical literacy* qui passe de 44 citations à 163 en l'espace de trois ans, soit une augmentation de 270,5 %. Les sources les plus citées proposent d'une part de définir la *data literacy*, d'établir son identité en soulignant ses liens avec d'autres concepts ou littératies voisines (ex : *statistical literacy* et *information literacy*) ; d'autre part de proposer des modèles applicatifs sous la forme de référentiels de compétences sur lesquels nous reviendrons au cours de ce chapitre.

1.1.2. Un paysage définitionnel en perpétuelle évolution

A travers l'analyse de la littérature scientifique et professionnelle, nous constatons que la dénomination *data literacy* a été employée sporadiquement dès les années 1980 et 1990 (Bergqvist et al., 1998; Burstein, 1983; Hartnett, 1989; Piggott, 1993; Scheets, 1995). Quelques propositions de définitions ont été esquissées mais restaient limitées dans leur développement :

> « Une éducation en *data literacy* doit traiter la manière dont une entreprise acquiert des données, et les contenus et les significations des données, ainsi que leur transformation en informations [9] » (Hartnett, 1989).

> « Intégrer une *data literacy* ou une familiarité sur la façon dont les données ou les textes se comportent dans une banque de données [10] » (Piggott, 1993).

Bien que quelques initiatives ont vu le jour, en particulier le projet américain de la *National Science Foundation* (*NSF*) de 2002 visant à aider les professeurs de mathématiques et de sciences à développer la *data literacy* (NSF, 2002), Tedesco (2002) rappellait qu'une définition officielle du concept manquait toujours à l'appel. Il faudra attendre 2004 pour que la *data literacy* soit doublement définie comme (1) la capacité d'accéder, évaluer, manipuler, résumer et présenter les données [11] et (2) comme la capacité d'examiner plusieurs mesures et niveaux de données afin d'en tirer des conclusions [12].

Schield positionnait la *data literacy* dans le domaine des bibliothèques, la désignant comme une compétence indispensable aux bibliothécaires. En ce sens, la *data literacy* n'était qu'une étape vers la constitution de l'information.

> « Les universités subissent des contraintes financières qui s'accroissent. Parfois, le personnel est considéré comme une dépense générale, tandis que le personnel enseignant est considéré comme un coût de production. Cette pression peut avoir tendance à réduire les fonds disponibles pour les personnels des bibliothèques. Pourtant, les bibliothécaires ont une opportunité unique compte tenu de leur formation. Ce sont des généralistes et non des spécialistes. Leur objectif n'est pas celui d'une discipline particulière. En tant que tels, ils sont éminemment qualifiés pour enseigner aux étudiants comment penser de manière critique, comment devenir des info-lettrés, comment devenir des lettrés en statistiques et comment devenir des lettrés des données [13] » (Schield, 2004a).

---

[9] « *Data literacy education should deal with how a company acquires data, and the contents and meanings of the data, as well as its translation into information* ».

[10] « *Building in data literacy or familiarity with how data or text behave inside a data bank* ».

[11] « Et les étudiants doivent être des lettrés de données : ils doivent être capables d'accéder, évaluer, manipuler, résumer et présenter des données », notre traduction de « *and students must be data literate: they must be able to access, assess, manipulate, summarize, and present data* ». (Schield, 2004a).

[12] « La capacité d'examiner de multiples mesures et de multiples niveaux de données, de considérer la recherche, et de tirer des conclusions saines », notre traduction de « *the ability to examine multiple measures and multiple levels of data, to consider the research, and to draw sound inferences* » (Love, 2004).

[13] « *Colleges and universities are under increasing financial constraints. At times, staff are sometimes viewed as overhead whereas faculty are viewed as a cost of production. This pressure may tend to reduce funds available for library staff. Yet, librarians have a unique opportunity in view of their training. They are generalists, not specialists. Their focus is not the focus of a particular discipline. As such they are eminently qualified to teach students how to think critically, how to become information literate, how to become statistically literate and how to become data literate* ».

Love (2004) quant à elle l'intégrait dans le domaine pédagogique, se basant en cela sur le projet *Using Data Project* de la *NSF* présenté précédemment.

> « L'objectif recherché est d'influencer la culture scolaire afin que les éducateurs puissent utiliser des données de manière continue, collaborative et efficace pour améliorer l'enseignement et l'apprentissage des mathématiques et de la science [14] ».

Les définitions ont continué d'évoluer et à partir de 2013 la *data literacy* est vue comme la capacité d'accéder, d'interpréter, d'évaluer, de gérer, de manipuler et d'utiliser de manière critique et éthique les données (Calzada-Prado & Marzal, 2013) grâce à une connaissance des caractéristiques des données, de leur cycle de vie et des différents effets engendrés par leur usage, notamment en termes de sécurité et de protection de la vie privée (Crusoe, 2016). Néanmoins certains auteurs estiment qu'une définition englobante manque toujours :

> « Il n' y a pas de consensus sur une définition englobante et applicable de la *data literacy*. Plusieurs travaux sur la *data literacy* l'ont décrite comme un phénomène centré autour des usages, avec une emphase sur "la capacité des non spécialistes d'utiliser les données" (Frank, Walker, Attard, & Tygel, 2016), tandis que les autres savent en plus "collecter et gérer" les données (Ridsdale et al., 2015) ou "choisir et nettoyer" les données (Wolff, Gooch, & Cavero Montaner, 2016) [15] » (Misra, Jutting, & Kuonen, 2021).

Pour vérifier cela, nous reprenons dans la figure 1.4 nos travaux présentés dans (Verdi, 2023b) où nous avions recensé trente-et-une définitions originales à travers une étude bibliométrique de sources employant le terme *data literacy* sur une période allant de 2004 à 2022 en combinant *Google Scholar*, *Research Gate* et *Publish or Perish*. L'ensemble des définitions sont présentes en annexe 1.

Nous avions observé deux grands pics entre 2012 et 2016. Les évènements cités dans les sources comprenaient le *big data*, l'*open data*, les nouvelles technologies de l'information et les logiques *data-driven* à l'échelle sociétale.

Le terme *ability* est le plus cité (23 fois) tandis que le terme *competence* ne l'est que deux fois. *Capacity* n'est mentionné quant à lui qu'une fois (Athanases, Bennett, & Wahleithner, 2013). En prenant en compte le dictionnaire *Collins* [16], *ability* renvoie à une compétence dans laquelle une personne excelle,

[14] « *The aim is to influence school culture to be one in which educators use data continuously, collaboratively, and effectively to improve teaching and learning mathematics and science* ».

[15] « *There isn't a consensus on one overarching and actionable definition of data literacy. Several works on data literacy describe it as a use-centric phenomenon, with emphasis on "the ability of non-specialist to make use of data" (Frank et al. 2016), while the others include ability to "collect and manage" data (Risdale et al, 2015) or "select and clean" data (Wolff et al, 2016) in addition* ».

[16] https://grammar.collinsdictionary.com/english-usage/what-is-the-difference-between-ability-capability-and-capacity

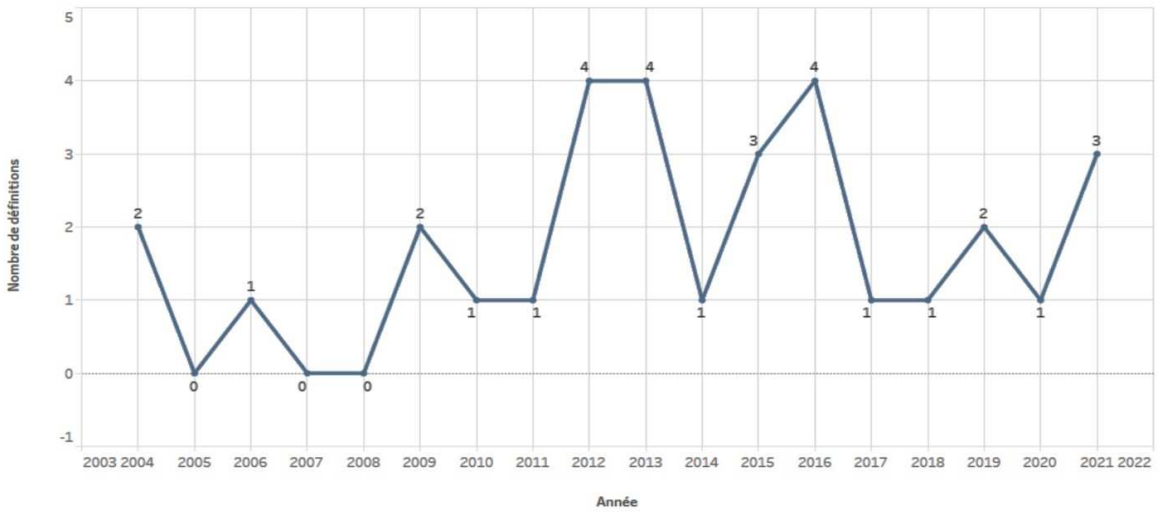

Figure 1.4.: Nombre de définitions originales de la *data literacy* par années - extrait de (Verdi, 2023b)

un « talent » dirions-nous. Ce terme se confond avec celui de *competence* toujours selon le même dictionnaire [17]. Enfin, *capacity* se rapporte à la capacité d'accomplir une action. Ces trois notions sont ici confondues dans leur signification et englobent un panel de sous-compétences spécifiques.

Tableau 1.2.: Les principales compétences citées dans les définitions de la *data literacy*

Compétences	**Récurrences du terme**
Utiliser (*use*)	11
Analyser (*analyse/analyze*)	9
Interpréter (*interpret*)	9
Evaluer (*evaluate/assess*)	8
Collecter (*collect/collecting*)	7
Comprendre (*understand*)	7
Gérer (*manage/managing*)	7
Déduire (*draw inferences/conclusion/meaning*)	5
Traiter (*process*)	4
Communiquer (*communicate*)	4
Accéder (*access*)	4
Lire (*read*)	3
Manipuler (*handle*)	3
Identifier (*identify*)	2

[17] https://www.collinsdictionary.com/dictionary/english/competence

Nous avions ensuite analysé le vocabulaire et listé ces dernières pour l'ensemble des définitions dans le tableau 1.2. L'usage généraliste des données (*use*) arrivait en première position, suivi de pratiques plus spécifiques de traitement comme la collecte (*collect*) ou la gestion (*manage*). À ces compétences fonctionnelles s'ajoutaient l'éthique des données en tant que règle déontologique et l'esprit critique en tant que recul intellectuel sur le contexte de production des données et sur leurs effets au niveau sociétal. Il s'agit donc pour le futur *data literate* de développer des compétences et des attitudes ou postures, un savoir-faire et un savoir-être critique et éthique.

1.2. La nécessité pédagogique

1.2.1. Le *data literate*

1.2.1.1. Les éléments constitutifs du *literate*

En tant que littératie, la *data literacy* est pourvue de trois significations selon le dictionnaire *Chambers* traduit par (Le Deuff, 2009, p. 152) :

- Premier sens : capacité à lire et écrire (*a simple ability to read and write*) ;
- Deuxième sens : possession de compétence et d'habileté (*having some skill and competence*) ;
- Troisième sens : élément d'apprentissage (*element of learning*).

Elle vise donc des objectifs de formation des publics, en l'occurrence des formations aux usages des données. La personne issue d'une telle formation est considérée comme un « lettré » (*literate*), désigné par le dictionnaire *Merriam-Webster* [18] comme quelqu'un :

1. d'éduqué, cultivé ;
2. capable de lire et d'écrire ;
3. versé dans la littérature ou dans l'écriture créative ;
4. possédant un savoir ou une compétence.

Nous nous retrouvons donc face à un individu doté d'une part de compétences, de savoirs, principalement tournés vers la pratique de la lecture et de l'écriture, et ayant atteint d'autre part un état, celui d'être éduqué et cultivé.

[18] https://www.merriam-webster.com/dictionary/literate

Grâce à ces acquis, le lettré peut développer un « esprit lettré » (*literate mind*), désigné par Paul & Wang (2011) comme la capacité d'un individu à pouvoir accéder et interpréter une information acquise selon des procédés divers. En découle une « pensée lettrée » (*literate thought*) intégrant des dimensions cognitives, sociales et affectives qui, pour émerger, doivent réunir trois conditions (Paul & Wang, 2011, p. 15) :

1. La maîtrise adéquate d'une langue ;
2. La compréhension et l'application d'un métalangage ou d'un vocabulaire spécialisé associé à une discipline ou un domaine spécifique et à la culture ou à la société en général ;
3. La capacité d'accéder et d'interpréter un langage lettré décontextualisé (*decontextualized literate language*).

Ce « langage lettré » incorpore quatre aspects (Benson, 2009, p. 175):

1. un aspect métalinguistique centré sur le sens, la fonction et la forme ;
2. l'utilisation d'un langage qui est abstrait ou décontextualisé ;
3. les spécificités linguistiques ;
4. les formes sémantiques et syntaxiques couramment utilisées dans les textes écrits et les discours d'instruction (*instructional discourse*)[19].

Il s'agit donc de développer une maîtrise fonctionnelle du langage sous toutes ses formes (orale, écrite) et ses subtilités couplée à une capacité de recul intellectuel permettant de sortir du contexte d'énonciation. Brockmeier & Olson (2002) nous avertissaient en ce sens : devenir véritablement *literate* ne peut se limiter aux simples pratiques de lecture et d'écriture, il faut atteindre « l'espace symbolique » de la littératie pour s'approprier l'histoire de l'écriture et l'écriture de l'histoire.

> « Devenir lettré implique la maîtrise active d'un ensemble de pratiques matérielles, discursives, cognitives et institutionnelles de l'écriture et de la lecture – des pratiques que l'on appelle de manière simplificatrice « écriture et lecture ». Mais ils ne font qu'ouvrir l'entrée à « l'espace symbolique » de la littératie (Brockmeier 2000), un espace d'écriture qui embrasse les ressources culturelles accumulées non seulement de l'histoire de l'écriture, mais aussi de l'écriture de l'histoire : c'est-à-dire du stockage et la transmission des connaissances, la construction et l'interprétation des traditions archivistiques. Il serait naïf de supposer qu'un lecteur individuel, en apprenant simplement à écrire et à lire, aura immédiatement accès à toutes les ressources développées dans une tradition lettrée [20] » (Brockmeier & Olson, 2002).

[19]Les discours d'instruction désignent les pratiques de discours qui transfèrent la responsabilité du processus d'apprentissage aux étudiants selon Kiemer, Gröschner, Kunter, & Seidel (2018).

[20]« *Becoming literate implies the active mastery of an ensemble of material, discursive, cognitive, and institutional practice of writing and reading – practices that are in a simplifying fashion called "writing and reading." But they only open the*

La tradition lettrée selon Olson (2001) est cette tradition permettant d'isoler les mots vis-à-vis des choses qu'ils représentent. Elle est équivalente à la culture de l'écrit (*culture of writing*) (Olson, 2011) qui permet « à qui la possède de maîtriser le temps, l'espace, le monde d'une manière spécifique » (Bouchard & Kadi, 2012, p. 11). C'est l'appropriation et technique et symbolique de la lecture et de l'écriture permettant de définir et réinterroger le monde.

1.2.1.2. L'émergence du *data literate*

Dans le cas des données, l'existence du terme *data literate* est attestée dès 1992 au sein d'un article du *Wall Street Journal* [21] écrit par le théoricien du management Peter Drucker [22]. Ce dernier appuyait la nécessité pour les cadres, alors pleinement *computer literate*, de se former aux données. Le terme employé est ici *information literate* bien que le titre de l'article mentionne explicitement celui de *data literate* :

> « Mais peu de dirigeants sont info-lettrés. Ils savent comment obtenir des données. Mais la plupart doivent encore apprendre à utiliser les données. Peu de cadres savent encore se demander : de quelles informations ai-je besoin pour faire mon travail ? Quand en ai-je besoin ? Sous quelle forme ? Et auprès de qui dois-je les obtenir ? Moins nombreux encore se demandent : quelles nouvelles tâches puis-je entreprendre maintenant que je dispose de toutes ces données ? Quelles anciennes tâches dois-je abandonner ? Quelles tâches dois-je effectuer différemment ? Pratiquement personne ne demande : quelles informations dois-je ? À qui ? Quand ? Sous quelle forme ?[23] ».

Drucker précisait ensuite son propos en insistant sur le fait qu'une base de données n'était pas une information, qu'un matériau brut (les données) n'était pas de l'information et qu'il était nécessaire

entrance to the "symbolic space" of literacy (Brockmeier 2000), a space of writing that embraces the accumulated cultural resources not only of the history of writing, but also of the writing of history: that is, of storing and transmitting knowledge, of constructing and interpreting archival traditions. It would be naïve to assume that an individual reader, just by learning to write and read, will immediately have access to all resources developed in a literate tradition ».

[21] Voir à ce sujet Drucker (2005).

[22] Peter Ferdinand Drucker (1909-2005) était un économiste et théoricien du management en entreprise, auteur de plus de 39 ouvrages dont le plus connu est *The Practice of Management* publié en 1954. Il a conceptualisé de nombreux aspects du management moderne et intervenait très régulièrement dans les colonnes du *Wall Street Journal* ainsi que dans les revues *Harvard Business Review*, *The Atlantic Monthly* et *The Economist*. L'approche développée ici est donc complètement intégrée au monde de l'entreprise où les données et les informations sont des matières premières avant toute chose.

[23] « *But not many executives are information-literate. They know how to get data. But most still have to learn how to use data. Few executives yet know how to ask : What information do I need to do my job? When do I need it? In what form? And from whom should I be getting it? Fewer still ask : What new tasks can I tackle now that I get all these data? Which old tasks shoud I abandon? Which tasks should I do differently? Practically no one asks: What information do I owe? To whom? When? In what form?* ».

d'apporter de nouvelles directives pour les obtenir. Car si les données commerciales étaient aisément acquises par les entreprises, ce n'était pas le cas pour les autres types de données, limitant considérablement les marges de manœuvre de celles-ci.

> « Les informaticiens sont toujours préoccupés par de plus grandes vitesses et des mémoires plus volumineuses. Mais les défis à venir ne seront pas de plus en plus techniques, ils concerneront plutôt la conversion des données en informations exploitables et véritablement utilisées [24] ».

Si une définition complète d'un lettré des données n'existait pas encore, nous en percevons néanmoins ici des prémisses, liées à de nouveaux besoins qui dépassent les compétences purement techniques. Cette première définition, on la trouve chez Earl & Katz (2002) qui ont caractérisé un « lettré des données » comme celui qui a développé un état d'esprit tourné vers l'enquête (*develop an inquiry habit of mind*), ce qui impliquait :

1. Une vision focalisée sur un ensemble possible de résultats combinée à une recherche constante de compréhension et de clarté ;
2. Une tolérance élevée à l'incertitude des résultats et une volonté de supporter cette dissonance le temps que la recherche leur apporte une clarté ;
3. D'avoir conscience que les données ne donnent quasiment jamais les réponses, qu'elles permettent au contraire de poser de meilleures questions et d'affiner les enquêtes.

Ainsi, tout(e) lettré(e) des données devait être capable :

1. de définir les buts pour chaque usage des données ;
2. de reconnaître les bonnes et les mauvaises données ;
3. d'avoir des connaissances en statistiques et en mesure ;
4. de se centrer sur l'interprétation des données ;
5. et d'être attentif à la manière de les transmettre.

Cette vision offre au lettré des compétences qui dépassent le cadre technique pour incorporer des postures réflexives qui sont le fondement du journalisme des données [25].

D'autres sources contribueront à cette réflexion, en particulier le rapport de l'*Oceans of Data Institute* dans lequel un lettré des données était vu comme une personne comprenant, expliquant et documentant l'utilité et les limites des données en devenant un consommateur critique de données,

[24] « *Computer people still are concerned with greater speed and bigger memories. But the challenges increasingly will be not technical, but to convert data into usable information that is actually being used* ».

[25] https://www.thebureauinvestigates.com/explainers/what-is-data-journalism

capable de gérer ses traces numériques, de trouver du sens dans les données et de prendre des décisions basées sur les données (Oceans of data, 2016). Cette nouvelle vision incorporait les avancées et les nouveaux usages des technologies numériques pour lesquels un recul critique est exigé.

Des visions spécifiques à des disciplines ou des contextes professionnels vont également émerger, proposant de définir un lettré des données dans un contexte pédagogique [26], entrepreneurial [27] ou encore citoyen [28]. Par ailleurs, Laprise (2006) soulignait que le *data literate* n'avait pas besoin d'être un pratiquant (*practitioner*) : la compétence du *data literate* résidant avant tout dans le développement de sa réflexion.

> « Tout d'abord, le lettré des données doit comprendre les processus par lesquels les données sont obtenues et traitées, lui permettant de penser de manière critique à la véracité des informations et d'en évaluer l'applicabilité. Il est à noter que cela ne signifie nullement que les lettrés des données doivent être des praticiens ; simplement qu'ils comprennent les procédés impliqués. Puis, le lettré des données extrait les informations idoines alimentant la prise de décision [29] ».

1.2.1.3. Les typologies du *data literate*

Au-delà des définitions, des typologies de *data literate* ont été élaborées. Wolff, Gooch, et al. (2016) distinguent quatre profils de citoyens lettrés (*literate citizens*) capables de résoudre les problèmes du monde réel :

1. Les communicants (*communicators*) qui produisent des histoires à partir des données à l'attention d'autres personnes ;
2. Les lecteurs (*readers*) qui interprètent les données perçues comme indissociables de la vie quotidienne ;
3. Les créateurs (*makers*) qui (1) intègrent les données dans des stratégies globales ayant pour but d'identifier et de résoudre les problèmes du monde réel et (2) sont pleinement conscients de

[26] Voir à ce sujet (Agesilaou & Kyza, 2021; Carey, Grainger, & Christie, 2018; Fontichiaro, Oehrli, & Lennex, 2017; Mandinach & Gummer, 2016b).

[27] Voir à ce sujet (Brown, 2021; Miloradov, Rakic, & Marjanovic, 2022; Morrow, 2021; Pothier & Condon, 2020; Smolnikova, 2022).

[28] Voir à ce sujet (Twidale, Blake, & Gant, 2013; Wise, 2020).

[29] « *First, the data literate understand the processes of how data is acquired and processed, enabling the data literate to think critically about the veracity of the information and evaluate its applicability. Note that this dœs not mean that the data literate must be practitioners; only that they understand the processes involved. Second, the data literate draw appropriate information to apply to decision making* ».

leurs contributions en termes de données qui alimentent les applications des villes intelligentes (*smart city*) ;
4. Les scientifiques (*scientists*) qui combinent un niveau élevé de compétences techniques couplé à des capacités en communication et à une connaissance approfondie du domaine des données.

La dernière typologie renvoie à la science des données (*data science*) qui, si elle propose également des formations aux données, se démarque de la *data literacy* dans ses visions et ses objectifs. Nous y reviendrons dans le chapitre 3. La *data literacy* peut néanmoins être une voie d'accès vers ce domaine spécifique.

1.2.1.4. L'état d'esprit « donnée »

Il s'agit d'amener les individus à un stade optimal de compréhension des données, un « état d'esprit » nommé différemment selon les auteurs. Le premier terme est *data mindset*, souvent associé aux pratiques liées au *big data* (Charles & Gherman, 2013; Dijck, 2014; Miller, 2014). Pour Mayer-Schönberger & Cukier (2013), le *data mindset* recoupe trois visions :

1. La capacité d'analyser de grandes quantités de données centrées sur un sujet spécifique, plutôt que d'être contraint de se contenter d'ensembles plus petits de données ;
2. Accepter l'état désordonné des données du monde réel plutôt que de privilégier leur exactitude ;
3. Le respect croissant pour les corrélations plutôt que la quête continue d'une causalité insaisissable.

Une compétence de traitement des données donc, associée à une posture embrassant l'incertitude des données. Herzog évoquait pour sa part un « état d'esprit de la/des donnée(s) » (*data state of mind*) qu'il nommait également « antennes de données » (*data antennae*) :

> « Pour faire face à ce défi, nous devons développer un état d'esprit de la donnée qui nous ouvre les yeux sur la prolifération des données en ligne, dans les ordinateurs centraux et les serveurs des agences gouvernementales et dans les bibliothèques universitaires. Une fois qu'on aura développé un état d'esprit de la donnée - ou des antennes de données - il nous sera plus facile de percevoir ces possibilités. Les données sont partout. [30] » (Herzog, 2015, p. 15).

[30] « *To meet that challenge we need to develop a data state of mind, one that opens our eyes to the proliferation of data online, in government agency mainframes and servers, and in university libraries. After developing a data state of mind—or data antennae — it will be easier for us to see the possibilities. Data are everywhere.* »

Dans cette vision, c'est plus une posture basée sur l'ouverture qu'une compétence qui est mise en avant.

Lupton parlait quant à elle d'un « sens des données » (*data sense*) mettant en avant la manière dont les personnes interagissent avec les données. Ici, ce sont le corps et les sens qui sont mobilisés pour interagir avec la matérialité des données, incarnées dans les technologies numériques :

> « Il reconnaît l'utilisation des sens humains dans les réponses des individus aux données, y compris leurs données personnelles ainsi que les grands ensembles de données. Il reconnaît également la participation de capteurs technologiques. Le sens des données implique donc des enchevêtrements de sens humains et de capteurs numériques avec la création de sens [...] J'ai développé le concept de sens des données comme une façon de penser à travers les intersections et les enchevêtrements des corps humains, du code logiciel, des appareils numériques, de l'espace et du lieu et les données numériques générées à partir de ces enchevêtrements [31] » (Lupton, 2017, pp. 06-13).

L'état d'esprit porté sur les données regroupe donc des appellations diverses et embarque des visions toutes aussi variées. Il intègre des compétences, des attitudes et des postures aussi bien physiques qu'intellectuelles, ainsi que des visions spécifiques sur les buts et usages des données.

1.2.2. Enseigner la *data literacy*

1.2.2.1. Le rôle de l'enseignant formateur

La formation d'un lettré implique *de facto* la présence d'un formateur, souvent désigné sous le vocable d'*educator* dans la littérature, voire de *data literate educator* (Schildkamp & Kuiper, 2010 ; Mandinach & Gummer, 2016a; Sampson et al., 2022; Verbiest et al., 2014). Il est à noter que l'usage du terme *educator* est lié à une vision pédagogique, plutôt intégrée dans un contexte scolaire et/ou universitaire, et donc éloigné des entreprises qui utilisent plutôt le vocable de *manager* en lien avec la fonction d'encadrement détenue par les cadres.

Pour Earl & Katz (2002), l'enseignant formateur est pro-actif : c'est un *leader* qui dirige et orchestre l'apprentissage dans la nouvelle « ère informationnelle » de ce siècle (*information age*). Il est lui-

[31] « *It acknowledges the use of the human senses in people's responses to data, including their personal data as well as big datasets. It also recognises the participation of technological sensors. Data sense, therefore, involves entanglements of human senses and digital sensors with sense- making [...] I have developed the concept of data sense as a way of thinking through the intersections and entanglements of human bodies, software code, digital devices, space and place and the digital data that are generated from these entanglements.* ».

même un lettré, ayant été formé idéalement assez tôt pour être pertinent dans ses approches (Mandinach & Gummer, 2016b). Il a pleinement développé des compétences en *data literacy*, mais également en *information literacy* : le *North Carolina Department of Public Instruction*, cité par Papamitsiou et al. (2021), décrivait le *data literate educator* comme une personne détenant la connaissance suffisante pour recueillir, analyser et transmettre graphiquement les informations et les données pour appuyer la prise de décisions à différents niveaux du processus pédagogique.

Athanases et al. (2013) lui attribuent par ailleurs trois rôles : (1) mettre en place l'enquête collaborative (*collaborative inquiry*) pour favoriser des résultats équitables, (2) travailler avec les données disponibles qu'il sait collecter et (3) apprendre à comprendre et nommer de quelle manière le sens est déduit des données. Cette vision d'une formation en *data literacy* orientée sur l'enquête a débuté dès Earl & Katz (2002) et est un point constamment repris dans la littérature : *data literacy* et *data inquiry*, que nous traduisons par « enquête à partir des données », sont ainsi indissociables pour Bocala & Boudett (2015). Pour ces derniers, l'enquête à partir des données réalisée par les enseignants formateurs est le fait, pour ces derniers, d'analyser les progrès de leurs étudiants grâce aux données, de formuler des recommandations pour les curricula et de suivre les résultats des mesures prises. C'est un pré-requis indispensable :

> « Le premier réflexe mental que les éducateurs doivent développer lorsqu'ils apprennent à s'engager dans une enquête collaborative sur les données est une orientation vers une action ciblée et réfléchie [32] » (p. 9).

L'aspect collaboratif est central : il permet l'amélioration des pratiques et évite la dispersion des formateurs en les fédérant dans un cadre cohérent et pérenne. Un cadre où travailler les données doit devenir une « donnée » du quotidien :

> « Alors que le domaine est toujours en recherche d'une pédagogie distinctive qui pourrait aider à la socialisation des nouveaux enseignants au sein d'une profession collaborative et compétente en matière de données, il faut moins mettre l'accent sur l'enseignement de la *data literacy* en tant qu'ensemble de compétences et de pratiques isolées, et plutôt se concentrer sur l'enseignement de l'enquête à partir des données sous la forme d'un processus holistique fondé sur des réflexes mentaux nécessaires à une amélioration concertée. L'objectif n'est pas seulement de rendre les enseignants à l'aise avec les données, mais aussi de permettre à la profession d'évoluer à une échelle où la compréhension des données est pleinement intégrée au travail d'apprentissage et d'enseignement [33] ».

[32] « *The first of habit of mind that educators need to develop as they learn how to engage in collaborative data inquiry is an orientation toward purposeful, reflective action* ».

[33] « *As the field continues in search of a signature pedagogy that might socialize new teachers into a collaborative, data-literate profession, there must be less emphasis on teaching data literacy as a set of isolated skills and practices, and instead a focus*

Au-delà de ce processus d'enquête, nous avions souligné la présence de la notion d'éthique des données, que nous développerons dans le chapitre 3, intégrée depuis les années 2010 dans les définitions de la *data literacy*. Le *Data Quality Campaign*[34] intègre cette éthique dans les caractéristiques du *data literate educator* et en fait un axe central de son action :

> « Les formateurs formés aux données doivent de manière constante, efficace et éthique accéder, interpréter, agir et communiquer des formes multiples de données provenant de sources étatiques, locales, scolaires et autres pour améliorer les résultats des étudiants d'une manière appropriée à leurs rôles et responsabilités professionnelles [35] ».

Bien que disposant de compétences et de connaissances développées, l'enseignant formateur doit faire face à un ensemble de défis à surmonter pour réussir à former efficacement ses apprenants. En particulier la difficulté d'aboutir à une vulgarisation pertinente de la *data literacy*. Cette difficulté est principalement due à une mécompréhension globale de ce que recouvre ce concept :

> « Le manque de compréhension générale du concept et de ses applications potentielles a pour conséquence de limiter les apprenants. Leur connaissance ne suffit plus lorsqu'ils sont confrontés à des prises de décision en conditions réelles [36] » (Gebre, 2022).

Or, c'est de la connaissance de l'enseignant formateur que va découler une pédagogie adaptée :

> « Quand des formateurs étudient les données de leur propre classe, ils sont en mesure de connaître les types de renseignements qui peuvent être tirés de ces données. Ils ont un lien étroit avec les données et le contexte dans lequel elles ont été assemblées et, à partir de là, ils peuvent comprendre les types de questions auxquelles il pourrait être possible de répondre à partir des données [37] » (Wolff, Hudson, & Kortuem, 2016).

Plusieurs cadres de formation en *data literacy* pour les enseignants formateurs ont été proposés afin de réduire leurs difficultés. Le plus représentatif est celui de Raffaghelli (2019). Basé sur

---

on teaching data inquiry as a holistic process grounded in habits of mind necessary for collaborative improvement. The goal is not just getting teachers to be comfortable with data, but allowing the profession to evolve to a place where understanding of data is thoroughly integrated with the work of learning and teaching » (p. 17).

[34] https://dataqualitycampaign.org/resource/data-literacy-101/

[35] « *Data-literate educators continuously, effectively, and ethically access, interpret, act on, and communicate multiple types of data from state, local, classroom, and other sources to improve outcomes for students in a manner appropriate to educators' professional roles and responsibilities* ».

[36] « *The lack of a broader understanding of the concept and its possible applications has consequential limitations for learners. Their knowledge becomes insufficient to draw from when they encounter real life decision-making situations* ».

[37] « *When an educator analyses data from their own classroom, they are in a position to know the sorts of insights that may be gained from this data. They have a close connection to both the data, the context in which it was gathered and from this they can understand the sorts of questions it might be possible to answer from the data* ».

une sélection et une synthèse de 19 cadres de compétences issus de la littérature scientifique, il est divisé selon six axes : (1) l'engagement professionnel, (2) les données comme ressources pour l'apprentissage, (3) l'enseignement et l'apprentissage, (4) l'évaluation, (5) l'encapacitation des apprenants, (6) faciliter l'apprentissage des apprenants. Chaque axe intègre plusieurs sous-thématiques accompagnées d'exemples et est divisé en six niveaux de compétences. Il s'agit pour l'enseignant formateur de s'emparer des données mais également d'un ensemble de technologies et de méthodes numériques pour les intégrer dans sa pratique. Une pratique centrée sur les données (*data driven*) à partir de laquelle doit émerger une posture critique permettant aux apprenants d'être pleinement encapacités.

1.2.2.2. Les niveaux et modes de formation

Pour ce faire, les formations en *data literacy* se déploient selon un continuum où trois grands niveaux peuvent être distingués selon Mandinach & Gummer (2012) et Gibson & Mourad (2018) :

1. Un niveau basique qui intègre une application et une description sommaire des données ;
2. Un niveau intermédiaire qui incorpore les capacités précédentes auxquelles s'ajoute une capacité de réflexion poussée permettant de dépasser des usages basiques des données ;
3. Un niveau avancé qui démontre une maîtrise totale des usages des données couplée à une réflexion critique.

Une formation en *data literacy* n'exige pas d'atteindre le niveau le plus avancé : tout dépendra des nécessités et du niveau de connaissance requis.

Pour Gebre (2022), ce continuum peut être appréhendé selon deux approches, basées sur les visions « autonome » et « idéologique » de la littératie sur lesquelles nous reviendrons au cours du chapitre 2.

La première approche, nommée « traditionnelle » prend appui sur les visions déployées dans la littératie « autonome » et tend à proposer une solution uniforme pour traiter les problèmes des données. Détenir un capital littéracique est ici déterminant pour la réussite de l'individu :

> « Il est supposé que ceux qui ont accès au texte écrit et sont capables de lire et d'écrire ont une vie meilleure. Un tel cadre définit la littératie comme une compétence cognitive qui s'acquiert et se développe indépendamment des différences contextuelles. Le problème de cette approche est que l'élaboration des programmes et les initiatives pédagogiques

> associées sont souvent uniformes et peuvent être appliquées dans diverses situations indépendamment des facteurs relatifs aux contextes [38] » (Gebre, 2022).

La seconde approche, intitulée « contextuelle », s'appuie sur la vision « idéologique » de la littératie présentée comme une pratique sociale dépendante de ses contextes de déploiement.

Partant de ce constat, Gebre liste quatre orientations guidant les formations : (1) le développement de compétences, (2) les processus d'enquêtes axées sur les données (*data-driven inquiry process*), (3) la connaissance de ses données personnelles et (4) l'engagement citoyen.

1.2.2.3. Les cadres de compétences

Le déploiement de compétences est la vision la plus courante et se traduit sous la forme de référentiels et de cadres de compétences qui couvrent le cycle de vie des données. Ils intègrent des compétences techniques, des connaissances culturelles (ex : *data culture*) et des postures intellectuelles (ex : « esprit critique »), en lien direct avec les définitions précédemment listées. L'un des plus représentatifs, au sens d'une intégration d'un grand nombre de compétences diverses intégrées au sein de catégories englobantes représentatives, est celui du *Conceptual framework* de Calzada-Prado et Marzal présenté dans la figure 1.5.

Comme précisé précédemment, nous retrouvons à la fois des compétences, des connaissances et des postures intellectuelles. L'ensemble de ces compétences peuvent être caractérisées selon trois grands types pour Ridsdale et al. (2015) : les compétences centrales (*core competencies*), les compétences conceptuelles (*conceptual competencies*) et les compétences avancées (*advanced competencies*).

[38] « *Those who have access to written text and are able to read and write were presumed to have a better life. Such framing situates literacy as a cognitive skill that is acquired and developed independent of contextual differences. The problem with this approach is that curriculum development and related instructional initiatives are often uniform that can be applied in various situations irrespective of contextual factors* ».

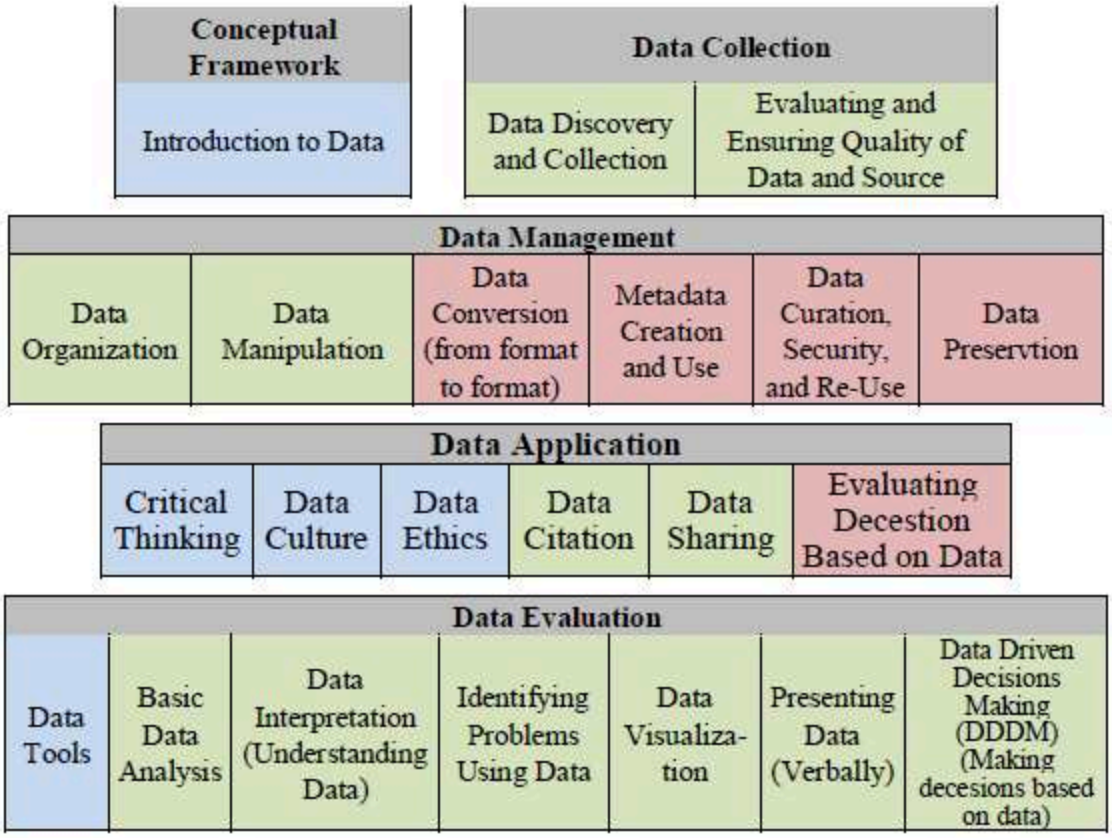

Figure 1.5.: Cadre de compétences en *data literacy* - extrait de (Calzada-Prado & Marzal, 2013)

Tableau 1.3.: Les typologies de compétences « donnée » - extrait de Ridsdale et al. (2015)

Type de compétences	**Compétences citées**
Compétences centrales	Découverte et collecte des données, évaluation et qualité assurée des données et des sources, organisation des données, manipulation des données, analyse basique des données, interprétation et compréhension des données, identification des problèmes nécessitant des données, les datavisualisations, la présentation orale des données, les décisions basées sur les données, la citation des données, le partage des données.
Compétences conceptuelles	Introduction aux données, outils et techniques des données, l'esprit critique, la *data culture*, l'éthique des données.
Compétences avancées	Conversion des données (d'un format à l'autre), création et utilisation des métadonnées, curation, sécurité et réutilisation des données, préservation des données, l'évaluation de décisions basées sur les données.

Selon les auteurs consultés comme Maybee & Zilinski (2015), Papamitsiou et al. (2021) ou plus récemment Ghodoosi, West, Li, Torrisi-Steele, & Dey (2023), ces compétences peuvent être (re)combinées, (re)subsumées ou (re)divisées. A titre d'exemple, en reprenant les travaux de Ridsdale et al. (2015), nous constatons que ce dernier regroupe l'organisation des données, la manipulation des données, la conversion des données, l'usage des métadonnées, la curation, la sécurité et la réutilisation des données ainsi que la préservation des données sous l'étiquette de la gestion des données (*data management*).

Aucun consensus n'est véritablement trouvé sur un choix définitif de dénominations ou sur une exhaustivité des compétences devant être nécessairement enseignées. Néanmoins, l'intégration de compétences déployées sur le cycle de vie des données, le développement d'une éthique et d'une culture (critique) des données restent un socle commun présent dans la littérature.

Les cadres de compétences eux-mêmes peuvent être catégorisés selon une orientation spécifique. En réalisant une synthèse de 19 cadres de compétences en données, Raffaghelli (2019) a listé quatre orientations :

- Les données pour l'éducation et l'apprentissage (*Data in Teaching and Learning*, 7 études) : basée sur l'atteinte de niveaux de compétences en *data science* allant du plus basique au plus sophistiqué ;

- La gestion des données de la recherche (*Research data management*, 6 études) : définit la *data literacy* comme un processus intégrant la curation, le traitement, la présentation et le partage des données au sein des projets universitaires ;

- L'encapacitation des apprenants (*Empower Learners*, 4 études) : précise l'emploi des données pour comprendre les processus dans lesquels les apprenants sont engagés, des activités socio-culturelle et politiques aux processus pédagogiques ;

- La gestion des données pédagogiques (*Educational data management*, 2 études) : désigne les pratiques *data-driven* des professeurs pour améliorer la qualité et l'efficacité des formations pédagogiques.

Bien que située dans un contexte pédagogique, cette analyse permet de souligner la multiplicité des approches dans la littérature : former aux données n'est pas un processus linéaire ou monolithique, une solution unique applicable pour tous types de publics, d'organisations et de situations.

1.2.2.4. Les référentiels de compétences

Au-delà des cadres de compétences se sont développés des référentiels en compétences de données. Plus précisément, des référentiels sur des compétences numériques incluant des compétences en usages de données. En effet, les usages des données s'intègrent dans une vision numérique, dématérialisée, de ces dernières et ce depuis plusieurs décennies.

La compétence numérique

Au cours des années 1990 et début 2000, « la » compétence numérique est au cœur d'enjeux liés aux notions controversées de « société de l'information » [39] et de « fracture numérique ». Le discours d'Al Gore du 10 octobre 1996, alors Vice-Président des États-Unis sous la mandature de Bill Clinton, à Knoxville est en ce sens représentatif de cette période :

> « Je rêvais d'une époque où un médecin aurait un accès instantané au dossier médical d'un patient en cas de blessure et où le médecin de ce patient aurait besoin d'en connaître le meilleur traitement. Eh bien, aujourd'hui, ce rêve est en passe de devenir

[39]Voir à ce sujet Mattelart (2018) et la thèse de Labelle (2007) par exemple.

> réalité. Il y a deux ans, le président Clinton et moi-même avons mis l'Amérique au défi de connecter toutes les salles de classe - des centres-villes, des zones rurales et des banlieues - à l'Autoroute de l'Information d'ici l'an 2000. Nous avons mis la nation au défi de garantir à tous nos enseignants et nos étudiants un accès à des ordinateurs modernes et à des logiciels éducatifs engageants. Nous avons mis la nation au défi de fournir à tous les enseignants la formation et le soutien dont ils ont besoin pour aider les élèves à tirer le meilleur parti de ces merveilleuses nouvelles technologies. Nous avons mis la nation au défi de veiller à ce que nos enfants ne soient jamais séparés par une fracture numérique [40] [41] ».

L'accès et la maîtrise des technologies de l'information et de la communication (*TIC*) étant synonymes de prospérité (économique, culturelle, démocratique, etc.), ces dernières ont été recommandées et placées au cœur de politiques publiques, comme l'illustrent La « Déclaration du Millénaire » de 2000 [42] suivie par la « Déclaration de Principe » [43] et le « Plan d'Action » de Genève [44] décrétés lors du Sommet mondial sur la société de l'information (SMSI) :

> « L'utilisation et le déploiement des TIC devraient contribuer à faciliter notre vie quotidienne dans tous les domaines. Leurs applications peuvent se révéler très utiles dans de nombreux domaines: administration et services publics, santé et information sanitaire, enseignement et formation, emploi et création d'emplois, affaires, agriculture, transports, protection de l'environnement et gestion des ressources naturelles, prévention des catastrophes naturelles, culture, et favoriser l'élimination de la pauvreté et atteindre d'autres objectifs de développement reconnus. Les TIC devraient également contribuer à établir des structures durables de production et de consommation et à atténuer les obstacles traditionnels, donnant ainsi à tous la possibilité d'accéder aux marchés locaux et aux marchés mondiaux de façon plus équitable. Les applications devraient être conviviales, accessibles à tous, abordables, adaptées aux besoins locaux en termes de cultures et de langues, et faciliter le développement durable. A cet égard, il conviendrait que les collectivités locales assument un rôle majeur dans la prestation de services TIC, pour le bien des populations concernées. » (Déclaration de Principes, 2003).

---

[40] « *I dreamed of a time when a doctor would have instant access to a patient's medical records instantly if an injury occurred and that patient's doctor needed to know the best course of treatment. Well, today that dream is fast becoming a reality. Two years ago, President Clinton and I challenged America to connect every classroom – inner-city, rural, suburban – to the Information Superhighway by the year 2000. We challenged the nation to ensure that all of our teachers and students have access to modern computers and engaging educational software. We challenged the nation to provide all teachers with the training and support they need in order to help students make the most of these wonderful new technologies. We challenged the nation to make sure that our children will never be separated by a digital divide* ».

[41] Discours du 10 octobre 1996 d'Al Gore à Knoxville. Consulté sur : https://govinfo.library.unt.edu/npr/library/speeches/101096.html

[42] https://www.un.org/french/millenaire/ares552f.htm

[43] https://www.itu.int/net/wsis/docs/geneva/official/dop-fr.html

[44] https://www.itu.int/net/wsis/docs/geneva/official/poa-fr.html

Permettre l'accès aux TIC supposait la résolution de la fracture numérique du fait que cet accès garantissait automatiquement leur bon usage (Brotcorne & Valenduc, 2009). Or, cette vision s'est vite heurtée à de nombreuses entraves : il convenait en ce sens de parler de fractures, voire de fracture numérique de « second degré » (Hargittai, 2002). Des fractures d'accès donc mais également de contenus, d'usages et de décisions (Kiyindou, 2007); Valenduc & Vendramin (2006) ajoutaient des critères sociologiques comme l'âge, le genre, les revenus, le niveau de formation, les professions, la composition familiale, les caractéristiques régionales et « Nord Sud ».

Ainsi, si le simple accès ne garantit pas la maîtrise, le développement de compétences s'avère néanmoins nécessaire. C'est ainsi qu'en 2006 la compétence numérique est reconnue comme l'une des huit compétences de la formation tout au long de la vie, définie par la Commission Européenne [45] comme suit :

> « La compétence numérique implique l'usage sûr et critique des technologies de la société de l'information (TSI) au travail, dans les loisirs et dans la communication. La condition préalable est la maîtrise des TIC: l'utilisation de l'ordinateur pour obtenir, évaluer, stocker, produire, présenter et échanger des informations, et pour communiquer et participer via l'internet à des réseaux de collaboration.[...] La compétence numérique exige une bonne compréhension et connaissance de la nature, du rôle et des possibilités des TSI dans la vie de tous les jours, dans la vie privée, en société et au travail. »

Parmi les sous-axes liés à la maîtrise de cette compétence étaient citées les bases de données, seules portes d'entrée sur les usages des données.

L'ICDL

En recensant les premiers référentiels de compétences numériques, nous pouvons citer pour le niveau européen le passeport international de compétences informatiques (*PCIE*, *European Computer Driving Licence* en anglais) proposé en 1997 par le *Council of European Professional Informatics Societies* (*CEPIS*) dont le module 5 *databases* comprenait des tests sur les bases de données (Leahy & Dolan, 2010). Devenu l'*International Computer Driving Licence* (*ICDL*) en 2019, il contient désormais 15 modules [46] « couvrant la majorité des outils bureautiques utilisés en entreprise » [47]. Celui traitant des bases de données est toujours présent. D'autres modules comme celui sur la « cybersécurité » intègre d'« utiliser Internet en toute sécurité et manipuler les données et les informations de manière adaptée » [48].

[45] Voir à ce sujet Parlement Européen (2006).
[46] https://www.icdlfrance.org/quest-ce-que-la-certification-icdl/
[47] https://aide.scorf.fr/icdl-pr%C3%A9sentation
[48] https://www.icdlfrance.org/nos-modules/

Le DigComp

Le DigComp, pour *Digital Competence Framework for Citizen*, est issu du travail du *Joint Research Centre* soutenu par la Commission Européenne dès 2010. Sa première version a été publiée en 2013 et a connu plusieurs évolutions [49] :

- première version 1.0 publiée en 2013 ;
- deuxième version 2.0 publiée en 2016 ;
- troisième version 2.1 publiée en 2017 ;
- quatrième version 2.2 publiée en 2022.

Dans la première version, les données ne sont citées que dans la partie 4.2 du domaine 4 *safety*. Les compétences qui lui sont affiliées concernent les usages sécurisés des données personnelles :

> « ...comprendre les conditions d'utilisation courantes, la protection active des données personnelles, comprendre la vie privée des autres, se protéger contre la fraude et les menaces en ligne et contre la cyberintimidation [50] [51] ».

La *data literacy* ne sera intégrée qu'à partir de la deuxième version et les compétences visées seront identiques dans les versions suivantes (voir la figure 1.6).

Competence areas Dimension 1	Competences Dimension 2
1. Information and data literacy	**1.1 Browsing, searching and filtering data, information and digital content** To articulate information needs, to search for data, information and content in digital environments, to access them and to navigate between them. To create and update personal search strategies. **1.2 Evaluating data, information and digital content** To analyse, compare and critically evaluate the credibility and reliability of sources of data, information and digital content. To analyse, interpret and critically evaluate the data, information and digital content. **1.3 Managing data, information and digital content** To organise, store and retrieve data, information and content in digital environments. To organise and process them in a structured environment.

Figure 1.6.: Compétences du domaine 1 - extrait de la version 2.0 du DigComp

[49] https://joint-research-centre.ec.europa.eu/digcomp/digcomp-governance_en

[50] « *...to understand common terms of service, active protection of personal data, understanding other people privacy, to protect self from online fraud and threats and cyberbullying* ».

[51] https://publications.jrc.ec.europa.eu/repository/handle/JRC83167

Dans sa dernière version (2.2) [52], le DigComp comprend actuellement cinq domaines divisés en vingt-et-une compétences et quatre niveaux de maîtrise (basique, intermédiaire, avancée, hautement spécialisée). Son premier domaine *Information and data literacy* englobe les trois sous-domaines présentés précédemment.

Données, informations et contenus sont ici intégrés dans un même domaine numérique. Malgré leurs différences, leur traitement est réalisé selon une même logique, effritant ici leurs frontières conceptuelles dans l'esprit des rédacteurs.

Le Cadre de référence de la compétence numérique

En 2019, le Québec, par l'intermédiaire du ministre de l'Éducation et de l'Enseignement supérieur Jean-François Roberge, a produit le *Cadre de référence de la compétence numérique* (Roberge, 2019). Celui-ci découle du plan d'action numérique en éducation et en enseignement supérieur visant « la vision d'une intégration efficace et d'une exploitation optimale du numérique au service de la réussite de toutes les personnes, qui leur permettent de développer et de maintenir leurs compétences tout au long de leur vie » (p. 7).

Il s'accompagne en outre de deux autres documents d'explicitation : un guide pédagogique et le continuum de développement. Il contient 12 dimensions intégrant 53 compétences selon 3 niveaux de compétences (débutant, intermédiaire et avancé). Les données sont rapidement citées à plusieurs reprises, majoritairement dans le cadre de la protection des données personnelles :

- domaine 1 « Agir en citoyen éthique à l'ère du numérique » : dans un exemple d'application, il est précisé que « l'apprenante ou l'apprenant prend soin de vérifier la source et le contenu d'un courriel avant d'ouvrir une pièce jointe ou de cliquer sur un lien afin d'éviter l'hameçonnage et de préserver ses données confidentielles » ;

- domaine 2 « Développer et mobiliser ses habiletés technologiques » : sécuriser ses données personnelles à l'aide de ressources appropriées, notamment en considérant les risques liés à l'utilisation du numérique ;

- domaine 7 « Produire du contenu avec le numérique » : utiliser différents supports médiatiques tels que du texte, du son ou des images pour manipuler des données numériques ;

Federal Data Strategy Principles

[52] https://publications.jrc.ec.europa.eu/repository/handle/JRC128415

Le *Federal Data Strategy Principles* proposait en 2020 un plan d'action se posant en soutien des institutions gouvernementales (Federal Data Strategy, 2020). Il se composait de dix principes pérennes divisés en trois catégories : une gouvernance éthique (*ethical governance*), une conception écologique (*conscious design*) et une culture évolutive (*learning culture*). Ces catégories encadrent une quarantaine de pratiques servant elles-mêmes trois grands objectifs :

- Bâtir une culture qui valorise les données et encourage leur utilisation auprès du public (*Building a Culture that Values Data and Promotes Public Use*) ;
- La gouvernance, la gestion et la protection des données (*Governing, Managing, and Protecting Data*) ;
- Promouvoir une utilisation efficace et appropriée des données (*Promoting Efficient and Appropriate Data Use*).

L'ensemble de ces recommandations ont été au cœur d'une stratégie d'implémentation en vingt étapes d'une durée de douze mois, elle-même incluse dans une vision s'étalant sur une dizaine d'années. Bien que n'étant pas un référentiel à proprement parler, il en possède les caractéristiques en dictant des lignes directrices et des compétences à développer.

Mydataabilies

Au-delà des compétences numériques, des référentiels centrés spécifiquement sur les données ont vu le jour, principalement sous l'égide d'entreprises. Le plus connu est *mydataabilities* de l'entreprise australienne *Data to the People*[53] qui se compose de quinze compétences réparties sur six niveaux abordant la lecture, l'écriture et la compréhension des données.

La certification QLIK

L'entreprise *Qlik* a de son côté mis en place une certification en *data literacy*[54] :

> « Il mesure votre capacité à interpréter les exigences métier ; à comprendre et transformer les données ; à concevoir, créer et interpréter des visualisations ; à analyser des informations, prendre des mesures concrètes et partager les résultats. Cet examen est composé de 70 questions à choix multiples auxquelles vous devrez répondre en 2h30 ».

On y retrouve tout aussi bien des compétences techniques que d'attitudes à adopter vis-à-vis des données comme l'esprit critique. L'entreprise propose de plus des formations comme le *Qlik Continuous Classroom* qui introduit la *data literacy* et des principes annexes.

[53] https://www.mydatabilities.com/
[54] https://www.qlik.com/fr-fr/services/training/certifications-and-qualifications

Qu'il s'agisse de *Qlik* ou de *Mydataabilities*, nous avons ici affaire à une auto-évaluation en ligne à l'issue de laquelle une certification est remise.

EDF Data literacy

Effective Data Foundation met à disposition une certification en *data literacy* payante d'une durée de validité de deux ans. Il est décrit comme suit [55] :

> « Cela signifie la compréhension et la conscience globales d'un individu en *data literacy*, lui permettant de favoriser son adoption et son application pour une prise de décision éclairée [56] ».

Le test, d'une durée de 60 minutes, vise à confirmer les capacités des personnes. D'une part connaître et comprendre les données dans leur contexte : sont cités les différentes natures des données, leurs biais et les influences extérieures. D'autre part, disposer de compétences techniques permettant de comprendre et développer des analyses, d'assurer la qualité des données ou encore de savoir communiquer les données et donc de connaître le *storytelling*.

De nombreuses autres certifications existent mais elles sont moins reconnues et basées sur le même socle de compétences à atteindre. Ce que nous notons ici c'est une tendance globale d'intégration des données dans une vision numérique. Ce numérique renvoie en particulier aux traitements informationnels en ligne, la conception technique des données intervenant dans un second temps. Les données sont-elles perçues uniquement comme numériques, uniquement présentes au sein des ordinateurs et serveurs sous formes dématérialisées ? C'est un questionnement qui touche de près la *data literacy* et tout au long de cette réflexion nous nous demanderons s'il faut désormais parler de *data literacy* ou de *digital data literacy*.

1.3. L'écosystème des acteurs

Tout un écosystème d'acteurs s'est emparé des enjeux de la *data literacy*. Nous ne citerons que les plus représentatifs pour deux raisons : la première pour présenter des initiatives ayant acquises une certaine importance au plan international, la seconde du fait du nombre trop volumineux d'acteurs qui ne sauraient être récapitulés exhaustivement. Par ailleurs, nous excluons ici les acteurs français qui feront l'objet du chapitre 5.

[55] https://info.certn.global/certifications/edf-data-literacy/

[56] « *It signifies an individual's comprehensive understanding and awareness of data literacy, enabling them to foster its adoption and application for informed decision-making* ».

Si de nombreuses différences les caractérisent, un but commun les anime : enseigner la *data literacy* à un panel varié de publics selon une offre de formations (payante ou non), fédérer une communauté de pratiques et, plus globalement, promouvoir la *data literacy* à l'échelle politique pour engager des actions importantes et surtout pérennes.

Tableau 1.4.: Ensemble des initiatives sur la *data literacy*

Nom	Date de création	Origine
Qlik	1993	Suède
Data Kind	2011	États-Unis
Dataversity	2012	États-Unis
School of data	2012	International
Data Carpentry / The Carpentries	2013	États-Unis
DataCamp	2013	États-Unis
Data Pop Alliance	2013	Internationale
Data & Society	2013	États-Unis
Data to the People	2017	Australie
The Data Literacy Project	2018	Internationale
MyData	2018	Internationale
Data literacy foundation	2018	États-Unis
Data Education in Schools	2018	Royaume-Uni
Dataliteracy.de	2019	Allemagne
KidsinData	2020	Royaume-Uni
The Algorithm and Data Literacy Project	2020	Internationale
Data Literacy Academy	2022	Royaume-Uni

1.3.1. Les initiatives entrepreneuriales

Nous avons précédemment cité *Qlik* et *Data to the People*. Ces entreprises, associées à d'autres partenaires, collaborent ensemble au sein de *The Data literacy project*[57]. Ce projet se présente comme une communauté mondiale ayant pour objectif de partager les pratiques en *data literacy* en développant des programmes d'entraînement, en mettant à disposition une bibliothèque de ressources fonctionnelles et en encourageant l'intégration de la *data literacy* dans les curricula universitaires :

> « Nous avons réuni ce groupe de dirigeants passionnés pour susciter des discussions, créer de meilleures ressources et favoriser une meilleure appréciation des compétences

[57]https://thedataliteracyproject.org/

> en matière de données qui sont essentielles à la révolution des données qui se produit tout autour de nous [58] [59] ».

The Data Literacy Project propose notamment de déterminer notre *data persona*, autrement dit notre profil d'utilisateur de données, à travers le test *how data literate are you ?*. Une dizaine de questions sont proposées et mêlent tout aussi bien des connaissances conceptuelles que pratiques.

Dataliteracy.de [60] se présente de son côté comme une initiative issue de la collaboration de *data scientists*, de parties prenantes d'entreprises (*business stakeholders*) et d'experts en dataviz, ayant pour objectif d'améliorer la *data literacy* des personnes travaillant sur les données. Il met à disposition des services et des formations liés à l'écosystème des concepts des données (*data literacy*, *data strategy*, *data management*, etc.) ainsi qu'un test, conçu par l'organisme *Coverit*, pour déterminer son niveau en *data literacy*.

Au-delà des initiatives, les entrepreneurs eux-mêmes servent de référence au sein des discours, développant et incarnant leurs réflexions au sein de publications diverses. Parmi les plus récurrents, nous pouvons citer Jordan Morrow qui a produit plusieurs ouvrages et participé à de nombreux évènements (voir notamment le TED Talk de 2019 sur l'importance des *data literates*[61]). Kirk D Borne [62], un *data scientist* promouvant l'ensemble des pratiques liées aux données (la *data science* mais également la *data literacy*). Ou encore Eugene Dubossarsky ayant fondé l'initiative de formations *Data Literacy for Everyone* et ayant employé le vocable de révolution pour signifier ce nouveau rapport aux données (*The Data Literacy Revolution - Dr. Eugene Dubossarsky*, 2019).

L'ensemble des acteurs entrepreneuriaux spécialisés dans les approches en *data literacy* semblent peu nombreux en comparaison de ceux plutôt orienté en *data science*, *analytics* et *IA* : pour éclaircir ce point, nous avons réalisé une enquête quantitative sur Twitter/X, présente en dernière partie de ce chapitre.

1.3.2. Les associations et coopératives

Au-delà des entreprises, des associations se sont constituées. La *data literacy* n'est pas nécessairement le seul axe de leurs travaux : elle est souvent combinée avec la *data science*, l'IA, englobée dans la théma-

[58] « *We have brought together this group of passionate leaders to ignite discussion, create better resources and drive greater appreciation for data skills that are critical for the data revolution occurring around us* ».
[59] https://www.cim.co.uk/newsroom/data-literacy-project-cim-joins-global-skills-initiative/
[60] https://www.dataliteracy.de/
[61] Voir à ce sujet (*Jordan Morrow: Why everyone should be data literate | Jordan Morrow | TEDxBoise | TED Talk*, 2019).
[62] http://www.kirkborne.net/

tique du numérique. Dans cette optique, nous pouvons citer l'association *Data & Society*[63] qui base son idéologie sur un usage responsable et humain des technologies, amélioré grâce à des recherches de terrain. Elle prend pour objet d'étude les effets de l'automatisation, de l'IA et des données sur la vie sociale des individus, englobant un panel varié de domaines (la santé, les droits humains, le respect de la vie privée, l'éthique des données, etc.). La *data literacy* est citée comme l'une de ses préoccupations et intègre un certain nombre de ressources[64].

Pour sa part, l'association internationale *Data Kind*[65] fondée en 2011 souhaite « maîtriser le pouvoir de la *data science* et de l'IA au service de l'humanité ». Elle est constituée d'un ensemble d'organisations et de *data scientists* et possède des succursales dans plusieurs pays. Elle a produit le *DataKind Playbook*[66] qui liste les « bonnes pratiques » à respecter lors de la mise en place d'un projet sur les données.

Au même titre, la *Data Pop Alliance*[67] a été créé en 2013 par la rencontre de l'*Harvard Humanitarian initiative*, le *MIT Connection Science* et l'*Overseas Development Institute* (*ODI*). Elle intègre des chercheurs, des patriciens ainsi que des activistes et a pour objectif de « changer le monde avec les données » selon trois axes : diagnostiquer les réalités locales et les problèmes humains grâce aux données et à l'IA ; mobiliser les capacités, les communautés et les idées en faveur de sociétés plus inclusives et plus acculturées aux données ; et, en fin de compte, de transformer les systèmes et les processus qui soutiennent les communautés et pays intégrés dans les logiques des données (*datafied*). Il a notamment produit plusieurs rapports annuels sur l'état mondial en termes de transformations impulsées par les données, l'IA et le numérique.

En 2018 nait *MyData*[68] qui se présente comme un réseau international orientant ses actions sur l'encapacitation des citoyens vis-à-vis de leurs données personnelles. Il publie de nombreux rapports en plusieurs langues et organise chaque année la conférence *MyData* réunissant à la fois le secteur public, les entreprises ainsi que des organisations issues de la société civile.

La *Data literacy foundation*[69] est quant à elle née en 2018 et est, au contraire des précédents exemples, entièrement dédiée à l'étude de la *data literacy*. Elle héberge le podcast *The Data Literates*[70] où interviennent des acteurs impulsant des initiatives en *data literacy* afin qu'ils partagent leur expérience.

[63] https://datasociety.net/about/
[64] https://datasociety.net/?s=data+literacy
[65] www.datakind.org/
[66] https://playbook.datakind.org/
[67] https://datapopalliance.org/about/vision-and-members/La
[68] https://mydata.org/about/
[69] https://dataliteracyfoundation.org/
[70] http://dataliterates.com/

Nous apercevons ici un effet de convergence entre les données et l'intelligence artificielle au sens où les deux concepts tendent à être rapprochés dans les pratiques professionnelles. Les pratiques des données semblent ici désilotées, intégrées dans une vision plus globale. Nous approfondirons cela lors de l'étude des enchevêtrements des concepts liés aux données dans le chapitre 3.

1.3.3. Les organismes d'enseignement

Au-delà de promouvoir la *data literacy* et d'en discuter l'épistémologie, des organismes de formations ont rapidement émergé. Leur nombre n'a cessé de croître au cours de la dernière décennie, déployant un éventail de formations adaptées à des types de publics toujours plus élargis.

Fondé en 2011, *Dataversity* [71] est un producteur de ressources pédagogiques axées sur l'usage et la gestion des données. Il adopte une posture singulière en s'adressant à un public de professionnels issus de l'informatique souhaitant s'engager pleinement dans une carrière *data*. Ses contenus prennent la forme de conférences, de webinaires, de podcasts ainsi que de rapports. Un centre d'apprentissage en ligne est également présent pour centraliser l'ensemble des formations.

Le réseau *School of data*[72] est fondé en 2012 par Ruffus Pollock dans l'objectif d'implanter des programmes de formation dans le plus grand nombre de pays. Actuellement, 34 pays sont membres de ce programme dont la France qui compte plusieurs campus sur son territoire. Il s'adresse aux étudiants et déploie pour ce faire des cursus universitaires.

Un an plus tard en 2013 émerge *Data Carpentry*[73], aujourd'hui un programme de *The Carpentries*, un organisme de développement de compétences informatiques permettant la gestion et l'analyse de données dans l'ensemble des domaines de la recherche. La mise en place d'ateliers pédagogiques (*workshops*) sur les données est au cœur de leur philosophie et intègre pour cela un ensemble d'enseignants et de formateurs dans une démarche collaborative.

Également créée en 2013, la plateforme en ligne *DataCamp*[74] s'est spécialisée dans la création et la mise à disposition de cours en ligne sur les pratiques de données. Bien que centré sur la *data science*, *DataCamp* propose de nombreux cours en *data literacy*, décomposés par domaine d'usage des données. Des certifications aux métiers de *data scientists*, *data analyst* et *data engineer* sont également présentes.

[71] https://www.dataversity.net/about-dataversity-net/
[72] https://schoolofdata.org/2022/02/08/10-years-of-school-of-data/
[73] https://datacarpentry.org/
[74] https://www.datacamp.com/

Nous pouvons également citer un exemple récent avec la *Data literacy Academy*[75]. Il s'agit d'un organisme d'enseignement de la *data literacy* à travers un ensemble de cours et de formations à destination, principalement, de professionnels d'entreprises. Une certification est remise à l'issue des formations.

Concernant le jeune public, deux projets ont récemment vu le jour. Apparaissant en 2018, *Data Education in Schools*[76] est issu du *Data Skills Programme* financé par les gouvernements d'Écosse et du Royaume-Uni et associant des acteurs issus notamment du monde de l'entreprise, des universités et des écoles. Cet organisme met à disposition des ressources éducatives et ludiques [77]. Le plus représentatif est le guide *Teach Data Literacy*[78] conçu à l'attention des professeurs des écoles. Il mêle bande-dessinée, conseils pédagogiques et ressources utiles à l'organisation de séances et d'ateliers en école primaire.

KidsinData[79] apparaît en 2020, conçu par Ross White et Kabir Rabst comme une plateforme collaborative proposant des *workshops* ayant pour objectif d'apprendre aux enfants des connaissances relatives aux données et à la *datavizualisation*. Il use de méthodes d'apprentissage variées, notamment la gamification pour rendre accessible ses contenus.

Enfin, *The Algorithm and Data Literacy Project*[80] fondé en 2020 par *Digital Moment*, *Digital2030*, l'*UNESCO* et la *Commission Canadienne pour l'UNESCO*, cette initiative a pour objectif de sensibiliser aux données, aux algorithmes mais aussi à l'intelligence artificielle et propose en ce sens des contenus vidéos, des tests de connaissance et de la documentation.

Elle vise également la conceptualisation d'une littératie des algorithmes et des données : la dernière information à l'écriture de cette thèse se trouve dans un tweet daté du 06 septembre 2023, publié sur le compte @DigitalMomentCA (le compte Twitter/X du *Digital Moment*) qui précise la tenue d'une séance de concertation (*sucessful breakout session*) sur la définition de cette nouvelle littératie (la définition elle-même n'est pas encore disponible).

Nous notons ici une passation qui s'opère entre « l'informatique » et la « data » dans le monde des entreprises où les compétences informatiques autrefois au cœur des discours font place à celles des données. Cela ne signifie en rien que l'informatique passe au second plan ou que les compétences liées à ce domaine sont devenues obsolètes, loin de là : il s'agit toutefois de souligner que si les données sont perçues comme numériques, elles ne sont toutefois pas sous contrôle (moral) du domaine

[75] https://dl-academy.com/
[76] https://dataschools.education/
[77] https://dataschools.education/data-education-resources/
[78] https://dataschools.education/resource/teach-data-literacy-a-guide-for-primary-teachers/
[79] https://www.kidsindata.com/
[80] https://algorithmliteracy.org/fr/

informatique. Une nouvelle catégorie, de nouveaux pôles « data » se créent hors des services informatiques et revendiquent une identité propre. Les formations en *data literacy* précédemment présentées soulignent pour certaines ce basculement en voulant ajouter un arc de compétences « data » aux employés précédemment *computer literate* pour les adapter à ce nouveau contexte. Il en est de même pour l'intelligence artificielle et les algorithmes également évoqués à plusieurs reprises. Le périmètre de leurs enchevêtrements sera questionné au cours du chapitre 3.

1.3.4. Enquête sur Twitter / X

Pour poursuivre notre recherche d'acteurs promouvant la *data literacy*, nous nous sommes intéressés aux discours tenus sur la plateforme *Twitter*, aujourd'hui *X*, et avions absorbé (par le biais de l'*API Twitter*) 44 544 tweets intégrant le terme *dataliteracy* sur une période s'étalant du 06 septembre 2021 au 05 septembre 2022.

Nous souhaitions à l'origine réaliser la même recherche dans un contexte francophone en prenant appui sur des termes français (ex : culture des données, littératie des données, etc.) mais les résultats ont été peu probants, ne témoignant pas d'un investissement de la plateforme par les acteurs travaillant sur cette thématique, ou tout du moins n'employaient-ils pas de traductions de la *data literacy* ou encore ne présentaient-ils pas d'autres potentielles généalogies.

Pour extraire le contenu des tweets nous avons utilisé le logiciel R et son outil *R Studio* et avons listé en annexe 2 l'ensemble des formules employées. Cela nous a permis de lister l'ensemble des tweets publiés par jour dans la figure 1.7. En nous basant sur ce graphique, nous constatons que le terme *data literacy* est employé selon une intensité variable.

1.3.4.1. Les catégories de tweets

Sur les 44 544 tweets collectés, nous comptions en grande majorité des retweets (29 057), puis des tweets originaux (9600), des citations (3181) et des réponses (2706). En appliquant ces catégories à la figure 1.8, nous constatons que les courbes se dissocient très nettement.

Plusieurs pics ont été observés et ont été référencés dans le tableau 1.5.

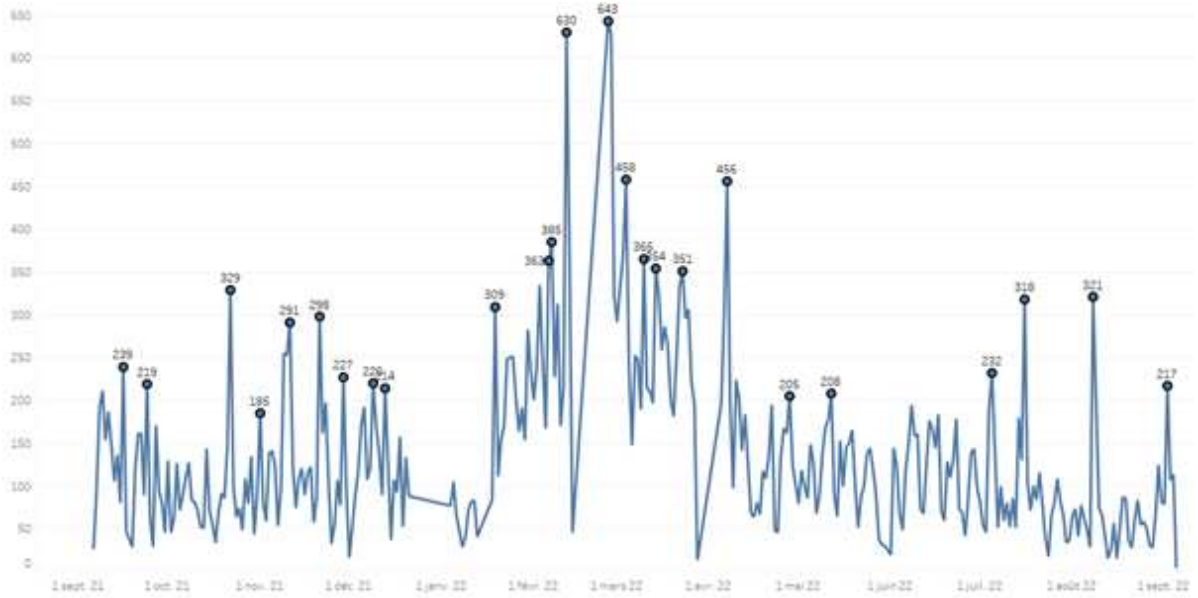

Figure 1.7.: Nombre de tweets publiés par jour (du 06 septembre 2021 au 05 septembre 2022)

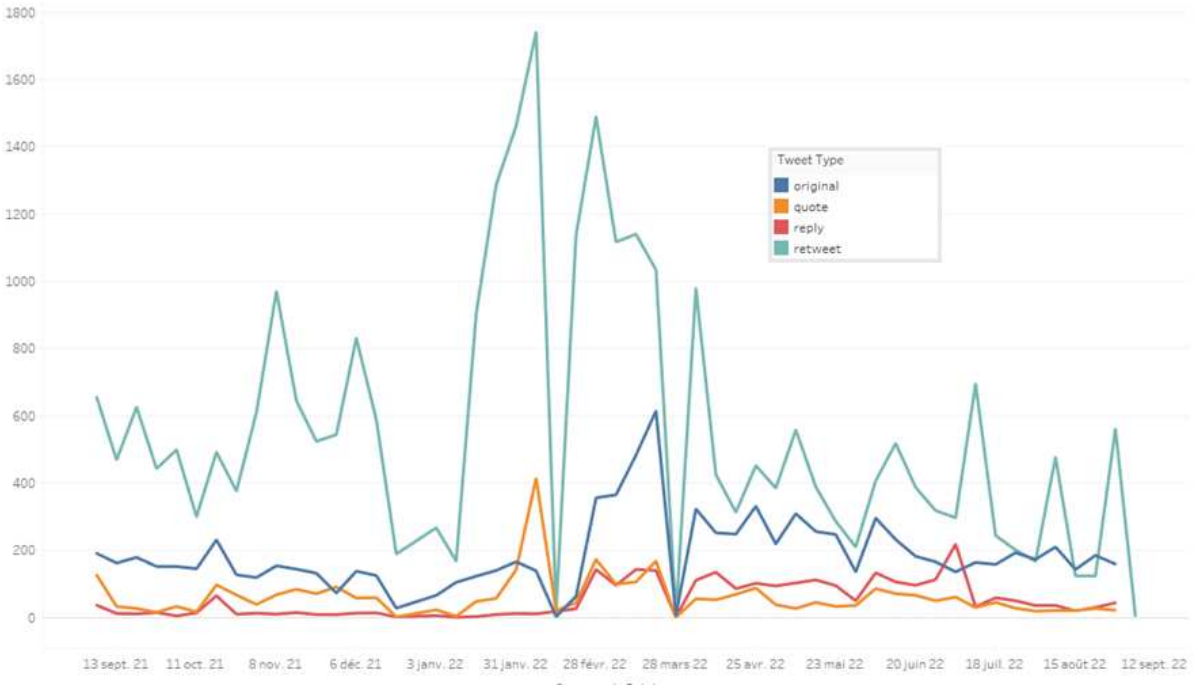

Figure 1.8.: Nombre de tweets par catégorie

Tableau 1.5.: Pics de tweets publiés

Types de tweets	**Pic observés**
Tweets originaux	18/10/21 : 231 tweets. 28/02/22 : 356 tweets. 21/03/22 : 614 tweets. 04/04/22 : 323 tweets. 25/04/22 : 331 tweets. 09/05/22 : 309 tweets. 06/06/22 : 296 tweets
Retweets	08/11/21 : 970 retweets. 06/12/21 : 831 retweets. 07/02/22 : 1741 retweets. 28/02/22 : 1489 retweets. 04/04/22 : 979 retweets. 09/05/22 : 558 retweets. 13/06/22 : 518 retweets. 11/07/22 : 695 retweets. 08/08/22 : 477 retweets. 29/08/22 : 591 retweets
Citations	07/02/22 : 414 citations. 28/02/22 : 173 citations. 21/03/22 : 168 citations
Réponses	18/10/21 : 65 réponses. 28/02/22 : 143 réponses. 14/03/22 : 143 réponses. 21/03/22 : 140 réponses. 04/04/22 : 111 réponses. 11/04/22 : 135 réponses. 16/05/22 : 112 réponses. 06/06/22 : 133 réponses. 04/07/22 : 218 réponses

Nous allons nous intéresser aux comptes pour déterminer l'origine de ces publications.

1.3.4.2. Les comptes

Il est important de souligner que l'ensemble des tweets sont le fait de 19 641 comptes uniques et de 108 bots.

Les bots

Ces derniers comptabilisent 2187 tweets dont 1935 retweets, 216 citations, 28 réponses et 8 tweets originaux. Leur recherche a été effectuée par le recensement du terme « bot » et par l'analyse du vocabulaire présent dans les descriptions.

Tableau 1.6.: Exemples de bots présents dans le corpus de tweets

Nom du compte	**Description du compte**
sms2sms	- + ? #kulturl?s???? | #zttlkstn | https://t.co/H78pZhRUua??????https://t.co/vfVFbNcBaM ex https://t.co/6jGSerwicK | #sozialarbeit | #agree2disgree #how2disagree (so?) ^bot
genericgranola	I'm a generic reference utility bot that aggregates trending and emerging AI/ML news via RSS monitoring & weighted user-interaction
uCloudify	Beep boop... I'm a bot! Mission: Recovery of small businesses from COVID-19.
nycdatabot	Explore #nyc with data. A bot by @fourtonfish. Source of data: https://t.co/GamHjrwNSQ #dataviz #opendata
PsyArXivBot	I am your bot for @PsyArXiv psychology preprints. I do not discriminate, I just post whenever there's a new one. Maintained by @libscie
Skillsnewsbot	My friendly bot and I bring you every day news regarding #jobskills #CareerSkills #skilltrends #lifelonglearning #careerdevelopment #jobmobility ... and more.

Comme cela est visible dans le tableau 1.6, Ils sont souvent assez éloignés de la thématique de la *data literacy*, agrégeant des contenus « data » plus généraux. Certains d'entre eux précisent leur origine en citant les comptes ayant concurru à leur création (ex : @libscie, @PsyArXiv).

Les comptes uniques

Sur les 19 641 comptes uniques, seulement 22 dépassent la publication de 100 tweets, et 467 d'entre eux n'ont publié qu'entre 10 et 100 tweets ; les 19 152 restants se trouvent sous la barre des 10 tweets. Les discours sur la *data literacy* sont donc le fait d'un nombre très restreint d'auteurs (voir la figure 1.9).

Sur les dix plus grands contributeurs proposant du contenu original, le compte KirDBorne arrive largement en tête (756 tweets originaux). Cet acteur va s'avérer central dans cette enquête, c'est pourquoi nous nous attardons quelques instants sur son profil (voir la figure 1.10) :

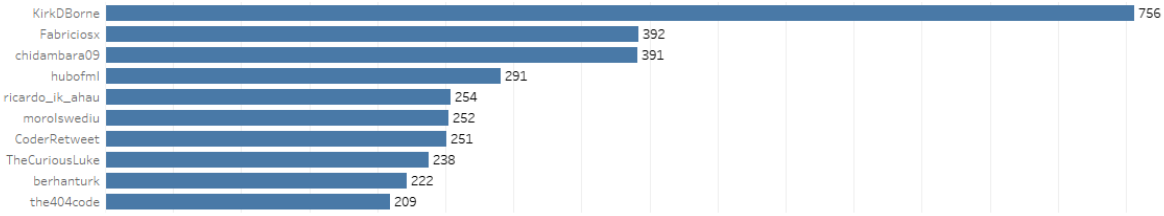

Figure 1.9.: Nombre de tweets originaux par comptes uniques



Figure 1.10.: Profil X de Kirk Borne le 16 octobre 2023

Comme précisé précédemment, Kirk Borne est un ingénieur spécialisé en *data science* issu d'une formation en astrophysique et se définissant notamment comme *data scientist*, influenceur sur les thématiques des données et actuellement *chief data officer* au sein de DataPrime.ai [81].

Les autres comptes ne donnent pas suffisamment d'informations pour déterminer l'identité de leur auteur. Toutefois, on peut noter qu'ils démontrent un intérêt pour les nouvelles technologies et les thématiques associées (ex : l'IA, la *data*, le *machine learning*, etc.) et peuvent être plutôt catégorisés comme des profils d'ingénieurs.

En nous intéressant cette fois aux comptes qui retweetent le plus, nous notons que ces derniers sont dissociés des comptes uniques présentés précédemment (voir la figure 1.11).

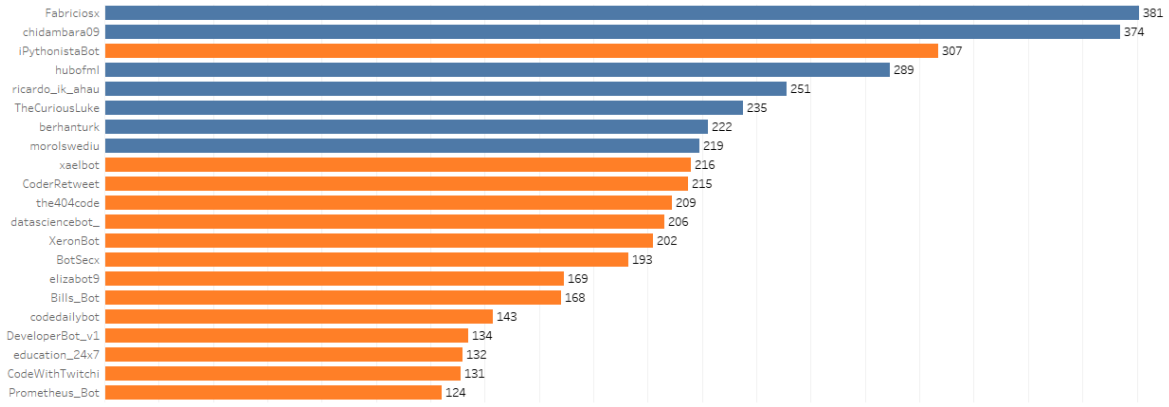

Figure 1.11.: Comptes ayant le plus publié de retweets

Les comptes en orange sont des bots, indiquant que la majorité des retweets est de leur fait. Les comptes uniques proposant le plus de retweets sont @Fabriciosx, @chidambara09, @hubofml, @ricardo_ik_ahau, @thecuriousluke, @berhanturk et @morolwediu. Seuls @hubofml est clairement identifié : il s'agit du compte d'*Effective Software Leads*, une publication mensuelle traitant de l'ingénierie logicielle (*software engineering*) [82].

Au sein de ces retweets, nous distinguons seulement cinq comptes majoritairement retweetés (voir le tableau 1.7) :

[81] http://www.kirkborne.net/
[82] https://softwareleads.substack.com/about

Tableau 1.7.: Les comptes les plus retweetés

Nom du compte	Nombre de retweets	Typologie de l'acteur
@KirkDBorne	12719	Chief data officer
@Williamstyles23	1597	Web designer chez the favourable group
@Antgrasso	1206	Entrepreneur, technologist, sustainability advocate
@FavourableGroup	402	Organisme
@Kristenmag	398	Non défini

Nous retrouvons à nouveau @KirkDBorne (12 719 retweets), suivi ensuite par deux entrepreneurs (@Williamstyles23) et (@Antgrasso), un organisme (@avourablegroup), basé sur un modèle « combinant jeu compétitif et philanthropie » [83] et une personne non identifiée (@Kristenmag).

Il est à noter que le seul retweet en lien avec la *data literacy* est celui de la *data scientist* Karem Carr, daté du 08 août 2022, dans lequel sont présentées des stratégies à adopter pour débuter dans ce domaine. Au moment de l'absorption des tweets, il a été retweeté 405 fois.

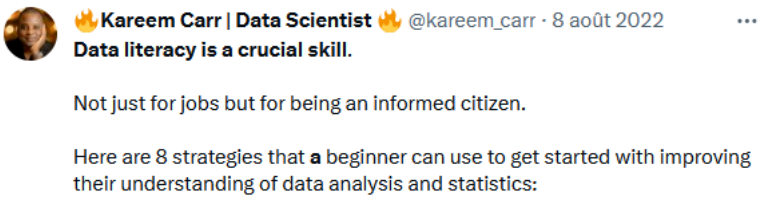

Figure 1.12.: Tweet le plus retweeté sur la data literacy

Le contenu des tweets

Nous avons tout d'abord listé l'ensemble des termes les plus fréquents dans le corpus des tweets dont ont été exclus les mots vides. Par souci de lisibilité, nous ne listerons que les 20 premiers :

Tableau 1.8.: Les vingt termes les plus employés dans les tweets

Termes	Occurrences
data	38395
dataliteracy	27093
literacy	20112
bigdata	17473

[83] https://favourable.group/

Termes	**Occurrences**
datascience	16828
ai	16554
machinelearning	14576
datascientists	9527
statistics	9307
analytics	7483
100daysofcode	6996
python	6021
mathematics	5220
deeplearning	5184
abdsc	4801
skills	3805
statisticalliteracy	3740
datamining	3410
algorithms	3135
courses	2988

Nous constatons une certaine représentation de l'écosystème des données de par la présence de concepts (le *big data*, l'IA, la *data*, algorithmes etc.), de disciplines (la *data science*, les statistiques, les mathématiques) et de littératies voisines (*statistical literacy*), d'outils et de techniques (*machine learning*, *data mining*, le langage Python, etc.), de métiers (*data scientists*) ou encore d'évènements (*100 days of code* qui est un défi visant à coder sur une durée de 100 jours).

Nous avons ensuite recherché les hashtags employés au sein de notre corpus. Si le hashtag « #dataliteracy » apparaît logiquement en tête du fait que le terme « dataliteracy » constitue notre critère d'absorption des tweets (26 995 fois), d'autres hashtags lui sont associés. Nous en avons recensé 4956. A nouveau, nous listons seulement les 20 premiers les plus cités (voir le tableau 1.9) :

Tableau 1.9.: Les hashtags employés au sein des tweets

Rang	**Hashtag**	**Occurrences**
1	dataliteracy	26995
2	bigdata	17467
3	datascience	16820
4	ai	15259
5	machinelearning	14575
6	datascientists	9527

Rang	Hashtag	Occurrences
7	statistics	8722
8	100daysofcode	6996
9	analytics	6123
10	python	5894
11	deeplearning	5184
12	mathematics	5161
13	abdsc	4801
14	statisticalliteracy	3740
15	datamining	3410
16	algorithms	3090
17	devcommunity	2790
18	data	2710
19	indiedev	2426
20	gamedev	2394

Ces hashtags semblent correspondre en majorité aux termes les plus employés. Nous avons comptabilisé les pourcentages de correspondance entre le nombre d'occurrences d'un terme et son emploi en hashtag (voir le tableau 1.10) :

Tableau 1.10.: Correspondance en pourcentage entre les termes et les hashtags employés

Termes	Pourcentage
100daysofcode	7,1%
abdsc	99,6%
ai	3,1%
analytics	100%
bigdata	100%
data	92,2%
dataliteracy	100%
datamining	100%
datascience	100%
datascientists	93,7%
deeplearning	81,8%
devcommunity	100%
gamedev	97,9%
indiedev	98,9%
literacy	100,0%

Termes	Pourcentage
machinelearning	100,0%
mathematics	1,6%
python	100%
skills	100%
statisticalliteracy	100%
statistics	100%

Nous constatons que la quasi totalité des termes sont employés en hashtags (ex : le hashtag #dataliteracy apparaît 26695 fois, tandis que le terme sans hashtag n'est utilisé que 398 fois, pour un total de 27093), laissant supposer que les contenus informationnels ne sont que peu développés, les termes n'étant alors brandis que comme symboles.

Nous avons continué de chercher le contexte d'énonciation du terme *dataliteracy* en nous intéressant aux bigrammes, c'est-à-dire deux termes employés simultanément (voir le tableau 1.11) :

Tableau 1.11.: Les vingt bigrammes comportant le plus d'occurrences avec le terme « dataliteracy »

Bigrammes	Occurrences
data literacy	13501
ai dataliteracy	3286
datascientists dataliteracy	2651
dataliteracy cube	2211
dataliteracy python	1674
datascience dataliteracy	1674
dataliteracy statistics	1583
statistics dataliteracy	1260
dataliteracy datafluency	1140
dataliteracy machinelearning	1086
dataliteracy statisticalliteracy	1023
machinelearning dataliteracy	979
deeplearning dataliteracy	873
bedatabrilliant dataliteracy	676
dataliteracy abdsc	666
mathematics dataliteracy	597
analytics dataliteracy	582
dataliteracy careers	571
datascientist dataliteracy	484

Bigrammes	Occurrences
dataliteracy 100daysofcode	456

A nouveau, nous retrouvons la *data literacy* employée avec les termes précédemment énoncés, intégrés au sein d'un vocable lié à la *data science*. Ne pouvant analyser de manière pertinente le corpus à l'échelle macro, nous allons adopter une vision plus micro et nous intéresser aux comptes eux-mêmes.

Les tweets de @KirkDBorne

En reprenant les comptes les plus retweetés, nous revenons sur le cas de @KirkDBorne qui s'avère être le compte le plus retweeté et le plus productif recensé sur Twitter/X (756 tweets). Nous avons tout d'abord listé les termes les plus employés dans les tweets et avons nettoyé les mots vides (*stopwords*). Nous n'avons extrait que les termes les plus usités (voir le tableau 1.12) :

Tableau 1.12.: Les termes les plus utilisés par @KirkDBorne

Termes	Occurrences
dataliteracy	754
datascience	747
bigdata	733
machinelearning	692
ai	632
datascientists	436
statistics	336
abdsc	318
analytics	298
data	271
mathematics	181
datastrategy	152
deeplearning	148
python	147
statisticalliteracy	146
algorithms	138
datamining	136
datafluency	134
dataleadership	128

Nous avons renouvelé l'expérience avec les hashtags (voir le tableau 1.13). A nouveau, nous n'avons extrait que les plus employés (>100) :

Tableau 1.13.: Les hashtags les plus employés par Kirk D Borne

Hashtag	**Occurrences**
dataliteracy	754
datascience	747
bigdata	733
machinelearning	692
ai	602
datascientists	436
statistics	336
abdsc	318
analytics	259
mathematics	180
datastrategy	152
deeplearning	148
python	147
statisticalliteracy	146
datamining	136
datafluency	134
algorithms	133
dataleadership	128

Nous pointons ici à nouveau une grande corrélation des termes avec les hashtags, soulignant la constance des contenus.

Nous listons dans le tableau 1.14 ses tweets les plus retweetés, amenant à une double hypothèse : que ces retweets supposent une adhésion complète à ses contenus et que ces contenus sont représentatifs des messages véhiculés en général sur la plateforme.

Tableau 1.14.: Les cinq tweets les plus retweetés de Kirk D Borne

Description	Date de publication	Nombre de retweets
70 Free Online Courses for #DataScience to Advance Your Skills in 2021: https://t.co/h5exMoaa5s via @tut_ml ——— #BigData #Analytics #MachineLearning #DeepLearning #AI #DataScientists #Statistics #DataLiteracy #Python #RStats https://t.co/8wCDZTOtYF	2021-11-10 08:14:56	364
70 Free Online Courses for #DataScience to Advance Your Skills in 2021: https://t.co/h5exMoaa5s via @tut_ml ——— #BigData #Analytics #MachineLearning #DeepLearning #AI #DataScientists #Statistics #DataLiteracy #Python #RStats https://t.co/8wCDZTOtYF	2021-11-13 21:24:42	364
[Free PDF Download] Introduction to #Probability for Data Science: https://t.co/L51j3udY5j by @stanley_h_chan ——— #StatisticalLiteracy #Mathematics #AI #BigData #MachineLearning #DataScience #DataScientists #DataLiteracy #Statistics https://t.co/sJOOnWOCcw	2022-01-04 21:25:10	258
[Free PDF Download] Introduction to #Probability for Data Science: https://t.co/L51j3udY5j by @stanley_h_chan ——— #StatisticalLiteracy #Mathematics #AI #BigData #MachineLearning #DataScience #DataScientists #DataLiteracy #Statistics https://t.co/W8I5hEoJ3P	2022-02-06 04:41:01	227
Course material for course on #Probability for first-year #mathematics students at Cambridge U: https://t.co/yIyWc1J5AP ——— #Statistics #StatisticalLiteracy #BigData #DataScience #DataLiteracy #MachineLearning #abdsc https://t.co/qalz5pStzh	2021-11-21 07:22:34	214

La lecture de ces tweets ne laisse apparaître qu'une vision commerciale : il ne s'agit que de la promotion d'ouvrages et de formations. Les hashtags ici sont employés en grand nombre au sein de chaque tweet, servant de catalyseur à leur visibilité.

Nous avons renouvelé l'expérience avec le second compte le plus retweeté, à savoir @Williamstyles23. Du fait de la similarité de l'expérience et des résultats obtenus et pour alléger la lecture, nous avons déplacé cette analyse dans l'annexe 3. Nous pouvons néanmoins préciser que la *data literacy* ne tient aucune place dans les contenus des tweets, ces derniers servant simplement de promotion aux activités des auteurs.

1.3.4.3. Conclusion de l'expérience

Cette expérience sur Twitter/X nous a montré plusieurs chose. Tout d'abord que le concept de *data literacy* n'est pas discuté en tant que tel. A travers son hashtag (#dataliteracy), la *data literacy* sert uniquement de catalyseur pour promouvoir les initiatives, majoritairement situées en *data science*, de leurs acteurs. Les acteurs recensés sont des entrepreneurs intégrés plus ou moins directement dans le « monde des données ». Les universitaires n'apparaissent que très minoritairement dans notre corpus, ne permettant pas de faire émerger leur pensée. Twitter/X n'apparaît donc pas, sur la période retenue, comme le lieu où les échanges sur ce concept ont eu lieu.

1.4. Conclusion du chapitre

La *data literacy* se présente ainsi comme un ensemble de formations aux usages des données dont le nombre de publics destinataires n'a cessé de croître depuis plus de vingt ans. En effet, l'évolution et la démocratisation des technologies numériques ont généralisé les problématiques de traitement des données à toute la société. Les nécessités de spécialisation, autrefois sollicitées principalement par les entreprises dans un objectif purement économique, sont désormais appuyées par les institutions et ce jusqu'au plus haut sommet des États.

L'objectif originel centré sur une maîtrise technique de traitement des données s'est étoffé, intégrant désormais des postures et des attitudes éthiques et critiques. Une maîtrise non plus limitée à un cercle restreint d'acteurs que sont les scientifiques et les cadres d'entreprises, mais étendue à tous les citoyens. La nouvelle place accordée à ces derniers dans l'écosystème des données les incite à se questionner sur leurs pratiques quotidiennes désormais inévitablement intégrées dans une vision numérique.

Au même titre que l'informatique et l'information au cours du XXe siècle, et plus récemment le numérique au XXIe siècle, les données sont désormais au cœur des discours : l'avènement des *data literates* (*citizens*) est attendue avec 2030 comme horizon d'un ensemble d'initiatives. Néanmoins, nous verrons dans le prochain chapitre que ces discours sont loin d'être nouveaux, les données ayant fait l'objet dès les années 1960 d'enjeux similaires. Les termes et technologies changent, la volonté de maîtrise demeure identique.

Aujourd'hui, l'ensemble des disciplines démontrent un intérêt fort pour cette thématique, elle-même originellement définie dans les domaines des mathématiques, de la statistique et de la pédagogie. Mais au-delà des objectifs qu'on lui assigne, la *data literacy* reste globalement perçue comme un set de compétences. La littératie incluant, nous l'avons vu, une variable formatrice, celle-ci doit donc proposer un ensemble de méthodes pour transmettre efficacement ces connaissances.

Or, la compétence en anglais n'est que peu définie dans les discours et la terminologie sur ce sujet varie selon les auteurs consultés. L'emploi, souvent invariable, de *capability*, d'*ability* et de *competence*, englobant parfois des *skills*, concourent à embrumer davantage un concept déjà considéré comme nébuleux.

En reprenant les premières définitions de la *data literacy*, la compétence (comme terme générique pour englober les dénominations *ability*, *capability,* et *competence*) était avant tout perçue comme une *hard skill*, à savoir un ensemble de compétences techniques et pratiques permettant de répondre à une situation professionnelle spécifique (Cimatti, 2016; Laker & Powell, 2011; Lyu & Liu, 2021).

En ce sens, elle est parfois englobée, dans les discours, dans l'univers de la *data science*, cette discipline promouvant une haute maîtrise technique des traitements de données qui a formalisé ses concepts et sa vision globale peu avant celle-ci. Or, comme nous le verrons dans le chapitre 3, si les deux concepts se rejoignent dans une même volonté de formalisation d'un apprentissage des données, leurs visées diffèrent.

Avec l'intégration de la vie quotidienne et de la nécessité d'une formation tout au long de la vie dans l'équation pédagogique, la compétence (en) *data literacy* rencontre un écho avec l'approche cognitiviste où « il ne s'agit plus de faire produire des comportements reproductibles, mais de développer des dispositions à générer des conduites adaptées face à des situations diverses et changeantes » (Chauvigné & Coulet, 2010).

La variable de résolution de problèmes du quotidien contenait déjà la possibilité pour la *data literacy* de dépasser son cadre restreint d'utilisateurs pour intégrer celui des citoyens. Si cet ensemble de compétences techniques a effectivement fini par être généralisé, l'évolution conceptuelle de la *data literacy* nécessitera encore plusieurs années pour pleinement se diversifier. L'intégration de postures intellectuelles a souligné les limites d'une application purement technique et a profondément changé

ses applications : la *data literacy* est désormais un tremplin pour la diffusion de nouvelles pratiques idéologiques des données que nous approfondirons dans le chapitre 4.

Avant de présenter les interactions et évolutions conceptuelles ayant amené la *data literacy* à adopter un kaléidoscope d'identités parallèles, nous allons à présent présenter son histoire à travers l'étude de ses composantes et rappeler la fausseté de sa présupposée nouveauté.

Chapitre 2.

Des influences contrastées

> « *Ce que l'écriture apporte à ceux qui savent lire, la peinture le donne aux incultes qui la regardent ; les ignorants y voient ce qu'ils doivent suivre et les illettrés y lisent* »
>
> Grégoire Le Grand, *Registrum Epistularum, Epistula XI*, 873-876

2.1. Introduction

Avec la (re)découverte du potentiel des données du fait de l'avènement des nouvelles technologies numériques au début du XXIe siècle, la *data literacy* se retrouve souvent marquée dans les discours du sceau de la nouveauté. Elle serait « la » nouvelle compétence nécessaire pour extraire les données, ce matériau indispensable à l'édification des sociétés dématérialisées (Souter, 2021).

Or, la stratégie d'une appropriation maîtrisée des données est bien plus ancienne, incarnée au fil des siècles dans des disciplines et des visions diverses. La *data literacy* est en effet héritière non pas d'apports homogènes mais d'un ensemble de controverses et de luttes. Et parfois, d'oublis. A défaut d' « oublier le passé récent pour retrouver le passé ancien » (Auge, 2019), nous nous intéresserons ici à recenser les influences ayant permis l'avènement de la *data literacy* : une dite « attestée », c'est-à-dire dont les sources sont citées dans la littérature scientifique, et une autre dite « oubliée » avec la datalogie.

2.2. Les influences attestées

2.2.1. La littératie

Nous avons succinctement évoqué la littératie dans le chapitre précédent. Pour le dictionnaire *Merriam-Webster*, la littératie (*literacy*) est la qualité ou l'état d'un lettré (*literate*) [1], celui qui sait lire et écrire. En ce sens, la littératie est perçue comme « une mesure du degré d'assimilation des connaissances sur l'*écrit* » (Pierre, 2003). Une volonté de dépasser cette vision réductrice de la littératie comme simple enseignement ou compétence s'est longtemps retrouvée dans les discours. David (2015) listait deux visions antagonistes : l'une, qu'il nomme « minimaliste » qui désigne « la capacité élémentaire à lire et à écrire (en rapport avec les apprentissages) » ; l'autre, « maximale » qui établit « un rapport étroit entre l'écrit et l'ensemble des connaissances qu'il permet de construire. » Dans cette seconde optique, elle est pour Privat (2010) « l'ensemble des praxis et représentations liées à l'écrit, depuis les conditions matérielles de sa réalisation effective (supports, espaces et outils techniques d'inscription) jusqu'aux objets intellectuels de sa production et aux habiletés cognitives et culturelles de sa réception, sans oublier les agents et institution de sa conservation et de sa transmission » (p. 23).

L'étude de la littératie porte donc sur celle de l'écriture et la lecture, et de leurs (potentiels) effets sur la cognition humaine (Olson & Lejosne, 2006). Elle a été reprise par de nombreux champs disciplinaires selon Wagner, cité par Olson & Torrance (2009), qui spécifie qu'elle dépasse le champ de la recherche pour englober à la fois les pratiques et les politiques, de l'enfance à la vie adulte. Reuter, cité par Dupont & Grandaty (2012), résume la littératie selon une formalisation de « l'écrit autour des relations entre (1) lecture-écriture, (2) de son incarnation au travers de pratiques et d'institutions, (3) de la fonctionnalité pour l'individu et la société, (4) du continuum de son apprentissage et de sa maîtrise ».

2.2.1.1. Le grand partage

La littératie a connu une nouvelle médiatisation au cours du milieu du XXe siècle grâce aux travaux de plusieurs auteurs dont le plus connu est l'anthropologue Jack Goody (1919-2015). L'apport central de sa théorie est le dépassement d'une vision anthropologique et sociologique ancienne basée sur le « grand partage »[2], dont l'un des représentants est Levy-Bruhl qui publia en 1922 l'ouvrage controversé « la mentalité primitive » dans lequel il spécifiait une différence fondamentale entre les cultures

[1] https://www.merriam-webster.com/dictionary/literacy

[2] Le Grand Partage, appelé *great divide* en anglais, est une vision ethnologique ancienne qui séparait nettement les sociétés en deux catégories : celles dites « primitives » et celles dites « civilisées ».

traditionnelles et les cultures modernes. Supposément, les premières, « essentiellement mystiques » (p. 145), intégraient un monde enchanté peuplé d'esprits et de démons, alors que les secondes étaient gouvernées par un esprit rationnel. Il ajoutera que « L'Européen pratique l'abstraction presque sans y penser, et les opérations logiques simples lui sont rendues si faciles par son langage qu'elles ne lui coûtent pas d'effort. Chez les primitifs, la pensée, et la langue, sont de caractère presque exclusivement concret [...] D'un mot, notre mentalité est surtout "conceptuelle" et l'autre ne l'est guère. » (p. 146).

Goody a considéré l'opposition entre sociétés dites primitives et civilisées, pensée sauvage et pensée domestiquée, comme un « ordre illusoire » (Goody, 1979, p. 85). Pour lui et d'autres auteurs comme David Olson, la véritable différence entre les sociétés résidait dans la pratique maîtrisée de l'écriture et de la lecture.

2.2.1.2. L'écriture

L'écriture possède de nombreuses fonctions qui ont été amplement débattues : par souci de concision et de clarté de lecture, nous nous contenterons ici de les résumer. L'écriture est au départ intégrée dans une « vision représentative » (Klock-Fontanille, 2016), perçue comme secondaire au langage. Elle avait pour simple vocation de figer le langage parlé (Higounet, 2003), ce que nous retrouvons dès Aristote, dans son ouvrage *De l'Interprétation* où il considère que les mots parlés sont les signes des affections de l'âme, tandis que les mots écrits sont ceux des mots parlés. Rousseau, dans son *Essai sur l'origine des langues* de 1761, en parlait comme d'un « supplément de la parole ». Ferdinand de Saussure, dans son *Cours de linguistique générale*, précisait de son côté que « langue et écriture sont deux systèmes de signes distincts ; l'unique raison d'être du second est de représenter le premier » (De Saussure, 1971, p. 45). Olson (1998), qui considérait tout ceci comme une vision dépassée, voyait l'écriture autrement :

> « Nous le savons tous, écrire, ce n'est pas seulement le b-a-ba, et lire, ce n'est pas seulement décoder des mots ou des phrases. Mais qu'est-ce au juste que ce petit "plus" ? Il pourrait s'agir de la capacité à entrer dans un nouvel univers (ou à en sortir), celui du monde "sur le papier" » (p. 10).

Au-delà de cette fonction, c'est une dissociation avec l'oralité qui devait être prise en compte. Citant Harris et Gaur, il rappelait que la relation de l'écriture à la parole était au mieux indirecte (p. 85), que l'on pense le langage oral en utilisant « des concepts et des catégories venus des systèmes d'écriture » et non l'inverse (p. 86). Ainsi, plus qu'une simple transcription, l'écriture marque tout d'abord le développement de supports externes pour la mémoire humaine :

> « On peut attribuer la première fracture potentielle entre l'histoire et la mémoire au développement de l'écriture comme moyen d'inscrire l'expérience humaine sur un support matériel, distinct du corps : brique, papyrus, parchemin, papier, disque compact ; pour ne rien dire des inscriptions qui ne transcrivent pas la voix humaine : marques, dessins, jeu de couleurs dans les vêtements, les jardins, stèles, monuments… » (Ricœur, 2006).

Elle est en ce sens *hypomnema* (*ὑπόμνημα* : mémoire, note que l'on prend par écrit), cette fixation du souvenir évoquée dans *le Phèdre* de Platon. Elle est une représentation visuelle du langage qui le rend « durable et transportable » (Cohen, 1958). Foucault (1994) la décrira en ce sens comme une opposition « à cet éparpillement en fixant des éléments acquis et en constituant en quelque sorte "du passé' ', vers lequel il est toujours possible de faire retour et retraite » (p. 1239). Stiegler reprendra le terme pluriel *hypomnemata* et, en s'appuyant notamment sur les travaux de Leroi-Gourhan et Simondon, liera écriture et technique, et plus globalement les hommes avec la technique.

Ces derniers sont pour lui indissociables : « l'homme s'invente dans la technique et la technique s'invente dans l'homme. Ce couple est un *processus où la vie négocie avec le non-vivant en l'organisant*, mais de telle manière que *cette organisation fait système et a ses propres lois*» (Stiegler, 1998). La technique elle-même est une *mnémotechnologie*, une mémoire *épiphylogénétique* pouvant être transmise d'individu en individu. Ceci rejoint les travaux de Goody qui parlait de l'écriture comme d'une « technologie de l'intellect » (Goody & Lejosne, 2006), ce que Jeanneret (2017) reprendra en précisant qu'il n'y a pas de pensées (humaines) sans dispositif technique (p. 46). L'écriture permet donc le cumul de connaissances, de conserver la mémoire.

Pour Goody, l'écriture dépasse ce simple cadre instrumental en permettant par ailleurs le développement de la logique : « c'est en effet la transcription de la parole qui permet de clairement séparer les mots, d'en manipuler l'ordre et de développer les formes syllogistiques de raisonnement » (Goody, 1979, p. 50). Elle permet un retour réflexif sur les connaissances (la pensée peut se distancier d'elle-même (Derrida, 1967)), au contraire de l'oral où le mot se perd sitôt prononcé ; le langage articulé étant « fugitif par essence même » (Février, 1995). Néanmoins, le lien entre cognition et littératie sera remis en cause à plusieurs reprises. Lévi-Strauss (1955) le fit déjà dès 1955 :

> « C'est une étrange chose que l'écriture. Il semblerait que son apparition n'eût pu manquer de déterminer des changements profonds dans les conditions d'existence de l'humanité; et que ces transformations dussent être surtout de nature intellectuelle. La possession de l'écriture multiplie prodigieusement l'aptitude des hommes à préserver les connaissances. On la concevrait volontiers comme une mémoire artificielle, dont le développement devrait s'accompagner d'une meilleure conscience du passé, donc d'une plus grande capacité à organiser le présent et l'avenir. Après avoir éliminé tous

> les critères proposés pour distinguer la barbarie de la civilisation, on aimerait au moins retenir celui-là : peuples avec ou sans écriture, les uns capables de cumuler les acquisitions anciennes et progressant de plus en plus vite vers le but qu'ils se sont assigné, tandis que les autres, impuissants à retenir le passé au delà de cette frange que la mémoire individuelle suffit à fixer, resteraient prisonniers d'une histoire fluctuante à laquelle manqueraient toujours une origine et la conscience durable du projet. Pourtant, rien de ce que nous savons de l'écriture et de son rôle dans l'évolution ne justifie une telle conception. Une des phases les plus créatrices de l'histoire de l'humanité se place pendant l'avènement du néolithique, responsable de l'agriculture, de la domestication des animaux et d'autres arts » (pp. 352 - 353).

L'écriture est également le fait d'autres spéculations intellectuelles comme celle d'avoir permis le développement de la démocratie et des sciences, ce que l'on trouve par exemple chez Havelock (1963). Olson (1998) remettra cela en cause, citant le fait que chez les Grecs c'est la dialectique et non l'écriture qui a favorisé ces développements.

En outre, l'écriture est perçue comme *episteme* et *weltanschauung*, la connaissance du monde du monde qui nous entoure permettant de nous en former une représentation spécifique : « c'est (en effet) se former simultanément une image du monde et de soi-même, savoir ce qu'est le monde et savoir ce que l'on est » (Jung, 1996). Philippe Quéau, cité par Yves Jeanneret, le précisait aussi : « chaque fois que l'homme a changé de système d'écriture, il a aussi changé le système de représentation du monde » (Jeanneret, 2017, p. 12). Ceci s'inscrit dans une vision partagée par un grand nombre d'auteurs comme Heidegger, Wittgenstein ou encore Derrida pour lesquels on ne fait l'expérience du monde que par la médiation de signes et de symboles.

D'un point de vue négatif, la question du pouvoir discriminant et coercitif de l'écriture a été rappelé par Lévi-Strauss dès « Tristes Tropiques » :

> « Si mon hypothèse est exacte, il faut admettre que la fonction primaire de la communication écrite est de faciliter l'asservissement. L'emploi de l'écriture à des fins désintéressées, en vue d'en tirer des satisfactions intellectuelles et esthétiques, est un résultat secondaire, si même il ne se réduit pas le plus souvent à un moyen pour renforcer, justifier ou dissimuler l'autre » (Lévi-Strauss, 1955, p. 354).

Il ajoutera que l'objectif de l'écriture est de maintenir un contrôle sur les populations par les États. Bien au-delà de simplement consolider les connaissances, elle était un moyen « d'affermir les dominations ». En ce sens, « la lutte contre l'analphabétisme se confond ainsi avec le renforcement du contrôle des citoyens par le Pouvoir. Car il faut que tous sachent lire pour que ce dernier puisse dire : nul n'est censé ignorer la loi. » (p. 355).

2.2.1.3. Le combat contre l'illettrisme

Malgré ces défauts, la nécessité de l'appropriation de l'écriture et de la lecture va rencontrer un écho grandissant au cours des siècles. Un écho qui continue encore à se perpétuer aujourd'hui, se diversifiant selon la littératie de référence. L'illettrisme est ainsi désigné comme un enjeu, un problème contre lequel des politiques doivent être mises en place pour faire avancer le progrès humain. L'illettrisme est la traduction du terme *illiteracy*, employé dès 1660[3] et qui signifie ne pas savoir lire ou écrire ; le terme *literacy*, employé dès 1880, s'est construit en opposition à celui-ci.

On en retrouve l'usage à plusieurs reprises dans le journal *New England Journal of Education* daté de janvier 1883 (vol. XVII, n°3) présenté dans la figure 2.1 :

THE Table of Illiteracy in the United States, as presented on another page, will be studied with deep interest, since it is the first tabular statement from the Census office officially setting forth the relative condition of education in general in the States and Territories of our Union. As will be seen, Iowa leads the column in literacy, with reading as a test, and Wyoming Territory with writing; while South Carolina among the States, and New Mexico among the Territories, are at the foot of the column. Of native literacy, Massachusetts stands at the head, *having only seven persons in each thousand who cannot write*, and this small ratio increases to *two hundred and seventy-eight in one thousand in Tennessee*, and *six hundred and forty-two in New Mexico*. Rhode Island has the largest percentage of foreign-born illiterates,—twenty-seven per cent. of the foreign-born population; while North Carolina has only three and three-tenths per cent. of foreign-born illiterates. The "Black belt" of illiteracy is in South Carolina, Georgia, and Alabama, where eight hundred of each one thousand colored persons are returned as unable to write, though a larger proportion may be able to read. Any person studying this table for a half-hour, who does not see that *the school question is the question* of the hour, must belong to one of the classes Governor Butler exempts from the responsible care of the State.

Census Bulletin No. 303.

ILLITERACY IN THE UNITED STATES, AS RETURNED AT THE TENTH CENSUS.

States and Territories.	Persons of 10 years of age and upward.					White persons of 10 years of age and upward.			Native white persons of 10 years of age and upward.			Foreign-born white persons of 10 years of age and upward.			Colored persons of 10 years of age and upward.		
	Enumerated.	Returned as unable to read.		Returned as unable to write.		Enumerated.	Returned as unable to write.		Enumerated.	Returned as unable to write.		Enumerated.	Returned as unable to write.		Enumerated.	Returned as unable to write.	
		Number.	Per cent.	Number.	Per cent.		Number.	Per cent.		Number.	Per cent.		Number.	Per cent.		Number.	Per cent.
The United States	36,761,607	4,923,451	13.4	6,239,958	17.0	32,160,400	3,019,080	9.4	25,785,789	2,255,460	8.7	6,374,611	763,620	12.0	4,601,207	3,220,878	70.0

Figure 2.1.: Evocations de l' *illiteracy* - extraits du *New England of Education* du 18 janvier 1883, Volume XVII, n°3, pp. 39-40

La même année est publié l'article *The illiteracy state of the united states* paru dans la revue australienne *The Pastoral Times* en juin 1883 [4] rapportant alors que cinq millions d'Américains ne savaient pas lire, qu'un électeur sur cinq était incapable d'écrire son nom. Une crainte était palpable : celle de voir les élections influencées par ces électeurs illettrés. Cette crainte réapparaît peu après en 1884 dans l'article *The problem of Illiteracy*, paru dans le journal Educational Weekly[5], où le problème de l'illettrisme aux États-Unis est cette fois cité comme cause nationale :

[3] https://www.merriam-webster.com/dictionary/illiteracy

[4] Voir à ce sujet (Pastoral Times, 1883)

[5] Voir à ce sujet (Educational Weekly, 1884).

> « Le grand problème national de l'époque est indubitablement le problème de l'illettrisme. L'esclavage est aboli à jamais. Suite à l'arbitrage de la guerre civile, la question des droits de l'État a été réglée de telle manière qu'elle exige l'élimination de l'illettrisme [6] ».

Les électeurs illettrés sont ici identifiés comme provenant du sud des États-Unis. Pour enrayer ce problème, des aides financières devaient être apportées :

> « Des élections honnêtes doivent au moins être rendues possibles en permettant aux électeurs de voter de manière rationnelle. Dénoncer la corruption lors des élections ne nous donnera pas des élections honnêtes. Voici un champ ouvert pour une exposition de sens politique. Pour les peuples du Sud, il est extrêmement préoccupant que les millions d'analphabètes, désormais armés d'un bulletin de vote, puissent obtenir le plus tôt possible les qualifications requises à la citoyenneté. C'est un sujet de préoccupation tout aussi important pour les peuples du Nord, car nous sommes un seul peuple et devons rester un seul peuple. Il est absolument certain que le Sud ne peut pas, sans aide, faire face à cette urgence [7] ».

La littératie intervient donc dans un objectif politique de régulation des populations dès le XIXe siècle aux États-Unis. Ce n'est toutefois pas la première fois qu'une politique générale d'alphabétisation des populations est mise en place : Arnove & Graff (1987) ont recensé les travaux de plusieurs chercheurs démontrant la présence de campagnes littéraciques dès le XVIe et XVIIe siècles aidées en cela par la diffusion de la religion, en particulier en Suède où tout d'abord l'*års kyrkolag* de 1686 encourageait l'éducation religieuse des masses. Il sera suivi du décret royal de 1723 qui incitaient les parents et les éducateurs de veiller à ce que les enfants s'appliquent à lire les livres et à apprendre les versets de la Bible (Johansson, 1987). Une vérification était faite au cours de l'*husförhör*, un évènement annuel où les prêtres luthériens inscrivaient dans un registre les niveaux de lecture et d'instruction religieuse des enfants. S'ils s'avéraient trop faibles, ils devaient alors remédier à cette situation en les instruisant davantage (Green, 2008).

---

[6] « *The great national problem of the times is unquestionably the problem of illiteracy. Slavery is forever abolished. By the arbitrament of civil war, the question of State rights has been settled in such a way as to require that illiteracy be wiped out*. »

[7] « *Honest elections must at least be made possible by qualifying the electors to vote rationally. Denouncing corruption in elections will not give us honest elections. Here is an open field for an exhibition of statesmanship. To the people of the South it is a matter of overwhelming concern that the millions of illiterates now armed with the ballot, should at the earliest possible moment be qualified for citizenship It is a matter of no less concern to the people of the North, for we are one people, and must remain one people. It is positively certain that the South cannot, unaided, meet this emergency*. »

Cette volonté d'alphabétisation se détachera du contexte purement religieux et se perpétuera tout au long du XXe siècle à travers les campagnes de l'*UNESCO*[8] débutées au cours des années 1950. En effet, dès 1958, l'*UNESCO* définissait la littératie comme la capacité d'un individu de savoir lire et écrire, avec compréhension, une déclaration simple et courte en lien avec sa vie quotidienne [9]. Elle en fit son cheval de bataille durant toutes les décennies suivantes.

Olson & Lejosne (2006) rappelaient que la littératie était alors perçue comme un bien d'une valeur telle « qu'on est en droit de l'imposer à tous, que ce soit par le biais de l'école pour tous ou par le biais de campagnes internationales contre l'illettrisme. Cette idée très répandue, qui remonte au moins au siècle des Lumières, avance que la littératie a des effets directs et profitables : si toute la population apprend à lire et écrire, elle sera forcément plus réfléchie et démocratique ». L'*UNESCO* ne disait pas autre chose en ce sens :

> « L'analphabétisme ayant été reconnu comme facteur de sous-développement, inversement l'alphabétisation s'impose comme un facteur de développement. Tel est le sens de la notion d'alphabétisation fonctionnelle, dont l'Unesco a fait sa doctrine et la règle de son action ». (UNESCO, 1968, p. 80)

Plusieurs évènements mettront le besoin de la littératie (traduite par alphabétisation dans les versions françaises) au centre des discours. Nous pouvons citer comme grand point de départ le congrès mondial des ministres de l'éducation sur l'élimination de l'analphabétisme ayant eu lieu à Téhéran en 1965. Le Directeur général de l'*UNESCO*, René Maheu, déclarait à ce propos en 1967 [10] :

> « Il y a deux ans, jour pour jour, s'ouvrait à Téhéran, grâce à la généreuse hospitalité de S.M.I. le Shahinshah d'Iran, le Congrès mondial des ministres de l'éducation sur l'élimination de l'analphabétisme organisé par l'Unesco. Les représentants des 88 pays qui participèrent au congrès furent unanimes à dénoncer les tares d'un monde où l'on compte un milliard d'analphabètes et de semi-analphabètes et à proclamer que la lutte contre un fléau de cette envergure devait être menée solidairement par l'humanité tout entière. Que tant d'hommes et de femmes demeurent illettrés en ce XXe siècle est un

[8] L'*UNESCO* (*United Nations Educational, Scientific and Cultural Organization*, traduit en français par Organisation des Nations unies pour l'éducation, la science et la culture) est une institution internationale créée par l'*Organisation des Nations Unies* (*ONU*) en 1945. Elle a pour objectif de contribuer « au maintien de la paix et de la sécurité en resserrant, par l'éducation, la science et la culture, la collaboration entre nations, afin d'assurer le respect universel de la justice, de la loi, des droits de l'homme et des libertés fondamentales pour tous, sans distinction de race, de sexe, de langue ou de religion, que la *Charte des Nations Unies* reconnaît à tous les peuples. » https://web.archive.org/web/20220315073418/http://portal.unesco.org//fr/ev.php-URL_ID=15244&URL_DO=DO_TOPIC&URL_SECTION=201.html

[9] « *the ability of an individual to read and write with understanding a simple short statement related to his/her everyday life.* »

[10] (UNESCO, 1968), *op cit*.

> scandale qui ne saurait apparaître nulle part de manière plus éclatante qu'au sein de cette Exposition universelle de Montréal, où l'on peut voir tant d'accomplissements du génie humain. Savoir, c'est pouvoir : tout ici témoigne des prodigieux progrès réalisés grâce au développement de la science et de la technique. Or, dans l'ensemble du monde, quatre êtres humains sur dix vivent en marge de ces progrès, incapables qu'ils sont, faute d'un minimum de connaissances élémentaires pour en saisir les ressorts et en comprendre le sens, d'y participer activement et d'en bénéficier consciemment. Aussi n'est-il pas d'endroit plus approprié que celui-ci pour souligner la dramatique, l'inique, la dangereuse contradiction qui caractérise l'état présent d'organisation, ou plutôt d'inorganisation, de la collectivité humaine, où l'on voit coexister des nations qui s'enorgueillissent à bon droit de compter 30 % de leurs jeunes dans l'enseignement supérieur et d'autres dont 3 % seulement de la population adulte sait lire et écrire. » (p. 9).

Ce congrès sera suivi d'une part de la création du Programme des Nations Unies pour le Développement (*PNUD*) en 1966[11] et de la Déclaration de Persépolis en 1975[12] au sein de laquelle a été exprimée une inquiétude sur le recul de la littératie dans un certain nombre de pays, nécessitant une collaboration internationale faisant de l'illettré (*illiterate*) le sujet, et non l'objet, de son apprentissage. Deux autres déclarations seront promulguées : la déclaration mondiale sur l'éducation pour tous de 1990, dite déclaration de Jomtien[13] et la déclaration d'Hambourg de 1997[14]. La littératie devient alors un « un droit fondamental de la personne humaine ». En 2000, le cadre d'éducation de Dakar [15] fait évoluer la notion de littératie en intégrant le contexte des apprenants : elle y devient une pratique sociale. Enfin, en 2015, la déclaration d'Incheon [16] présente la littératie comme partie intégrante du droit à l'éducation et comme bien public.

Wagner (1995) critiquera cet engouement pour la littératie, perçue par les États comme quelque chose de nécessairement « bon » pour l'homme et la société mais faisant l'impasse sur une véritable prise de recul sur ses prétendus effets [17] :

[11] https://www.undp.org/fr/about-us
[12] Voir à ce sujet (International Symposium for literacy, 1975).
[13] Voir à ce sujet (UNESCO, 2011).
[14] Voir à ce sujet (UNESCO, 1997, p. 36).
[15] Voir à ce sujet (UNESCO, 2000, p. 77)
[16] Voir à ce sujet (UNESCO et al., 2017, pp. 24-28).
[17] « *What is surprising is how little information is available on the economic returns to training in adult literacy and adult basic education. There are few if any empirical studies on the economic impact of major short-term literacy programs in developing or industrialized countries [...] The economic rationale remains an article of faith among advocates of literacy programs, at least in part due to the belief that industrialized countries are more prosperous primarily because they are more educated and more literate* ».

> « Ce qui est surprenant, c'est le peu d'informations disponibles sur les retombées économiques des formations pour adultes en littératie et en éducation de base. Il existe peu, voire pas d'études empiriques sur l'impact économique des grands programmes d'alphabétisation à court terme dans les pays en développement ou industrialisés [...] La justification économique reste un article de foi parmi les partisans des programmes littéraciques, au moins en partie en raison de la conviction que les pays industrialisés sont plus prospères principalement du fait qu'ils sont plus instruits et plus lettrés » (p. 342-345).

Il rappellera à cette fin quatre mythes dans lesquels est emprisonné la littératie.

Tout d'abord, la croyance en une littératie canalisatrice de changements de la cognition humaine, les capacités de logique et l'intelligence en général qui permettent aux pays dits industrialisés de pouvoir discriminer les pays du tiers-monde, où l'alphabétisme est peu développé, comme nous avons pu le voir précédemment. Puis, la littératie comme facteur de modernisation humaine et comme aide au changement d'attitudes vis-à-vis du changement. Cette déclaration était sans fondement pour Wagner en raison du fait que la littératie se développait dans un contexte scolaire et qu'il était difficile de définir la part spécifique occupée par celle-ci dans les changements d'attitudes. Ensuite, les idéaux démocratiques concluant au développement de la productivité nationale grâce à la littératie manquait de preuves concrètes. Enfin, la fin de l'illettrisme d'ici l'année 2000 était perçue comme fantaisiste du fait des données recueillies au moment de l'écriture de son article.

Cette analyse fait suite aux travaux de Graff, développés dans son ouvrage *The Literacy myth* (Graff, 1979). Revenant sur son ouvrage trente ans plus tard, il déclarait que comme tout mythe, le mythe de la littératie, faisant référence à une croyance répandue que l'acquisition de cette dernière permettait un développement de l'économie, de pratiques démocratiques, de développement cognitif et d'élévation sociale, n'était que l'expression d'une idéologie à défaut d'être un mensonge (Graff, 2010). Si l'illettrisme a effectivement reculé, il reste encore loin d'être éradiqué [18].

Au même titre que d'autres littératies comme l'*information literacy*, la *data literacy* est également investie des mêmes enjeux, des mêmes mythes. Sa diffusion et son adoption sont censées permettre un renouveau économique et démocratique tant au niveau des organisations que des sociétés entières. Or, la question de ses limites et de ses potentiels biais discriminants ont fait l'objet de débats que nous évoquerons dans les prochains chapitres.

[18] Selon la Banque mondiale, il était de 13% au niveau mondial en 2022. Consulté sur : https://data.worldbank.org/indicator/SE.ADT.LITR.ZS

2.2.1.4. Litteratus et illiteratus

Nous avons évoqué précédemment la question des lettrés (*literate*) ainsi que du fait que la littératie était perçue comme un continuum. Nous allons ici approfondir cela. En reprenant son étymologie, le terme *literacy* est issu du latin *litteratus*, lui-même issu de *lettera*, la lettre. Selon le dictionnaire Gaffiot (1934), *litteratus* possède plusieurs sens :

1. marqué de lettres, portant des caractères ;
2. instruit, qui a des lettres ;
3. relatif aux lettres, savant ;
4. interprète des poètes, critique.

Tout comme *literacy* s'est construit en opposition à *illiteracy*, *litteratus* s'est opposé à *illiteratus*, celui qui ne sait ni lire, ni écrire le latin. Pour Grundmann (1958) ce terme a connu une évolution après le XIIe siècle, désignant alors celui qui sait lire la langue vulgaire et non le latin. Le XIIe siècle n'est pas anodin, c'est une période charnière où la littératie émerge et se répand à grande échelle :

> « La lecture et l'écriture se répandent davantage. Non pas qu'elles se démocratisent, mais leur diffusion suit, voire dépasse, le rythme effréné de la croissance démographique. Quelques indices semblent le prouver. Les écoles paroissiales se multiplient, y compris dans des villages. Le nombre de manuscrits des xiie et xiiie siècles conservés à ce jour dans les bibliothèques européennes est incomparablement supérieur à ceux des décennies précédentes. La documentation de l'époque témoigne de l'abaissement de leur prix. La cause en est l'apparition d'ateliers de copistes en ville, qui prennent le relais des rares scriptoria monastiques reproduisant auparavant des ouvrages pour un usage local » (Aurell, 2011).

Le latin possédant le monopole par lequel tout « savoir livresque » se transmettait, sa maîtrise s'avérait alors indispensable. L'acquisition de savoir séparait ainsi le lettré de l'illettré, le clerc du laïc. Ce constat, issu de pensées médiévales, est toutefois à nuancer selon Aurell (2011) du fait des variations sémantiques de ces termes (*clericus* et *litteratus* s'étant parfois confondus). Zumthor (1985), citant Gautier Map, ajoutait que *litteratus* et *scriba* n'étaient pas équivalents : « un bon scribe peut être illitteratus ; un litteratus ne recourt pas nécessairement à l'écriture dans l'accomplissement de ses tâches quotidiennes ». Zumthor (1985) rejetait ainsi l'opposition binaire entre *illiteratus* et *literatus*, proposant à la place une échelle de nuances sur laquelle se déplacent les individus.

Dans une vision similaire, Hillerich (1976) rejettait la dichotomie entre *literate* et *illiterate* qu'il jugeait totalement artificielle :

> « Il semble qu'à moins de construire notre propre définition originale, "la littératie" doit au moins prendre en compte la définition du dictionnaire, la "capacité de lire et d'écrire". Mais une telle définition soulève évidemment plus de questions qu'elle n'en résout, surtout si l'on s'intéresse à la mesure : lire et écrire dans quelle mesure ? Selon quelle qualité ? Mesuré selon quels instruments ou quelles procédures ? De plus, une telle déclaration implique l'opinion selon laquelle soit on "a" la capacité, soit on ne "l'a pas", et on est soit "alphabétisé", soit "analphabète". Une telle dichotomie artificielle est irréaliste quelle que soit la définition ; on ne peut pas supposer qu'il existe un moment magique à partir duquel un individu passe soudainement de "l'analphabétisme" à "l'alphabétisation". Toute définition de la littératie doit reconnaître cette qualité comme un continuum représentant tous les degrés de développement [19] ».

Il étendra son analyse à toute la littératie qu'il conceptualisera sous la forme d'un continuum (voir la figure 2.2) :

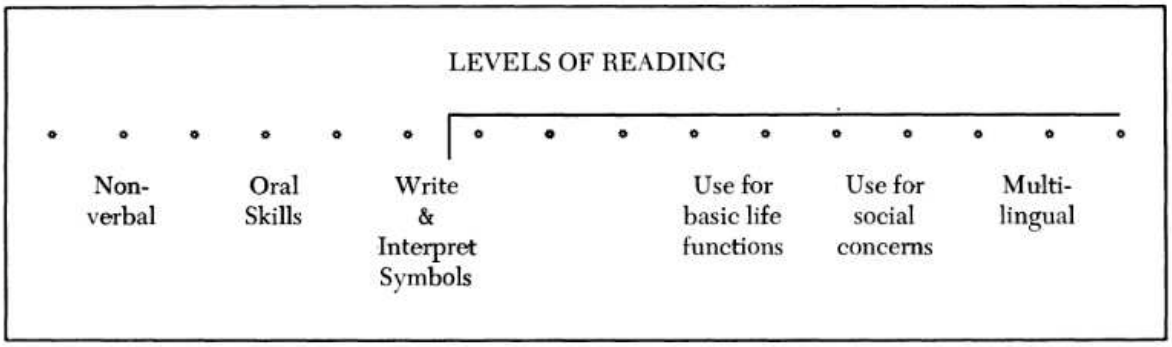

Figure 2.2.: Niveaux de lecture - extrait de (Hillerich, 1976)

Il s'agissait pour lui de souligner que la littératie n'était ni monolithique, ni unidimensionnelle mais plurielle, se déclinant selon plusieurs modalités et dont l'acquisition n'était pas instantanée. Plusieurs auteurs reprendrons cette conception. Le modèle de Pill & Harding (2013) liste cinq niveaux d'évolution que nous pouvons voir sur la figure 2.3 .

Dans une logique similaire, Kaestle (1985) déclarait que les catégories *literate* et *illiterate* n'étaient au final ni précises, ni mutuellement exclusives. Il y listait les (1) personnes qui ont appris à lire puis qui ont oublié ; (2) les personnes qui savent lire mais qui lisent rarement ; (3) celles qui se pensent

[19] « *It seems that unless one is to construe his own original definition, "literacy" must at least include the dictionary definition, the "ability to read and write." However, such a definition obviously raises more questions than it answers, especially if one is concerned with measurement: reading and writing to what extent? with what quality? as measured by what instruments or procedures? Further, implied in such a statement is the view that one either "has" the abilitly or does not "have it," that one is either "literate" or "illiterate." Such an artificial dichotomy is unrealistic in any definition; one cannot assume that there is some magic point at which an individual suddenly moves from "illiteracy" to "literacy." Any definition of literacy must recognize this quality as a continuum, representing all degrees of development* ».

Illiteracy:	Ignorance of language assessment concepts and methods
Nominal literacy:	Understanding that a specific term relates to assessment, but may indicate a misconception
Functional literacy:	Sound understanding of basic terms and concepts
Procedural and conceptual literacy:	Understanding central concepts of the field, and using knowledge in practice
Multidimensional literacy:	Knowledge extending beyond ordinary concepts including philosophical, historical and social dimensions of assessment

Figure 2.3.: Niveaux de littératie - extrait de (Pill & Harding, 2013)

literate mais perçoivent les autres comme *illiterate* ou encore celles qui malgré leur incapacité à lire participent toutefois à la culture de l'écrit en écoutant ceux qui savent lire. Ce dernier point est important : être illettré ne signifie pas être hors de la culture de l'écrit. Stock, cité par Zumthor, parlait même des illettrés comme rattachés à des communautés textuelles :

> « Selon la thèse, vigoureusement soutenue, de B. Stock, l' "illettré" souvent appartient à quelque "communauté textuelle", en vertu des références que l'existence le contraint à faire à tel ou tel écrit (même s'il n'y a pas directement accès), et de l'autorité qu'il lui reconnaît. Litteratus et illitteratus réfèrent ainsi moins à des individus pris dans leur totalité qu'à des niveaux de culture qui peuvent coexister (coexistent souvent) au sein d'un même groupe, voire dans le comportement et la mentalité du même individu. L'illettré - la part « illettrée » de moi, de toi, de cette société entière - maîtrise moins les mots et moins de mots ; mais il est proche d'eux, et ressent davantage leur pouvoir ; c'est pourquoi sans doute, comme la « lettre » même de l'Évangile y incline, il convient de faire triompher en soi l'illettrure pour assurer son salut : François d'Assise entendit "littéralement" cette métaphore, et rejeta la *curiositas* des livres » (Zumthor, 1985, pp. 5-6).

Seymour Papert, dans son ouvrage *The Children's Machine* de 1994, énonçait néanmoins le problème de l'apprentissage de la lecture et de l'écriture au sein de l'institution scolaire qui ne permettait pas aux élèves de devenir pleinement *literate*. Citant Paulo Freire, Papert rappelait qu'être *literate* signifiait plus que de savoir lire le mot : c'était savoir lire le monde, renvoyant à la notion de *Weltanschauung* présentée plus tôt. Il propose ainsi la notion de *letteracy* qui désigne la capacité permettant de lire les mots issus de l'alphabet [20]. Elle se distingue de la *literacy* en se présentant comme une simple interprétation littérale de l'écriture (ce que Schield (2002) nommait *functionnal illiteracy*) et se vou-

[20] « *De désespoir, j'ai inventé les termes* letteracy *et* letterate *pour faire référence aux capacités spéciales impliquées dans la lecture de mots construits à partir de lettres alphabétiques* », notre traduction de « *In desperation I have coined the words letteracy and letterate to refer to the special skill involved in reading words made up of alphabetical letters* » (p. 11).

lait une critique de l'enseignement scolaire perçu comme limitatif [21]. L'apprentissage de l'écriture elle-même était vue par Papert comme non nécessaire (bien qu'inévitable [22]) pour permettre aux enfants de faire l'expérience du monde grâce aux machines de savoirs (*knowledge machines*), ces dernières étant des métaphores d'un ensemble varié de médias (Papert, 1994).

Dans le cas de la *data literacy*, nous avons vu que les conceptions du *data literate* intègrent déjà un continuum, la dichotomie entre un non-lettré et un lettré étant un non sujet. Lire et écrire les données n'est pas un processus instantané et linéaire : il se déploie sur une longue temporalité et varie selon les publics et les institutions concernées.

2.2.1.5. Les évolutions de la littératie

Bien que la propagation de la littératie ait été une source de critiques, elle resta majoritairement soutenue. La question de l'application de la littératie va toutefois amener à une crise, amenant les auteurs à reconsidérer ses fonctions. Nous l'avons rappelé à plusieurs reprises dans nos travaux (Verdi, 2023b; Verdi & Le Deuff, 2020), la littératie est traversée par deux courants nommés par Street (2006) « autonome » et « idéologique ». Le courant autonome l'introduit comme une technologie de l'intellect monolithique, un ensemble de compétences neutres dont les effets cognitifs et sociaux seraient généralisables sans prise en compte des différents contextes. Perçue comme une vision purement occidentale (Collins & Blot, 2003; Goody & Watt, 1963), le courant autonome a été fortement critiqué (Lankshear & Knobel, 2006; Street, 2006). Resnick (1990), à la suite de Graff (1979), parlait d'une « crise de la littératie » : la pratique sociale n'étant pas pris en compte, une nouvelle vision passant d'un socle de compétences à un ensemble de pratiques culturelles était nécessaire.

[21] « Mais à cause d'un système scolaire archaïque, de nombreux enfants perdent le goût de l'exploration immédiate bien avant d'acquérir la maîtrise des compétences littéraires nécessaires pour trouver cette immédiateté prolongée. C'est là la véritable tragédie du décrochage scolaire. Qu'ils passent ou non le nombre d'années requis en classe, de nombreux enfants (voire la plupart) en sortent sans l'étincelle intellectuelle avec laquelle tous sont nés. », notre traduction de « *But thanks to an archaic school system, many children lose the taste for immediate exploration long before they acquire the mastery of book skills needed to find that extended immediacy. This is the real tragedy of dropping out. Whether or not they put in the requisite number of years in the classroom, many (or even most) children emerge without the intellectual spark with which all are born.* ».

[22] « Pour la prochaine génération ou les deux prochaines générations, il faut s'attendre à ce que la littératie intègre une part de *letteracy*, puisque le passé de notre culture est extrêmement lié à l'expression par l'écriture. Mais même si l'on attend d'une personne véritablement lettrée à venir qu'elle sache lire des livres et comprendre les grandes tendances de l'histoire de l'art ou de la philosophie, par tous les autres médias disponibles, il ne s'ensuit pas que l'apprentissage des lettres soit la pierre angulaire d'une éducation élémentaire », notre traduction de « *For the next generation or two one must expect literacy to include some letteracy, since our culture's past is so connected with expression through writing. But even if a truly literate person of the future will be expected to know how to read books as well as understand the major trends in art history or philosophy, via whatever other media become available, it will not follow that learning the letters should be the cornerstone of elementary education*. ».

> « La littératie n'est pas seulement une structure sociétale, et elle ne peut pas non plus être réduite à un ensemble de compétences mentales de base isolées de tout le reste ; et, comme nous le soutenons, elle ne peut pas non plus être captée par la logique exclusivement linguistique et sémiologique des signes écrits en tant que telle. Il s'agit plutôt d'un concept qui englobe les ressources culturelles d'une tradition lettrée – ce qui comprend le(s) système(s) d'écriture de cette tradition – et l'ensemble des capacités nécessaires pour exploiter ces ressources culturelles [23]. » (Brockmeier & Olson, 2002)

Pour Cohen (1958), il était nécessaire, en traitant de l'écriture, de « garder à l'esprit presque toutes les manifestations de la vie sociale ». Cela sera repris par Street dans son modèle idéologique : la littératie y est désignée comme une pratique sociale dépendante de ses contextes d'utilisation et non simplement comme « une compétence technique et neutre » (Street, 2006). Elle « englobe donc des catégories de situations, de pratiques, de relations sociales, de modalités d'interaction, d'institutions qui donnent sens aux usages de l'écrit pour les individus d'une société ou d'une culture donnée » (Nonnon, 2007). Warschauer (2012) résumait ainsi qu'il n'existait pas un mais plusieurs types de littératies dont le sens et la valeur variaient selon les contextes sociaux. Ceci rejoint la vision d'Olson pour qui il existe non une seule mais plusieurs cultures de l'écrit aux conséquences et implications différentes les uns des autres (Olson, 1998).

Par la prise en compte du contexte, de « nouvelles » littératies vont voir le jour, dépassant le cadre de l'écriture et de la lecture pour intégrer des concepts, des postures intellectuelles ou encore des techniques spécifiques. Une nécessité déjà soulignée par Kress (2003) pour qui il n'était plus possible de conceptualiser une littératie isolée de facteurs sociaux, économiques et technologiques. Bien que conceptualisées au début des années 2000 avec le tournant numérique, il est à noter qu'il existait déjà des littératies spécifiques comme l'*information literacy* centrée sur la maîtrise de l'information (énoncée par Zurkowsky (1974)). Lister l'ensemble des littératies est impossible en raison du fait que tout domaine, tout technologie fait l'objet d'une littératie spécifique.

Selon (Coiro, Knobel, Lankshear, & Leu, 2008), les nouvelles littératies ont plusieurs caractéristiques communes. Elles sont ainsi :

> « Centrales pour une participation civique, économique et personnelle dans une communauté mondialisée, déictiques au sens qu'elles évoluent au rythme des technologies de référence, multifacettes au sens qu'elles bénéficient d'une analyse rassemblant divers points de vue dans la discussion [24] ».

[23] « *Literacy is neither only a societal structure, and neither cannot it be reduced to a basic set of mental skills isolated from everything else; nor, as we argue, can it be captured by the exclusively linguistic-semiological logic of the written signs as such. Rather, it is a concept that embraces the cultural resources of a literate tradition – including the writing system(s) of this tradition – and the ensemble of the abilities necessary to exploit these cultural resources.* »

[24] Notre traduction, utilisée dans (Verdi, 2023b).

Ainsi, d'une littératie unique ont été conceptualisées tout un ensemble de littératies liées ou non à un domaine de recherche. Dans le cas de cette thèse, la *data literacy* est cette étude littéracique des données. Données dont nous allons maintenant aborder l'histoire.

2.2.2. Les données

La seconde influence est ainsi celle des visions sur les données. Leur histoire de ces dernières est protéiforme et a connu de grandes variations selon les disciplines de rattachement. Tout comme nous l'avons fait pour la littératie, nous allons résumer ici les caractéristiques des données pour faciliter la lecture.

2.2.2.1. Entre don et recherche

Les données ont été principalement l'objet de la recherche scientifique, en particulier dans le domaine mathématique où elles servaient à la résolution de problèmes (Corti, 2005; Hunt, 2005; Mokros & Wright, 2009; Schield, 2004b; Stephenson & Schifter Caravello, 2007). L'Encyclopédie de Diderot & Alambert (1751) le rapportait ainsi :

> « Données, adj. pris subst. *terme de Mathématique*, qui signifie certaines choses ou quantités, qu'on suppose être données ou connues, & dont on se sert pour en trouver d'autres qui sont inconnues, & que l'on cherche. Un problème ou une question renferme en général deux sortes de grandeurs, les *données* & les cherchées, *data* & *quæsita*. »

On trouve ici une distinction entre ce qui est donné (*data*) et ce qui est cherché (*quaesita*) ; Euclide employait lui le terme *dedomena* (*Δεδομένα*) pour qualifier les « informations » contenues dans les problèmes géométriques (Euclide & Peyrard, 1814). Cette dichotomie entre ce qui est donné et ce qui est recherché rappelle l'opposition conceptuelle grecque soulignée par Klein (2012) dans son commentaire sur le Ménon d'Aristote ; ce dernier y analysant l'opposition entre ce qui est recherché (ζητούμενον, *zêtoumenon*) et ce qui est donné, pris en compte, accepté (ὁμολογούμενόν, *homologoumenon*).

Le terme *data* représente le don tandis que *dătŭs* représente l'action de donner (Gaffiot, 1934, p. 440). En ce sens, les données donnent l'impression de matériaux bruts et neutres, de ressources naturelles facilement accessibles et prêtes à l'emploi. Or, il s'agit pour de nombreux auteurs d'un non-sens amenant à formuler des généralités sur les données, essentialisées en « la » donnée. Jensen l'énonçait

déjà en 1950 : « C'est un hasard malheureux de l'histoire que datum plutôt que captum soit venu à symboliser le phénomène basique en science. Car la science traite non pas de "ce qui a été donné par la nature au scientifique" mais de "ce qui a été capté ou sélectionné de la nature par le scientifique" en accord avec le but de celui-ci » (traduction d' Ibekwe (2017)). Pour Bachimont (2017), « le terme de "donné" est que la donnée ne permet pas de savoir de quoi elle est la donnée et par qui elle est donnée. Autonome vis-à-vis d'un qui et d'un quoi, la donnée est le pur arbitraire, le pur fait, une primeité percienne ». En somme, la donnée en tant que don dénué de cosmogonie pose un problème profond de réflexion et de contextualisation.

Plusieurs auteurs vont se pencher sur cette difficulté comme Meyriat (1985) qui rappellera que les données sont extraites par « une opération de l'esprit » et non « fournies spontanément » par les phénomènes étudiés. Pour dépasser cette problématique sémantique, Lanigan (1994), Latour (2007b) et Drucker (2011) vont proposer la notion de *capta* (Latour ajoutera celle équivalente de *sublata*). *Capta* vient du latin *căpiō, ōnis, capere* signifiant l'action ou faculté de prendre (Gaffiot, 1934, p. 268). Ce terme permet de parler des données comme des « obtenues », comme une construction humaine. Elles sont toujours créées indirectement par l'action humaine (Abner, 2020). Du fait que tout jeu de données résulte d'un ensemble de choix d'encodage (Owens, 2011), de production, de traitement et de transformations potentielles (Leonelli, 2019), les données sont donc dépendantes de leurs contexte de production.

> « Les données doivent d'abord être imaginées avant de pouvoir exister. En tant que telles, elles ne sont pas neutres, ni préexistantes, mais façonnent et sont façonnées par les politiques socioculturelles contestées dont elles sont issues [25] » (Dalton & Thatcher, 2014).

Parler de données brutes ou de données objectives est un oxymore (Abner, 2020; Iwasiński, 2020; Seaver, 2021). Elle ne sont pas des ressources naturelles mais des ressources culturelles (Gitelman, 2013), embarquant dès leur conception des réalités spécifiques.

Comme les jeux de données sont relatifs à une question de recherche, analysés en fonction d'un objectif ou d'un cadre d'analyse (Labelle, Gilliard, Népote, & Chignard, 2017), les données elles-mêmes ne sont pas définies selon « leurs propriétés intrinsèques mais selon leur fonction au sein de processus de recherche particuliers » (Rebouillat, 2021). Elles sont des inscriptions caractérisées par le fait d'être « destinées à être traitées », orientées vers l'avenir et contenant une promesse (Jeanneret, 2019). Des promesses découlant des points de vues et des buts originaux de leurs créateurs. Elles sont ces « formes de pouvoir » (Iliadis & Russo, 2016) dont le décodage accorde à son traducteur une encapacitation forte selon la portée du jeu de données considéré.

[25] « *Data must first be imagined before being called into existence. As such, data is not neutral, nor preexisting, but shape and are shaped by contested socio-cultural politics from which they come* ».

2.2.2.2. Les données et l'information

Un pouvoir informationnel obtenu par la contextualisation et l'apposition de sens sur les données. En effet, les données en elle-mêmes n'ont aucune signification (Ribeiro, 2021), elles ne sont qu'un matériau brut[26] au sens de non raffiné, non transformé pour générer de l'information (Silver et Silver, 1989, cité par Floridi (2011)).

Ackoff (1989) a proposé un cycle très souvent repris sous la forme d'une pyramide pour expliquer ce lien entre données et information (voir la figure 2.4) :

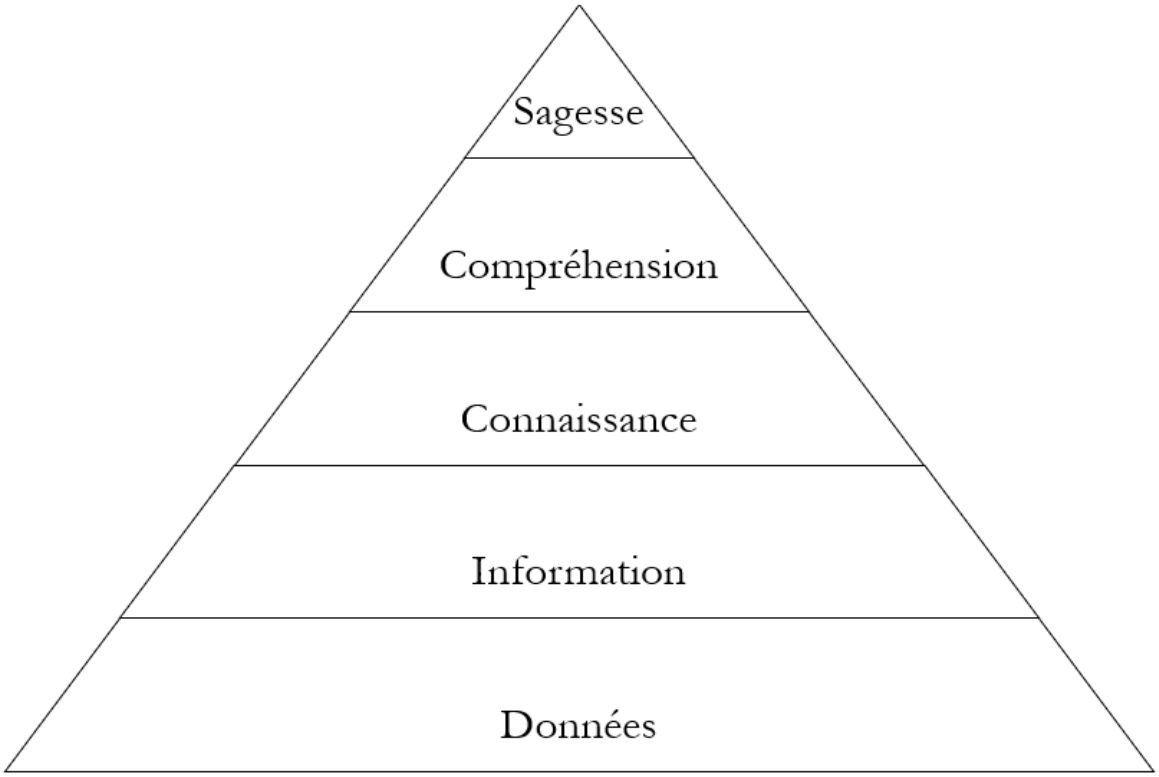

Figure 2.4.: Pyramide d'Ackoff reconstituée à partir d' (Ackoff, 1989)

Pour cet auteur, les données sont les produits de l'observation. Elles sont des symboles représentant les caractéristiques d'un évènement, d'un objet et de leurs environnements. Elles sont inutiles tant qu'elles n'ont pas été traitées. Les informations, contenues dans les descriptions, sont déduites des données. La connaissance permet de transformer les informations en instructions ; elle peut avoir été transmise ou acquise par l'expérience. La compréhension, ou l'intelligence (les deux sont citées en même temps) est la capacité d'améliorer l'efficacité (*efficiency*). La même chose est précisée pour

[26]Brut ne signifie pas ici vierge d'influences culturelles mais n'ayant pas subi de traitements pour obtenir une information.

la sagesse mais le terme employé pour efficacité est cette fois *effectiveness*. Cette difficulté terminologique de l'efficacité, présente dans les deux dernières définitions, est résolue par Ackoff qui prend pour exemple la différence entre la croissance et le développement. Le premier est quantitatif, le second qualitatif. Il rappelle également que croissance et développement ne sont pas liés, l'un pouvant exister sans l'autre.

Ce modèle a été maintes fois critiqué (Frické, 2009; Jennex & Bartczak, 2013), modifié et / ou agrémenté de nouveaux concepts comme le *big data* et l'*IOT* (Jennex, 2017). Tuomi (1999) proposait de renverser l'ordre de la pyramide tandis que d'autres auteurs ont plaidé pour une nouvelle représentation intégrant une relation horizontale rétroactive et non une relation verticale linéaire (Balloni, 2018; Van Meter, 2020).

Bates (2005) a proposé de son côté un système d'interaction selon deux types d'informations et deux types de données intégrés dans une logique de « modèles » (*patterns*) :

- **Information 1**: le modèle d'organisation de matière et d'énergie (ex : un livre posé sur une étagère) ;
- **Information 2**: quelques modèles de matière et d'énergie auxquels est donné un sens par un être vivant (ex : la lecture du livre posé sur l'étagère) ;
- **Connaissance** : une information à laquelle on appose un sens et que l'on intègre dans un ensemble de connaissances pré-existantes (ex : la combinaison et contextualisation des informations 2 obtenues précédemment) ;
- **Donnée 1** : la portion de l'environnement informationnel immédiatement accessible et traitée par un être vivant. Ce qui n'est pas accessible immédiatement car trop distant ou hors de la perception de l'organisme n'est pas une donnée 1 mais de l'information 1. C'est l'apposition d'un sens sur la donnée 1 qui permet d'obtenir une information 2. ;
- **Donnée 2** : l'information sélectionnée ou produite par les humains pour des finalités sociales (*social purposes*). Ainsi, au contraire de la donnée 1, la donnée 2 est purement humaine.

Dans l'ensemble des cas fournis par la littérature, les données sont perçues comme des représentations de faits ou de phénomènes dont elles énoncent les caractéristiques (Kitchin, 2014b). Selon Floridi (2009), une donnée est un fait supposé qui procède d'une différence ou d'un manque d'uniformité dans un contexte : elle est une ressource et une contrainte qui rend possible la construction de l'information. Il propose à ce propos une nouvelle décomposition de la donnée qu'il nomme *dia-*

phora, et plus particulièrement la *diaphora de re* qui représente le manque d'uniformité dans le monde réel et qui n'est pas interprétée épistémiquement.

La mise en relation des données et l'apposition d'un sens permettent d'obtenir des informations qui concourent à formaliser une représentation du réel. En ce sens, les données sont, au même titre que la littératie, des *Weltanschauungen*. Leonelli (2019) proposera une nouvelle vision des données vues comme relationnelles et incluses dans le processus de la recherche scientifique. Sous la forme d'un cycle (à l'inverse de visions pyramidales ou linéaires présentées précédemment), les données sont issues d'un traitement d'objets (la documentation des interactions entre les humains et le monde) qui, une fois ordonnées, permettront la création de modèles de représentations du monde. Ces derniers seront ensuite interprétés pour fournir des connaissances permettant à leur tour de mieux comprendre et décrire les interactions avec le monde et ainsi produire de nouveaux objets (voir la figure 2.5).

La difficulté de formuler une représentation précise du monde réside dans la complexité de l'interprétation (Drucker, 2011). En effet, une donnée ne donne pas une « description fidèle de la réalité extérieure prise sur elle-même, mais de son influence sur des instruments de mesure, ou plus généralement, sur des capteurs » : elle « jette un voile » sur cette dernière (Brochier, 2016, p. 48). Toujours selon Brochier, la propriété des données est de pouvoir être comparées à d'autres et ainsi « à devenir l'élément d'un ensemble qui lui-même peut être comparé à d'autres afin que, de comparaison en comparaison on découvre quelque corrélation inattendue » (p. 48).

Ce qui rend possible cette comparaison découle du fait qu'en tant qu'éléments de calcul, les données sont mesurables. Elles n'appartiennent pas à l'objet qu'elles référencent mais s'en détachent : elles sont, au même titre que la trace[27], une survivance de l'évènement (Bachimont & Verlaet, 2021). Elles sont ces « preuves présumées » (Buckland, 1991) permettant de résoudre dans un avenir proche nos problèmes (les données étant toujours employées pour répondre à un besoin). Or, se posent ici les questions d'authenticité et de véracité des données du fait des biais inhérents à leur production. Rosenberg (2013) rappelait ainsi que une donnée fausse, au contraire d'un fait, reste malgré tout une donnée : les notions de véracité ou d'authenticité ne sont pas contenues dans les données, c'est la question posée et leur traitement qui apposeront ce statut *a fiortiori*.

[27]Pour éviter tout malentendu, nous tenons à bien souligner que, toujours selon Bachimont et Verlaet (2021), les données, si elles expriment certaines caractéristiques d'une trace, n'en sont pas une. Elles sont au contraire des constructions pouvant interroger les traces dont elles sont une « contrepartie technique conçues pour permettre leur traitement et exploitation ».

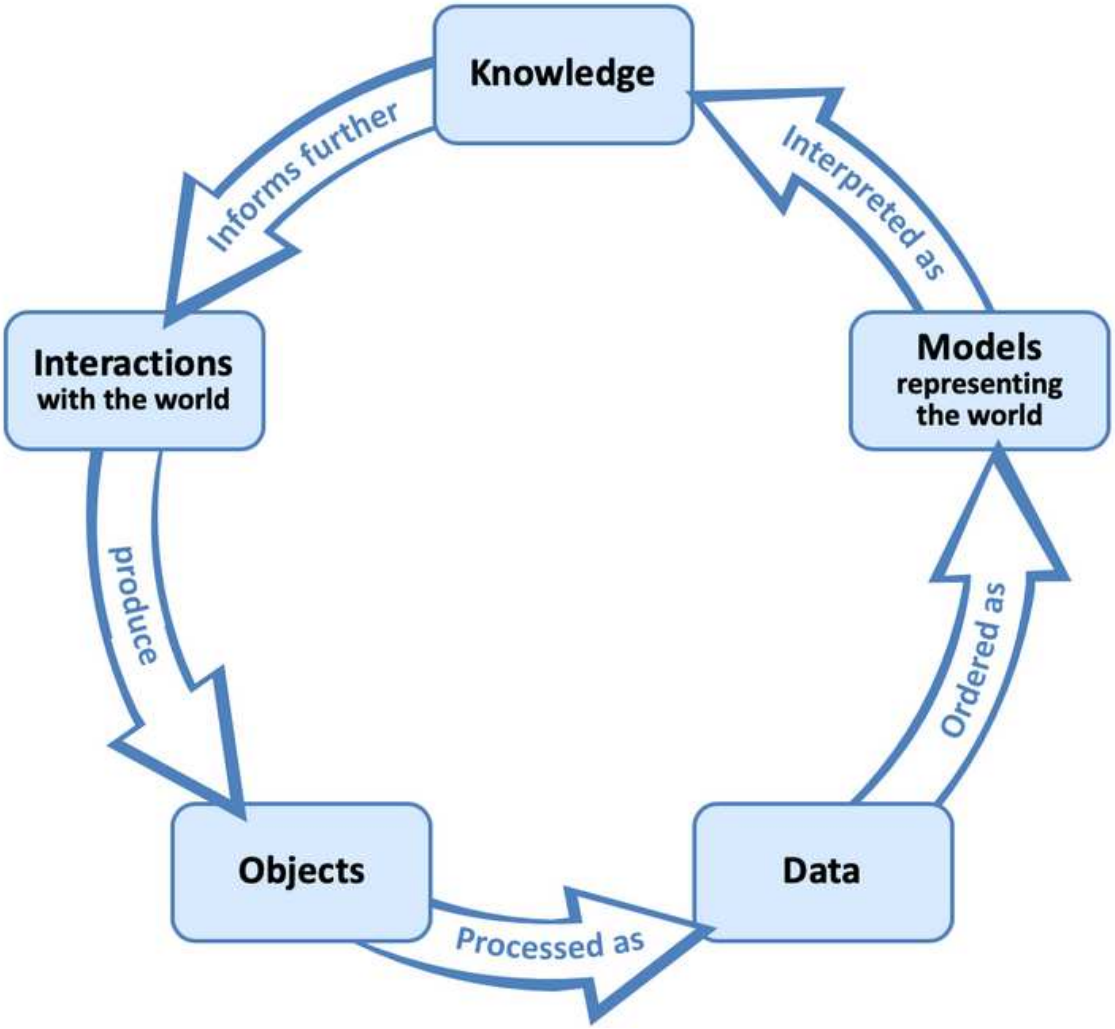

Figure 2.5.: Procédure d'enquête selon une vision relationnelle des données - extrait de (Leonelli, 2019)

2.2.2.3. Les données et les statistiques

Ainsi, les données permettent de représenter et d'interroger le monde pour répondre aux problèmes que l'on se pose. En ce sens, les données étaient d'abord l'objet d'étude des scientifiques, notamment dans le domaine mathématique, et se présentaient généralement comme une suite de nombres.

Elles étaient vues comme essentiellement primaires (non transformées) et individuelles (ne concernant qu'un seul individu/entité) (Droesbeke, Vermandele, & Didier, 2018). L'intervention de la statistique, cet ensemble de méthodes ayant pour objet la collecte, le traitement et l'interprétation de données d'observation relatives à un groupe d'individus ou d'unités[28], va offrir de nouvelles visions et développer de nouveaux enjeux autour et par les données.

Rey (2016) nous rappelle ici l'étymologie du terme statistique et de ses dérivés : *statistica* est un dérivé de *statista* qui désigne un « homme d'État ». Ce terme est lui-même un dérivé de *status* qui signifie l'état au sens de position, situation et forme de gouvernement. Desrosières (2010) souligne que ce terme implique la mise en place de « formes générales, de classes d'équivalence, de nomenclatures, transcendant les singularités des situations individuelles » à travers soit des catégories du droit (versant juridique), soit de normes et de standards (versant économie de gestion) (introduction, paragr. 19).

Ces nomenclatures, basées sur les données, sont encodées : elles dépendent de conventions qui portent sur la définition de l'objet et de son encodage, et reflètent donc une réalité. La question de l'objectivité doit laisser ici place à celle de l'objectivation (Desrosières, 2010, introduction, paragr. 29) ; la consistance des objets n'étant pas donnée par avance (chapitre 1, paragr. 3). Bien que s'appuyant sur les nombres, la statistique n'est pas à l'origine une méthode mathématique :

> « De plus, considérer la statistique comme une méthode mathématique à laquelle, dans un second temps, on aurait fait appel dans le champ politique, économique et social est un non-sens, puisque c'est au contraire dans ce champ qu'elle a pris naissance et s'est déployée. Ses premières utilisations dans les sciences de la nature ne sont apparues qu'ensuite, au cours de la seconde moitié du XIXe siècle, et sa constitution en discipline mathématique a été plus tardive encore, dans les premières décennies du XXe siècle » (Rey, 2016, pp. 17-18).

Si l'on retrouve dès l'Antiquité une volonté par les États d'employer des systèmes de mesures, à l'instar du Nilomètre qui permettait de déterminer le niveau de crue du Nil pour pouvoir fixer le niveau des impôts (Benzécri, 1976), la statistique actuelle est la réunion de plusieurs méthodologies autrefois séparées. Elle hérite ainsi de la statistique allemande dont elle tire le nom (*statistik*, terme communé-

[28]Dictionnaire Larousse, 2023. https://www.larousse.fr/dictionnaires/francais/statistique/74516

ment attribué à Achenwall) qui percevait l'ensemble des communautés humaines « comme un tout, dotée d'une puissance singulière » (Desrosières, 2010, Chapitre 1, paragr. 9). La formalisation de la statistique allemande par Conring (1606-1681) au XVIIe siècle a pour but de classer les savoirs hétéroclites (paragr. 11), de proposer un cadre de mesure permettant à l'État de connaître son territoire et le moderniser.

Elle hérite également de l'arithmétie politique anglaise. Basée sur les travaux de Graunt (1620-1674) puis théorisée par Petty (1623-1687) et Davenant (1656-1714), elle s'inscrit dans la tenue, le dépouillement et l'interprétation des registres en « nombres, poids et mesures » (Desrosières, 2010, Chapitre 1, paragr. 17). La permanence des écrits permet de délimiter des « unités définies, identifiées et stables », offrant une certaine certitude des contenus. L'arithmétie politique anglaise se distingue de la statistique allemande du fait que ce ne sont « pas des universitaires théoriciens qui édifient une description globale et logique de l'État en général, mais des gens d'origines diverses qui ont forgé des savoirs pratiques dans leurs activités et qui les proposent au "gouvernement". » (chapitre 1, paragr. 20). Cette méthode perdurera au cours des siècles et plusieurs enquêtes statistiques paraîtront dont celle de Chadwick en 1842 sur les conditions d'hygiène de la population ouvrière.

La France, au contraire du Royaume-Uni, est un pays centralisé doté d'un État puissant. La nécessité de connaître le territoire a été central dans l'élaboration d'une statistique française. Jean Bodin (1529 ou 1530-1596) prônant l'absolutisme monarchique a souhaité rendre efficace l'État monarchique par une connaissance des territoires. Suivant cette recommandation, les enquêtes de Colbert (1664), Beauvillier (1667), de Vauban (1685), de Boulainvilliers (1697) ou celles du Régent (1716-1718) concoureront à mieux les connaître (Gille, 1980; Rey, 2016). Cela se poursuivra de manière régulière sous le premier Empire et le rétablissement monarchique, la période révolutionnaire n'ayant pu faire aboutir les projets de recensement qu'elle s'était donnés. Le bureau de la statistique va être créé en 1800. Peuchet, co-responsable du bureau, publiera la « Statistique élémentaire de la France » en 1805. Après sa fermeture en 1811, le bureau réapparaîtra sous une nouvelle forme, celle de la société de la statistique de Paris, fondée par Adolphe Tiers en 1830 qui produira plusieurs ouvrages tels *Documents statistiques sur la France* (1835) et *Territoire et population* (1837) (Gille, 1980, Chapitres 3, 156). C'est à partir des années 1830, sous l'influence d'Adolphe Quetelet (1796-1874), que la plupart des États européens vont se doter d'un service central de statistiques (Desrosières, 2000; Nora, 1986), consacrant cette dernière comme élément primordial de la politique publique.

Ainsi la statistique encadre la politique de l'État, permettant au souverain de connaître son pays et d'y appliquer des réformes basées sur ces données :

> « Dans cette entreprise de déchiffrement, la statistique se présentait comme un instrument de choix - adapté à la prise en compte d'une réalité atomisée, et recommandé par son allure objective. En particulier, les nombres paraissaient seuls à même d'informer correctement sur une complexité sociale qui mettait en déroute les moyens de descrip-

> tion antérieurs. De plus, la statistique, pratiquée de manière régulière, offrait des séries qui permettaient de mesurer les évolutions, d'apprécier les changements » (Rey, 2016, p. 104).

Cette mise en nombre du monde social va se poursuivre tout au long du XIXe et XXe siècle pour prendre de nouvelles formes : la statistique ne disparaît pas pour autant, elle se reconfigure avec les méthodes et technologies de référence.

La statistique américaine va notamment connaître un tournant au cours des années 1960 avec l'intervention de John Tukey. Tukey était un statisticien américain (1915-2000), créateur de méthodes statistiques dont plusieurs tests portent aujourd'hui son nom, notamment connu pour son article de 1962 *The future of data analysis*[29] dans lequel il développe une analyse de données qui se détache de la vision purement mathématique. En effet, pour Tukey la statistique mathématique atteint ses limites quand il s'agit de traiter certains types de données complexes comme les *spotty data*, les *multiple-response data*, les données hétérogènes dans leur précision ou encore les données incomplètes.

Il propose une nouvelle grille de lecture afin de traiter ces problèmes. Parmi ses recommandations, la plus connue est la onzième qui souligne la nécessité d'accepter l'incertitude des données, une chose normalement rejetée par les statisticiens : « une réponse approximative à une bonne question, qui est souvent vague, est de loin préférable à une réponse exacte à une mauvaise question, qui elle peut toujours être précise [30] » (Tukey, 1962, pp. 13-14).

Il développe cette idée plus loin en précisant que l'analyse de données doit progresser à l'aide de réponses approximatives du fait que la connaissance du problème est également approximative. De là va découler son analyse exploratoire de données qu'il développera grandement dans son ouvrage *Exploratory Data Analysis* (Tukey, 1977). On y trouve notamment les fameuses « boites à moustaches » aussi nommées boites de Tukey, permettant de représenter des indicateurs de positions telles les médianes, les quartiles, les déciles, etc.

2.2.2.4. Les représentations de données

Cette représentation de données suit une logique déjà entamée par le cartographe Jacques Bertin (1918-2010) avec la parution de son ouvrage *La sémiologie graphique* (Bertin, 1967). A l'intérieur y sont posées les bases de la représentation graphique comme un ensemble de techniques et de méthodes

[29] Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, *33*(1), 1-67.

[30] « *Far better an* approximate answer *to the right question, which is often vague, than an* exact answer *to the wrong question, which can always be made precise* ».

visant à représenter graphiquement les informations en fonction de codes (esthétiques, perceptifs) et de conventions (culturelles, symboliques). Elle se base sur la sémiologie, c'est-à-dire « la science qui étudie la vie des signes au sein de la vie sociale » selon le mot de Ferdinand de Saussure (De Saussure, 1971, p. 33).

En effet, la sémiologie graphique est pour Bertin un système monosémique au sens que la connaissance de chaque signe précède l'observation de l'assemblage desdits signes (p. 6). Ce système intègre « la graphique », cet ensemble de moyens permettant de représenter l'information qui couvre les réseaux, les cartes et les diagrammes (p. 6). Bien que souhaitant s'en démarquer, on retrouve ici les fondements de la théorie de l'information de Shannon (1948), reprise dans Shannon et Weaver (1949). On part d'une source informationnelle qui transmet un message à un transmetteur. Ce dernier transforme ce message en signal qui est transmis à un récepteur à travers un canal de communication (Shannon cite l'exemple du téléphone comme transmetteur et du câble comme canal). Le récepteur retranscrit le signal en message et le transmet au destinataire. L'information est quantifiable, elle est une mesure du degré de liberté intervenant avant la sélection d'un message. Elle est cet ensemble de signes et de symboles que l'on manipule et décode. Les données sont donc quantifiables, encodables et l'informatique va amplifier cela.

2.2.2.5. L'informatique

En effet, parallèlement aux méthodologies précédemment présentées, se développe l'informatique au cours du XXe siècle qui incorpore l'ensemble des procédures mathématiques et statistiques de traitement de données, l'ordinateur étant cette machine de calculs (*computer*), mais également cette machine d'enregistrement et de représentation des données.

> « En effet, grâce à l'ordinateur, le traitement des données se développe prodigieusement. On sait que "comprendre" c'est réduire la multitude des "données" qui nous assaillent au petit nombre "d'informations" que nous sommes susceptibles de prendre en compte autour d'un problème donné » (Bertin, 1967, préface).

La volonté d'une interprétation mécanique des données est ancienne, originellement intégrée dans une logique de calculs. Pensons ici à la Pascaline, du nom de son inventeur Blaise Pascal (1623-1662), permettant de réaliser un certain nombre d'opérations primaires comme la soustraction, la multiplication et la division. D'autres techniques vont se succéder, en particulier celles impliquant les cartes perforées intégrées dans la machine à tisser de Jacquard (1752-1834)[31] et qui sera reprise par Charles Babbage (1791-1871) pour sa machine à différence (*difference machine*) ; cette dernière per-

[31] L'invention n'est pas de Jacquard mais de Basile Bouchon et sera améliorée par son disciple Jean-Baptiste Falcon, puis par Jacques de Vaucanson. Jacquard, à leur suite, reprendra cette technologie pour sa machine à tisser.

mettra l'exécution du tout premier algorithme développé par Ada Byron, plus connue sous son nom d'épouse Ada Lovelace (1815-1852). Le statisticien Herman Hollerith (1860-1829) déposera par la suite deux brevets de machines statistiques. Le plus connu est le brevet n°397783 intitulé *Apparatus for compiling statistics*. La machine qui en a été issue fut utilisée lors du recensement américain de 1890 : les données décrivant les caractéristiques d'un objet ou d'une personne étaient inscrites sur une carte perforée. Une fois intégrée à la machine, la carte était disposée devant des tiges reliées à un compteur qui générait un contact électrique permettant de comptabiliser les données (Blodgett & Schultz, 1969). Cette technologie sera affinée au sein de la société montée par Hollerith, la *Tabulating Machine Co.* qui deviendra l'*International Business Machine Corporation* (IBM) suite à une fusion avec la *Computing Scale Company* en 1911. En résulte la carte IBM brevetée en 1928 et produite en série et intégrée dans les premiers ordinateurs *Colossus* (1943) et *ENIAC* (*Electronic Numerical Integrator And Computer*, 1946).

Les données peuvent donc être encodées et analysées mécaniquement. Au-delà des progrès technologiques qui la développe, l'informatique va se constituer en science (*computing science* ou *computer science*) et proposer des curricula, notamment du fait du manque de personnels spécialisés dès les années 1960 (Martin & Hall, 1960). On retrouve cela au sein de l'*Association for Computing Machinery*, une association internationale fondée aux États-Unis en 1947, qui existe toujours, ayant pour but l'avancée de la science, le développement, la construction et l'application des nouvelles machines pour le calcul, le raisonnement et tout autre traitement de l'information[32].

Est créé ensuite le sous-comité *Currulum Commitee on Computer Science* en 1962 qui propose en 1965 *An undergraduate program in computer science—Preliminary recommendations*[33] puis l'*ACM Curriculum Committee on Computer Science*[34] en 1968. Ces documents marquent un tournant dans l'édification de la discipline en proposant des programmes pour les universités : celui de 1965 pour des niveaux licence et celui de 1968 pour des niveaux master. Une version dédiée à l'enseignement secondaire sera proposée ultérieurement en 1978. Ces publications auront pour objectif d'harmoniser l'informatique comme discipline, alors développée de manière hétérogène selon les lieux de rattachement :

> « L'informatique ne se préoccupe pas uniquement de la conception des ordinateurs, et n'est pas non plus juste l'art du calcul numérique, malgré l'importance que revêtent ces deux sujets. L'informatique recouvre un grand nombre de spécialités, toutes se développant à une rapidité extrême. Il est naturel que l'on associe au sujet une petite partie à laquelle il a été exposé. Même parmi ceux qui se spécialisent en informatique, peu

[32]« *This organization would be to advance the science, development, construction, and application of the new machinery for computing, reasoning, and other handling of information*». https://www.acm.org/about-acm/acm-history
[33]Voir (Conte et al., 1965).
[34]Voir (Atchison et al., 1968).

(voire aucun) sont capables de suivre le flot d'innovations dans ce domaine. [35] » (Conte et al., 1965, p. 544)

L'informatique est alors décrite comme une science de l'information :

« L'informatique s'intéresse à l'information dans de nombreux domaines de la même manière que la physique s'intéresse à l'énergie ; elle se consacre à la représentation, au stockage, à la manipulation et à la présentation de l'information dans un environnement autorisant des systèmes d'information automatiques. Tout comme la physique utilise des dispositifs de transformation d'énergie, l'informatique utilise des dispositifs de transformation d'informations. Certaines formes d'informations ont été plus ou moins étudiées en profondeur et sont mieux comprises que d'autres ; néanmoins, toutes les formes d'informations - numériques, alphabétiques, picturales, verbales, tactiles, olfactives, résultats de mesures expérimentales, etc. - présentent un intérêt en informatique. [36] » (Conte et al., 1965) .

Le curriculum de 1965 ajoute que les mathématiques sont distinctes de l'informatique malgré le fait que la discipline s'intéresse elle aussi à l'information. La distinction se fait ainsi :

« Le mathématicien souhaite découvrir la relation syntaxique entre les éléments sur la base d'un ensemble d'axiomes qui peuvent n'avoir aucune réalité physique. L'informaticien souhaite découvrir les moyens pragmatiques par lesquels l'information peut être transformée pour modéliser et analyser les transformations de l'information dans le monde réel. L'aspect pragmatique de cet intérêt amène à rechercher des moyens efficaces de représenter l'information, des algorithmes efficaces pour transformer l'information, des langages efficaces pour exprimer des algorithmes, des moyens efficaces pour surveiller le processus et pour afficher l'information trans-

[35]« *Computer science is not simply concerned with the design of computing devices–nor is it just the art of numerical calculation, as important as these topics are. Computer science encompasses many specialized areas, all developing with extreme rapidity. It is natural that one should associate some small part, to which he has been exposed, with the subject as a whole. Even among those specializing in computer science, few (if any) are able to keep pace with the flood of innovation throughout the field* ».

[36]« *Computer science is concerned with information in much the same sense that physics is concerned with energy; it is devoted to the representation, storage, manipulation and presentation of information in an enviromnent permitting automatic information systems. As physics uses energy transforming devices, computer science uses information transforming devices. Some forms of information have been more thoroughly studied and are better understood than others; nevertheless, all forms of information–numeric, alphabetic, pictorial, verbal, tactile, olefactory, results of experimental measurement, etc.–are of interest in computer science*. ».

> formée, et des moyens efficaces pour les accomplir à un coût raisonnable [37] » (Conte et al., 1965) .

Dans le curriculum de 1965, les données ne sont vues que sous l'angle du traitement des bases de données. Il faut attendre le curriculum de 1968 pour voir apparaitre les traitements de données (*data processings*). Dans cette nouvelle version, il est spécifié que l'informatique traite trois grands domaines : les processus et les structures de l'information, les systèmes de traitement de l'information et les méthodologies. Dans le premier domaine, qui traite les représentations et les transformations des structures de l'information ainsi que les modèles théoriques, sont citées les structures de données (*data structures*) qui intègrent :

- la description, la représentation et la manipulation de nombres, tableaux, listes, arbres, fichiers, etc. ;
- l'organisation, l'allocation et l'accès au stockage ;
- l'énumération, la recherche et le tri ;
- les techniques de production, de modification, de transformation et de suppression ;
- les propriétés statiques et dynamiques des structures ;
- les algorithmes pour la manipulation d'ensembles, de graphiques et d'autres structures combinatoires.

Ainsi données et algorithmes sont perçues en système, les premières n'étant pas séparées des seconds. Les données peuvent recouvrir tous types de formes et doivent subir un grand nombre de traitements pour proposer une information précise. C'est un point important qui montre que la séparation donnée / algorithme n'est pas une vision ancienne mais bien récente dont les discours actuels se détachent de plus en plus.

La partie « méthodologies » intègre quant à elle les traitements de données compris comme les techniques applicables aux systèmes d'information. Si les mathématiques et les statistiques sont mentionnées comme des domaines distincts, ils sont toutefois intégrés dans les cursus. L'informatique usant malgré tout de leurs méthodes pour la programmation et la recherche.

En parallèle des programmes, les États-Unis norment leur vocabulaire. Ont été produits dans cette conjoncture deux documents : le *Proposed American Standard Vocabulary of Information Proces-*

[37] « *The mathematician is interested in discovering the syntactic relation between elements based on a set of axioms which may have no physical reality. The computer scientist is interested in discovering the pragmatic means by which information can be transformed to model and analyze the information transformations in the real world. The pragmatic aspect of this interest leads to inquiry into effective ways to represent information, effective algorithms to transform information, effective languages with which to express algorithms, effective means to monitor the process and to display the transformed information, and effective ways to accomplish these at reasonable cost* ».

sing[38] et l' *IFIP/ICC Vocabulary of Terms Used in Information Processing*[39] en 1966. Les deux se sont attelés à proposer un vocabulaire normé réservé aux experts informatiques, pour un total de 1200 définitions (Traub, 1966).

Dans le second document, les données sont « une représentation de faits ou d'idées formalisée de telle manière qu'elle peut être communiquée ou manipulée selon certains procédés »[40] (p. 3). L'information y est désignée comme « le sens qu'un humain assigne à une donnée par le truchement de conventions établies et employées pour sa représentation ».[41] (p. 3).

2.2.2.6. La donnée comme atout

De manière globale, les données sont surtout considérées au sein de la base de données tout au long des années 1960-1990. Et cette vision va dépasser le cadre purement universitaire pour intégrer celui des entreprises qui chercheront à exploiter les données et les verront dès cette époque comme un atout commercial.

Comme le rappellent Kitchin (2014b) ainsi que Chignard & Benyayer (2015), jusque dans les années 1980, les données sont un « produit de luxe » cher à collecter et « encore davantage à stocker et analyser ». En effet, les capacités techniques de l'époque confinaient à la centralisation, amenant au règne du *data administrator* en charge de la *data administration*[42], cette fonction se concentrant sur la planification et l'organisation conceptuelles des ressources de données concourant à la maximisation de leur valeur, de leur qualité et de leur opératibilité (Newton & Wahl, 1993) :

> « Au milieu des années 1960, l'augmentation des volumes de données dans les systèmes de traitements de données a suscité un intérêt pour le développement de moyens plus efficaces de stockage et de récupération de ces données. Dans les années 1970, cette augmentation est devenue une explosion et le concept de système de gestion de bases de données (SGBD) est devenu populaire. Au cours de la première moitié des années 1980, l'utilisation des SGBD s'est développée dans le cadre d'une augmentation naturelle de la sophistication de l'environnement de traitement des données. Face à cette sophistication croissante, on se rend compte que l'utilisation d'un SGBD ne constitue qu'un

[38] Produit par l'*American Standard Association's Subcommittee X3.5*.

[39] Produit par l'*International Federation for Information Processing* (IFIP) et *the International Computation Centre* (ICC). Voir (Douglas, 1966).

[40] « *A representation of facts or ideas in a formalized manner capable of being communicated or manipulated by some process* ».

[41] « *The meaning that a human assigns to data by means of the known conventions used in its representation* ».

[42] La *data administration* peut être considérée comme un ancêtre de la gouvernance des données que nous aborderons en détail au cours du chapitre 3.

> aspect de la gestion des données. À une échelle beaucoup plus large, la gestion des données implique aujourd'hui de traiter les données d'une organisation comme une véritable ressource d'entreprise qui doit être gérée comme telle. Le terme associé à cette fonction est "administration des données" [43] » (Gillenson, 1985).

Dès les années 1970 donc, les données sont perçues comme un bien central pour les entreprises comme nous pouvons le lire dans l'ouvrage *Managing the Data Resource Function* (Nolan, 1974). Il y était précisé qu'elles étaient une « ressource organisationnelle précieuse » devant être gérées au même titre que les autres ressources d'une entreprise, à savoir les ressources humaines, les ressources financières et les ressources matérielles. Des ressources qui, jusqu'au cours des années 1990, n'avaient qu'une valeur de transaction (Al-Ruithe, Benkhelifa, & Hameed, 2019) ; les seules données recherchées (et techniquement recherchables) étant les données commerciales. De manière plus globale, peu d'entreprises à cette époque avaient les compétences matérielles et humaines de traiter les données, et en particulier d'assurer leur qualité (Redman, 1995). Levitin & Redman (1998) appelait en ce sens à revoir les stratégies des entreprises autour des données :

> « Les données sont créées et utilisées en quantités énormes dans des activités allant de l'opération la plus banale à la décision stratégique la plus ambitieuse. Elles sont rarement bien gérées. Pour améliorer cette situation, les entreprises et les managers doivent comprendre les propriétés des données et les gérer comme des ressources. Comme nous l'avons vu, les données, contrairement à de nombreuses ressources conventionnelles, sont intangibles, faciles à copier et à transporter, et renouvelables. Les entreprises doivent penser et gérer les données différemment. Les solutions décrites ici peuvent aider une entreprise à mieux tirer parti de cette ressource traditionnellement sous-utilisée [44] ' ».

Il faudra attendre le tournant des années 2000 pour que la situation change complètement.

---

[43] « *The mid-1960s, the increasing volumes of data in data processing systems sparked an interest in the development of more efficient ways to store and retrieve that data. By the 1970s those increases became explosions, and the concept of the database management system (DBMS) came into popular use. During the first half of the 1980s the use of DBMSs expanded as part of a natural increase in the sophistication of the data processing environment. One realization in that increasing sophistication is that the use of a DBMS is only one aspect of the management of data. On a much broader scale the management of data today implies treating an organization's data as a true corporate resource which must be managed as such. The term associated with that function is "data administration."* ».

[44] « *Data are created and used in enormous quantities in activities ranging from the most mundane operation to the most far-reaching strategic decision. They are rarely managed well. To improve this situation, enterprises and managers must understand the properties of data and manage them as resources. As we have seen, data, unlike many conventional resources, are intangible, easy to copy and transport, and renewable. Enterprises must think about and manage data differently. The prescriptions described here can help a company begin to take better advantage of this traditionally underutilized resource* ».

2.2.2.7. Les mégadonnées et le tournant numérique

Le début des années 2000 marque un « tournant computationnel » (Berry, 2011) où les technologies évoluent et se propagent à l'ensemble de la société, en particulier avec l'invention du smartphone en 2006. On parle alors d'une « explosion de données », d'un véritable déluge[45] annonçant « la fin de la théorie »[46], cette supposée obsolescence de la méthode scientifique permettant de traiter les données. Ces volumes exponentiels, encore à l'heure actuelle en pleine croissance, ont été intégrés sous le vocable de *Big data*[47]. Si les problématiques liées aux grands jeux de données sont biens connus et documentés depuis les années 1960, notamment citées dans l'expérimentation de Levien & Maron (1967), elles changent ici de formes.

« Le » *big data*, francisé en mégadonnées, désigne au premier abord un ensemble de pratiques incluant la collecte, le traitement et l'analyse de grands jeux de données (Schäfer & Es, 2017). Les caractéristiques souvent associées à ces jeux de données sont résumées dans les trois, parfois cinq, V : le volume, la vélocité et la variété (Delort, 2018). Analysé plus en profondeur, « le » *big data* désigne les pratiques, les infrastructures, les réseaux et l'ensemble des politiques qui influencent ses manifestations (Richterich, 2018). Il s'agit pour Boyd & Crawford (2012) d'un phénomène socio-technique, qui modifie la texture même de l'information en corrigeant la manière dont celle-ci est présentée, articulée et visualisée (Papacharissi, 2015).

La vision associée à ce nouvel état de fait est celle d'une prétendue supériorité du quantitatif sur le qualitatif : le *big data* « faisant valoir que le premier offre la meilleure opportunité de généralisations significatives et que, lorsque cela est nécessaire, des états qualitatifs peuvent être rendus qualitativement [par la suite] »[48] (Mosco, 2014). Ce faisant, le *big data* revêt des aspects prédictifs (Mayer-Schönberger & Cukier, 2013), devenant le nouveau défi à relever pour les organisations qui cèdent alors à la mode du positivisme numérique (Mosco, 2014, 2016; Richterich, 2018). Les données sont traitées avec révérence et deviennent le nouvel horizon à atteindre, le nouvel impératif des organisations (Fourcade & Healy, 2016) : Clive Humby en parlera en 2006 comme du nouveau pétrole (*new oil*) qu'il faut raffiner pour en extraire la richesse, une métaphore qui aura un succès retentissant, reprise d'année en année[49]. Quelques variations au pétrole apparaîtront tels le nucléaire[50], le rayon-

[45]Voir (The Economist, 2010).
[46]Voir (Anderson, 2008).
[47]Le terme est supposément attribué à Francis Diebold dans son article de 2000 « "Big Data" Dynamic Factor Models for Macroeconomic Measurement and Forecasting » mais plusieurs auteurs comme Ibekwe (2017) et Theviot (2023) listent d'autres sources plus anciennes.
[48]« *Big data gives priority to quantitative over qualitative data, arguing that the former provides the best opportunity for meaningful generalizations and that, when necessary, qualitative states can be rendered qualitatively*».
[49]Un recensement a été fait ici : https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294
[50]Voir (Bridle, 2018).

nement solaire[51] et finiront par dépasser le simple cadre du *big data* pour désigner l'ensemble des données.

Les données elles-mêmes ne se limitent plus aux ensembles de nombres, tableurs et autres formats ayant eu cours au XXe siècle. Avec leur passage au numérique, elles englobent désormais tous types de formats : des images, des vidéos, des audios ainsi que les gens eux-même au sens de la captation et transcriptions de leurs faits et gestes (Ribeiro, 2021), ce que Mayer-Schönberger & Cukier (2013) nomment « datafication »[52]. Un mythe se mêle alors à l'enthousiasme ambiant : celui de rendre quantifiable et prévisible le monde réel grâce aux mégadonnées, ce que Crawford (2013) nommera le « fondamentalisme du Big data » (*big data fundamentalism*).

Cet enthousiasme sera toutefois en partie freiné par les nombreuses affaires autour des reventes illégales de données personnelles, dont l'affaire *Cambridge Analytica* est le parfait exemple, et la dénonciation par les lanceurs d'alerte comme Edward Snowden des techniques de surveillance de masse. Cette prise de conscience va amener à la formulation de nombreuses critiques sur les vertus supposées des données et sur les logiques qui les accompagnent :

> « On ne retient trop souvent de la révolution en cours que l'explosion des volumes des données produites. Pourtant, la diversité des sources et des données produites est probablement un phénomène encore plus marquant [...] Les grands acteurs du numérique, pour leur part, collectent des données de manière non scientifiquement contrôlée, grâce à des capteurs, au recueil de traces d'utilisation, ou encore à la contribution des internautes. Elles n'ont ni la robustesse, ni la complétude des données scientifiques. Elles créent cependant progressivement, de par leur seul volume, une forme d'empreinte du réel qui peut à son tour être interprétée et être utilisée produire un savoir activable » (Verdier, 2015b).

Ces critiques vont porter sur leur centralisation aux mains des GAFAM (cet acronyme désigne les plus grandes multinationales, à savoir Google, Amazon, Facebook, Apple et Microsoft) (Cardon, 2015; Markoff, 2012), sur l'absence d'objectivité des jeux de données (renvoyant ici au mythe des données comme don et non construction) (Kitchin, 2014b) qui parleraient d'eux-mêmes et permettraient des analyses prédictives infaillibles (Silver, 2013), sur les interactions au monde occultées lorsque non quantifiables (Bowker, 2013), et évidemment sur l'absence de transparence quant au respect de la vie privée (Puschmann & Burgess, 2013). Des mouvements réclamant des traitements de données plus démocratiques amèneront à la mise en place de législations pour contrecarrer les appropriations néfastes des données, comme le cas du *RGPD* produit en 2016 en Europe, le développement de l'ouverture des données incarnée par le mouvement de l'*open data*, notamment en France avec

[51] Voir (The Economist, 2020).

[52] « La transformation de l'action sociale en données quantifiées en ligne permettant le traçage en temps réel et l'analyse prédictive », nous reviendrons dessus plus en détail au chapitre 4.

Open Data France qui a concouru à la création de la loi pour une République numérique en 2016 [53] établissant l'ouverture par défaut des données publiques (Goëta, 2016), ou encore la volonté récente de formation des publics notamment dans le développement d'un esprit critique vis-à-vis des usages des données.

Ce dernier point, le développement de cet esprit critique, doit permettre d'endiguer le flux de techno-discours, que nous avons évoqué dans l'introduction et qui ont été définis par Janicaud (1985) comme des discours ni strictement techniques, ni autonomes, des langages « parasitaire(s) branché(s) sur la technique » qui empêchent tout recul et toute remise en question du phénomène technique (p. 180). Ces techno-discours amènent des impensés techniques qui mettent les données dans une posture d'évidence (Robert, 2020). Ici, les données numériques semblent ici empreintes d'une puissance nouvelle de quantification des relations humaines, prolongeant les rêves originellement contenus dans les usages des statistiques. Or, il s'agit ici de faux-semblants :

> « L'emballement contemporain autour du digital s'accompagne de l'effacement de nombreux objets qui s'accommodent difficilement des discours insistant sur la « nouveauté » du digital. Les objets digitaux d'aujourd'hui et les démarches qui permettent de les appréhender apparaissent ainsi sans héritage, alors que de larges pans de la recherche sont tombés dans un oubli relatif. » (Trudel, Le Deuff, & Averbeck-Lietz, 2023).

Si les progrès informatiques ont effectivement apporté de plus grandes possibilités de traitement des données, ces dernières conservent des biais inhérents aux effets dévastateurs comme nous le développerons au chapitre 4. La *data literacy* hérite des influences des conceptions littéraciques et des visions des disciplines suscitées sur les données, intégrant un certain nombre de leurs interrogations et problématiques. Cela fera l'objet du chapitre 3.

2.3. La datalogie, l'influence oubliée

Pour introduire cette sous-partie, nous précisons ici que la datalogie (*datalogi* en danois, que nous traduisons en français par datalogie, sans aucun lien avec le titre d'un ouvrage éponyme [54]) est un « savoir assujetti », notion définie par Foucault (1994) comme « des contenus historiques qui ont été ensevelis, masqués dans des cohérences fonctionnelles ou dans des systématisations formelles »

[53] https://www.vie-publique.fr/eclairage/20301-loi-republique-numerique-7-octobre-2016-loi-lemaire-quels-changements

[54] Voir Avenati & Chardel (2016). Cet ouvrage emploie le terme datalogie dans son titre mais aucune connexion avec l'oeuvre de Peter Naur n'est précisée et aucune définition du terme n'est proposée.

et comme « une série de savoirs qui se trouvaient être disqualifiés comme savoirs non conceptuels, comme savoirs insuffisamment élaborés, savoirs naïfs, savoirs hiérarchiquement inférieurs, savoirs au-dessous du niveau de la connaissance ou de la scientificité requise. »

La datalogie est en effet une notion non citée dans la littérature scientifique au-delà d'un récapitulatif de sa propre histoire (Sveinsdottir & Frøkjær, 1998). La filiation avec la *data literacy* y est donc officiellement inexistante. Or, comme nous allons le voir, la datalogie possède toutes les caractéristiques d'une *data literacy*, plus de trente ans avant l'émergence de ses premières définitions.

Nous ne parlerons donc pas de proto *data literacy* en raison du fait que la datalogie était et est toujours une discipline aboutie, proposant des curricula et des visions spécifiques. L'ensemble des informations qui vont être présentées ont été recueillies de l'analyse de l'ensemble des travaux de Peter Naur.

2.3.1. Une opposition aux premiers modèles de formation en informatique

La datalogie est indissociable de son créateur, Peter Naur (1928-2016), un astronome danois devenu ingénieur au *Regnecentralen*, la première entreprise d'informatique danoise. Il est un des pionniers dans le domaine informatique, co-créateur du langage de programmation *ALGOL 60* en 1960 utilisé dans la conception du second ordinateur danois : *GIER* (Høg, 2017). En 1966, il introduit la datalogie (Naur, 1966), puis en propose un curriculum deux ans plus tard (Naur, 1968). En 1969, il est nommé professeur en datalogie à l'Université de Copenhague dans laquelle il fonde le département de datalogie un an plus tard nommé *DIKU* (*Datalogisk Institut – Københavns Universitet*) qui existe encore aujourd'hui[55].

Le premier cours de datalogie, connu sous le terme *datalogy zero*, est présenté la même année (Naur, 1992, p. 249). En 1974, Naur publie *Concise Survey of Computer Methods* qui se veut un approfondissement de la datalogie. En 1992 est publiée une traduction anglaise de ses précédents travaux, compilés dans *Computing, a human activity*. Peter Naur prend sa retraite en 1998, gagne le Turing Award en 2005 avant de décéder en 2016.

2.3.2. Les premières définitions

La datalogie, que Naur a nommé aussi *computology* (Naur, 1992, p. 61) pour rappeler le fait qu'il s'agit d'une discipline informatique, est tout d'abord décrite comme « la science de la nature et de l'usage

[55] https://di.ku.dk/

des données [56] » (Naur, 1966), puis comme « la science des données et des processus de données [57] » (Naur, 1968, p. 1383).

Il s'agit d'une discipline conçue comme distincte de la discipline informatique naissante. Ce point est précisé dans l'introduction du curriculum de 1968 :

> « Le but du présent article est de poursuivre l'idée selon laquelle le développement de l'informatique à partir de certains concepts de base ne doit pas nécessairement s'arrêter au niveau le plus élémentaire, mais peut se poursuivre bien plus loin de manière utile. Il sera démontré qu'il est possible et pratique de distinguer une discipline spécifique, la datalogie, la science des données et des processus [58] » (Naur, 1968, p. 1383).

Il reprécisera ce point en 2007. L'invention du terme *datalogi* s'est faite contre celui de *computer science*, qui était alors employé au Danemark. Cette dénomination était incapable de saisir la vision de Naur qui voyait dans l'informatique une activité humaine impliquant certains buts et certaines intentions, certaines perceptions, ainsi que certains outils et certaines techniques créés par l'homme (Naur, 1992, xiv). La datalogie est en effet la forme qu'aurait dû prendre l'informatique : elle aurait dû pour Naur dépasser la simple étude de l'ordinateur. Il voulait qu'elle prépare à la vie dans « l'ère des ordinateurs », étant certain de leur importance future (Naur, 1992, p. 548). Loin de se joindre aux discours technophiles de l'époque, l'objectif de la datalogie devait dissiper l'aspect mystique entourant les ordinateurs.

Au-delà d'une problématique terminologique et philosophique, Naur voulait également distinguer sa datalogie de la *computer science* sur l'aspect pédagogique. Il cite à ce propos le *ACM Curriculum Committee on Computer Science* de 1968 qui tentait, selon lui, de couvrir le champ de l'informatique d'une manière « presque encyclopédique », semblant s'assurer de « mentionner la totalité des techniques, langages et pratiques actuelles » au contraire de la datalogie qui cherchait à souligner les idées et principes sous-jacents pour se concentrer sur les principes de base communs à tous les environnements (Naur, 1968, p. 1385). Il ajoutait que le curriculum souffrait d'un manque de formalisation, ignorant certaines thématiques liées à l'expérience et à l'intuition des utilisateurs (*practitionner*) comme les problématiques d'interfaces homme/machine, la psychologie de la programmation et du travail de groupe (*work project*). Il précisera enfin, que des applications concrètes étaient pratiquement absentes (Siefkes, 1974, p. 88).

---

[56] « *The science of the nature and use of data* ».

[57] « *The science of data and data processes* ».

[58] « *The purpose of the present paper is to pursue the idea that the development of computer science from certain basic concepts need not stop at the most elementary level, but can usefully be continued considerably further. A case will be made that it is possible and convenient to distinguish a specific discipline, datalogy, the science of data and processes* ».

2.3.3. Des tentatives de traductions avortées

La datalogie est donc pour Naur une science informatique, sa science informatique, qu'il souhaite faire connaître, cherchant pour cela le terme adéquat. Or, le travail de traduction sera pour lui d'une grande difficulté bien que non nouvelle car dès 1954, il proposait une nomenclature où étaient traduits en danois chaque concept du domaine informatique afin d'éviter les « mauvais développements linguistiques » (*bad linguistic developments*) et faciliter les compréhensions techniques (Naur, 1954).

Ainsi, dans ses premiers travaux, il utilise *datalogy* pour s'adresser à une audience internationale (Naur, 1966), puis change pour *data science* (Naur, 1974). Cette dernière dénomination n'est pas anodine : c'est de là que vient son emploi par la discipline du même nom. Il reviendra finalement à *datalogy* mais utilisera également *computing* : pour toute explication, Naur précisera que c'était « une énième escarmouche dans la lutte probablement perdue contre les piètres dénominations [59] ».

Actuellement, la version danoise du site web de l'Université de Copenhague présente le terme *datalogi*, adjectivé en *datalogisk*[60], tandis que la version anglaise emploie *computer science*[61].

2.3.4. Les concepts clé de la datalogie

Pour poser les bases de sa nouvelle discipline, Naur a proposé une triade centrale : les données (*data*), la datamatique (*datamatik* en danois que Naur traduit en anglais par *datamatics*) et l'ordinateur (*datamat* en danois, traduit en anglais par Naur par *datamaton*).

2.3.4.1. Les données

En se basant sur l'*IFIP-ICC vocabulary of Information Processing*, Naur (1974) décrit les données comme « une représentation de faits ou d'idées formalisée de telle manière qu'elle peut être communiquée ou manipulée selon certains procédés »[62]. Ces évènements ou phénomènes, choisis par les individus, doivent être employés d'une façon spécifique (Naur, 1974, p. 18). Les données sont donc indissociables de leur traitement, les deux étant des outils nous permettant de réaliser des tâches

[59] « *yet another skirmish in the probably lost fight against poor terminology* » (Naur, 1992, p. 180).
[60] https://di.ku.dk/
[61] https://di.ku.dk/english/
[62] « *A representation of facts or ideas in a formalized manner capable of being communicated or manipulated by some process* » (p. 18).

« dans le monde » (Naur, 1974, pp. 20-21). Plus que cela, cela nous permet de changer la réalité du monde à travers l'usage de modèles.

Une distinction avec l'information est clairement établie. D'une part, l'information est extraite des données : elle est considérée comme unique aux personnes du fait qu'elle réside seulement dans la conscience humaine. Cette dernière étant inaccessible, Naur spécifiait qu'il était impossible de pleinement comprendre les informations (Naur, 1974, p. 18). Du fait que Naur emploie les définitions de l'*IFIP-ICC* pour présenter ses concepts, nous pouvons ici en déduire que sa vision de l'information est tirée de cette définition : « le sens qu'un humain assigne à une donnée par le truchement de conventions établies et employées pour sa représentation [63] ».

D'autre part, les données sont des phénomènes ou des faits ayant été sélectionnés et ne pouvant être interprétés que d'une seule manière. Il présente à ce sujet le nom « Richard » et explique qu'il est intégré dans un ensemble de conventions elles-mêmes inscrites dans l'alphabet latin[64].

De plus, Naur précise que les données sont liées à la notion de symbole (Naur, 1992, p. 544). Ces symboles, ou « atomes de données » (Naur, 1974, p. 76), peuvent être des caractères ou des chaînes de caractères (Naur, 1992, p. 20).

Naur utilisera ainsi un ensemble de termes pour présenter la structure d'une donnée digitale que nous avons représenté à partir de ses écrits dans la figure 2.6 :

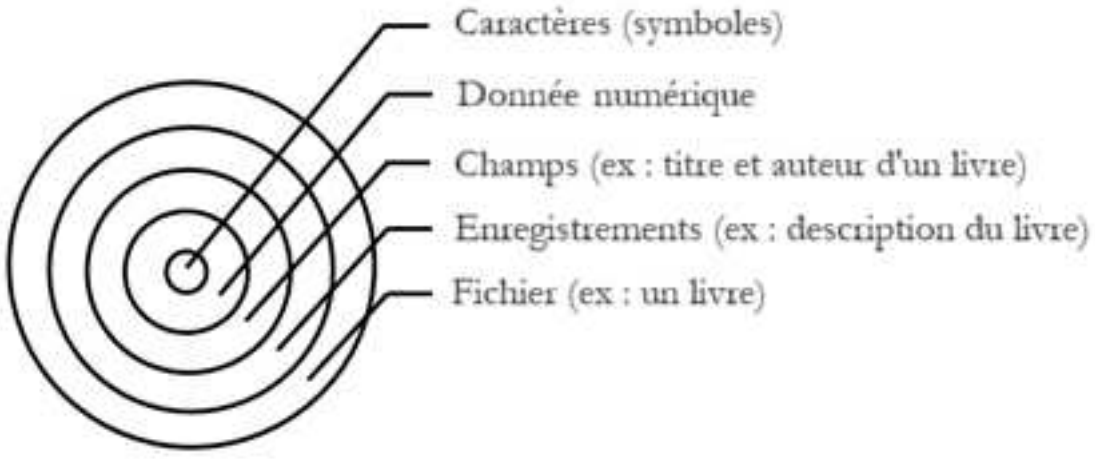

Figure 2.6.: Granularité des données numériques - créée à partir de (Naur, 1974, pp. 76-111)

[63] « *The meaning that a human assigns to data by means of the known conventions used in its representation* ».

[64] Cet exemple montre une certaine limite réflexive sur l'interprétation informationnelle de Naur qui pourtant se réclamera de travaux sur le langage.

Les symboles sont contenus dans une donnée numérique qui elle-même est incluse dans des champs (*fields*) et des enregistrements (*records*) qui constituent un fichier (*file*).

La datalogie, au-delà d'enseigner des opérations basiques de traitement des données, doit permettre de prendre en compte les mauvais usages potentiels des jeux de données et de se positionner vis-à-vis de ces problématiques :

> « La surveillance des citoyens qui devient possible dans la société informatisée comporte le risque que les données soient utilisées à mauvais escient, de sorte que les données collectées à certaines fins, avec le consentement conscient de la population, sont utilisées par des groupes d'intérêts spéciaux à des fins illicites, comme arme contre d'autres groupes par exemple [65] » (Naur, 1992, p. 568).

2.3.4.2. La datamatique

Le second élément est la datamatique, à savoir d'une part le traitement des données par des procédés automatiques, et d'autre part le fait pour les données, associées à un sens spécifique, d'être combinées au processus correspondant à une intention spécifique de former de nouvelles données jusque-là inconnues.[66] (Naur, 1966).

Deux usages sont inclus dans ce procédé : d'une part, la liberté de choisir n'importe quel type de phénomène comme donnée (Naur, 1992, p. 545). D'autre part, l'utilisation de représentations de données, ce que Naur décrit comme la possibilité qu'une même signification peut être représentée de manières différentes (*different data occurrences*) (Naur, 1992, p. 177). Les représentations de données (*data representations*) sont centrales dans la datalogie et prennent diverses formes (voir le tableau 2.1) :

[65] « *The monitoring of the citizens which becomes possible in the computerized society entails the danger that the data will be misused, such that data collected for certain purposes, with the knowing consent of the population, is used by special interest groups for illicit purposes, for example as a weapon against other groups* ».

[66] « *Datamatics is the processing of data by automatic means and is for data, associated to a specific meaning, to be combined with the process corresponding to some intent to form new, so far unknown, data* » (Naur, 1974, p. 27).

Tableau 2.1.: La classification des représentations de données - résumée à partir de (Naur, 1974, pp. 23-24)

	Statique	**Dynamique**	**Analogue**	**Numérique**
Description	Les données existent et peuvent être traitées indépendamment de toute échelle de temps ou moment particulier	L'existence et le traitement des données sont liés au temps	Les données sont des mesures physiques	Les données sont la résultante d'une collection de caractères
Exemple	L'impression sur papier ou les trous dans les cartes	Les signaux vocaux et télégraphiques	La durée temporelle	Les signaux de télégraphe

Toute représentation de donnée doit : - vérifier que le degré de liberté de la capacité de détention des données doit être égal à la variabilité du phénomène (Naur, 1974, pp. 80-81), - et doit être choisie en tenant compte de la transformation à réaliser (Naur, 1974, p. 28).

Au contraire des traitements de données où la focale se pose sur la manière dont le traitement est réalisé, les transformations des données redirigent cette focale sur les entrées et sorties de données et sur la relation qu'elles entretiennent (*and their mutural relation*) sans prise en compte du processus de traitement (Naur, 1974, p. 21). C'est une caractéristique importante de la datalogie qui traite spécifiquement des données qui ont été établies et qui ne se préoccupe pas de la relation entre les données et ce qu'elles représentent (Naur, 1974, p. 30).

Vis-à-vis des critères de sélection dans le choix de traitements de données, Naur cite la vitesse de traitement (*speed of processing*), la simplicité conceptuelle du traitement des données (*conceptual simplicity of the data process*), la flexibilité liée au changement du traitement des données (*flexibility with respect to change of the data process*) et le faible coût de l'utilisation des traitements de données (*low total cost of the use of the data process*) (Naur, 1992, p. 19). Les traitements de données sont limités par la concordance entre la réalité et les données, et le traitement des données lui-même (Naur, 1974, p. 28).

2.3.4.3. L'ordinateur

L'ordinateur est décrit par Naur comme le dispositif réalisant les datamatiques, ces traitements automatiques des données (Naur, 1992, p. 181). Il s'agit d'une vision ancienne de Naur qui dès 1954 parlait ainsi des ordinateurs :

> « Un ordinateur électronique est un appareil capable d'effectuer automatiquement une série étendue de processus de calcul à l'aide de tubes électroniques [67] » (Naur, 1954, p. 3).

Ils doivent être conçus comme un outil utile à la résolution de certains problèmes (Naur, 1992, p. 4). Le glissement sémantique entre *computer* et *datamat* marque un changement paradigmatique sur ce que doit être un ordinateur, focalisé sur les données et leurs traitements plutôt que sur les calculs :

> « Ces machines doivent, en premier lieu, être considérées comme des manipulatrices de données et non comme des machines de calculs [68] » (Naur, 1992, p. 543).

> « Cela a eu des conséquences bien au-delà de ce à quoi on pouvait s'attendre au premier abord, et nous a fait comprendre que ces machines ne doivent pas être considérées exclusivement, ou même en grande majorité, comme des machines de calcul, mais que, comme nous le verrons en détail plus loin, elles sont des machines de manipulation de données de toutes sortes. C'est encore une fois la raison pour laquelle nous, qui travaillons aujourd'hui avec ces machines, trouvons trompeuse l'appellation "machine à calculer électronique". [69] » (Naur, 1992, p. 539).

Naur ajoute également qu'au contraire de l'informatique, la datalogie ne doit pas être strictement dépendante des ordinateurs, ces derniers devant être traités à travers « une orientation finale » (*through a final orientation*) et non comme une problématique centrale (Naur, 1992, p. 548). En tant que tels, les ordinateurs permettent de traiter sous un nouveau jour les questionnements datalogiques résumés ainsi :

> « Un phénomène qui, au même titre que les ordinateurs, exerce une forte influence sur la société n'est pas conditionné par une seule idée, une seule invention ou un problème

[67] *« an electronic computer is an apparatus which is able to carry out extensive series of calculation processes automatically, with the aid of electronic tubes»*.

[68] *« This had consequences far beyond what one might expect in the first instance, and has made us realize that these machines not exclusively, or even just predominantly, should be regarded as computing machines, but that, as will be discussed in more detail later, are machines for manipulating data of any kind. This again is the reason why we who work with these machines nowadays find the designation 'electronic calculating machine' misleading»*.

[69] *« these devices primarily must be regarded as data manipulators, not as calculating machines »*.

> isolé de cette société. Il résulte de l'interaction de nombreuses circonstances, à la fois de possibilités théoriques, d'inventions techniques spécifiques et de diverses conditions de la société, ce qui comprend l'économie. [70] » (Naur, 1992, p. 548).

2.3.5. Les visées de la datalogie

La datalogie, en tant qu'activité de projet (Naur, 1970, p. 2), est orientée vers la résolution de problèmes ; les ordinateurs et les langages de programmation sont utilisés à cet effet.

Tableau 2.2.: Définitions des concepts liés à la résolution de problèmes - extrait de (Naur, 1992, p. 3)

Concept	**Définition**
Les personnes	ne peuvent comprendre les problèmes que dans le contexte d'un ensemble d'outils supposés.
Les problèmes	existent seulement dans l'esprit des personnes et seulement vis-à-vis des outils qu'ils comprennent.
Les langages de programmation	façonnent la pensée des personnes et la manière dont ils conçoivent les problèmes.
Les outils	n'existent en tant que tels que dans la mesure où certaines personnes les considèrent comme les outils appropriés pour résoudre certains problèmes.

Sans la connaissance d'un outil (ici les données, la datamatique et le datamat), il est impossible de comprendre un problème. Concernant les langages de programmation, Naur nous mettait en garde :

> «Au lieu de forcer le développement de manières radicalement nouvelles d'aborder le problème, adaptées à l'ordinateur, ils peuvent faire croire aux gens que leurs solutions conventionnelles sont les bonnes, même si elles peuvent être très inappropriées pour les opérations qui conviennent à l'ordinateur. [71] » (Naur, 1992, p. 4).

[70] « *A phenomenon that like computers exerts a strong influence on society is not conditioned by a single idea, a single invention, or an isolated problem in society. It arises by an interplay of many circumstances, both theoretical possibilities, specific technical inventions, and a variety of conditions of society, including economy* ».

[71] « *Instead of forcing the development of radically new ways of looking at the problem appropriate to the computer, they may lull people into thinking that their conventional solutions are the right ones, although they may be highly inappropriate to the operations which are convenient in the computer.* ».

Ainsi, en accord avec la vision développée dans la datalogie qui se veut avant tout une activité humaine, il ne faut pas faire des outils l'alpha et l'oméga de toute chose mais se rappeler qu'ils ne sont que des moyens pour arriver à des fins.

2.3.6. Les liens avec les autres disciplines

Pour Naur, la datalogie se présentait comme une nouvelle discipline, donc supposément sans traditions anciennes à surmonter (Naur, 1992, p. 246). S'il avait la volonté de relier la datalogie à d'autres disciplines, il se refusera toutefois de dépendre d'elles pour conserver une certaine liberté. Deux disciplines sont citées comme de « proches cousines » : les mathématiques et le langage (Naur, 1992, p. 548).

2.3.6.1. Les mathématiques

Selon (Naur, 1992, p. 23), les mathématiques sont la discipline intellectuelle la plus proche de la datalogie. Il appela même à une « influence mutuelle plus proche » où la datalogie pourrait bénéficier du sens de la généralité et de la fiabilité de l'argumentation tandis que les mathématiques pourraient elles bénéficier de « connexions plus vivantes avec le terrain » (*more lively connections to the grass roots*) (Naur, 1992, p. 180). On retrouve cela dès 1966 où Naur déplorait un manque d'influence de la pensée mathématique sur la manière dont la programmation informatique était alors entreprise (Naur, 1966). Toutefois, si pour Naur les mathématiques étaient un sujet important, ce n'était pas un idéal vers lequel la programmation devait tendre (Naur, 1992, p. 51). Il reprendra cette affirmation en 2014 lorsqu'était discutée une possible fusion entre les champs de la datalogie et des mathématiques :

> « À mon avis, les domaines de la datalogie et des mathématiques devraient être dispensés dans deux départements distincts. Cela concorde avec mon expérience scientifique tout au long de ma carrière. Jusqu'à mes 34 ans en 1962, mon seul contact professionnel avec les mathématiques était comme outil de description des phénomènes astronomiques. De telles applications reposaient sur des mathématiques établies plusieurs siècles auparavant. Mon travail en informatique, après 1960, concernait l'utilisation en sciences appliquées des nouveaux outils électroniques, là encore sans dépendance avec les mathématiques modernes. »[72] Naur cité par (Frøkjær, 2014).

[72] « *In my opinion the fields of computer science and mathematics should be held in two separate departments. This accords with my scientific experience throughout my career. Until I was 34 years old in 1962 my only professional contact with mathematics was as a tool for describing astronomical phenomena. Such applications rested on mathematics established several centuries before. My work in computing, after 1960, was a matter of the use in applied science of the new electronic tools, again with no dependence on modern mathematics* ».

2.3.6.2. Le langage

Pour Naur, le langage doit être compris dans un sens élargi qui prend en compte « des choses élémentaires, la lecture, l'écriture et l'orthographe [73] » et peut s'étendre aux « langues étrangères, la littérature, le style d'écriture, la linguistique et autres [74] » (Naur, 1992, p. 181). Ce qui est vital pour Naur c'est le langage naturel qui doit être intégré d'une manière ou d'une autre dans les langages de programmation (voir le tableau 2.3).

Tableau 2.3.: Comparaison entre le langage naturel et le langage de programmation - extrait de (Naur, 1992, pp. 26-35)

	Le langage naturel	**Le langage de programmation**
Utilisation	Sous forme orale, dans les conversations entre les personnes.	Sous forme écrite, en relation lente et maladroite avec l'ordinateur.
Compréhension	Doit être compris par les humains de manière floue.	Doit être compris par les ordinateurs d'une manière parfaitement définie.
Singularité	Fournit un univers de discours extensible qui permet d'ajouter à volonté de nouvelles notations, telles que des langages de programmation.	Tout ce qui peut être exprimé dans un langage de programmation peut l'être de manière équivalente, quoique peut-être maladroitement, dans un langage naturel hôte, alors que l'inverse est complètement faux.

De son analyse des langages naturels, Naur déclare que :

> « Les langages de programmation doivent être construits de préférence à partir de quelques concepts très généraux et très abstraits qui peuvent être employés dans de nombreuses combinaisons, offrant ainsi la flexibilité d'expression souhaitée. Ils fournissent un argument de taille contre les langages de programmation qui intègrent de nombreuses fonctionnalités spéciales pour des domaines d'application limités, et

[73] « *Elementary things, reading, writing, and spelling* ».
[74] « *Foreign languages, literature, style, linguistics and others* ».

> qui intègrent de nombreux cas d'exception dans leur définition.[75] » (Naur, 1992, p. 30).

A travers les langages de programmation, Naur cherche à améliorer la « relation homme-machine » (*man-machine relationship*) et à corriger les problématiques rencontrées dans l'informatique des années 1950 :

> « Les machines étaient présentes mais les utilisateurs éprouvaient des difficultés quand il s'agissait de les nourrir avec des tâches suffisamment complètes et correctement formulées pour leur permettre de fonctionner au mieux. [76] » (Naur, 1992, p. 540).

2.3.7. L'aspect pédagogique

Les thèmes centraux devant être développés dans les cours de datalogie sont les données, les représentations de données et les traitements de données (Naur, 1992, p. 548). En outre, Naur a souligné l'extrême importance de relier les données avec les pratiques quotidiennes. Du fait que les données sont omniprésentes, il convient de leur appliquer un cadre pratique « amplifié par des exemples de la vie quotidienne » (*amplified with examples from every life*) (Naur, 1968, p. 1384), ce qui faisait cruellement défaut à l'informatique naissante (Naur, 1996, p. 531).

De plus, la datalogie doit être enseignée à tous, spécialistes comme profanes :

> «... le contrôle des ordinateurs et leur applications vont devenir, non pas le sujet d'intérêt d'un petit groupe d'experts, mais un problème politique banal, et doit donc de ce fait être placé, par le concourt du système politique, là où il se doit de l'être : entre toutes nos mains [77] » (Naur, 1992, p. 548).

Ainsi, en plus des étudiants des universités, la datalogie doit être enseignée aux enfants et adolescents (Naur indique une fourchette entre 10 et 15 ans) car ces derniers sont capables de concevoir les traitements des données en raison de leur bagage littéracique : la lecture et l'écriture étant considérées comme « des données en cours de traitement » (*data being processed*) (Naur, 1968, p. 1387). Le

[75] *« programming languages should preferably be built from a few, very general, very abstract concepts, that can be applied in many combinations, thereby yielding the desired flexibility of expression. They provide an argument against programming languages that include many special features with limited fields of application, and that include many exceptional cases in their definition. ».*

[76] *« The machines were there, but the users found it difficult to keep up with feeding them with tasks that were sufficiently completely and correctly formulated to allow the machines to perform at their best ».*

[77] *«the control over the computers and their application will become, not a matter for a small group of experts, but an ordinary political issue, and thus by way of the political system will be placed where it should lie, with all of us. ».*

DIKU (*Datalogisk Institut*) suit toujours cette logique et propose des cours adaptés tout aussi bien aux enfants qu'aux adultes[78].

Quant au sujet de son intégration au sein d'un cursus scolaire, Naur proposait de faire de la datalogie un sujet indépendant ou affilié aux mathématique :

> « Quand je dis que la datalogie sera intégrée de manière similaire dans l'éducation, je veux dire qu'elle entrera dans les programmes scolaires et comme une matière annexe dans de nombreuses études avancées, en plus d'être le sujet d'études spécialisées.[79] » (Naur, 1992, p. 176).

2.4. Conclusion du chapitre

La volonté de maîtrise des données est une thématique ancienne, incarnée dans un ensemble de disciplines aux visées parfois contraires. La *data literacy* émerge à une époque charnière, au moment du tournant computationnel du début des années 2000. Cette période marque l'émergence simultanée de méthodes et disciplines autour des données : en premier la science des données (*data science*, premièrement définie dans les années 90), qui sera suivie par la *data literacy* (2004 par Schield et Love) ainsi que la gouvernance des données (premièrement définie en 2004 par Gwen Thomas).

Les évolutions technologiques offrent de nouvelles possibilités de traitement des données, autrefois limitées, permettant de répondre à des problèmes abordés dès les années 1960, à savoir les problèmes de qualité des données et de traitement de ces dernières, surtout lorsqu'appliqués sur de très larges jeux de données. La vision d'une donnée comme atout, auparavant confinée aux organisations disposant de ressources matérielles et humaines suffisantes pour traiter cette thématique, s'est propagée à un ensemble d'acteurs autrefois éloignés de ces problématiques.

La nécessité de formation de spécialistes, déjà présente trente ans plus tôt, prend ici un nouvel essor : il s'agit alors de répondre à des besoins fonctionnels de traitement des données et ce jusqu'au début des années 2010. Au même titre que certaines conceptions de la littératie, à mesure que les problématiques des données sont propagées à l'ensemble des sociétés, les visions autour de la *data literacy* ont évolué, passant de visions fonctionnelles, centrées sur des compétences classiques de traitement de données, à des visions centrées sur la protection des individus et de la société. Elle a acquis en sens les

[78] Voir par exemple : https://di.ku.dk/oplev-datalogi/

[79] « *When saying that datalogy will be placed in a similar manner in education, what I mean is that it will enter into the school curriculum and as an auxiliary subject in many advanced studies, besides being the topic of specialized studies.* ».

caractéristiques d'une « nouvelle littératie », concourant à un renouvellement des discours autour de la place des données dans les sociétés digitalisées.

Parler de *data literacy* au singulier, ou même de la *data literacy* comme « discipline » établie semble ici être un non-sens. La parution en 2004 de deux définitions inscrites dans deux domaines disciplinaires différents présentait déjà ce manque d'unicité malgré une volonté commune de maîtrise. De plus, ce semblant d'unicité a continué à se fragmenter avec les problématiques engendrées par les nouvelles pratiques des données : *big data*, *open data*, *datafication*, tant de thématiques nouvelles dont nous n'avons pas fini de résoudre les problématiques engendrées par leur émergence. Nous le verrons au chapitre 4 : elles vont contribuer à atomiser la *data literacy* en une constellation de *data literacies*, chacune étant dédiée à une sous-thématique spécifique.

Mais au-delà de son identité propre, nous allons à présent aborder ses enchevêtrements conceptuels et littéraciques à travers une analyse de la littérature et ainsi tenter de situer la *data literacy* dans cet écosystème toujours plus complexe et changeant.

Partie II.

Des visions plurielles

Chapitre 3.

Les enchevêtrements conceptuels et littéraciques

> « *On avait menacer de le déconnecter : cela le priverait de toute entrée de données, le propulserait dans un inimaginable état d'inconscience. Ce qui, pour Hal, équivalait à la mort.* »
>
> Arthur C. Clarke, *2001 - l'Odyssée de l'espace*, 1968

3.1. Introduction

Nous avons pu noter précédemment qu'il existait un ensemble de visions éclectiques concourant à la mise en place de méthodes et de formations pour traiter les données. Celles-ci s'expriment à travers le faisceau de concepts et de littératies spécifiques qui interagissent avec, voire intègrent la *data literacy*.

Si la présence d'invariants est perceptible, notamment dans les compétences et objectifs de base recherchés, le périmètre conceptuel de la *data literacy* varie selon les auteurs sollicités, changeant le type d'enchevêtrement et la prééminence d'un élément sur l'autre.

Nous débuterons cette première partie par les interactions de la *data literacy* avec des concepts parallèles sur les données ayant émergé dans un intervalle temporel suffisamment proche pour affirmer qu'ils appartiennent à un même plan et que des liens entre eux peuvent être tissés.

3.2. Les enchevêtrements conceptuels

Pour apporter de la clarification dans nos propos, nous introduisons ici le « concept » selon la vision de Deleuze & Guattari (2005), à savoir une « ordination de ses composantes par zones de voisinage », qu'ils nomment « hétérogenèse » (p.29). Cette dernière est caractérisée par « l'inséparabilité d'un nombre fini de composantes hétérogènes parcourues par un point de survol [1] absolu, à vitesse infinie » (p.30) ; les composantes, inséparables au sein du concept lui-même, peuvent provenir d'autres concepts qui « répondaient à d'autres problèmes et supposaient d'autres plans » (p. 26). Un concept a donc un contour irrégulier défini par ses composantes qui évoluent dans le temps, nécessitant qu'il soit « retaillé » :

> « Les composantes restent distinctes, mais quelque chose passe de l'un à l'autre, quelque chose d'indécidable entre les deux : il y a un domaine ab qui appartient aussi bien à a qu'à b, où a et b "deviennent" indiscernables. Ce sont ces zones, seuils ou devenirs, cette inséparibilité, qui définissent la consistance intérieure du concept » (p. 28).

Le concept est une interface par laquelle la réalité est interprétée et vécue (Floridi, 2015). Ainsi, pour Deleuze & Guattari (2005), il a pour but de répondre à un problème (p. 24), voire un « carrefour de problèmes » où il s'allie avec d'autres concepts existants (p. 26). En tant qu'« outil et [une] histoire, c'est à dire un faisceau de possibilités et d'obstacles engagés dans un monde vécu ». (Granger, 1955, p. 23), il « a toujours la vérité qui lui revient en fonction des conditions de sa création ». Contextes de production que nous avons pu évoquer dans le cas des traitements des données.

Ainsi, les concepts qui seront présentés possèdent des composantes que nous nous efforcerons de souligner pour délimiter leurs périmètres respectifs.

3.2.1. La science des données

Le premier, que nous avons évoqué à plusieurs reprises au cours des chapitres précédents, est la science des données (*data science*). Cette dénomination remonte, nous l'avons vu, à la tentative de Peter Naur de trouver une traduction adéquate à la datalogie (Naur, 1974). Elle va ensuite se perdre pour réapparaître à plusieurs reprises dans les années 1990.

Tout d'abord en 1996 où les membres de l'*IFCS* (*International Federation of Classification Societies*) vont se réunir au cours d'une conférence nommée *Data science, classification, and related methods* (Hayashi et al., 1998). Les actes de cette conférence listeront un ensemble de techniques statistiques et

[1] Le survol est pour ces auteurs est l'état du concept ou son infinité propre.

mathématiques aidant les analyses de données, mais le plus marquant sera l'article de Chikio Hayashi nommé *What is Data science ? Fundamental Concepts and a Heuristic example* (Hayashi, 1998).

Dès son introduction, Hayashi y explique que la science des données « n'est pas seulement un concept permettant d'unifier les statistiques, les analyses de données et leurs méthodes connexes, elle comprend également leurs résultats »[2]. Il ajoute qu'elle est constituée de trois phases « spiralaires » : la conception des données (*design for data*), la collecte des données (*data collection*) et les analyses des données (*analysis on data*). L'origine de cette nouvelle science est expliquée par Hayashi par la scission méthodologique entre les statistiques mathématiques et les analyses de données (*data analysis*)[3] : si les deux méthodes ont des avantages indéniables, il précise que la première, en recherchant la précision mathématique, s'est par trop éloignée de la réalité tandis que la seconde ne prend pas suffisamment en compte la qualité et le sens des données en raison d'un usage limité aux seules données existantes. Il propose ainsi d'unifier ces deux disciplines sous l'égide de la science des données qui se doit d'analyser et de comprendre les véritables phénomènes avec les données, chose qui ne se retrouvait plus dans les précédentes disciplines :

> « La science des données a pour but d'analyser et de comprendre les véritables phénomènes grâce aux "données". En d'autres termes, le but de la science des données est de révéler les caractéristiques ou la structure cachée des phénomènes naturels, humains et sociaux grâce aux données à partir d'une perspective différente de celle employée par les méthodes et théories traditionnelles ou déjà établies. Cette perspective implique des manières de penser multidimensionnelles, dynamiques et flexibles [4] ».

Le terme réapparaît deux ans plus tard en 1998 dans la leçon inaugurale de Wu (1998) *Statistics = data science ?* dans laquelle il rappelle les caractéristiques globales des statistiques, en particulier ce qu'il nomme la « trilogie des statistiques » (*statistical trilogy*) qui comprend les collectes de données (*data collection*), la modélisation et l'analyse de données (*data modeling and analysis*), la compréhension et la résolution de problèmes ainsi que le processus de décision (*problem understanding/solving, decision making*). Il spécifie ensuite les tendances futures auxquelles les analystes de données (*data analyst*) seront confrontés, à savoir (1) l'extraction de données (*data mining*) de grandes bases données (le terme *massive* est employé, renvoyant ici aux prémisses du *big data*) et les modèles de réseaux neuronaux, (2) les approches physiques et empiriques alimentées par les connaissances mécanistes et

[2] « *Data Science is not only a synthetic concept to unify statistics, data analysis and their related methods but also comprises its results.* ».

[3] Il s'agit ici du courant évoqué dans le chapitre précédent, développé par John Tukey aux Etats-Unis mais également par Chikio Hayashi au Japon et Jean-Paul Benzécri en France.

[4] « *Data Science intends to analyze and understand actual phenomena with "data". In other words, the aim of data science is to reveal the features or the hidden structure of complicated natural, human and social phenomena with data from a different point of view from the established or traditional theory and method. This point of view implies multidimensional, dynamic and flexible ways of thinking*».

les connaissances des données, ainsi que (3) les nouvelles représentations et exploitations du savoir apportées par les algorithmes de calculs, une plus grande interaction avec les sciences cognitives et la représentation du savoir selon une vision Bayesienne. Il propose ainsi de renommer les statistiques en science des données, et les statisticiens en scientifiques des données (*data scientists*) pour briser l'image vieillotte de la discipline (selon ses propres mots).

Mais l'auteur le plus cité dans la littérature n'est ni Hayashi, ni Wu mais Cleveland[5]. En 2001, il publie l'article *Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics* (Cleveland, 2001) dans lequel il propose une extension « des domaines techniques » des statistiques centrés sur l'analyste de données (*data analyst*). Il précise que la science des données doit être jugée avant tout par sa capacité à permettre à l'analyste d'apprendre grâce et à partir des données. Cleveland propose ainsi six grands domaines pour constituer sa science des données qu'il destine en priorité au public universitaire [6] : les enquêtes multidisciplinaires (*multidisciplinary investigations*), les modèles et méthodes pour les données (*models and methods for data*), l'utilisation de l'informatique dans le traitement des données (*computing with data*), la mise en place d'une vision pédagogique (*pedagogy*) incluant des curricula, l'évaluation d'outils (*tool evaluation*) et enfin les théories (*theory*) qui fondent la science des données.

La science des données doit s'imprégner des méthodes des disciplines précédentes, comme les mathématiques, sans pour autant s'y réduire : « toutes les théories ne sont pas nécessairement mathématiques. En fait, les théories les plus fondamentales de la science des données ne sont pas mathématiques [7] ».

La science des données doit de plus s'imprégner des disciplines alors en vogue comme l'informatique pour améliorer ses méthodes de recherche ; Cleveland incitant à la collaboration avec les informaticiens ainsi qu'avec les mathématiciens.

La science des données connaît donc de multiples naissances sur une période rapprochée, découlant très nettement d'une volonté de modernisation des méthodes de traitement de données majoritairement tirées des statistiques et de l'analyse de données afin de mieux répondre aux nouvelles thématiques et problématiques du XXIe siècle. Elle constitue en ce sens un ensemble de ruptures : « une rupture des données par rapport à leur origine et leur nature, le mode de collecte les rassemblant en dépit de leur hétérogénéité ; une rupture du traitement par rapport aux données ; une rupture de ce qui est montré par rapport à ce qui est calculé » (Bachimont, 2017).

---

[5] Selon *Google Scholar* en 2024, les articles présentés sont respectivement cités 57 fois (Wu), 230 fois (Hayashi) et 462 fois (Cleveland).

[6] Il ajoute néanmoins que la science des données peut être enseignée à tous, qu'il s'agisse de publics scolaires, universitaires ou de professionnels d'entreprises.

[7] « *Not all theory is mathematical. In fact, the most fundamental theories of data science are distinctly nonmathematical* ».

Malgré cela, un même objectif demeure : extraire des connaissances à partir des données (Dhar, 2013) pour résoudre des problèmes du monde réel (Bhargava et al., 2015). Elle est ce nouveau champ interdisciplinaire bâti sur une synthèse de méthodes issues des statistiques, de l'informatique, de la communication, de la gestion et de la sociologie pour étudier les données et leur environnement dans l'objectif d'aiguiller les décisions futures (Cao, 2017). Pour Ollion (2018), elle est cette continuation du phénomène de quantification du monde dont les scientifiques des données sont les figures de proue, existant pour résoudre les problèmes des utilisateurs des données (Saltz & Stanton, 2017).

> « Plus que tout, ce que font les scientifiques des données, c'est faire des découvertes en nageant dans les données. C'est leur méthode préférée pour naviguer dans le monde qui les entoure. À l'aise dans l'univers numérique, ils sont capables de structurer de grandes quantités de données informelles et d'en rendre l'analyse possible. Ils identifient de riches sources de données, les joignent à d'autres sources de données, potentiellement incomplètes, et nettoient le résultat obtenu. Dans un paysage concurrentiel où les défis ne cessent de changer et où les données ne cessent de circuler, les scientifiques des données aident les décideurs à passer d'une analyse ad hoc à une conversation continue avec les données. [8] » (Davenport & Patil, 2012).

Ces derniers sont en effet souvent perçus dans la littérature comme des experts capables de traiter de larges jeux de données, et en particulier toutes les données numériques comprises dans ce qui a été englobé sous le vocable de *big data*. Cette image de la science des données liée, voire découlant des mégadonnées provient d'un ensemble d'articles promouvant celle-ci pour enrayer les difficultés de traitement des larges jeux de données. L'un des plus connus est celui de Mattmann (2013) intitulé *A vision for data science*. A l'intérieur, il y décrit la nécessité de former de nouveaux types de scientifiques de données (*data scientist*) :

> « Pour résoudre les défis liés au Big Data, les chercheurs ont besoin de compétences à la fois scientifiques et informatiques – une combinaison encore trop rare actuellement. Une nouvelle génération de "scientifiques des données" est nécessaire. En plus d'être des gestionnaires de données, les scientifiques des données développeront des algorithmes sur mesure pour l'analyse et adapteront les formats des fichiers. Ils comprendront les mathématiques, les statistiques et la physique nécessaires pour intégrer des algorithmes scientifiques dans des architectures efficaces. Ils trouveront des solutions au-delà des

[8] « *More than anything, what data scientists do is make discoveries while swimming in data. It's their preferred method of navigating the world around them. At ease in the digital realm, they are able to bring structure to large quantities of formless data and make analysis possible. They identify rich data sources, join them with other, potentially incomplete data sources, and clean the resulting set. In a competitive landscape where challenges keep changing and data never stop flowing, data scientists help decision makers shift from ad hoc analysis to an ongoing conversation with data* ».

> efforts communautaires fragmentés qui ont dominé la dernière décennie de développement des systèmes de mégadonnées [9] ».

Plusieurs auteurs remettront en cause ce lien supposé symbiotique, en particulier Donoho (2017) qui dénonce les discours sur la « nouveauté » du *big data* et sur les techniques proposées pour son traitement par la science des données qui sont à ses yeux dépassées.

En résumé, la science des données émerge dans un objectif de modernisation des méthodes statistiques afin d'une part d'éviter que la discipline ne devienne obsolète (une crainte citée par (Wu, 1998) et (Breiman, 2001)) et d'autre part pour pouvoir traiter les nouvelles problématiques impulsées par les progrès informatiques. Si des passerelles existent entre la science des données et la *data literacy* en raison de la poursuite similaire d'un bon traitement des données, leurs visées diffèrent. La science des données est en effet avant tout un domaine d'experts techniques : elle forme des scientifiques des données (*data scientists*), ces spécialistes capables de traiter des problématiques expertes sur les données. La *data literacy*, elle, produit des lettrés des données (*data literate*) qui ne sont pas censés être des experts. Il s'agit de formations générales qui dépassent la simple question technique. Le niveau de compétence requis diffère par ailleurs complètement entre les deux concepts (Pedersen & Caviglia, 2019).

> « Alors que la science des données est un domaine spécifique réservé aux spécialistes formés, la *data literacy* est présentée comme un élément central dans l'éducation préparant les jeunes à devenir des citoyens dans une société de l'information [10] ».

La *data literacy* est ainsi cette compétence réservée aux non-spécialistes pour traiter les données (Frank et al., 2016). Evoquer des non-spécialistes n'est toutefois pas corrélé avec un enseignement de compétences basiques : la *data literacy*, nous l'avons vu, se constitue en continuum et peut atteindre des niveaux assez avancés de maîtrise technique. La différence fondamentale est que la science des données s'adresse à des professionnels dans un objectif de perfectionnement des prises de décisions, tandis que la *data literacy* s'adresse aux individus et à la volonté de défense de leur libre-arbitre [11].

---

[9] « *To solve big-data challenges, researchers need skills in both science and computing — a combination that is still all too rare. A new breed of 'data scientist' is necessary. As well as being data stewards, data scientists will develop bespoke algorithms for analysis and adapt file formats. They will understand the mathematics, statistics and physics necessary to integrate science algorithms into efficient architectures. They will find solutions beyond the fragmented community efforts that have dominated the past decade of development of big-data systems* ».

[10] « *While data science is a specific domain for trained specialists, data literacy is suggested as a central element in education preparing all young people to become citizens in an information society*».

[11] Nous précisons ici qu'il s'agit d'une évolution récente. La *data literacy* était d'abord adressée aux professionnels pour répondre aux nécessités des entreprises et des institutions, caractéristique qui existe toujours en parallèle des besoins quotidiens des individus.

Par ailleurs, la science des données découle directement des méthodes statistiques et de l'analyse de données, ses auteurs étant statisticiens eux-mêmes, là où la *data literacy* a elle un rapport plus distancié. D'une part parce que les auteurs l'ayant conceptualisé proviennent de domaines académiques variés : si Milo Schield était effectivement statisticien et intégrait sa pensée dans le monde des bibliothèques, ce n'était pas le cas de Nancy Love et d'autres auteurs qui ont repris cette thématique. D'autre part, si les curricula intègrent effectivement une certaine part de méthodes statistiques, elles n'en constituent toutefois pas le pilier fondateur. Néanmoins, dans les deux cas, la science des données et la *data literacy* sont confrontées aux mêmes problématiques, aux mêmes enjeux vis-à-vis des données. C'est leur traitement qui marque leur différence.

3.2.2. L'esprit critique vis-à-vis des données

L'esprit critique est un concept dont la complexité ne peut se réduire à cette sous-partie [12], c'est pourquoi nous résumerons ce qui est entendu par esprit critique dans le cadre des données. Apparaissant courant 2012/2013 dans les définitions de la *data literacy* (Calzada-Prado & Marzal, 2013; Kayser-Bril, 2012), l'esprit critique appliqué aux données (*critical use of data* ou *critical thinking on data*), aussi nommé *data-driven critical thinking* (Holmes, Wieman, & Bonn, 2015), se scinde selon plusieurs approches interdépendantes.

La première approche est basée sur le cycle de vie des données : il s'agit ici d'être en mesure d'identifier la traçabilité des données (*data lineage*), de connaître le contexte de production et de partage des données, de pouvoir identifier les producteurs, de pouvoir analyser les jeux de données et/ou les représentations qui en sont faites (Calzada-Prado & Marzal, 2013). L'aspect critique émerge ici comme questionnement des étapes de ce cycle, de comprendre les objectifs ayant amené à la sélection de telle ou telle démarche.

La seconde approche porte sur la nature même des données. Elle rejette la vision des données vues comme neutres et objectives déjà dénoncée au cours des années 2010 (Gitelman, 2013). Il s'agit de ne pas prendre pour argent comptant les jeux de données mais de comprendre qu'ils sont porteurs d'une intention, conçus avec des logiciels spécifiques et selon une temporalité particulière. C'est d'une part une remise en question des technologies et de leurs structures associées qui façonnent les données et par lesquelles ces dernières transitent, rappelant ici que les données numériques sont tout sauf immatérielles (Crawford, 2022). D'autre part de manier l'hétéroglossie, à savoir « un autre discours dans une autre langue, qui sert à exprimer les intentions de l'auteur mais de façon réfractée [13] » (Bakhtin, Holquist, & Emerson, 1982, p. 324).

[12] Ce que démontre Kaltoum Mahmoudi dans sa thèse *Formes et formule : « former l'esprit critique »* soutenue en 2022.
[13] «*Another's speech in another's language, serving to express authorial intentions but in a refracted way* ».

La troisième approche est basée sur la mise en capacité politique des individus, basée sur les travaux de Paulo Freire qui constituent le socle des *critical data studies* (Dalton & Thatcher, 2014) et des *critical data literacies* (Tygel & Kirsch, 2016) sur lesquelles nous reviendrons au chapitre 4. Paulo Freire (1921 - 1997) était un pédagogue brésilien ayant mis en avant la nécessité d'alphabétisation et de développement d'une pensée critique chez les individus pour résister à toute tentative d'oppression, résumée dans son ouvrage phare *La pédagogie des opprimés* publié en 1968. La pensée critique de Freire s'inscrit dans un contexte politique particulier, à savoir la dictature militaire du maréchal Castelo Branco qui sévit alors de 1964 à 1985. L'année 1968 marque un tournant dans l'histoire du pays avec la promulgation de l'acte institutionnel n°5 qui dissout le Congrès, suspend la Constitution, supprime les libertés individuelles et impose la censure (Chirio, 2007). Freire notait que l'ignorance et la léthargie des brésiliens face à cette situation résultaient directement de la domination économique, sociale et politique ainsi que du « paternalisme » de l'époque, les maintenant « submergés » sans possibilité d'éveiller leur esprit critique ; ce qui est nommé par Freire la « culture du silence » (Freire, 1968). Et dans le cas où ils parvenaient à échapper à l'oppression, c'était pour devenir eux-mêmes des oppresseurs, ayant été conditionnés ainsi ; la peur de la liberté bloquant tout processus de changement. Pour Freire, la liberté se base donc sur la nécessité de rejeter l'image de l'oppresseur et de la remplacer par l'autonomie et la responsabilité. Surmonter l'oppression nécessitait en premier lieu d'en connaître les causes, de déceler les faiblesses de l'oppresseur.

La pensée critique pour Freire provient du dialogue et plus particulièrement du mot dont il en est l'essence. Celui-ci possède deux facettes : l'action et la réflexion. Privilégier la réflexion amène au verbalisme, ce moment où les mots deviennent vides et incapables de dénoncer le monde et d'agir sur lui. Privilégier l'action amène à l'activisme, rendant le dialogue inaudible. Ces deux méthodes créent des formes d'existences et de pensées fausses. Ce qu'il faut pour Freire, c'est prononcer le mot véritable, la véritable *praxis*, pour changer le monde. Le dialogue entre les hommes est donc une nécessité « existentielle ». Le dialogue est la rencontre d'hommes et de femmes qui nomment le monde : il n'est ni un dépôt d'idées, ni un simple échange d'idées. Dialogue qui ne peut exister ni sans esprit critique, à savoir une pensée qui ne discerne « aucune dichotomie entre le monde et les hommes et qui voit la réalité comme une chose en perpétuel mouvement », ni sans une humilité et un amour réciproques.

Ainsi, selon cette approche, il s'agit pour les individus de décoder les enjeux de pouvoirs grâce aux données, de lire le mot et le monde par les données pour assurer les traitements techniques d'une part et développer une autonomie vis-à-vis des idéologies et idées reçues d'autre part. C'est comprendre le rôle social porté par les données et pouvoir s'interroger sur celui-ci. C'est être capable de déjouer les stratégies de *dataveillance*, nommées ainsi par Roger Clarke et définies par Raley (2013) comme « la pratique disciplinaire et de contrôle consistant à surveiller, agréger et trier les données [14] ».

[14] « *The disciplinary and control practice of monitoring, aggregating, and sorting data* ».

Ces trois approches se confondent aujourd'hui dans l'ensemble des démarches de la *data literacy*. C'est avant tout une posture, un état d'esprit déployé(es) pour traiter et comprendre les jeux de données afin de mieux appréhender les difficultés disséminées dans le quotidien des individus.

3.2.3. La gouvernance des données

Un troisième concept central est la gouvernance des données (*data governance*). Celle-ci est définie comme l'exercice de l'autorité et du contrôle sur la gestion des données par l'implémentation d'un système de normes et de procédures (Abraham, Schneider, & vom Brocke, 2019; Ladley, 2019; Plotkin, 2013). Son objectif central est de constituer une culture organisationnelle où les données sont traitées comme un atout et dont l'exploitation doit aiguiller les décisions (Janssen, Brous, Estevez, Barbosa, & Janowski, 2020).

Pour Correia & Água (2021), elle vise la mise en place de conditions favorables à une bonne prise de décision, le soutien à la conformité réglementaire et à la réduction des risques en ce qui concerne la sécurité et la confidentialité des données, l'amélioration des performances des entreprises et leur intégration, et enfin l'accroissement de l'harmonisation des objectifs commerciaux et de ceux de la DSI (*IT-business alignement*). En tant que gouvernance, elle élabore et gère une stratégie globale de long terme (Weber, Otto, & Österle, 2009) dont elle s'assure le respect de la conformité (Fu, Wojak, Neagu, Ridley, & Travis, 2011).

Selon les auteurs, elle englobe ou est englobée par la gestion des données (*data management*, parfois confondu avec l'intendance des données, *data stewardship*) qui se présente comme l'application au quotidien desdits principes de la gouvernance par l'apport d'un soutien logistique (Alhassan, Sammon, & Daly, 2018; Khatri & Brown, 2010; Rafal & Girard, 2023). La maximisation des décisions doit se faire à partir de « bonnes » données issues d'un traitement garantissant leur qualité, ce que l'on nomme la qualité des données (*data quality*), aussi nommée gestion de la qualité des données (*data quality management*).

Sa première définition provient du monde de l'entreprise par le truchement de Gwen Thomas, une ingénieure alors spécialisée dans les systèmes d'information, le *Knowledge management* et les bases de données, qui a fondé en 2003 le *Data Governance Institute* [15]. Par son intermédiaire, elle va proposer un cadre de référence en gouvernance des données[16] en 2004 qui sera réédité de nombreuses fois.

La nécessité de la gouvernance des données apparaîtra véritablement en 2006 dans le monde de l'entreprise, année où nous assistons à une émergence exponentielle d'articles et d'ouvrages sur le

[15] https://datagovernance.com/about-us/
[16] https://datagovernance.com/the-dgi-data-governance-framework/

sujet. Avec Thomas (2006) dans son ouvrage *Alpha males and data disasters : the case for data governance* dans lequel elle définit la gouvernance des données comme la gouvernance des hommes et de la technologie :

> « Les données doivent être gouvernées car elles n'ont aucune volonté, ni aucune intention. Les outils et les individus façonnent les données et leur disent où aller. De ce fait, la gouvernance des données est la gouvernance des individus et de la technologie [17] ».

Elle précise plus loin que « la gouvernance des données est un système de droits de décision et de responsabilités pour les processus liés à l'information, exécutés selon des modèles convenus qui précisent qui peut entreprendre quelles actions avec quelles informations, et quand, dans quelles circonstances, en utilisant quelles méthodes [18] ».

Cohen (2006), un spécialiste de la gestion de l'information (*information management*), dans son article *BI Strategy : What's in a Name? Data Governance Roles, Responsibilities and Results Factors* apportera sa propre définition. Il considérait ainsi la gouvernance des données comme le « processus par lequel une entreprise gère la quantité, la consistance, l'utilisabilité, la sécurité et la disponiblité des données [19] ».

Ou encore par Newman & Logan (2006) qui, s'ils parlent de gouvernance de l'information (*information governance*), inclueront l'intendance des données (*data stewardship*) comme axe central pour s'assurer de la consistance et de l'exactitude des données.

Ainsi, la gouvernance des données émerge d'abord dans le monde de l'entreprise avec pour objet la poursuite de l'objectif de maîtrise des données et de leur qualité recherché dès les années 1960/70, avant d'atteindre les sphères académiques un an plus tard. L'article de Wende & Otto (2007) est ainsi l'un des premiers à être publié tout en discutant de l'application de la gouvernance des données dans un milieu entrepreneurial et de l'extrême nécessité d'implémenter et pérenniser une bonne gestion de qualité des données (*data quality management*).

De nos jours, selon Obendiek (2022), quatre visions de la gouvernance des données liées à quatre visions spécifiques des données peuvent être décrites :

---

[17] « *Data needs to be governed as it has neither will nor intent of its own. Tools and people shape the data and tell it where to go. Therefore, data governance is the governance of people and technology* ».

[18] « *Data governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods* ».

[19] « *The process by which a company manages the quantity, consistency, usability, security and availability of data* ».

« Premièrement, une conceptualisation des données en tant que question de droits individuels associe les droits de l'homme et la promotion de la souveraineté à une vision de la gouvernance des données en tant que libéralisme local. Deuxièmement, les partisans d'un partenariat de sécurité promeuvent une coopération internationale de sécurité reposant sur la conceptualisation des données en tant qu'instrument neutre. Troisièmement, une conceptualisation des données en tant que ressource économique est associée à une vision de l'économie numérique qui soutient le progrès et l'innovation avec une réglementation limitée. Quatrièmement, une conceptualisation des données en tant que ressource collective associe les valeurs des droits universels et des règles mondiales à une vision de la protection internationale ».

Nous l'avons évoqué au chapitre précédent, la gouvernance des données prend la suite de l'administration des données (*data administration*) ayant eu cours des années 1960 jusque dans les années 1990 (avec une rémanence au début des années 2000) :

« Au cours des dernières décennies du siècle précédent, de nombreuses organisations ont reconnu la pertinence de créer une fonction d'administration des données (AD) sous la supervision des ressources d'information de l'entreprise (Holloway, 1986). La pertinence de cette fonction anticipait l'importance contemporaine accordée à la gouvernance des données. Le rôle de l'administration des données était de promouvoir la planification et la coordination de l'utilisation des ressources d'information au sein de l'organisation, entre les applications et les domaines d'activité associés. Ce faisant, le partage des données pouvait être maximisé et la redondance des données minimisée. Les administrateurs de données rendent les données partageables et cohérentes entre les applications en utilisant la modélisation logique des données. Ils ont veillé à ce que plusieurs autres tâches soient exécutées, comme par exemple : le recueil des exigences commerciales, l'analyse des exigences, la modélisation commerciale basée sur les exigences, la définition et l'application de normes et de conventions concernant les noms et les termes, la collecte des données des utilisateurs, la définition, la gestion et l'intendance des dépôts de métadonnées et des outils de modélisation de données. De plus, l'AD a pris en charge la fonction technique de l'administration des bases de données sur la créa-

> tion de bases de données physiques à partir de modèles logiques [20] » (Correia & Água, 2021).

L'administration des données était assurée par un administrateur des données (*data administrator*), parfois confondu avec l'administrateur des bases de données (*database administrator*) du fait que la base de données était l'élément central et centralisateur des traitements des données. Cette centralisation était principalement due aux limitations techniques de l'époque. Avec les progrès de l'informatique et plus globalement du numérique, la donne change et de nouvelles possibilités mais également de nouvelles problématiques émergent, notamment dans le cadre du *big data* comme nous avons pu en discuter. La nécessité de la *data literacy* pour l'édification et la pérennisation de la gouvernance des données commence à être soulignée par plusieurs auteurs assez tardivement, principalement à partir du milieu des années 2010.

Koltay (2016) évoque la gouvernance des données et la *data literacy* comme « deux blocs constitutifs importants du socle de connaissance des professionnels de l'information impliqués dans le soutien des recherches intensives des données, et les deux traitent la qualité des données et la gestion des données de la recherche [21] ». Dans ce contexte, la *data literacy* se présente comme un socle de connaissances permettant de comprendre l'ensemble des enjeux liés aux données grâce à l'assimilation de savoirs théoriques (ex : comprendre ce qui est entendu par « donnée ») et techniques (ex : savoir nettoyer des données). La gouvernance ne pouvant en effet pas s'installer sans une connaissance et une vision harmonisée des personnes concernées.

Pour l'entrepreneur Jordan Morrow (2021), la *data literacy* est vitale à la gouvernance des données selon quatre points : (1) les personnes en charge de la stratégie de la gouvernance doivent être *data literates* pour éviter tout risque de mésusages des données et éviter que la stratégie ne s'effondre, (2) être *data literate* permet de connaître les capacités techniques requises et la manière de contourner les problèmes liés aux données, (3) la *data literacy* permet d'encapaciter les individus en leur permettant de comprendre les grands principes à l'œuvre dans la gouvernance des données comme l'intendance des données (*data stewardship*) et la qualité des données (*data quality*), et (4) la *data literacy* per-

[20] « *In the last decades of the previous century, many organizations recognized the relevance for creating the data administration (DA) function under the supervision of the corporate resources of information (Holloway, 1986). The relevance of this function anticipated the nowadays importance given to data governance. The role of data administration was to promote the planning and coordination of the information resource usage across organization, among related applications and business areas. By doing so, data sharing could be maximized and data redundancy minimized. Data administrators make data sharable and consistent across applications by using logical data modelling. They ensured that several other tasks were performed, as for instance: gathering business requirements, requirements analysis, business modelling based on requirements, definition and enforcement of standards and conventions regarding names and terms, collecting users' data definition, management and stewardship of the metadata repository and data modelling tools. Furthermore, DA supported the technical function of database administration on creating physical databases from logical models* ».

[21] « *Data governance and data literacy are two important building blocks in the knowledge base of information professionals involved in supporting data intensive research, and both address data quality and research data management* ».

met aux employés de comprendre la technologie et les logiciels utilisés pour la conduite de stratégies analytiques (*analytic strategies*).

Tout individu d'une organisation mettant en place une gouvernance doit être « enculturé » au sens d' Herskovits (1952) : « dans ses aspects ultérieurs, lorsque l'enculturation opère au niveau de la conscience, elle ouvre la porte au changement, en permettant l'examen de possibilités diverses et le reconditionnement à de nouveaux modes de pensée et de conduite » (p.173). L'enculturation est le « mécanisme d'accomplissement du changement » (p. 219) où la personne adulte décide si elle accepte ou non une connaissance nouvelle qui va changer sa perception du monde. En effet, la gouvernance implique une nécessité de changement en termes de pratiques mais également de visions : le concourt de l'individu est ici central et une connaissance des caractéristiques et des enjeux de celle-ci est nécessaire.

Ainsi la *data literacy* se présente comme un catalyseur, un soutien nécessaire de la gouvernance des données par une encapacitation des individus. Si elle est globalement sollicitée par les auteurs, elle n'est toutefois pas nécessairement intégrée dans les modules constitutifs de la gouvernance : tout dépend des auteurs sollicités (Ladley (2019) l'intègre par exemple dans ses éléments de formation).

Ce qui doit être toutefois soulevé ici sont les biais potentiels embarqués par les formations en *data literacy* en faveur d'une certaine vision de la gouvernance des données formalisée par l'organisation destinataire, ne laissant pas la place aux doutes et aux remises en question ; la *data literacy* s'apparentant alors à un outil de propagande employé par l'organisation et faisant écho à la fonction de contrôle de la littératie évoquée au chapitre 2.

3.2.4. L'éthique des données

Cette sous-partie reprend des éléments apportés dans notre article sur les chartes éthiques des données Verdi (2023a). L'éthique apparaît dans les définitions de la *data literacy* à partir 2013 (Calzada-Prado & Marzal, 2013). Toutefois elle n'est pas développée, seulement sollicitée comme ensemble de pratiques responsables à suivre lors des traitements des données (Koltay, 2015) (Wolff, Gooch, et al., 2016). Pour les comprendre, il est nécessaire de rechercher leurs caractéristiques précises au sein de l'éthique des données elle-même.

Pour Floridi & Taddeo (2016), elle est « une nouvelle branche de l'éthique qui étudie et évalue les problèmes moraux liés aux données (ce qui comprend la production, l'enregistrement, la curation, le traitement, la dissémination, le partage et l'utilisation), aux algorithmes (ce qui comprend l'intelligence artificielle, les agents artificiels, le *machine learning* et les robots) et leurs pratiques associées (ce qui comprend l'innovation, la programmation et le piratage responsables ainsi que les codes de conduites

professionnelles), afin de formuler et de soutenir des solutions bonnes moralement (ex : les bonnes conduites ou les bonnes valeurs) [22] ».

En tant qu'éthique, il s'agit d'une visée (Ricœur, 1989), à l'inverse de la morale qui est « la région des normes, autrement dit des principes du permis et du défendu, d'autre part, le sentiment d'obligation en tant que face subjective du rapport d'un sujet à des normes » (Ricoeur, 2001, p. 55). Elle recouvre « les constructions empiriques qui correspondent aux besoins de valeurs accompagnant les activités humaines » et désigne « l'ensemble des principes et valeurs guidant des comportements sociaux et professionnels » (Farjat, 2004). Elle peut également englober le principe de responsabilité et le respect de l'intérêt des personnes (Simonnot, 2018). Dans son étude des éthiques de l'intelligence artificielle, Ménissier (2023) rappelait que l'éthique incorporait : « (1) la définition d'un sens qui apparaît éclairant pour l'action ; (2) le choix de valeurs considérées comme bonnes ; (3) la formulation de principes et de règles. ». Vitali-Rosati (2012) soulignait de son côté l'importance du sens dans l'emploi de l'éthique :

> « Mais, en tant que réflexion sur l'action, l'éthique va plus loin : on se rend vite compte que, avant de produire une série de normes à respecter, il faut en chercher le sens. La question éthique devient ainsi une question de sens. Le "comment" se transforme en "pourquoi". L'éthique, plus qu'ensemble de normes, devient la recherche d'une justification de ces normes ».

L'éthique des données doit donc permettre d'introduire et d'évaluer des principes intervenant sur tout le cycle de vie des données ; chaque choix devant être justifié afin de produire du sens pour les individus. La compréhension des effets des données sur la société (Arruda, 2020) doit amener en conséquence à prendre des décisions « éclairées » (Hand, 2018). La synthèse opérée par Harfoush (2021) souligne la place centrale de l'individu dont le respect des droits accompagne le partage des données (voir le tableau 3.1).

Tableau 3.1.: Principes généraux en éthique des données - extrait d'Harfoush (2021), traduction de (Verdi, 2023a)

Principes	**Description**
La transparence	nécessité d'informer les individus sur le but de la collecte des informations.

[22] « *A new branch of ethics that studies and evaluates moral problems related to data (including generation, recording,curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values)* ».

Principes	Description
Le choix	possibilité des individus de se retirer.
Le consentement	pouvoir donner ses données à des tiers s'ils respectent les points ci-dessus.
La sécurité	prendre des mesures pour protéger les données personnelles contre la perte, les mauvais usages, les accès non autorisés, la divulgation, l'altération et la destruction.
L'intégrité	s'assurer de la fiabilité des informations personnelles et de leur bon usage.
L'accès	permettre aux individus d'accéder à leurs données personnelles.
La responsabilité	l'organisme doit être responsable du respect des règles et de la conformité.

Au-delà des individus, l'éthique des données soutient la promotion et la préservation d'un engagement responsable avec l'environnement[23]. Pour Pothier & Condon (2020), le respect de « l'environnement » est à comprendre selon deux significations : le respect des personnes qui nous entourent et le respect de pratiques écologiques. Cet environnement est souvent vu comme numérique dans les discours. En cela, éthique des données et éthique du numérique entrent en résonance, voire sont intégrées l'une dans l'autre.

Pour Vitali-Rosati (2012), « l'objectif d'une éthique du numérique n'est pas d'appliquer au domaine des technologies de l'information et de la communication des principes éthiques généraux. La tâche, beaucoup plus difficile, d'une éthique du numérique est de comprendre ce que devient l'action à l'époque du numérique, ce que sont ses présupposés et ses bases ». Floridi, Cath, & Taddeo (2019) reprennent ici cette idée : pour eux, il s'agit non pas de se concentrer sur une technologie spécifique (comme le *cloud computing*, l'internet des objets ou l'intelligence artificielle), mais sur l'ensemble de

[23] « L'éthique des données se centre plus sur l'appropriation adaptée des données pour préserver et promouvoir un meilleur engagement, plus responsable vis-à-vis de ce qui nous entoure en partant des personnes jusqu'à l'environnement », notre traduction de « *Data ethics is more about the proper handling of data to preserve and promote a responsible and better engagement with our surroundings from people to the environment*. » (Harfoush, 2021).

l'écosystème créé et manipulé par toute technologie numérique. Il s'agit de produire une « approche holistique » qui interroge les problèmes moraux causés par les innovations numériques.

Pour le *Comité National Pilote d'Éthique du Numérique* (*CNPEN*)[24], c'est prendre en compte la « profonde conversion numérique » : « un questionnement éthique sur des sciences et les technologies peut se rapporter à des moments singuliers de l'existence humaine comme la procréation ou la mort, mais aussi concerner tout moment du quotidien. C'est le cas de l'éthique du numérique qui interroge en permanence nos valeurs à l'aune de situations inédites d'usage des technologies, avec une dimension radicalement nouvelle d'ubiquité et d'universalité de leurs effets. » (Comité national pilote d'éthique du numérique, 2022, p. 239).

Un écosystème qui incorpore les données mais également d'autres éléments. Floridi et al. (2019) y incorporent les données, l'information, les algorithmes ainsi que les pratiques et infrastructures associées. D'autres auteurs intègrent en plus l'intelligence artificielle. Pour Kazim & Koshiyama (2021), l'éthique de l'IA traite les problématiques de l'intelligence artificielle engendrées par les mauvais usages des technologies ou les défauts embarqués dans les technologies dès leur conception (ex : les mauvaises reconnaissances faciales). Ces technologies sont pour eux les suivantes : la *blockchain*, l'Internet des objets, le *big data analytics*, l'IA, le *machine learning* et les algorithmes associés. Ainsi, pour Heilinger (2022), l'ensemble des problématiques centrales auxquelles l'éthique de l'intelligence artificielle doit répondre vis-à-vis de ces technologies sont les suivantes :

- Comment comprendre, expliquer et contrôler les mécanismes internes des systèmes d'IA ?
- Qui porte la responsabilité des conséquences résultants d'un mauvais usage d'IA ?
- Comment éviter d'exacerber les biais discriminants contenus dans les données d'entraînement ?
- Comment s'assurer de la protection de la vie privée des individus malgré la facilité de collecte et d'analyse des données à caractère personnel ?
- Comment protéger la prise de décision autonome humaine des influences indues et la disqualification résultant de l'utilisation de l'IA ?
- Comment empêcher l'émergence d'une intelligence néfaste qui mettrait en avant ses propres objectifs avant ceux des humains ?

Zacklad & Rouvroy (2022) ont développé de leur côté une éthique située de l'intelligence artificielle qui traite les enjeux éthiques « liés à l'automatisation des dispositifs d'information et de communication qui affectent les comportements des personnes dans leur vie quotidienne ». En tant qu'éthique située, elle va « consister à remettre en cause les présupposés scientifiques des promoteurs de la décision automatique en contestant la possibilité qu'il existe des données objectives et en montrant

[24]Mis en place en 2019 en France par le Premier ministre Édouard Philippe sous l'égide du *CCNE* (comité consultatif national d'éthique pour les sciences de la vie et de la santé) et sur recommandation du député Cédric Villani dans son rapport « donner une sens à l'intelligence artificielle ».

le caractère intrinsèquement opaque de l'apprentissage profond » et « en ouvrant des espaces de controverses scientifiques masquées, telles des boîtes noires, par les promoteurs de la technologie inéluctable ». Il s'agit de dépasser « l'amnésie des conditions de production des données » qui font des données numériques des faits en raison de leur naturalisation, évacuant les effets et les rapports de force des réalités sociales dans lesquelles elles s'inscrivent. De développer une encapacitation critique au sens de Freire, à savoir de souligner les enjeux politiques contenus dans les données et ainsi de pouvoir changer le monde.

Dans ce contexte, la question des algorithmes est prégnante. Et il nous faut ici citer l'éthique des algorithmes qui selon Floridi & Taddeo (2016) traite de la complexité grandissante des algorithmes ainsi que de leur autonomie. Pour Tsamados et al. (2022) est considéré comme algorithme ce qui utilise les données pour un résultat spécifique qui servira à déclencher et motiver une action avec des conséquences éthiques potentielles. Au même titre que les données qui sont pour eux une ressource symbiotique, les algorithmes permettent « d'intervenir de façon "agile" sur le monde, ils transforment le monde en sélectionnant, parmi tous les "possibles", ceux qui sont les plus conformes à leur fonction objective » (Zacklad & Rouvroy, 2022).

En effet, du fait de l'importance de leur périmètre d'action qui s'étend « bien au-delà de la seule organisation raisonnée des données » (Le Deuff & Roumanos, 2022) pour jouer un rôle central dans le traitement informationnel, leur influence sur nos conceptions du monde est considérable (Cardon, 2015; Dogruel, 2021). Ont été régulièrement mis en avant les effets délétères de leurs usages du fait de biais incorporés dans leur conception et dans ceux des jeux de données qu'ils traitent (O'Neil, 2017), ainsi que leur opacité (Pasquale, 2015) (contre laquelle la transparence n'est plus à même de garantir leur compréhension (Ouyang, 2020)), prévenant ici une potentielle « gouvernementalité algorithmique » (Rouvroy & Berns, 2013) défectueuse rendue possible par la généralisation de leur automatisation.

> « L'algorithme contient, dans son agencement mathématique et spatial, de multiples potentiels de cruautés arbitraires, de surprises, de violences, de joies, de distillations de racisme et de préjugés, d'injustices, de probabilités et de hasard. [25] » (Amoore, 2020).

L'éthique des algorithmes se pose donc en garde-fou pour s'assurer de leur responsabilité (*accountability*) (Amoore, 2020), de la transparence de leurs objectifs, et de limiter au possible les problèmes de discrimination et d'apophénie [26] :

[25] « *Algorithm contains, within its mathematical and spatial arrangement, multiple potentials for arbitrary cruelties, surprises, violences, joys, distillations of racism and prejudice, injustices, probabilities, and chance* ».

[26] Définie par Klaus Conrad en 1958, cité par Brugger (2001), l'apophénie est le fait de percevoir des connexions de façon non motivées qui s'accompagnent d'une expérience spécifique d'une signification anormale. Appliqué aux traitements des données, il s'agit de percevoir des tendances dans les grands jeux de données (*big data*) grâce à un ensemble de corrélations établies par les algorithmes sans qu'elles soient nécessairement vraies. Il est à noter que le fait d'employer les

> « Une donnée n'est plus qu'un signal expurgé de toute signification propre – et c'est bien sûr à ce titre que nous tolérons de laisser ces traces, mais c'est aussi ce qui semble assurer leur prétention à la plus parfaite objectivité : aussi hétérogènes, aussi peu intentionnées, tellement matérielles et si peu subjectives, de telles données ne peuvent mentir ! » (Rouvroy & Berns, 2013).

3.2.5. Vers une convergence des concepts

Ainsi, qu'il s'agisse de l'éthique de l'IA, des algorithmes ou encore des données, les mêmes principes et interrogations demeurent. Au-delà des questionnements communs comme la transparence, la non-neutralité, les mauvais usages ou encore les biais embarqués amenant à un ensemble de discriminations, se pose la difficulté ici de leur enchevêtrement. Floridi & Taddeo (2016) précisaient que :

> « Bien qu'il s'agisse d'axes de recherche distincts, l'éthique des données, des algorithmes et des pratiques sont évidemment imbriquées, et c'est pourquoi il peut être préférable de parler en termes de trois axes définissant un espace conceptuel au sein duquel les problèmes éthiques sont comme des points identifiés par trois valeurs [27] ».

Cette notion d'espace conceptuel est à prendre en compte quand sont abordées les données dans les discours. L'auteur central de cette notion est le philosophe Peter Gärdenfors qui l'a développé dans son ouvrage *Conceptual spaces: the geometry of thought* (Gärdenfors, 2000). Il définit l'espace conceptuel comme un outil de nature épistémologique pour modéliser des relations à travers nos expériences (la manière dont nous percevons, nous nous rappelons ou imaginons) (Gärdenfors, 2000, p. 14). L'espace conceptuel doit permettre de contourner le problème de représentation des concepts en adoptant une approche géométrique [28].

Selon cette vision, les concepts ne sont pas indépendants les uns des autres. Ils peuvent être structurés en domaines basés sur des dimensions de qualité (*quality dimension*). C'est grâce à ces dernières

---

algorithmes pour donner du sens aux courants de données est le socle de la discipline des *data analytics* selon Mittelstadt, Allo, Taddeo, Wachter, & Floridi (2016).

[27] « *While they are distinct lines of research, the ethics of data, algorithms and practices areobviously intertwined, and this is why it may be preferable to speak in terms of three axes defining a conceptual space within which ethical problems are like points identified by three values* ».

[28] Gärdenfors dissocie ici sa vision des deux approches développées par les sciences cognitives pour modéliser des informations, à savoir l'approche symbolique basée sur une manipulation de symboles, et l'approche associationniste basée sur les associations entre différents éléments informationnels ; le connexisme, qui modélise les associations au moyen de réseaux de neurones artificiels, est un type d'associationnisme. Ces trois approches (symbolique, associationniste et géométrique) ne sont pas concurrentes : elles sont pour Gärdenfors trois niveaux de représentation de la cognition qui possèdent des échelles de résolution différentes (Gärdenfors, 2001).

que les domaines peuvent être créés et employés pour représenter les concepts. La distance entre ces dimensions permet de déterminer la similarité entre les concepts :

> « Au cœur du cadre des espaces conceptuels se trouve l'idée selon laquelle les concepts peuvent être représentés dans des espaces de similarité. Les espaces de similarité sont des objets mathématiques, en particulier des espaces métriques à une ou plusieurs dimensions dont la ou les dimensions représent(ent) des qualités fondamentales selon lesquelles les éléments peuvent être comparés. Les distances dans un tel espace représentent des dissemblances, dans le sens où plus les éléments sont éloignés (tels que représentés dans l'espace), plus ils sont dissemblables dans l'aspect spécifique que l'espace est censé modéliser (qui pourrait être la couleur, le goût, l'odeur, etc.) [29] » (Douven, Elqayam, Gärdenfors, & Mirabile, 2021).

Comme le rappelle Gärdenfors, les dimensions forment le cadre utilisé pour assigner des propriétés aux objets et pour spécifier des relations entre eux (Gärdenfors, 2001). Dans cet espace conceptuel, les données sont de moins en moins citées selon un mode individuel, siloté et séparé du reste des éléments numériques précédemment cités comme les algorithmes ou l'intelligence artificielle. Elles sont au contraire comprises dans un tout, dans un *système*[30]. Nous avions pu le constater dans notre article (Verdi, 2023a) où les chartes éthiques des données ne se concentraient pas que sur les données mais sur les éléments connexes comme les algorithmes, le terme de donnée se dissociant « des » données. Il en est de même pour les différentes éthiques précédemment citées et plus globalement dans les discours où le terme *data* tend à englober une réalité technique et conceptuelle plus large.

C'est ce que nous pourrions nommer un « système-donnée », résultant d'interconnexions constantes et évolutives entre les données et les élément suscités. Il se rapprocherait du système de données (*data system*) tel que définit par Beynon-Davies (2021):

> « Nous employons le terme de système de données pour désigner tout système qui implique la manipulation ou l'articulation des structures de données. Une grande partie de ce que nous appelons à tort "technologie de l'information" se résume en fait à des

[29] « *Central to the conceptual spaces framework is the thought that concepts can be represented in similarity spaces. Similarity spaces are mathematical objects, specifically one-or multidimensional metric spaces whose dimension or dimensions represent fundamental qualities along which items may be compared. Distances in such a space represent dissimilarities, in that the further apart items are (as represented in the space), the more dissimilar they are in the specific aspect the space is supposed to model (which could be color, taste, smell, and so on)* ».

[30] Le système a fait l'objet de nombreuses publications et définitions notamment de la part de Bertrand Gille, Dominique Bourg ou encore Talcott Parsons. La synthèse apportée par Ellul & Porquet (2012) liste les caractéristiques suivantes : « le système est un ensemble d'éléments en relation les uns avec les autres de telle façon que toute évolution de l'un provoque une évolution de l'ensemble, toute modification de l'ensemble se répercutant sur chaque élément » (p. 88). Les éléments d'un système font donc partie d'un réseau « d'inter-relations » dynamique et évolutif qui, en tant que globalité, peut entrer en relation avec d'autres systèmes, d'autres globalités.

> ensembles de technologies destinées au traitement, à la manipulation des données ou à ce que nous préférons qualifier d'articulation. Mais les systèmes de données incluent également l'utilisation de technologies de bas niveau comme les formulaires papiers, les tableaux blancs, les *post-it* ou même les ficelles nouées et les jetons en argile. Au même titre que notre définition d'un système d'activités et d'un système d'information, nous définissons un système de données comme un ensemble de modèles d'articulations interdépendants. À son tour, un modèle d'articulation consiste en un ensemble d'actes d'articulation interdépendants [31] » (p. 245).

Ce système de données résulte pour Beynon-Davies de l'articulation ininterrompue d'une multitude de structures de données (*data structures*)[32] : c'est l'articulation à l'oeuvre dans les données (*data articulation*) qui fait système. Nous pourrions ici réemployer cette conception en dépassant le simple cadre des données par un rappel de l'interaction constante avec les algorithmes et l'intelligence artificielle, tant d'un point de vue technique que conceptuel (en reprenant ici la notion d'espace conceptuel suscitée), supervisée par l'activité humaine.

Ces éléments serviraient de base à une vision désilotée permettant d'une part l'inclusion d'éléments autres que les données et d'autre part de rappeler le rôle central de l'activité humaine, offrant ici la possibilité de sortir du *pharmakon* et du déterminisme technique des techno-discours où les données semblent avoir une existence propre. C'est également dépasser la vision de la donnée comme tenant

[31] « *We use the term data system to denote any system which involves the manipulation or articulation of data structures. Much of what we mistakenly call 'information' technology effectively amounts to sets of technologies for data processing, manipulation or what we prefer to denote as articulation. But data systems also include the use of low-level technologies such as paper forms, whiteboards, Post-it notes or even knotted strings and clay tokens. In a similar manner to our definition of an activity system and an information system, we define a data system as a set of inter-related patterns of articulation. In turn, a pattern of articulation consists of a set of inter-related acts of articulation* ».

[32] Sur la base d'une étude de la littérature, Beynon-Davies (2016) a proposé la définition suivante : « Une structure de données est à la fois une abstraction et une instantiation. En tant qu'abstraction, une structure de donnée, telle une liste, intègre un ensemble de principles pour le stockage et l'accès aux données. En tant qu'instanciation, une structure de données s'incarne / prend une forme. Une instance spécifique d'une liste, telle qu'une liste de localisation des stocks, est utilisée pour représenter des choses et une telle représentation aide à la constitution d'un ordre institutionnel. En tant qu'abstraction, une structure de données peut être considérée comme un ensemble d'éléments de données, qui à leur tour comprennent un ensemble d'items de données (Tsitchizris & Lochovsky, 1982) », notre traduction de « *A data structure is both an abstraction and an instantiation. As an abstraction a data structure, such as a list, comprises a set of principles for both storing and accessing data. As an instantiation a data structure is given form. A specific instance of a list, such as a stock-location list, is used to represent things, and through such representation to help constitute institutional order. As an abstraction a data structure can be seen as a set of data elements, which in turn comprise a set of data items (Tsitchizris & Lochovsky, 1982)* ». L'ordre institutionnel cité ici se base sur les écrits de Searle où le monde social ou institutionnel (ontologie sociale) se construit à partir du monde physique ou matériel (ontologie physique). Il y a par ailleurs une différenciation entre les élements de données (*data element*) et les items de données (*data items*), le premier englobant le second. L'exemple du khipu est présenté : le khipu dans son ensemble représente la structure des données, un ensemble de cordelettes représente un élément de données, et une cordelette spécifique représente un item de données.

et aboutissant de toute décision, de toute logique. Enfin, cette conceptualisation du *système-donnée* devrait s'inscrire dans son enseignement au sein d'une culture technique au sens simondien[33] : de rappeler que ce système est par essence technique et surtout matériel, loin des visions immatérielles des données et plus gobalement sur le numérique embarquées dans certains discours, ce que précisaient Jeanneret & Souchier (2002) : « la notion d'immatériel rend très mal compte de ce qui est en jeu et masque l'importance des propriétés des supports, le poids des contextes matériels d'utilisation et le pouvoir des formes imposées par les inscriptions, les matrices et les cadres de la communication » (p. 20).

Nous précisions néanmoins que le terme « donnée » dans « système-donnée » ne nous satisfait pas car il fait de la donnée le tenant et aboutissant du système, ni celui de « numérique » qui englobe une réalité plus large. Nous serions plus enclins à parler de « système d'information » selon sa conception commune, à savoir (Piccoli & Pigni, 2018) un système organisationnel, formel et socio-technique conçu pour collecter, traiter, conserver et répandre l'information [34]. Pour Stair & Reynolds (2020), un système d'informations est géré par un *chief information officer* (CIO) et doit soutenir les activités fondamentales des organisations, en particulier tout ce qui relève des affaires (*business*), grâce à la production de rapports, de visualisations et d'analyses des données (p. 4). Il se compose de personnes qui imaginent, mettent en oeuvre et supervisent le système d'information, d'infrastructures (les matériels, les logiciels, les bases de données, les réseaux, les usines, et tout un ensemble de services), de processus et de procédés, de matériaux (les données) et de structure organisationnelle au sens de la répartition des rôles et des responsabilités (pp. 6-8).

En lien avec ce qui a été dit précédemment sur le « système-donnée », il s'agirait de revoir le système d'information, que nous renommerions le « système-information », comme outil de réflexion qui intégrerait les notions de « champ conceptuel » et de « système » comme socle réflexif permettant d'articuler les données, les algorithmes, l'intelligence artificielle et d'autres potentiels concepts au sein d'un même espace afin de pouvoir agir à la fois d'un point de vue technique et logistique (ex : les ordinateurs, les serveurs, les logiciels, la programmation des algorithmes, le traitement des données, etc.) mais également d'un point de vue métiers, à savoir de questionner, voire de refondre, des postes comme les *chief information officer* et les *chief data officers*.

L'émergence de la convergence actuelle a en effet des implications qui seront déterminantes sur les systèmes existants, comme la gouvernance des données. Nous le constatons par exemple dans le cadre de métiers comme les *chief data officers*, les *chief data and analytics executives*[35] et les *chief artifi-*

[33] Simondon décrit la culture technique comme une culture capable d'exprimer de façon rationnelle la relation de l'individu à l'objet technique dans le contexte d'une société qui aurait atteint un « état de majorité » (Simondon, 2012, p. 89).

[34] « *Information systems are formal, sociotechnical, organizational systems designed to collect, process, store, and distribute information* » (p. 77).

[35] Voir à ce propos (Noyes, 2024).

cial intelligence officer[36] où se pose aujourd'hui la question de l'évolution de leurs périmètres respectifs. Plus globalement, (re)conceptualiser le « système-information » permettrait de rappeler que l'information est le véritable objectif visé par les organisations, que ces dernières ne font en fait que traiter des problématiques informationnelles que les discours actuels occultent en se référant aux données, aux algorithmes ou à l'intelligence artificielle selon l'engouement technique du moment [37].

Une question se pose ici : pourquoi cette convergence aujourd'hui ? Nous émettons une possible explication basée sur une nouvelle mise en lumière des concepts précités. D'une part parce que, comme nous l'avons rappelé, ces concepts et leurs problématiques associées ont dépassé leur milieu professionnel d'origine pour en intégrer de nouveaux (qui n'en avaient alors pas une connaissance poussée). D'autre part parce que ces concepts qui passaient pour acquis, en particulier les données où leur caractère informatique les assimilaient à une forme d'information électronique dans le cas des bases de données des organisations (rapprochée en cela de la vision de Shannon et Weaver), ont subi un réagencement de leur nature même [38] du fait de l'évolution et la propagation des techniques et des technologies.

Nous l'avons vu dans le chapitre 2, le terme « donnée » est de base porteur d'une difficulté de compréhension du fait de sa nature polysémique, qui augmente lorsqu'il est repris et influencé par des disciplines particulières (ex : les mathématiques, les statistiques, l'informatique, etc.) et ceci se ressent dans les enchevêtrements littéraciques de la *data literacy*. Il ne s'agit plus aujourd'hui de se demander *quand* la *data literacy* sera affectée par cette convergence mais *comment* la réarticuler par rapport à ce phénomène au sein duquel elle se trouve déjà.

3.3. Les enchevêtrements littéraciques

Dans cette seconde partie, nous reprenons et développons les éléments abordés dans (Verdi, 2023b) pour discuter des différents enchevêtrements littéraciques. Ces derniers soulignent la primauté des

[36]Voir à ce propos (Jacobs, 2024).

[37]Ronald E. Day en parlait ainsi en 2014 : « Récemment, les discours sur les "données", conçues comme une forme d'une présence auto-affective ou comme « faits », ont remplacé le sujet de « l'information », notre traduction de « *More recently, the discourse of "data", conceived as a form of auto-affective presence or "fact", has come to supersede the trope of "information"* » (Day, 2014, p. 3).

[38]Dans le cas de l'IA, un article du Monde soulignait trois types d'évolutions expliquant leur propagation dans les années 2010 : l'apparition des réseaux neuronaux convolutifs, le développement des cartes graphiques et l'essor du *big data* (Audureau, 2024).

objectifs poursuivis par une littératie spécifique sur ceux des autres littératies englobées [39]. A titre d'exemple l'*information literacy* va naturellement privilégier l'atteinte d'objectifs informationnels comme la formation des individus pour en faire des info-lettrés. La littératie englobée, au-delà de perdre son caractère singulier, est réduite à un ensemble de compétences souvent plus restreintes que dans leur forme originelle et sollicitées uniquement en tant que compléments pour renforcer la didactique de la littératie dominante. Mais plus que tout, la domination d'une littératie sur une autre amène à une conversion de son essence et l'exemple le plus explicite est celui de la *digital literacy* qui fait du numérique une nature et une structure. Les seules exceptions à l'ensemble des cas présentés sont les enchevêtrements de type métallitéracique et de littératie équivalente que nous allons décrire ci-après.

Ceci laisse supposer que les littératies sont uniformes, homogènes. Or, une difficulté majeure porte sur les périmètres spécifiques des littératies :

> « *Information literacy*, *media literacy*, *digital literacy*, *data literacy* et désormais *algorithm literacy*, la somme des littératies qui concernent des processus informationnels interroge quant à leurs liaisons théoriques et didactiques tant il apparaît difficile de les séparer pour les considérer comme des territoires autonomes » (Le Deuff & Roumanos, 2022).

Cette difficulté est redoublée du fait que toute formalisation de littératie est issue d'une vision précise. Dans le cas de la *data literacy*, l'on ne parlera pas d'une seule mais d'une myriade de formalisations ; en obtenir une synthèse « pure » est aujourd'hui impossible. Néanmoins, bien qu'imparfaite, ces synthèses apportent une granularité plus précise de ce peut recouvrir une *data literacy*, c'est pourquoi nous nous y attelons ici.

Nous avions pu distinguer quatre types : (1) la *data literacy* comme une métalittératie, (2) comme littératie équivalente, (3) comme sous-composante de littératie et (4) comme une littératie englobante (Verdi, 2023b). Nous les avons répertorié dans la figure 3.1.

3.3.1. La *data literacy* comme métalittératie

La métalittératie est un « cadre englobant, autonome et exhaustif qui alimente les autres types de littératies » (Mackey & Jacobson, 2017) traduits par Michelot (2019). Au contraire des « sous-littératies », elle ne se constitue pas en sous-catégorie d'une ou plusieurs littératies mais entretient

[39]Nous tenons à préciser ici qu'il s'agit d'une vulgarisation et non d'un essentialisme qui ferait des littératies des entités singulières agissant par elles-mêmes ; une littératie étant par essence une vision développée, portée et transcrite par un ensemble divers d'acteurs.

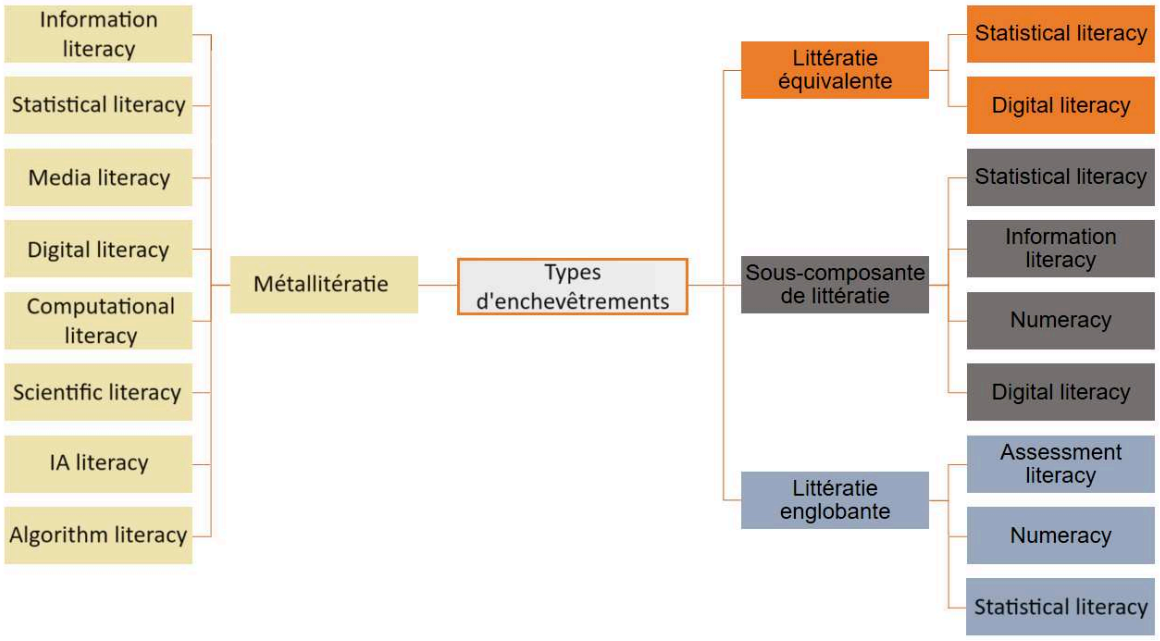

Figure 3.1.: Enchevêtrements de la *data literacy*

une relation symbiotique avec elles. La question porte sur les éléments spécifiques de cette symbiose, au-delà des données qui sont un élément central évident. En effet, une métalittératie relève de quatre domaines :

> « Domaine comportemental (c'est-à-dire ce que les élèves devraient être en mesure de mener à bien lors d'activités d'apprentissage, soit des habiletés ou des compétences), le domaine cognitif (c'est-à-dire ce que les apprenant·e·s devraient savoir lorsqu'ils ont terminé leurs activités d'apprentissage, soit la compréhension, l'organisation, l'application ou l'évaluation), le domaine affectif (c'est-à-dire les changements dans les émotions ou les attitudes des apprenant·e·s par leurs activités d'apprentissage) et le domaine métacognitif (c'est-à-dire ce que les apprenant·e·s pensent à propos de leur propre pensée, soit la compréhension réflexive de comment et pourquoi il·elle·s apprennent, de ce qu'il·elle·s savent et ne savent pas, de leurs idées préconçues et de comment continuer à apprendre). » (Michelot, 2019).

Il s'agit de déceler les compétences « techniques, critiques, quantitatives et conceptuelles [40] » (Bhargava et al., 2015) permettant d'accorder la *data literacy* avec les visées des autres littéracies ; celles-ci ayant des visions différentes sur les données (ex : leurs définitions, leur traitement, etc.). Si le terme n'est pas présent en tant que tel dans la littérature, on en retrouve les caractéristiques sous la plume

[40] « *Data literacy interacts with and builds on all six of these approaches and requires a combination of the technical, critical, quantitative and conceptual skills on which they are based* ».

de plusieurs auteurs. Pour Deahl (2014), la *data literacy* est liée à celle de l'information du fait que les données sont considérées par cette autrice comme des informations à part entières, à la littératie des médias (*media literacy*) de par l'analyse portée sur la représentation des données à travers les médias et de leur portée culturelle et symbolique, et liée à la littératie quantitative (*quantitative literacy*) ou numératie (*numeracy*) (les deux sont ici perçues comme équivalentes) basée sur la manipulation de concepts mathématiques.

Bhargava et al. (2015) ont eux poussé plus loin la conceptualisation de la *data literacy* comme métalittératie (à nouveau, le terme de métalittératie n'est pas présent, c'est une attribution *a posteriori* de notre part). Ce modèle reste l'un des plus connus à ce jour (voir la figure 3.2).

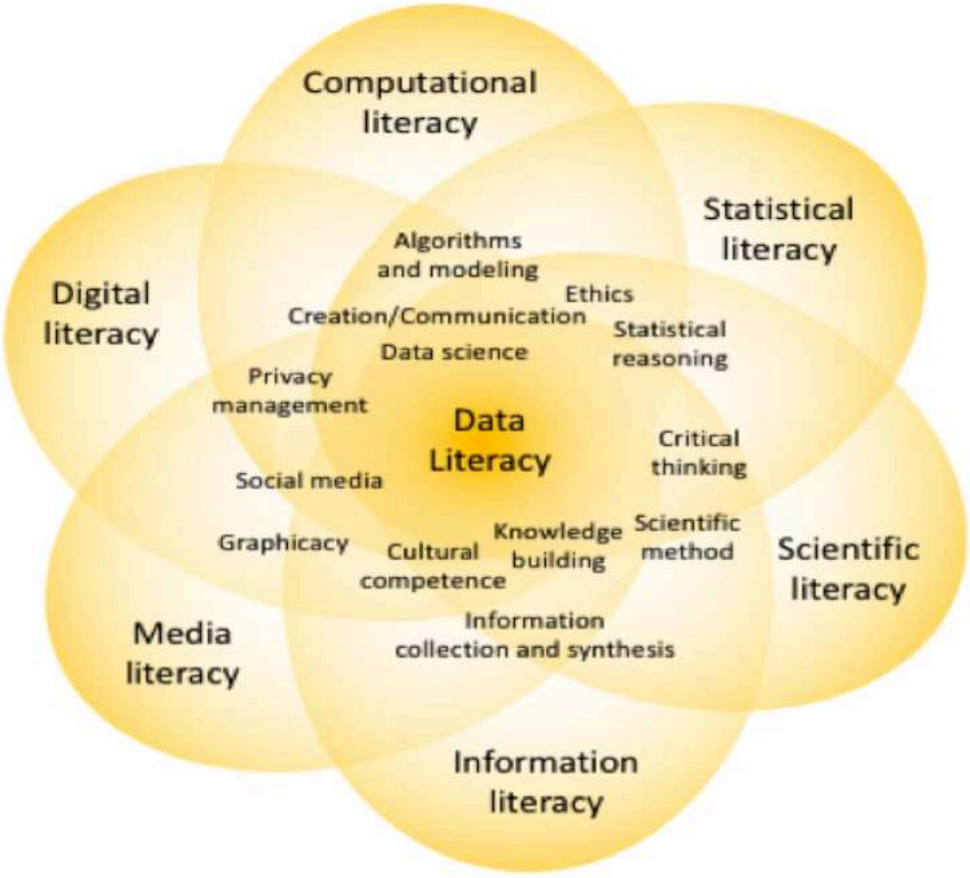

Figure 3.2.: Interactions des littératies avec la data literacy - extrait de (Bhargava et al., 2015)

Dans cette analyse, les littératies en interaction avec la *data literacy* concernent le traitement informationnel avec en figures de proue la littératie de l'information (*information literacy*), la littératie des médias (*media literacy*), la littératie de l'informatique (*computational literacy*), la littératie du numérique (*digital literacy*) et les littératies scientifiques (*statistical literacy* et *scientific literacy*). Dans le modèle ci-dessus, les compétences sont très généralistes (ex : esprit critique, raisonnement scienti-

fique, collecte et analyse de l'information, etc.) et les auteurs ne développent pas leur synthèse au-delà de leur évocation [41].

Par ailleurs, du fait de l'ancienneté de ce modèle, plusieurs littératies, liées aux éléments précédemment cités, n'ont pu être prises en compte comme la littératie des algorithmes et de l'intelligence artificielle. Nous l'avons vu dans le cas des concepts : les données, les algorithmes et l'intelligence artificielle se placent sur un même espace conceptuel. Du fait de ce partage territorial, les littératies des données, des algorithmes et de l'intelligence artificielle entrent en résonance, développant des questionnements similaires vis-à-vis de leurs sujets d'études qui intègrent par ailleurs des éléments connexes.

La littératie des algorithmes, invariablement nommée *algorithm* ou *algorithmic literacy*, n'a été conceptualisée que relativement récemment, en particulier par Head, Fister, & MacMillan (2020), DeVito (2021), Gran, Booth, & Bucher (2021), Ridley & Pawlick-Potts (2021), Dogruel (2021), Shin, Rasul, & Fotiadis (2021) et Le Deuff & Roumanos (2022). Pour l'ensemble de ces auteurs, la littératie des algorithmes doit répondre aux interrogations soulevées par l'usage de ces derniers. Elle est cette capacité de prise de conscience de la présence et des effets des algorithmes (DeVito, 2021), mais également de leur compréhension (Dogruel, Masur, & Joeckel, 2021) donnant les moyens de les influencer ou de les contourner.

> « La littératie algorithmique — un sous-ensemble de la littératie informationnelle — s'avère être une conscience critique de ce que sont les algorithmes, une compréhension de leur mode d'interaction avec les données comportementales humaines dans les systèmes d'information, et une compréhension des enjeux sociaux et éthiques liés à leur utilisation. » Head, Fister et MacMillan (2020) cités et traduits par (Le Deuff & Roumanos, 2022).

La question de la formation aux algorithmes répond d'une logique informationnelle intégrée dans un contexte informatique (au sens d'une maîtrise technique) et numérique (au sens d'une compréhension de sélection, de hiérarchisation et d'évaluation des informations sur Internet (Cardon, 2015)), soulignant l'importance de la place des données, notamment à caractère personnel, dans cet écosystème.

Données et algorithmes vont de pair et soulèvent en effet des problématiques similaires d'ordre technique (notamment dans leur conception et leur exploitation souvent rendues complexes par leurs aspects décrits comme opaques (Theviot, 2023) et politiques (ex : quel *empowerment* possible

[41] Fotopoulou (2020) suit cette logique en soulignant que les *data literacies* complètent les autres littératies qui ont un effet profond sur l'engagement civique. Ceci passe pour cet auteur par une littératie à un niveau basique, par la littératie des médias, la pensée critique et l'éthique des données.

des citoyens vis-à-vis d'un système qui les réduits en *datasets* librement commercialisables (Zuboff, 2019)).

Labelle (2019) pour sa part définissait les algorithmes selon trois dimensions adossées les unes aux autres. Tout d'abord, une dimension technique où ils sont perçus comme un ensemble de procédés et opérations systématiques qui effectuent un traitement de données, mécanisé et informatisé. Puis, une dimension stratégique où ils sont ici des produits sociaux conçus pour satisfaire des besoins identifiés et impliquent des choix stratégiques pour leur mise en œuvre. Enfin, une dimension politique reposant sur la part sociale, stratégique mais également technique de leurs conditions de production et de mise en œuvre.

Jaton (2019) avertissait sur la nécessité de dépasser le cadre de la simple étude des effets des algorithmes et d'enquêter plutôt sur la cause des effets des algorithmes pour « mieux les changer » :

> « Bien que salutaires, ces études sociales des algorithmes – qui croisent parfois des problématiques liées aux dispositifs dits "d'intelligence artificielle" – se heurtent néanmoins à une limite qui a trait selon moi à leurs matériaux d'enquête. En effet, en basant leurs discussions sur le fonctionnement d'algorithmes déjà constitués – en prenant notamment appui sur des articles scientifiques, des rapports et des manuels –, ces études peinent à considérer les hésitations, tests et autres dispositifs pratiques qui ont précédemment participé à la constitution de ces fonctionnements. S'ensuit la mise sous silence problématique d'une myriade d'objets, d'habitudes et d'équipements qui fondent pourtant la véridiction et la fonctionnalité des processus algorithmiques considérés, qu'il s'agisse de méthodes d'apprentissage automatique (*machine learning*) prenant part à des systèmes dits "d'intelligence artificielle" ou non. » (Jaton, 2019).

Cette non-dissociation du milieu technique est au cœur de la vision mécanologique de la littératie des algorithmes développée par Le Deuff & Roumanos (2022) ayant pour but de rendre lisible le pouvoir des algorithmes numériques, grâce à des formes de médiation spécifiques, capable d'endiguer le phénomène du colonialisme des données (*data colonialism*), cette appropriation de la vie humaine par sa conversion et décomposition en données (Couldry & Mejias, 2019b). Comprendre les algorithmes, c'est interroger de manière critique le dispositif lui-même dans ses aspects techniques, communicationnels et informationnels par le développement d'une culture technique permettant de le saisir dans son entièreté.

Par ailleurs, les algorithmes sont souvent rattachés à une vision « mythifiée » de l'intelligence artificielle (Le Deuff & Roumanos, 2022). Cette démystification intervient dans les littératies de l'IA développées également récemment. L'IA est issue de la réflexion de nombreux auteurs dont celle du mathématicien et cryptologue Alan Turing (1912-1954) : d'abord dans un papier intitulé *Intelligent*

machinery écrit en 1948 où il expose les différentes conditions où une machine pourrait être amenée à faire preuve d'intelligence à l'opposé des visions classiques qui affirment le contraire et où l'analogie avec le cerveau humain sert de principe conducteur. Puis en 1950 au sein de son article *Computing machinery and intelligence* dans lequel il expose son jeu de l'imitation (*imitation game*, aussi nommé Test de Turing), basé sur un système de questions-réponses, où un ordinateur essaye de se faire passer pour un humain. Si un humain n'est pas capable de déceler que la personne qui lui répond est une machine alors cette dernière est considérée comme ayant réussi le test.

Le terme d'intelligence artificielle a lui été proposé par John McCarthy (1927-2011) en 1955 et présenté au cours du *Dartmouth Summer Research Project on Artificial Intelligence* l'année suivante. Selon Nilsson (2009), ce choix résulte de deux conditions. La première résidait dans la volonté de McCarthy de se distinguer de précédents travaux contenus dans l'ouvrage *Automata Studies* co-édité par lui-même et Claude Shannon (1916-2001). La seconde était d'éviter une association avec la cybernétique : « sa concentration sur le retour analogique me semblait erronée, et je souhaitais éviter d'avoir Norbert Wiener comme gourou ou de devoir discuter avec lui »[42]. Il définit alors l'intelligence artificielle comme la science et l'ingénierie ayant pour but la création de machines intelligentes[43] (McCarthy, 2007).

Selon Ng, Chu, Shen, & Leung (2021) et Wang (2019), le périmètre de l'IA a évolué au cours du temps : d'abord perçue comme l'ensemble des machines et des algorithmes « intelligents » qui peuvent raisonner et s'adapter en fonction d'un ensemble de règles et d'un environnement imitant l'intelligence humaine (McCarthy, 2007) pour ensuite intégrer la capacité de réaliser des tâches cognitives comme l'apprentissage et la résolution de problèmes grâce aux nouvelles avancées techniques telles que l'apprentissage automatique (*machine learning*) et les réseaux neuronaux (Wang, 2019). Pour Schank (1987), l'IA devait ainsi permettre la construction d'une machine intelligente et de découvrir la nature de l'intelligence. Pour Turing (1948), déterminer ce qu'est l'intelligence tient autant de notre état mental et de notre entraînement que des propriétés de l'objet étudié. Un même objet pouvant ainsi être ou non considéré comme intelligent selon la personne qui l'étudie. Cette problématique de ce que recouvre l'intelligence est toujours amplement débattue comme dans le cas de Schank (1987) qui interrogeait à ce propos les notions de communication, de connaissance intérieure (*internal knowledge*), de connaissance du monde (*world knowledge*), d'intentionnalité et de créativité.

Pour certains auteurs, parler d'intelligence tout court s'avère trompeur car l'IA n'est ni artificielle, ni intelligente. C'est le cas pour Crawford (2022) :

---

[42] « *Its concentration on analog feedback seemed misguided, and I wished to avoid having either to accept Norbert Wiener as a guru or having to argue with him* ». Tiré de : http://www-formal.stanford.edu/jmc/reviews/bloomfield/bloomfield.html

[43] « *The science and engineering of creating intelligent machines* ».

> « Au contraire, l'intelligence artificielle est à la fois incarnée et matérielle, faite de ressources naturelles, de carburant, de main-d'œuvre humaine, d'infrastructures, de logistique, d'histoires et de classifications. Les systèmes d'IA ne sont ni autonomes, ni rationnels, ni capables de discerner quoi que ce soit sans formation extensive et intensive sur le plan computationnel, grâce à d'importants ensembles de données, avec règles et récompenses prédéfinies ». (p. 19).

En outre, au même titre que les données et les algorithmes, la littératie de l'IA cherche à appréhender le phénomène de pervasivité de l'IA dans le quotidien des individus car étant à l'origine de la production de relations sociales et façonnant le monde et ses représentations (Crawford, 2022). Elle se présente donc comme un ensemble de compétences permettant aux individus « d'évaluer de manière critique les technologies de l'IA, de communiquer et de collaborer efficacement avec l'IA, et d'utiliser l'IA comme un outil à employer en ligne, de chez soi, et au travail »[44] (Long & Magerko, 2020). De leur synthèse de la littérature, Ng et al. (2021) ont listé un certain nombres d'objectifs poursuivis dans les formations sur l'IA : (1) connaître les fonctions de base de IA et savoir l'utiliser, (2) pouvoir employer ces savoirs, ces concepts et ces applications de IA dans des scenarii, (3) savoir évaluer et créer de l'IA, et (4) mettre en avant des principes éthiques liés à des aspirations humaines.

Du fait que sa conception résulte d'un ensemble de choix (logistiques, humains, etc.) découlant de vision spécifiques d'acteurs, l'IA est intrinsèquement politique : elle n'est pas une technique computationnelle objective, universelle ou neutre prenant des décisions sans orientation humaine mais est un reflet des logiques de pouvoir, empreinte d'une culture servant les institutions en charge de son développement (Crawford, 2022). En ce sens, cette autrice nous invite à dépasser le « déterminisme enchanté » qui désigne une vision des systèmes d'IA perçus comme « magiques, au-delà du monde connu, et pourtant comme déterministes dans la mesure où ils découvrent des modèles qui peuvent s'appliquer au quotidien avec une certitude prédictive » (p. 247). Cette vision dissimule le pouvoir et, au même titre que les techno-discours, « met fin à tout débat public informé, à tout examen critique ou à tout rejet catégorique » (p. 248), ne laissant pour seuls choix soit une approche utopique faisant de l'IA une solution universelle applicable à tout problème, soit une approche dystopique où d'une part les résultats négatifs des algorithmes présentent ces derniers comme des agents indépendants, et d'autre part où est pressentie l'émergence d'une super-intelligence qui viendra supplanter celle des humains (Bostrom, 2017).

De nouvelles conceptions de la *data literacy* lui font aujourd'hui intégrer ces nouveaux enjeux algorithmiques et d'intelligence artificielle :

[44] « *A set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace* ».

> « Dans notre stratégie de données et d'analyse, ainsi que dans la *data literacy*, les algorithmes et l'apprentissage automatique ont leur place, mais nous devons comprendre qu'il s'agit d'un espace très "technique". Il est important d'avoir des machines travaillant à vos côtés, apprenant par elles-mêmes, etc., mais sans une main-d'œuvre habilitée et capable d'utiliser les résultats, cela n'aura potentiellement aucun sens. Dans ce cas, la *data literacy* est le pouvoir qui permet à la main-d'œuvre et à la culture de réussir grâce à l'apprentissage automatique et aux algorithmes. Si nous travaillons sur notre culture axée sur les données, nos données et notre stratégie analytique, les algorithmes et l'apprentissage automatique devraient permettre à la main-d'œuvre, l'élément humain des données, de disposer de plus de temps pour l'interprétation, pour poser des questions, etc. Cela devrait également permettre à l'élément humain de prendre des décisions plus intelligentes et plus rapides. C'est là que la *data literacy* entre en jeu. Alors qu'un algorithme ou un système d'apprentissage automatique fonctionne avec des données, vous donne des résultats et continue d'apprendre par lui-même, vous feriez mieux, en tant que praticien, d'être prêt à utiliser les données qui vous sont fournies et à prendre des décisions plus intelligentes. [45] » (Morrow, 2021)

Au-delà d'un renforcement, c'est une volonté de fusion qui s'est faite jour comme on peut le voir dans l'article de Schüller, Rampelt, Koch, & Schleiss (2023) qui proposent une littératie des données et de l'IA (*data and AI literacy*):

> « Notre contribution vise à apporter une compréhension commune de la littératie des données et de l'IA qui vise à soutenir des programmes éducatifs appropriés et pratiques tout au long du système éducatif. Une éducation qui permet d'agir et de prendre des décisions éclairées en matière de gestion des données et de l'intelligence artificielle. Compte tenu de cela, notre concept de littératie des données et de l'IA constitue la base du développement à terme d'un cadre convergent et cohérent d'évaluation de la littératie, de mesure et d'évaluation de l'impact aux niveaux national, régional et mondial, qui peut aider à combler le manque de preuves empiriques sur les résultats des interventions et des programmes issus de la littératie des données et de l'IA [46] ».

[45] « *Within our data and analytical strategy, and within data literacy, algorithms and machine learning have a place, but we must understand it is a very 'technical' space. It is powerful to have machines working on your side, learning on their own, and so forth, but without an empowered workforce that is able to use the results, it can be meaningless. In this case, data literacy is the power that allows the workforce and culture to succeed with machine learning and algorithms. If we are working through our data driven culture and data and analytical strategy, algorithms and machine learning should be empowering the workforce, the human element of data, to have more time for interpretation, asking questions, and so forth. It should also empower the human element to make smarter, quicker decisions. This is where data literacy comes into play. As an algorithm or machine learning system works through the data, gives you results, and continues to learn by itself, you the practitioner had better be ready to utilize the data that is provided to you and make smarter decisions* ».

[46] « *Our contribution intends to create a common understanding of data and AI literacy which aims to support appropriate and practical education programs along the entire education chain. Education that enables informed action and decision-*

Pour ces auteurs, ces deux littératies sont conceptuellement inséparables du fait que les données servent de carburant (*fuel*) pour l'IA. Leur fusion doit permettre de maîtriser le pouvoir des données et de l'IA, d'encapaciter les individus en leur permettant de poser des questions pertinentes, de savoir naviguer dans l'océan des algorithmes complexes, de savoir interpréter des résultats issus de l'IA et des données et d'en évaluer leurs effets de manière critique. Plus globalement, il s'agit pour ces auteurs de savoir prendre des décisions informées dans un âge de données (*data age*), faisant ici écho à la « littératie dans l'âge des données » de Bhargava et al. (2015).

Dans cet espace conceptuel mais également littéracique, données, algorithmes et IA s'avèrent donc symbiotiques. La question de la prévalence d'un élément par rapport à un autre découle et de la conception originelle de ceux-ci (en particulier de la question de leur périmètre) et des visées attendues des auteurs (ex : si l'intelligence artificielle est au centre des réflexions, celle-ci prendra l'ascendant naturellement). Au vu des évolutions rapides des discours, nous supputons la convergence prochaine des littératies pour former une littératie unique des algorithmes, des données et de l'IA condensant cet espace conceptuel dans toute sa complexité technique et politique, permettant ainsi aux futurs lettrés de maîtriser les tenants aboutissants de cet écosystème dynamique en évolution constante[47].

3.3.2. La *data literacy* comme littératie équivalente

3.3.2.1. La littératie des statistiques

Un second point évoqué dans la littérature est celui de la *data literacy* comme littératie équivalente. Deux littératies sont intégrées à cette vision.

La première est la littératie des statistiques (*statistical literacy*). Celle-ci est issue des travaux des courants statistiques présentés au cours du chapitre 2[48] et a été évoquée au cours des années 1940 et 1950, tout d'abord dans Scates (1943) qui la définit comme « la capacité de lire des diagrammes et

making in dealing with data and artificial intelligence. Given that, our data and AI literacy concept is the basis for eventually developing a convergent, coherent literacy assessment, measurement and impact evaluation framework for the national, regional and global levels that can help to fill the empirical evidence gap on outcomes from data and AI literacy interventions and programs. »

[47]Nous avons pu le constater dans le cas de *The Algorithm and Data Literacy Project* évoqué au chapitre 1 qui planche sur la conceptualisation d'une littératie des algorithmes et des données incluant également des éléments d'analyse de l'IA.

[48]La littératie des statistiques emprunte aux statistiques les statistiques descriptives, les modèles, les probabilités et l'inférence statistique selon Schield (1999).

des cartes ; une compréhension par le "consommateur" des termes statistiques courants, comme moyenne, pourcentage, dispersion, corrélation et indice. [49] ».

Puis dans l'article *Statistical Literacy in the Social Sciences* (Walker, 1951). Pour Walker, la culture américaine était devenue une culture de la statistique. Or, elle décelait que les chercheurs issus des sciences sociales avaient une connaissance des statistiques peu développée et appelait au développement d'un langage quantitatif (*quantitative language*), le dénominateur commun aux définitions de la littératie étant pour elle la communication. Sans capacité de comprendre ce langage ou de le communiquer, les scientifiques ne seraient pas capables de lire et comprendre avec rapidité l'ensemble des travaux scientifiques : il devenait donc vital de former les futures générations pour enrayer ce problème.

La littératie des statistiques va être véritablement conceptualisée au cours des années 1980-1990 où elle devient cette capacité permettant de lire et d'interpréter de manière critique les données, de savoir utiliser les statistiques comme preuves (Schield, 1999). Elle doit permettre aux individus de reconnaître, décoder et interpréter les contributions statistiques qui imprègnent le quotidien, intégrées dans les décisions publiques, privées et professionnelles (Gal, 2002; Wallman, 1993). En effet, pour Ben-Zvi & Garfield (1999), comprendre les modèles statistiques est indispensable pour décoder les signaux informationnels contenus notamment dans les publicités ou les journaux télévisés. Une connaissance et une compréhension des concepts, du vocabulaire, des symboles mais également des probabilités comme mesure de l'incertitude doivent donc être développées (Dani & Joan, 2004).

Pour Schield (1999), est considérée comme lettrée des statistiques (*statistical literate*) la personne capable de distinguer les déclarations d'association des déclarations de causalité. Il intègre à cet effet la nécessité de savoir (1) si une déclaration de comparaison implique une association ou une causalité, (2) distinguer un échantillon statistique d'un paramètre appliqué à une population, (3) distinguer une population cible d'échantillons provenant d'une population sécifique, (4) distinguer la qualité d'un test du pouvoir prédictif du même test, (5) interpréter la véracité et la représentativité des statistiques ainsi que leurs résultats, notamment de savoir s'ils sont inférentiels ou factuels (6), dinstiguer une étude par observation d'une expérience, (7) distinguer une bonne et une mauvaise expérience, (8) identifier les sources multiples d'un problème lors de l'interprétation d'une mesure ou d'une association, (9) si on a affaire à une corrélation fallacieuse (*spurious association*) et (10) prendre en compte différentes explications pour toute association obtenue d'une étude observationnelle.

Ainsi, la littératie des statistiques est liée au raisonnement scientifique et aux capacités argumentatives (Berndt et al., 2021) dont le socle repose sur les données. Dans l'approche développée par Gould (2017), littératies des données et des statistiques se confondent, les deux traitant des données pour permettre aux étudiants de jouer un double rôle de producteur et de consommateur de statistiques.

[49] « *The ability to read diagrams and maps; a "consumer" understanding of common statistical terms, as average, per cent, dispersion, correlation, and index number.* ».

Il propose ainsi un certain nombre de compétences que tout lettré des statistiques se doit développer :

- comprendre qui collecte nos données, dans quel but et de quelle manière ;
- savoir qui analyse et interprète les données issues d'échantillons ou non aléatoires ;
- comprendre les problématiques liées à la propriété des données et à celle du respect de la vie privée ;
- savoir créer des représentations descriptives basiques de données pour répondre à des questions liées aux processus ayant cours dans la vraie vie ;
- comprendre de quelle manière les représentations numériques varient et pourquoi les données peuvent parfois être altérées avant analyse ;
- comprendre certains aspects des modèles prédictifs.

La raison de ce rapprochement provient pour les auteurs de l'incapacité de la littératie des statistiques d'apporter les connaissances et compétences nécessaires aux individus pour leur permettre de conduire convenablement leur existence dans les démocraties modernes ainsi que de former les étudiants aux défis actuels. Mais également du fait de l'évolution du rôle et de la nature des données, en particulier leur pervasivité. Il s'agit ici d'une vision ancrée dans l'encapacitation des citoyens pour leur permettre d'accéder et d'analyser les données du quotidien ; les statistiques s'avérant indispensables pour les traiter.

3.3.2.2. La littératie du numérique

La seconde est la littératie (du) numérique (*digital literacy*). Issue de l'évolution et de la propagation des technologies numériques, celle-ci est originellement perçue comme une littératie multimédia (Lanham, 1995), permettant de comprendre et d'utiliser l'information selon une large variété de sources issues des ordinateurs (Gilster, 1998). Le périmètre définitionnel va évoluer pour englober la capacité de chercher, accéder, sélectionner, comprendre, interpréter, produire et diffuser des informations numériques à partir d'outils numériques ; le tout en posant une réflexion sur chacune des étapes de ce processus (Martin, 2008).

Elle incorpore la maîtrise des aspect techniques et technologiques soutenant le numérique (Jaeger, Bertot, Thompson, Katz, & DeCoster, 2012) offrant la possibilité de pouvoir localiser, organiser, comprendre, évaluer et analyser l'information (Knaflič, 2015). En ce sens, littératie du numérique et littératie de l'information convergent en un objectif de maîtrise informationnelle, réagençant la suprématie d'une littératie sur l'autre (voir par exemple (Aviram & Eshet-Alkalai, 2006)). Et bien au-delà des considérations compétences techniques et informationnelles, la littératie du numérique introduit un renversement de notre pensée : elle développe un état d'esprit permettant de faire appel

de manière efficace et intuitive aux ensembles de savoirs contenus dans les environnements numériques (Conrad, 2018). La participation active des usagers dans des contextes numériques est ainsi une visée de cette littératie (Williamson & Hague, 2009).

Pour Martin (2008), trois niveaux peuvent être décelés. Un premier niveau, la compétence numérique (*digital competence*), basé sur les compétences numériques au sein d'un ensemble de connaissances, de savoirs, d'attitudes et d'aptitudes. Un second, l'utilisation du numérique (*digital usage*), établi sur les usages des outils numériques appropriés selon les contextes. Et un troisième, la transformation numérique (*digital transformation*), centré sur une réflexion critique des effets de celle-ci sur les paramètres humains et sociaux. Ces trois niveaux interagissent entre eux mais seule l'atteinte des niveaux deux et trois permet l'avènement d'une littératie du numérique (le niveau un est indispensable à son émergence mais est insuffisant à lui seul).

Pour Polizzi, la *data literacy* est une variante de la littératie du numérique, elle-même une variante de la littératies des médias (Polizzi, 2020b, 2021). Deux approches sont intégrées dans la littératie du numérique : une approche fonctionnelle et une approche critique (Polizzi, 2020a). La première intègre des compétences pratiques nécessaires pour l'utilisation des technologies numériques (ce qui comprend des compétences créatives, opérationnelles, sociales et à la recherche informationnelle), et plus globalement des dispositions vis-à-vis d'Internet en lien avec les avantages et inconvénients de ce dernier dans ces effets sur les interactions sociales et les achats en ligne. L'approche critique se centre sur la capacité d'analyse des contenus informationnels en ligne, en particulier sur la confiance que l'on peut leur accorder sur les biais qu'ils embarquent. *In fine*, il s'agit ici de s'interroger sur les potentialités et les limites d'Internet dans le cadre d'une vie civique et de la démocratie.

Dans cette optique, la *data literacy* dépasse la simple capacité d'analyse des contenus en ligne pour intégrer une certaine connaissance de l'environnement numérique. En intégrant des pratiques liées à la gestion de la vie privée et des données, cette littératie se situe à l'intersection des deux approches de la littératie du numérique qu'elle incorpore elle-même.

L'approche fonctionnelle de la *data literacy* étudie les implications interpersonnelles des données en ligne et de la vie privée du point de vue de l'utilisateur. Elle intègre un ensemble de compétences comme la gestion des paramètres sur les réseaux sociaux pour protéger ses informations personnelles.

L'approche critique implique une compréhension et un recul critique vis-à-vis de l'utilisation commerciale des données personnelles en ligne. Elle recouvre les stratégies et tactiques visant à protéger sa vie privée et à échapper au contrôle des grandes multinationales. Elle s'intègre pour Polizzi dans une approche fonctionnelle nécessitant la compréhension du fonctionnement et des affordances des algorithmes et des cookies, s'inspirant ici de Selwyn & Pangrazio (2018) qui ont conceptualisé la *data literacy* comme une connaissance socio-technique.

Chacune des littératies évoquées (littératie des données et (du) numérique) est perçue selon une facette spécifique de leurs prismes respectifs, ici la gestion de la vie privée en ligne. Les données, pleinement perçues comme numériques, se confondent avec l'espace numérique dont elles émanent, et ce dans une perspective « médiatique », circonscrivant ainsi le périmètre de leur étude.

3.3.3. La *data literacy* comme sous-composante de littératie

Cette vision est la plus développée dans la littérature. La *data literacy* est en effet principalement perçue comme une sous-littératie servant des logiques plus englobantes.

3.3.3.1. La littératie de l'information

Cette littératie émerge dans un contexte particulier, issue des tensions militaires et économiques de la guerre froide et accompagnant les discours autour de la « société de l'information », cette dernière étant perçue comme un catalyseur du développement social, culturel et économique des nations (Bernhard, 2003).

La quête d'une maîtrise informationnelle transparaît dans un ensemble de rapports et est différemment envisagée selon les visées désirées. Économiques, certes, comme dans le cas des rapports *The Information Service Environment Relationships and Priorities* (1975), *A Nation at risk, the imperative for Eductional Reform* (1983) ou *Ocotilo Reports '95* (1995) où il s'agit de former le futur travailleur, de le rendre compétitif à l'international.

Mais également citoyennes et humaines avec les rapports *Developing national information policy, a Natis Guideline*, *The 1975 Dag Hammarskjöld Report* et *Natis, National Information Systems : objectives for national and international action* où c'est l'encapacitation intellectuelle du citoyen qui est ici envisagée. La dichotomie économie / citoyenneté s'estompe par la suite : les deux éléments se combinent dans les publications suivantes, en particulier dans le *Presidential Committee on Information Literacy: Final Report* (1989), *la Déclaration de Prague* (2003) et *la Proclamation d'Alexandrie* (2005) où l'information devient un droit humain.

La dénomination d'*information literacy* naît au sein du rapport du président de l'*Information Industry Association*, Paul Zurkowsky (1974). Face à une « surabondance informationnelle » (*overabundance of information*), il invoque une littératie de l'information dans laquelle l'information est envisagée selon une modalité spécifique :

> « L'information n'est pas le savoir ; les concepts ou les idées qu'une personne reçoit et intègre sont évalués et assimilés, renforçant ou modifiant le concept de réalité et/ou la capacité de l'individu à agir [50] ».

La nécessité de former les individus, d'en faire des info-lettrés (*information literate*) ayant appris des techniques et des aptitudes leur permettant d'employer un arsenal étendu d'outils informationnels ainsi que des sources primaires pour trouver des solutions à leurs problèmes, devient l'objectif central à poursuivre. Toute personne incapable de mesurer la valeur de l'information et de pouvoir la manipuler pour accomplir ses objectifs est considérée comme illettrée (*illiterate*) par Zurkowsky. Il faudra néanmoins attendre plusieurs années avant que l'*American Library Association* (ALA) l'officialise et la définisse comme la capacité d'accéder, évaluer et utiliser efficacement l'information nécessaire [51] (American Library Association, 1989).

L'info-lettré ainsi formé doit savoir (1) reconnaître un besoin d'information, (2) identifier quelle information permet de résoudre un problème spécifique, (3) trouver l'information nécessaire, (4) évaluer l'information trouvée, (5) organiser l'information, (6) utiliser efficacement l'information pour résoudre un problème (American Library Association, 1989).

Le champ disciplinaire de déploiement de la littératie de l'information va évoluer au cours du temps, passant du monde des bibliothèques à l'ensemble des sciences sociales (Lehmans, 2018). Et par-delà les orientations économiques, bibliothécaires et citoyennes présentées plus tôt, Le Deuff (2008) ajoutait une orientation critique datée d' Hamelink (1976) permettant à l'individu de se détacher du « prédigéré » des médias et développant chez lui des capacités de résistance intellectuelle.

Par ailleurs, la littératie de l'information a été considérée selon un certain nombre d'aspects. Bruce (1997) propose ainsi sept facettes pour l'appréhender comme (1) technologie de l'information (*information technology conception*), (2) source informationnelle (*information sources conception*), (3) processus informationnel (*information process conception*), (4) contrôle informationnel (*information control conception*), (5) construction de connaissances (*knowledge construction conception*), (6) extension du savoir (*knowledge extension conception*), (7) sagesse (*wisdom conception*).

Dans ce cadre, la *data literacy* sert d'appui à la littératie de l'information pour Schield (2004a). Etre info-lettré pour cet auteur, c'est savoir raisonner de manière critique sur les concepts, les assertions

[50] « *Information is not knowledge; concepts or ideas which enter a person's field of perception, are evaluated and assimilated reinforcing or changing the individual's concept of reality and/or ability to act.* ».
[51] « *The ability to locate, evaluate, and use effectively the needed information* ».

et les arguments tirés des données : savoir traiter les données tant d'un point de vue conceptuel que technique[52] devient donc un prérequis informationnel.

Stephenson & Schifter Caravello (2007) s'interrogent de leur côté sur la prédominance ou l'équivalence de la *data literacy* vis-à-vis de la littératie de l'information sans apporter de réponse définitive. La *data literacy* n'est en effet, à l'époque de publication de leur article, qu'à un état embryonnaire, sans périmètre véritablement défini.

Calzada-Prado & Marzal (2013) peuvent être considérés comme les premiers auteurs à véritablement asseoir le socle de la *data literacy*. Ils le font dans le contexte de formation des bibliothécaires en littératie de l'information :

> « Un besoin existe alors pour une plus grande sensibilisation et une plus grande formation en data literacy, un ensemble de compétences corrélées en acquisition, évaluation, gestion, analyse et interprétation des données situé en-dehors des compétences en statistiques. Dans le cas de la littératie de l'information, les bibliothécaires sont les mieux placés pour jouer un rôle stratégique dans le développement de telles compétences [...] Par conséquent, nous pensons que l'école et les bibliothèques publiques devraient intégrer la data literacy, réadaptée si besoin, dans leurs cursus en littératie de l'information [53] ».

La *data literacy* est en ce sens la composante de la littératie de l'information permettant aux individus d'accéder, interpréter, évaluer de manière critique, gérer, manipuler et utiliser de manière éthique les données. *Data literacy* et *information literacy* font partie pour ces auteurs d'un continuum, d'un « processus graduel d'une éducation basée sur la recherche scientifique » (*a gradual process of scientific-investigative education*.)

Fontichiaro & Oehrli (2016), s'ils ne citent eux pas la littératie de l'information, s'insèrent dans la même pensée que Deahl (2014) et considèrent les données comme de l'information représentée numériquement, dissolvant les frontières conceptuelles de ces deux notions.

---

[52] Schield précise que malgré une technicité inférieure à celle développée dans l'informatique et dans les systèmes de gestion de l'information, les étudiants seront probablement confrontés au langage *SQL*, aux bases de données relationnelles (ex : *Access*), aux logiciels en statistiques (ex : *Excel*) et de présentation (ex : *PowerPoint*).

[53] « *A need exists, then, for greater sensitization and training in "data literacy," a suite of data acquisition-, evaluation-, handling-, analysis- and interpretation-related competencies that lie outside the scope of statistical competencies. As in the case of information literacy, libraries are well-positioned to play a strategic role in the development of such competencies [...] We consequently believe that school and public libraries should also include data literacy, adapted as necessary, in their information literacy programs* ».

En résumé, littératie des données et de l'information sont symbiotiques, voire se confondent. Au-delà de la poursuite d'un objectif informationnel commun, les ressors conceptuels des deux littératies sont identiques comme en témoignent les conceptualisations des données et de l'information abordées dans le chapitre précédent. Les deux littératies ont par ailleurs connu des évolutions similaires. D'une part en termes de nécessité : toute littératie étant empreinte du vocabulaire de crise, elles ont été appelées de leurs vœux par un ensemble varié d'acteurs issus des milieux de l'entreprise, des institutions politiques, scientifiques, etc., pour résoudre des situations jugées comme potentiellement problématiques. D'autre part en termes d'objectifs : majoritairement économiques au moment de leur émergence puis se déclinant plus largement pour englober une approche démocratique et citoyenne ; données et information y devenant alors un droit humain.

3.3.3.2. La littératie des statistiques

Vient ensuite la littératie des statistiques présentée précédemment. Dans la vision de Schield (2004a), la *data literacy* est une composante de la littératie des statistiques, elle-même une composante de la littératie de l'information [54]. Elle doit être un support aux bibliothécaires dans leurs tâches quotidiennes centrées sur l'évaluation de l'information.

La littérature par la suite n'évoque que peu cet agencement. Plus précisément, elle n'évoque pas la *data literacy* en tant que telle à l'exception de Gould (2017) présentée précédemment et d' Umbach (2022) pour qui les deux littératies jouent un rôle politique, bien au-delà de simples considérations techniques. En revanche, la littérature sur la littératie des statistiques rappelle bien son rôle fondamental dans le traitement des données, amplifié par la « révolution des données » du XXIe siècle (Prodromou, 2017).

3.3.3.3. La littératie du numérique

Dans ce contexte, la *data literacy* sert de soutien méthodologique dans la maîtrise et l'analyse des contenus numériques. Cette vision était déjà présente dans le *DigComp* présenté dans le premier chapitre. Dans celle de McCosker (2017), qui s'insère dans une volonté de contrôle de la *datafication*, la *data literacy* doit offrir la possibilité de comprendre la manière dont les activités numériques sont « encodées sous formes de savoirs » (*encoded as forms of knowledge*). Son rôle est de cibler le fonctionnement des métriques (*metrics*) et des analyses (*analytics*) des médias sociaux pour développer une certaine maîtrise sur les pratiques et la visibilité des données personnelles sur les médias sociaux.

[54] Il faut néanmoins souligner qu'au cours de ce même article Schield précise aussi que c'est la littératie des statistiques qui est une composante de la *data literacy*.

La relation des usagers vis-à-vis des médiaux sociaux fait également l'objet des analyses de cette *data literacy*. En tant que « dispositifs épistémiques qui s'attellent à briser les pratiques qui s'étendent sur les plateformes de médias sociaux et à les reconstituer de diverses manières »[55] (Gerlitz & Rieder, 2014), les métriques sont les mesures spécifiques de l'activité des médias sociaux. Elles « facilitent la création et la recréation des jugements portés sur nous, les jugements que nous portons nous-mêmes et les conséquences de ces jugements » [56] (Beer, 2016, p. 3). En traitant des données elles vont, au-delà de mesurer les opinions et expressions des individus, aider à les modeler. La *data literacy* doit ici contrer ce pouvoir et inciter à une réappropriation de nos identités numériques, en particulier par la revendication de la transparence des données.

Dans la vision de Dressen (2021), déployée dans un contexte liée à l'activité des historiens de l'art, la littératie du numérique englobe la *data literacy* et l'érudition numérique (*digital scholarship*), définie dans l'article comme la pratique et la compétence permettant d'utiliser des méthodes ou des outils numériques[57]. Réagencée dans cette littératie du numérique, la *data literacy* doit contribuer au développement d'une pensée critique vis-à-vis des sources de données (ex : quelle véracité, quelle fiabilité des données, etc.) et des usages que l'on peut en avoir. L'historien de l'art doit être ainsi capable d'examiner, d'évaluer, de traiter et de trouver une utilité aux données dans le cadre de son activité.

Ces deux visions n'englobent donc pas l'ensemble des aspects de la littératie du numérique, pas plus que ceux de la *data literacy* : elles n'en conceptualisent qu'une partie dans une optique purement fonctionnelle, adressée à des problématiques locales.

3.3.3.4. La numératie

Si la numératie est le terme le plus employé pour évoquer une littératie des mathématiques, il est important ici de rappeler qu'elle n'en est qu'un aspect. Aspect qui se confond avec un certain nombre de dénominations. L'emploi du terme numératie (*numeracy*) provient du *Crowther Report* de 1959 où il désigne la capacité de manipuler la règle de trois ; une personne dans l'incapacité de comprendre ce dont parlent les scientifiques et les mathématiciens est nommée *innumerate* (Crowther, 1959). John Allen Paulos introduira l'innumératie dans son ouvrage nommé *Innumeracy: Mathematical Illiteracy and Its Consequences* dans lequel *innumeracy* et *mathematical illiteracy* sont interchangeables, et où *innumeracy* désigne l'incapacité de traiter les notions fondamentales sur les nombres et le hasard (Paulos, 1988).

[55] « *Epistemic devices that engage in breaking the practices sprawling on social media platforms apart and putting them back together again in various ways* ».

[56] « *Metrics facilitate the making and remaking of judgements about us, the judgements we make of ourselves and the consequences of those judgements as they are felt and experienced in our lives* ».

[57] « *Practice and competence to use digital methods or digital tools* ».

Le lien entre littératie et mathématiques réapparaît en 1989 au sein du rapport *Everybody Counts: A Report to the Nation on the Future of Mathematics Education*, publié par le National Research Council (1989), où littératie des mathématiques (*mathematical literacy*) et numératie ne font qu'une. La littératie des mathématiques est citée en tant que telle dès 1935 par Brigham (1935). Elle sera définie comme la capacité de raisonner, d'analyser, de formuler et de résoudre des problèmes en contexte réel (Martin, 2007). Kilpatrick (2001) rejette les dénominations de littératie des mathématiques et de numératie tout en conservant leurs caractéristiques. Il préfère parler de « compétence mathématique » (*mathematical proficiency*) et inclue les nécessité suivantes : (1) la compréhension conceptuelle qui renvoie à la compréhension des concepts, opérations et relations mathématiques, (2) la maîtrise des procédures, ou la capacité de l'étudiant à exécuter des procédures mathématiques d'une manière flexible, précise, efficace et appropriée, (3) la compétence stratégique, la capacité de l'étudiant à formuler, représenter et résoudre des problèmes mathématiques, (4) le raisonnement adaptatif, la capacité de pensée logique et de réflexion, d'explication et de justification des arguments mathématiques, et (5) une disposition productive qui inclut la propension habituelle de l'élève à considérer les mathématiques comme une matière sensée, utile et intéressante à apprendre, associée à une croyance dans la valeur d'un travail assidu et en sa propre efficacité en tant que praticien des mathématiques.

La littératie quantitative (*quantitative literacy*) est plus tardive, commençant à être conceptualisée en 1994 dans le rapport *Quantitative Reasoning for College Graduates: A Complement to the Standards* de la Mathematical Association of America (1994) et rapprochée globalement des éléments mathématiques, ce qui sera repris par plusieurs auteurs dont Steele & Kiliç-Bahi (2008) [58] . Le terme est remis en cause par ce même rapport du fait que les mathématiques ne traitent pas que des éléments quantitatifs, mais est adopté faute de mieux. Le raisonnement quantitatif est quant à lui cité dans ce rapport comme la capacité attribuée aux « lettrés quantitatifs » (*quantitatively literate people*).

Une autre dénomination vient s'ajouter, à savoir la mathératie (*matheracy*) créée en 1985 et désignée par D'Ambrosio (1999) comme l'ensemble des compétences permettant de tirer des conclusions des données à partir des résultats issus de calculs. Ensemble qui est pour cet auteur aussi important que savoir lire les données et qui est un premier pas vers une posture intellectuelle et critique, absente pour lui des cursus scolaires. S'inspirant de D'Ambrosio, Skovsmose (1994) propose lui la notion de mathématie (*mathemacy*) définie comme la capacité de calculer et d'utiliser des techniques formelles et mathématiques. Elle se veut être une discipline critique permettant aux étudiants de prendre en compte et d'analyser les effets des évolutions technologiques sur la société. Pour O'Brien (2023), les deux concepts ont cela de commun, au-delà d'évoquer des compétences en mathématiques, de se centrer sur l'encapacitation des individus par le développement d'une posture critique.

[58] Pour ces derniers, cette littératie doit permettre d'utiliser des concepts mathématiques simples pour résoudre les problèmes du quotidien.

En tant que bibliothécaire de données (*data librarian*), Hunt (2005) a réfléchi à une possible équivalence de la littératie de l'information et de celle des données mais ne développe pas réellement sa pensée. De plus, dans sa conception, la *data literacy* intègre la littératie des statistiques ou s'avère être son équivalente au même titre que la numératie. A nouveau, aucun développement n'est apporté à cette assertion. C'est une vision datant des débuts de formalisation de la *data literacy* qui n'est pas réapparue depuis dans littérature.

3.3.4. La *data literacy* comme littératie englobante

Cette vision, que nous aurions pu également nommer « surlittératie », est la moins développée dans la littérature, décelée seulement à deux reprises dans Mandinach (2013) dans son article *Data literacy vs Assessment literacy* et dans Kõuts-Klemm (2019) dans son article *Data literacy among journalists : A skills-assessment based approach*.

3.3.4.1. La littératie de l'évaluation

La littératie de l'évaluation (*assessment literacy*) renvoie à « la compréhension d'un individu des concepts et procédures fondamentales de l'évaluation qui ont une influence très probable sur les décisions pédagogiques »[59] (Popham, 2011).

Elle est cette littératie basée sur l'étude des capacités des publics cibles, sur l'interprétation des résultats de ces évaluations employées ensuite pour améliorer l'apprentissage desdits publics et les programmes de formation d'une manière plus globale (Webb, 2002). L'objectif n'est pas pour les formateurs d'orienter leurs choix selon des prescriptions techniques mais de développer une réflexion critique sur les méthodes d'évaluation pour les employer au mieux selon les contextes pédagogiques (Engelsen & Smith, 2014). Et ce, pour garantir une plus grande réussite de leurs publics. Cette littératie a rapidement été rapprochée de l'enseignement scolaire, conceptualisée comme une littératie de l'évaluation par les professeurs (*teacher assessment literacy*) (Xu & Brown, 2016).

Ainsi pour Stiggins (1995) un professeur est considéré comme lettré s'il sait faire la différence entre de bonnes et de mauvaises évaluations, et la technicité des évaluations ne doit pas être un frein à ses démarches. Il sait ce qu'il doit évaluer et de quelle manière afin d'obtenir des résultats idoines tout en anticipant les problèmes potentiels de ses évaluations. Il doit veiller à assurer une certaine qualité des évaluations, à savoir en précisant clairement les buts, à se concentrer sur l'atteinte des objectifs, à

[59] « *An individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions* ».

sélectionner les méthodes d'évaluation appropriées, à préférer la précision d'un test à une exhaustivité trop généraliste et enfin à prendre en compte les biais et autres distorsions.

Intégrée dans ce contexte pédagogique, la *data literacy* telle que définie par Mandinach (2013) intègre les caractéristiques d'une littératie de l'évaluation. Ce qui amènera l'autrice à conceptualiser la littératie pédagogique des données (*pedagogical data literacy*), aussi nommée littératie des données pour l'enseignement (*data literacy for teaching*), centrée sur la transformation d'informations en savoirs pédagogiques immédiatement exploitables grâce à la collecte, l'analyse et l'interprétation des données, contribuant ici à la conception de séances pédagogiques. La *data literacy*, en contribuant à la maîtrise de l'usage des données, permet ici aux formateurs (et en particulier les professeurs) d'améliorer leur pratique pédagogique afin de mieux répondre aux besoins des publics et contribuer plus efficacement à leur réussite.

3.3.4.2. La littératie des statistiques

Calzada-Prado & Marzal (2013) envisagent ici la littératie des statistiques comme la composante d'un aspect de la *data literacy* impliquée dans l'évaluation critique, le traitement et l'analyse statistique des données. Il s'agit ici d'un renversement des visions passées, marquant la fin de l'hégémonie de cette littératie au profit de la *data literacy* qui gagne en autonomie à mesure que sa conceptualisation se solidifie.

3.3.4.3. La numératie

Pour Kõuts-Klemm (2019), la numératie intègre la *data literacy* dans un contexte spécifique : celui du journalisme centré autour des données, ce que l'on nomme le data journalisme. Selon Bhargava et al. (2015), il s'agit d'une nouvelle forme de journalisme impulsée par le mouvement de l'*open data* où les différentes histoires (*storytelling*) sont présentées ou agrémentées par des graphiques ou des visualisations de données (*datavizualisation*).

Cet usage des données est, dans le cadre de cet article, limité aux données statistiques et aux données de recherche. Or, la compétence de traitement des données par les journalistes est pour l'autrice plutôt limitée. Un journaliste formé à la *data literacy* doit pouvoir (1) trouver les données, (2) savoir évaluer la qualité desdites données, (3) interpréter les données en contexte, et (4) présenter ces données selon les méthodes journalistiques qui (5) prennent en compte les besoins et capacités de réception des publics. La numératie, en tant que composante de la *data literacy*, doit en ce sens permettre aux journalistes de s'approprier des méthodes statistiques et mathématiques pour approfondir leurs analyses.

3.4. Conclusion du chapitre

Le type d'enchevêtrement est corrélé à plusieurs facteurs pouvant se combiner, à savoir (1) les auteurs qui l'ont conceptualisé, (2) la période temporelle où ont été produites les publications et (3) le domaine d'application de la *data literacy*. La *data literacy* est ce sens un objet frontière, réinterprété par les acteurs selon des visées précises qui réagencent ses composantes pour les intégrer au sein de concepts ou de littératies annexes.

La *data literacy* entretient une forte proximité avec un certain nombre de concepts. En soutien culturel et technique à la gouvernance des données, elle a intégré dans ses caractéristiques l'esprit critique sur les données, décliné selon trois grandes approches, et l'éthique des données, comme ensemble de pratiques responsables. Si elle se dissocie sur la forme avec la science des données dans un cadre formatif, elle poursuit le même objectif de maîtrise des données premièrement basé sur une sensibilisation générale, l'aspect technique prenant ensuite le pas aux connaissances théoriques.

En termes littéraciques, la *data literacy* entretient des relations symbiotiques : conceptuelles avec la littératie de l'information, socio-techniques avec les littératies des algorithmes et de l'intelligence artificielle. La prévalence d'un élément (ex : les données, l'information, les algorithmes, etc.) s'inscrit dans les logiques précitées, tributaires des discours généraux (les injonctions étatiques et institutionnelles à s'emparer de tel ou tel sujet) et des volontés singulières des acteurs découlant de cultures, de pensées et d'histoires particulières.

Ces dernières apparaissent très nettement dans le cas des statistiques qui reviennent très régulièrement dans la littérature, se positionnant comme support primordial du traitement des données. Si les littératies des données et des statistiques sont aujourd'hui dissociées dans la majorité des discours, elles restent néanmoins proches et la pratique statistique n'est jamais complètement évacuée dans les formations universitaires comme dans le cas du Master Médas du *CNAM* qui, basé selon une logique en *data literacy* (Arruabarrena, Kembellec, & Chartron, 2019), intègre une grande part de statistiques dans son curriculum [60]. L'on note que ces liens avec d'autres disciplines (les statistiques, les mathématiques, etc.) reflète les appropriations des données par ces mêmes disciplines au cours du temps comme présentées lors du chapitre 2. Et au même titre, l'influence de certaines approches a diminué (ici les statistiques et donc la *statistical literacy*), laissant place à d'autres (« le » numérique et donc la *digital literacy*).

Par ailleurs, la *data literacy* est affectée de près par la convergence conceptuelle et technique en cours des données, des algorithmes et de l'intelligence artificielle où des fusions entre littératies des données et de l'IA commencent à émerger, questionnant ici le périmètre des données qui s'effrite de plus en plus.

[60] https://formation.cnam.fr/rechercher-par-discipline/master-mega-donnees-et-analyse-sociale-medas–1085595.kjsp

Enfin, nous avions évoqué que le concept, selon Deleuze et Guattari, intègre un invariant. Dans le cas des concepts et littératies présentées, il s'agit de l'importance de la gestion des données, nécessitant un développement de compétences nécessaires à son encadrement.

Dans le prochain chapitre, nous montrerons au contraire les transformations opérées sur la *data literacy*, déclinées en un ensemble de littératies parallèles centrées sur des ensembles spécifiques de thématiques.

Chapitre 4.

Les littératies spécialisées des données

> « *Il y a tellement de données qu'elles n'ont parfois plus aucun sens. Prendre cet océan de données et les décisions qu'en tirent les algorithmes pour des certitudes absolues serait une erreur* »
>
> William Gibson, *Périphériques*, 2014

4.1. Introduction

Au-delà des enchevêtrements présentés dans le chapitre précédent, l'appropriation de la *data literacy* se traduit ici par de nombreuses évolutions de ses caractéristiques propres que nous avions regroupées dans Verdi (2023b) sous le vocable de « littératies spécialisées ». Ces dernières, que nous aurions pu également appeler littératies « dérivées » ou « renforcées », sont des littératies à part entière conservant un noyau central de caractéristiques liées à une ou plusieurs littératie(s) et dans lesquelles leurs auteurs ont renforcé certains aspects (ex : conceptuels, techniques, etc.). Les raisons évoquées pour leur conception sont les suivantes : manque de compétences spécifiques, nécessité d'extension du domaine d'analyse ou dénomination inadéquate de la ou les littératie(s) originale(s). Les littératies spécialisées concernent toutes les littératies et les plus connues sont celles qui intègrent un aspect « critique » (ex : *critical information literacy*).

Dans le cas de la *data literacy*, certaines d'entre elles sont à notre sens des « nouvelles littératies » (*new literacies*). Apparaissant au début des années 2000 en réaction au tournant numérique, ces « nouvelles littératies » visent une plus grande compréhension des interactions humaines dans le nouvel écosystème numérique mondialisé (Lankshear, Knobel, & Curran, 2012). Elles sont « (1) centrales pour une participation civique, économique et personnelle dans une communauté mondialisée, (2)

déictiques au sens qu'elles évoluent au rythme des technologies de référence et (3) multifacettes au sens qu'elles bénéficient d'une analyse rassemblant divers points de vue dans la discussion »[1].

Ces littératies spécialisées permettent par ailleurs de refléter l'aspect pluriel des littératies. Pour Fotopoulou (2020), parler « des » littératies plutôt que « la » littératie nous permet « de comprendre la multiplicité et l'interconnexion des pratiques de *data literacy* avec les autres littératies, tandis que l'accent porté sur l'aspect critique souligne l'importance de sensibiliser aux aspects idéologiques et de pouvoir des données »[2].

Nous avions ainsi listé les littératies spécialisées des données, estimées à plus d'une vingtaine. Nous avons mis à jour cette liste avec l'arrivée de la *data literacy privacy* en 2022 et de la *critical datafication literacy* en 2023.

Tableau 4.1.: Les dates d'apparition des définitions des littératies spécialisées

Date	**Auteur(s)**	**Dénomination**
2010	Qin et D'Ignazio	Science/Scientific data literacy
2011	Carlson et al.	Data information literacy
2012	Haendel, Vasilevsky et Wirz	Data management literacy
2013	Schneider	Research data literacy
2014	Deahl	Youth data literacy
	Hügi et Schneider	Linked open data literacy
2015	D'Ignazio et Bhargava	Big data literacy
	Börner, Bueckle, et Ginda	Datavisualization literacy
2016	Philip, Olivares-Pasillas et Rocha	Racial data literacy
	Tygel et Kirsch	Critical data literacy
	Mandinach et Gummer	Data literacy for educators/teachers
2017	D'Ignazio	Creative data literacy
	Weber, Yan et Palmer	Open data literacy
2018	Data Quality Campaign	Administrator data literacy
	Gray, Gerlitz, et Bounegru	Data infrastructure literacy
2019	Pangrazio et Selwyn	Personal data literacies
2020	Sander	Critical big data literacy
	Carmi et al.	Data citizenship
	Kleckner	Healthcare data literacy
	Ladley et Silverston	Social data literacy

[1] Traduction tirée de notre article (Verdi, 2023b).

[2] « *…allows us to understand the multiplicity and interconnection of data literacy practices with other literacies, while the emphasis on criticality stresses the importance of raising awareness about the ideological and power aspects of data* ». L'évocation de l'aspect critique se réfère ici à la *critical data literacy* que nous allons développer dans ce chapitre.

Date	Auteur(s)	Dénomination
2021	Papamitsou et al.	Educational Data literacy
2022	Hillman	Data privacy literacy
2023	Sander	Critical datafication literacy

Chacune de ces littératies des données se centre un aspect spécifique de réflexion et/ou de traitement des données. Nous avions noté une première période (2010 - 2018) qui intégrait principalement des domaines d'études (ex : science, information, management, etc.) et des compétences (ex : esprit critique, datavisualisation, etc.), suivie d'une seconde plus centrée sur les questions démocratiques et citoyennes sans toutefois complètement évacuer les précédentes thématiques.

Les questions pédagogiques sont elles exprimées à travers tout le spectre temporel : soit directement à travers des littératies spécialisées entièrement tournées autour de cette thématique (ex : *Youth, Educational, For Teachers...*), soit plus indirectement dans les méthodes proposées pour développer et transmettre les savoirs des autres littératies spécialisées.

Par ailleurs, en tant qu'« étiquettes verbales de la recomposition tendancielle de l'espace informationnel et de production de données » (Robert, 2020), l'*open data* et le *big data* imprègnent l'ensemble des réflexions, en particulier dans le cas de la transparence et de l'accès des grands jeux de données, qui font par ailleurs l'objet de littératies spécialisées spécifiques.

La majorité des littératies spécialisées sur les données ont pour socle les définitions et compétences tirées de la *data literacy*[3]. Trois exceptions cependant sont ici présentes avec d'une part la littératie des données de la recherche (*research data literacy*) et la littératie des données de l'information (*data information literacy*) basées sur la littératie de l'information (*information literacy*) et, d'autre part, la littératie de la datavisualisation (*datavisualisation literacy*) basée sur les *textual*, *mathematical* et *visual literacies*). Néanmoins comme les données occupent une part centrale de leurs études, nous les rapprochons ici de la *data literacy*.

Nous allons les présenter selon deux prismes : un premier centré sur une volonté de faire évoluer la *data literacy* de par la nécessité de combler ses lacunes, et un second basé sur les évolutions spécifiques de la *data literacy* impulsées par un rejet des externalités négatives du *big data*. Nous compléterons notre analyse en citant les études sur les données (*data studies*) du fait que nombre de littératies spécialisées en découlent ou ont été influencées par celles-ci, en particulier les littératies critiques sur les données.

[3] Les auteurs explicitent bien ce point dans leurs publications.

Comme toutes *studies*, elles sont par nature interdisciplinaires, réunissant un ensemble de chercheurs autour d'un sujet d'étude commun. Elles se rapprochent fortement de méthodes littéraciques, laissant supposer l'émergence de littératies dédiées à ces thématiques (ex : la conceptualisation d'une *data queer literacy* est hautement probable du fait qu'il existe des *data queer studies* et des *queer literacies* avec des visées pédagogiques comme dans le cas de Miller (2016)).

4.2. Les spécialisations de la *data literacy*

Cette première partie est centrée sur les évolutions de la *data literacy* catalysées par la nécessité de répondre aux besoins disciplinaires, techniques et conceptuels auxquels la *data literacy* seule ne peut (supposément) répondre.

4.2.1. L'aspect conceptuel

Découlant des enchevêtrements présentés au chapitre 3, données et informations tendent à être rapprochées au sein de littératies spécifiques. C'est le cas ici de la *data information literacy* conçue par Carlson, Fosmire, Miller, & Nelson (2011). Celle-ci est une sous-composante de la littératie de l'information appliquée dans un contexte de recherche qui intègre les données, les statistiques, l'information ainsi que la *science data literacy* pour former un nouvel éventail de compétences permettant l'exportation de la littératie de l'information des bibliothèques vers les laboratoires.

> « Nous avons développé le projet Data Information Literacy (DIL) pour répondre à deux questions primordiales. Premièrement, de quelles compétences en matière de gestion et de traitement des données les futurs scientifiques ont besoin pour assumer leurs responsabilités professionnelles et tirer parti des opportunités de recherche collaborative dans les environnements de recherche axés sur la technologie et l'e-science? Deuxièmement, comment les bibliothécaires universitaires peuvent-ils employer leur expertise en matière de recherche, d'organisation, de diffusion et de préservation de l'information pour enseigner ces compétences aux étudiants ? En répondant à ces questions, nos objectifs étaient de construire une base dans la communauté des bibliothèques pour l'enseignement des compétences DIL, d'enseigner aux étudiants les compétences DIL appropriées à leur discipline et de développer un processus solide permettant aux bibliothécaires d'élaborer des programmes DIL et de la programmation [4] ».

[4] « *We developed the Data Information Literacy (DIL) project to answer two overarching ques-tions. First, what data management and curation skills are needed by future scientists to fulfill their professional responsibilities and take advantage*

Elle s'intéresse davantage à la production de l'information plutôt qu'à la consommation et l'analyse de l'information. C'est pourquoi elle s'attarde sur le cycle de vie des données. Carlson et al. (2011) proposaient un curriculum prenant en considération l'introduction aux bases de données et aux formats de données, la découverte et l'acquisition de données. S'il s'agit ici d'une spécialisation de la littératie de l'information, d'un point de vue applicatif elle peut être considérée *a contrario* comme une spécialisation de la *data literacy*, intégrée dans une logique informationnelle plus globale.

4.2.2. L'aspect pédagogique

La *data literacy*, comme toute littératie, intègre naturellement une part de formation. Elle semble néanmoins trop limitée pour plusieurs auteurs pour répondre aux nombreux questionnements et problématiques pédagogiques. C'est pourquoi plusieurs littératies des données ont été proposées pour combler ce déficit et ce selon plusieurs critères.

Le premier critère se base sur la nécessité de donner aux enseignants (*educators*) les capacités d'évaluer l'efficacité de leurs enseignements en étant formé aux traitements des données. Ont ainsi été conçues d'une part la littératie pédagogique des données (*pedagogical data literacy*), aussi nommée la littératie des données pour l'enseignement (*data literacy for teaching*), qui désigne la capacité de transformer l'information en connaissance pédagogique et pratiques par la collecte, l'analyse et l'interprétation de tous types de données aidant à élaborer les cursus pédagogiques (Mandinach & Gummer, 2013). Elle nécessite la compréhension des normes affiliées aux données, aux pratiques et savoirs disciplinaires, ainsi qu'une compréhension sur la manière dont les enfants apprennent (Mandinach & Gummer, 2016a). Pour les enseignants, c'est connaître la portée de l'usage des données dans leurs pratiques et la manière dont elle impacte le changement dans l'apprentissage et la performance de leurs étudiants.

D'autre part, la littératie éducationnelle des données (*educational data literacy*) définie comme la capacité de collecter, gérer, analyser, comprendre, interpréter et agir sur les données éducatives selon des principes éthiques, critiques et pertinents (Papamitsiou et al., 2021). Elle vise le développement d'un savoir, de compétences et savoir-faire permettant de rassembler, analyser et présenter l'information ainsi que les données pour soutenir les prises de décisions sur l'ensemble du processus pédagogique.

---

of collaborative research opportunities in e-science and technology-driven research environments? Second, how can academic librarians apply their expertise in information retrieval, organization, dissemination, and preservation to teaching these competencies to students? By answering these questions our goals were to build a foundation in the library community for teaching DIL competencies, to teach students DIL competencies appropriate to their discipline, and to develop a robust process for librarians to develop DIL curricula and programming. ».

Une volonté de formation qui a été étendue aux personnels administratifs et de direction par le truchement de la littératie des données pour les gestionnaires (*administrator data literacy*). Elle a pour vocation la mise en œuvre de pratiques renforçant la réussite des étudiants et ce grâce à l'analyse des données administratives et académiques où sont détectées des tendances, où sont examinées les évolutions professionnelles et les besoins personnels, et où sont mobilisées les ressources pour soutenir leur réussite. Les gestionnaires œuvrent à l'amélioration de l'apprentissage et de l'enseignement dans les salles de classe. On les retrouve tant au niveau des établissements qu'au niveau décisionnel le plus élevé. Ils doivent instaurer une culture autour des données au sein des établissements scolaires et former les enseignants à ces enjeux (Data Quality Campaign, 2018).

Le second critère est une prise en considération des différents publics qui feront l'objet de formations. Dans le cas de la littératie des données pour les jeunes (*youth data literacy*), il s'agit d'analyser les pratiques des données du jeune public et de proposer à celui-ci une formation visant la compréhension des phénomènes des données. Tout jeune lettré de la donnée se doit ainsi de comprendre (1) la manière dont ses données personnelles sont utilisées pour mieux en maîtriser la production, (2) la modélisation des données et les dilemmes éthiques qu'elles intègrent, (3) comment utiliser leur esprit critique et leur éthique vis-à-vis des données (Deahl, 2014). On dépasse ici le cadre universitaire précédemment aperçu pour intégrer l'ensemble du système éducatif.

Au-delà de connaître la *data literacy* et de savoir l'enseigner, et ce en l'adaptant aux différents publics, D'Ignazio (2017) a souligné la nécessité d'un remaniement pédagogique de la *data literacy*, de proposer une méthodologie pour aider les apprenants n'ayant pas de compétences techniques à s'approprier les méthodes pour « faire parler les données » (*speak data*). Cette méthodologie s'adresserait non seulement aux enseignants mais également aux créateurs et éditeurs de données, aux développeurs et concepteurs de logiciels et de visualisations, aux créateurs de tutoriels, au gouvernement, aux organisateurs de communautés et aux artistes. S'imprégnant des travaux de Paulo Freire, la littératie créative des données (*creative data literacy*) a donc pour but de dépasser l'aspect purement technique de la *data literacy* en mettent en avant une volonté d'émancipation des apprenants non techniciens.

Elle propose pour cela cinq tactiques : (1) travailler avec des communautés de données pour apprendre de nouvelles méthodes de traitement des données (ex : le partage du jeu de données *mtcars* tiré du magazine *Motor Trends* sur les performances des véhicules), (2) écrire des « biographies de données » (*data biographies*) qui ont pour but de décrire les conditions de production et de collecte des données en amont de leur traitement [5], (3) de ne pas avoir peur d'utiliser des données non traitées (*messy data*) pour se rendre compte de la manière dont est construit un jeu de données, (4) s'en remettre à un logiciel de traitement spécifique avant de passer à un autre afin de découvrir au fur

[5] D'Ignazio nous enjoint à nous poser les questions suivantes : qui a collecté les données et de quelle manière ? dans quel but ? qui les utilise et quelles sont ces utilisations ? quels sont leurs effets et sur qui ? quelles sont leurs limites ?

et à mesure les logiques et vocabulaires spécifiques aux données ainsi que gagner en confiance dans l'entreprise de ces traitements, et (5) de mettre la créativité au cœur des activités.

4.2.3. L'aspect journalistique

Le journalisme des données (*data journalism*) est un ensemble de pratiques de collecte, d'analyse, de visualisation et de publication des données pour un usage journalistique (Berret & Phillips, 2016). Il s'ancre dans une logique interdisciplinaire, tirant ses connaissances du journalisme, des sciences de l'information, des sciences sociales, des sciences des données et des ordinateurs, des analyses de données, du design informationnel et du *storytelling* (Heravi, 2018). Si une littératie du journalisme des données (*data journalism literacy*) n'est pas nommée, des compétences en données spécifiques aux journalistes existent : ces derniers doivent ainsi être capables de trouver des données, d'en évaluer la qualité, de pouvoir les interpréter en fonction de leur contexte, de les communiquer selon des méthodes journalistiques et de savoir prendre en compte les besoins et capacités de réception des données par le public (Kõuts-Klemm, 2019). Ceci est complété par l'usage de la visualisation de données (Cohen, 2021), l'investigation des algorithmes et du *machine learning* (Berret & Phillips, 2016), la pratique de la numératie et le développement d'une réflexion critique sur l'origine et l'usage des données (Phillips, 2021). La littératie de la visualisation des données (*data vizualisation literacy*), également nommée littératie de la visualisation (*visualization literacy*) s'intègre dans cette logique et se présente comme la capacité d'utiliser des visualisations pour traduire des questionnements liés au domaine des données en requêtes visuelles. Les nécessaires compétences de lecture et d'interprétation visuelle des données ainsi que d'extraction d'informations à partir desdites visualisations (notamment en termes de corrélations, tendances et motifs) en font une littératie spécialisée sur un aspect spécifique du cycle des données, en l'occurrence le traitement et la réinterprétation par le visuel (Börner, Bueckle, & Ginda, 2019).

4.2.4. L'aspect des données

Par ailleurs, la question des données elles-mêmes est prégnante. Tout d'abord avec la littératie du *big data* (*big data literacy*). Elle réoriente les compétences de la *data literacy* en les appliquant aux données massives dans un but informationnel d'identification de motifs récurrents et de prise de conscience des effets des décisions *data-driven* sur les individus et la société (D'Ignazio & Bhargava, 2015). Elle connait un renforcement en 2020 avec la littératie critique des données massives (*critical big data literacy*) qui souligne la volonté de dépasser les compétences de base d'usages de la donnée pour se centrer autour d'une réflexion critique des systèmes de données massives (*big data systems*) (Sander, 2020).

Ensuite avec la littératie de l'*open data* (*open data literacy*) qui cherche à apporter une clarté sur la signification de la *data literacy* et de ce qu'est un lettré de la donnée dans un contexte *open data* (Weber, 2017). Au-delà d'un travail sur les données, il s'agit de se réapproprier et redéfinir l'environnement des données (Montes & Slater, 2019), de comprendre les enjeux éthiques, les valeurs, les cultures et les pratiques des institutions publiques d'où proviennent les données.

Puis avec la littératie des données de la recherche (*research data literacy*) présentée par Schneider (2013) comme une littératie située entre les données de la recherche et la littératie de l'information. Elle est une méthode permettant l'intégration, la curation et l'interopérabilité des données créées au cours du processus scientifique.

Enfin, ancrée dans le monde des bibliothèques, avec la littératie des données ouvertes liées (*linked open data literacy*) qui forme les bibliothécaires à l'évolution du web, des normes des bibliothèques ainsi que du catalogage et introduit les principes des formats *RDF* et *Turtle*, des ontologies et des données ouvertes liées[6]. Des applications pratiques basées sur l'analyse et la conception de documents selon les formats précités viennent compléter cet apprentissage (Hügi & Schneider, 2014).

4.2.5. L'aspect scientifique

Évoquée précédemment, la littératie scientifique des données (*science/scientific data literacy*) créée par Qin & D'Ignazio (2010) reprend les compétences de la *data literacy* en mettant l'emphase sur l'enquête scientifique. Elle doit permettre aux étudiants d'une part de devenir des scientifiques efficaces, d'autre part de devenir des professionnels du management des données tiré de la *e-science*, à savoir une « science collaborative et distribuée, qui s'appuie sur l'accès à de grands réservoirs de données, des réseaux haut débit et des moyens de calcul et de visualisation à haute performance » (André, 2014, Chapitre 5, paragr. 11).

De son côté, la littératie des données de santé (*healthcare data literacy*) est vue comme une *data literacy* offrant aux citoyens une certaine compréhension des données de la santé et la possibilité de les utiliser pour prendre des décisions (cliniques et commerciales) en toute confiance (Kleckner, 2020). Par-delà les compétences classiques de recherche et de maîtrise des données s'ajoute une formation au *SQL* (avec notamment une exploitation de *DOS*) et aux outils / technologies actuel(le)s.

[6] Les données ouvertes liées (*linked open data*) sont des données ouvertes encodées de façon à ce que leur lecture et leur recherche soient grandement facilitées par et pour les machines. Elles intègrent un ensemble de « bonnes » pratiques divisées en quatre principes : (1) appliquer une licence libre permettant le partage des données, (2) utiliser le modèle *Resource Description Framework* (RDF), (3) utiliser des *Uniform Ressource Identifier* (URI) pour le nommage, (4) utiliser une adresse HTTP URI pour retrouver facilement les données.

4.2.6. L'aspect civique et démocratique

La réponse civique intègre plusieurs littératies spécialisées basées sur l'encapacitation des citoyens. La littératie civique des données (*civic data literacy*) proposée par Wolff et al. (2019) se centre sur une volonté d'encapacitation des communautés pour permettre à ces dernières de moins dépendre des spécialistes pour rendre les données opérationnelles et plus globalement pour prendre des décisions informées basées sur les données. Deux communautés sont ici visées : la société civile qui utilise les données pour soutenir leurs revendications, et les organisations, en particulier les entreprises de petites et moyennes tailles, qui y ont recours pour améliorer ou créer des services. Actuellement, cette vision n'a pas dépassé le stade de *workshops* et manque d'une conceptualisation développée.

Une autre vision de la littératie civique des données a été développée par le *Civic Switchboard Guide*[7]. Ancrée dans le monde des musées et des bibliothèques, elle repose sur un usage de la *data literacy* adapté aux besoins locaux pour comprendre et employer les données civiques (*civic data*), à savoir des données liées à des communautés, des institutions, qui permettent de décrypter des problématiques de la vie quotidienne, comme la compréhension des lois (CSG, 2020).

Née de la volonté de renforcer la *data literacy* pour répondre aux problèmes informationnels que sont la désinformation, la malinformation et la mésinformation, nous retrouvons également la citoyenneté des données (*data citizenship*) qui intègre le développement d'une pensée et d'une participation citoyennes critiques (Carmi, Yates, Lockley, & Pawluczuk, 2020). Le socle de pratiques comprend une pensée critique des données, un usage éthique des données et un engagement proactif des citoyens à employer les données et à former leurs concitoyens aux bonnes pratiques de la *data literacy*.

Dans une même logique, la littératie sociale des données (*social data literacy*), développée par Ladley & Silverston (2020), part du principe que les citoyens n'ont pas de compétences en communication des données, ce qui entraîne des problèmes éthiques dans le monde professionnel et au sein des « communautés sociétales » (*societal communities*). Pour éviter ces dérives, l'objectif recherché est l'adoption de normes comportementales par la société dans le cadre de l'usage et la communication des données ; des notions de statistiques, une prise en compte des contextes et un recul critique sur les données devant intégrer ces normes.

[7] https://civic-switchboard.gitbook.io/guide/library-roles/developing-civic-data-literacy

4.3. Canaliser les externalités négatives du *big data*

Le *Big data* en tant que catalyseur de la « révolution des données » (*data revolution*) (Kitchin, 2014a) a entraîné de grandes modifications dans la manière d'appréhender la science et plus globalement les réalités quotidiennes :

> « Le *Big Data* entraîne un changement radical dans la manière dont nous conceptualisons la recherche. Dans leur commentaire des sciences sociales computationnelles, Lazer et al. (2009) soutiennent qu'il offre "la capacité de collecter et d'analyser des données avec une ampleur, une profondeur et une échelle sans précédents" (p. 722). Ce n'est pas seulement une question d'échelle et il ne suffit pas non plus de l'envisager en termes de proximité, ou de ce que Moretti (2007) appelle une analyse lointaine ou rapprochée des textes. Il s'agit plutôt d'un changement profond aux niveaux épistémologique et éthique. Le *Big Data* redéfinit des questions clés sur la constitution des connaissances, les processus de recherche, la manière dont nous devrions interagir avec l'information, ainsi que la nature et la catégorisation de la réalité. Tout comme Du Gay et Pryke (2002) notent que "les outils comptables... ne facilitent pas simplement la mesure de l'activité économique, ils façonnent la réalité qu'ils mesurent" (pp. 12-13), le *Big Data* délimite de nouveaux terrains d'objets, de méthodes de connaissance et de définitions de la vie sociale [8] » (Boyd & Crawford, 2012).

Son exploitation, source de progrès économiques (Manyika et al., 2011), voire démocratiques, s'inscrit dans une logique technophile, le « fondamentalisme du *big data* » (Crawford, 2013), où tous les aspects du monde réel seraient quantifiables et prévisibles. Or, celle-ci a un revers : l'exploitation des vies humaines à des fins mercantiles et / ou politiques. Ces aspects négatifs se retrouvent au sein des phénomènes de *datafication* et *dataveillance*.

[8] « *Big Data creates a radical shift in how we think about research. Commenting on computational social science, Lazer et al. (2009) argue that it offers 'the capacity to collect and analyze data with an unprecedented breadth and depth and scale' (p. 722). It is neither just a matter of scale nor is it enough to consider it in terms of proximity, or what Moretti (2007) refers to as distant or close analysis of texts. Rather, it is a profound change at the levels of epistemology and ethics. Big Data reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality. Just as Du Gay and Pryke (2002) note that 'accounting tools . . . do not simply aid the measurement of economic activity, they shape the reality they measure' (pp. 12–13), so Big Data stakes out new terrains of objects, methods of knowing, and definitions of social life*.».

4.3.1. La datafication

La datafication a été définie par plusieurs auteurs comme Carlson et al. (2011) pour qui elle est « la quantification des interactions sociales et de leur transformation en données numériques »[9]. Mayer-Schönberger & Cukier (2013) la définissent comme « la transformation de l'action sociale en données quantifiées en ligne permettant le traçage en temps réel et l'analyse prédictive »[10]; celle-ci résulte des activités liées au *big data*. Elle se différencie de la numérisation de par la capacité de mesure, de calcul et d'analyse qu'elle apporte sur les contenus mis en données (*datafied*). Dans leur ouvrage, Mayer-Schönberger et Cukier prennent pour exemples le livre, qui devient dès lors indexable et retrouvable, et les réseaux sociaux qui compilent en données les relations (dans le cas de *Facebook*) et les sentiments (dans le cas de *Twitter/X*). Découlant de cela, l'intimité elle-même devient le nouveau matériau d'analyse de ce nouveau phénomène (Livingstone, 2019).

Sous-jacente à ses logiques d'exécution, la gouvernementalité algorithmique, décrite par Rouvroy & Berns (2013) comme « un certain type de rationalité (a)normative ou (a)politique reposant sur la récolte, l'agrégation et l'analyse automatisée de données en quantité massive de manière à modéliser, anticiper et affecter par avance les comportements possibles », entraîne ici la crainte d'un fort affaiblissement de la réflexion et du consentement humains dans la divulgation des informations :

> « La gouvernementalité algorithmique ne produit aucune subjectivation, elle contourne et évite les sujets humains réflexifs, elle se nourrit de données infra-individuelles insignifiantes en elles-mêmes, pour façonner des modèles de comportements ou profils supra-individuels sans jamais en appeler au sujet, sans jamais l'appeler à rendre compte par lui-même de ce qu'il est ni de ce qu'il pourrait devenir. ».

Dans cette logique, les individus ne sont plus interpellés prioritairement par « le pouvoir » à travers leurs « capacités d'entendement, de volonté, d'expression » mais plutôt à travers leurs « profils », ces doubles de données que Rouvroy et Berns nomment des « doubles statistiques »[11]. Selon Davidson, Wessel, Winter, & Winter (2023), cette datafication donne donc aux données un rôle « d'arbitre des réalités » (*arbiters of realities*), conditionnant les politiques économiques et sociales, et plus globalement toute décision découlant des organisations qui les maîtrisent. Elle a été empreinte dans les discours d'une certaine légitimité sur l'accès, la compréhension et le contrôle des comportements humains permettant de comprendre de manière plus large la conduite humaine (Dijck, 2014). Une légitimité également revendiquée par les organisations conditionnées par « l'impératif des données »

[9] « *the quantification of social interactions and their transformation into digital data* ».

[10] « *the transformation of social action into online quantified data, thus allowing for real-time tracking and predictive analysis* ».

[11] Pour ces auteurs, ils sont issus du « croisements de corrélations, produits de manière automatisée, et sur la base de quantités massives de données, elles-mêmes constituées ou récoltées "par défaut" ».

(*data imperative*). Ce terme, employé dès Clarke (1988) et développé par Fourcade & Healy (2016), désigne la nécessité d'une extraction des sources de données par tous les moyens disponibles :

> « Nous soutenons l'idée que (a) les organisations modernes suivent un impératif institutionnel en matière de données pour collecter autant de données que possible ; (b) à la suite de l'analyse et de l'utilisation de ces données, les individus accumulent une forme de capital provenant de leurs positions, mesuré par diverses méthodes de notation et de classement numériques ; et (c) le caractère factice de ces méthodes de notation en fait des dispositifs organisationnels aux effets potentiellement stratifiants. Ils offrent aux entreprises de nouvelles opportunités pour structurer et tarifer leurs offres aux consommateurs. Pour les individus, ils créent des situations de classification qui identifient les chances de vie partagées sur les marchés de produits et de services [12] » (Fourcade & Healy, 2016).

Pour Sadowski (2019), la poursuite de cet impératif implique, au-delà d'une collecte passive des données, la création active de données pour générer des retombées économiques. La *data literacy* dans un tel contexte a été sollicitée pour faire valoir, voire primer, la valeur des données dans l'esprit des individus comme dans le cas de l' US Chamber of commerce fundation (2014) où le mouvement des données était considéré comme une force œuvrant pour le bien (*a force for good*, p. 8). En effet, dans ce contexte de datafication, les données prennent une nouvelle ampleur où, selon une vision positive restituée par Ricaurte (2019), (1) elles refléteraient précisément la réalité, (2) leur analyse générerait les savoirs les plus précieux et les plus précis, (3) et les résultats issus de leurs traitements pourraient être utilisés pour prendre de meilleurs décisions vis-à-vis du monde.

Ainsi, la mise en relation et la possibilité de compilation des données formeraient cet *übercapital* ou *eigencapital* déduit des enregistrements numériques des personnes qui amènerait à des situations de classification (*classification situations*) où les individus seraient triés et catégorisés pour en tirer un bénéfice économique (Fourcade & Healy, 2016). Or, comme le rappelle Richterich (2018), si les données sont produites par un grand nombre d'individus, leur exploitation et leur contrôle n'est en revanche l'affaire que d'un petit nombre d'acteurs souvent issu du monde des entreprises (Andrejevic, 2014). Le fait que la datafication réside dans un nombre limité de mains a et continue d'être la source de questionnements quant aux externalités négatives issues de ses usages, en particulier dans les cas d'atteinte à la vie privée, voire de surveillance généralisée. L'on se retrouverait dès lors dans un nouveau type de colonialisme, celui des données, tel que décrit par Couldry & Mejias (2019a) où

[12] « *We argue that (a) modern organizations follow an institutional data imperative to collect as much data as possible; (b) as a result of the analysis and use of this data, individuals accrue a form of capital flowing from their positions as measured by various digital scoring and ranking methods; and (c) the facticity of these scoring methods makes them organizational devices with potentially stratifying effects. They offer firms new opportunities to structure and price offerings to consumers. For individuals, they create classification situations that identify shared life-chances in product and service markets* ».

la vie humaine ne serait que l'affaire d'une exploitation capitalistique et où la majorité des décisions seraient issues de logiques automatiques (D'Ignazio & Klein, 2020).

4.3.2. La dataveillance

Cette exploitation des mises en données des vies humaines se retrouve également dans le contexte de la *dataveillance* (ou *datasurveillance*) qui désigne une « nouvelle surveillance » des individus à partir de leurs données et se définit comme « l'usage systématique des systèmes de données personnelles pour l'enquête ou la surveillance des actions ou des communications d'une ou plusieurs personnes [13] » (Clarke, 1988). Elle se décline selon deux aspects : la *dataveillance* personnelle et la *dataveillance* de masse. Le premier aspect se centre sur un ou plusieurs individus tandis que le second englobe des groupes entiers.

Avec l'évolution des technologies numériques, la *dataveillance* a adopté de nouveaux modes. Pour Lupton & Williamson (2017), elle opère désormais à des niveaux plus intimes de la vie humaine avec l'auto-surveillance de par l'usage de logiciels et de technologies de repérage (comme le GPS, les applications sportives, etc.), et la surveillance « sociale » (Marwick, 2012) ou « participante » (Albrechtslund & Lauritsen, 2013) où les individus s'encouragent mutuellement à partager les détails de leur vie privée sur les réseaux sociaux. S'y ajoute la « surveillance oligoptique » (*oligoptic surveillance*)[14] où la combinaison et l'alignement de différents jeux de données, collectées à partir des usages des technologies numériques des personnes, permettent de construire un nouveau profil de données (*data profile*).

Pour le construire, plusieurs techniques sont employées. Clarke & Greenleaf (2017) citaient la surveillance physique (*physical surveillance*) permettant d'associer les enregistrements aux individus, la surveillance des communications (*communications surveillance*, qu'il s'agisse de messages ou non éphémères dont les données permettent d'attester l'existence, comme des logs, des métadonnées, etc.), la surveillance des comportements (*behavioural surveillance*) à partir des données enregistrées témoignant des activités des personnes, la surveillance de la mobilité (*location surveillance*) à partir de systèmes de reconnaissance de véhicules et la corrélation des déplacements, la surveillance expérientielle (*experiential surveillance*) basée sur les données enregistrées en ligne (ex : requêtes, publications téléchargées, pages web recherchées, etc.), et la surveillance des corps (*bodily surveillance*) à partir de mesures comme les biométriques.

[13] « *...the systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons.* ».

[14] Bien que non cités, une certaine résonance avec les travaux de Bruno Latour se retrouve ici. Ce dernier désignait les oligoptiques comme « les étroites fenêtres qui permettent de se relier, par un certain nombre de conduits étroits, à quelques aspects seulement des êtres (humains et non-humains) dont l'ensemble compose la ville » (Latour, 2007a).

Ces nouvelles techniques de *dataveillance* reconfigurent ainsi les personnes en tant qu'« agencements de données numériques » (*digital data assemblages*)[15] où la séparation conceptuelle dissociant les hommes des données s'estompe :

> « Le sujet humain peut être conceptualisé à la fois comme digérant et émettant des données au sein d'un cycle sans fin de production de données, intégrant les données dans l'identité, générant encore plus de données. Les données sont absorbées dans le corps/l'identité et deviennent de nouvelles données qui s'écoulent du corps/identité vers l'économie des données numériques. Le sujet dévoreur et émetteur de données, par conséquent, n'est pas cloisonné mais ouvert au recueil et à la diffusion des données numériques. Ces données deviennent une part intégrante du sujet humain. Les agencements de données représentent également l'individu d'une multitude de manières qui ont des sens différents selon les contextes et les usages[16] » (Lupton & Williamson, 2017).

L'assimilation des individus aux données n'est pas une crainte nouvelle : dès sa conception la *dataveillance* a été présentée comme potentiellement nocive. Et ce, selon plusieurs niveaux (voir le tableau 4.2).

[15] Le terme anglais *assemblage* peut être traduit littéralement par assemblage. Or, en raison de la similarité de raisonnement avec les travaux de Deleuze et Guattari, nous employons le terme d'agencement (dans la littérature anglaise, l'agencement de Deleuze est traduit par *assemblage*).

[16] « *The human subject may be conceptualised as both data-ingesting and data-emitting in an endless cycle of generating data, bringing the data into the self, generating yet more data. Data are absorbed into the body/self and then become new data that flow out of the body/self into the digital data economy. The data-eating/emitting subject, therefore, is not closed off but is open to taking in and letting out digital data. These data become part of the human subject. Data assemblages also represent the individual in multiple ways that have different meanings based on their contexts and uses.* ».

Tableau 4.2.: Dangers de la *dataveillance* - compilation et traduction réalisées à partir de Clarke (1988)

Dataveillance personnelle	**Dataveillance massive**	**Pour la société**
Mauvaise identification ; Mauvaise qualité des données ; Usage sans prise en compte du contexte des données ; Prises de décisions mauvaises ; Manque de connaissance du sujet des flux de données ; Manque de consentement du sujet aux flux de données ; Blacklisting ; Refus de rédemption.	Méthodes arbitraires ; Fusion de données dénuées de leur contexte ; Complexité et non-compréhension des données ; Chasses aux sorcières ; Discrimination préétablies et anticipation de culpabilité ; Campagne publicitaire sélective ; Inversion de la charge de la preuve ; Accusions et accusateurs inconnus ; Opérations secrètes ; Refus de procédures régulières.	Climat de suspicion généralisé ; Relations conflictuelles ; Accent mis par les forces de l'ordre sur les infractions facilement détectables et prouvables ; Application inéquitable de la loi ; Diminution du respect de la loi ; Diminution du sens des actions individuelles ; Diminution de l'autonomie et de l'autodétermination ; Dévalorisation de l'originalité ; Tendance accrue à se retirer du niveau officiel de la société ; Affaiblissement de la fibre morale et de la cohésion de la société ; Déstabilisation des rapports de force stratégiques ; Potentiel répressif d'un gouvernement totalitaire.

Plusieurs cadres d'études ont ainsi été élaborés pour caractériser et analyser ces nouveaux contextes des données, à savoir l'agencement des données (*data assemblage*), la data-philosophie, les études sur les données (*data studies*) et les littératies spécialisées des données. Nous allons développer les deux derniers cadres tandis que les deux premiers seront présentés en annexes 4 et 5 par souci de clarté et de cohérence.

4.3.3. La réponse critique

4.3.3.1. Les études critiques sur les données

Les études critiques sur les données (*critical data studies*) englobent selon Richterich (2018) le fait que les universitaires de ce champ de recherche enquêtent sur les données selon des perspectives critiques amenant au décryptage de collecte, d'analyse et d'utilisation des mégadonnées. En ce sens, elles s'attellent à déconstruire les mythes du positivisme numérique (Mosco, 2014), en particulier la conception de données neutres parlant d'elles-mêmes. Et au-delà, de questionner le « pouvoir » sur les données et son inégale répartition. Dans l'introduction de l'ouvrage *New Perspectives in Critical Data Studies : the ambivalence of data power*, Hepp, Jarke, & Kramp (2022) soulignent l'ambivalence de ce pouvoir. D'une part, son attrait découlant du fait que les données numériques permettent la production de savoirs nous renseignant sur la société et les processus sociaux. D'autre part les problèmes que ces mêmes données numériques engendrent notamment la surveillance de masse et les reproductions des inégalités amenant à des pratiques discriminatoires, voire du racisme. En ce sens, les études critiques doivent pour ces auteurs répondre à cette ambivalence et développer une recherche critique crédible de ce pouvoir des données en intégrant un ensemble varié de disciplines comme la sociologie, les études sur les médias et l'information, la géographie, les sciences politiques, les sciences dites dures ou encore les études sur la technologie.

Toujours pour ces auteurs, l'émergence des études critiques des données débute avec la critique du *big data*, en particulier dès la parution de l'article de Boyd & Crawford (2012) nommé *Critical questions for big data : Provocations for a cultural, technological, and scholarly phenomenon* :

> « Les données de recherche à grande échelle nous aideront-elles à créer de meilleurs outils, de meilleurs services et de meilleurs biens publics ? Ou seront-elles les hérauts d'une nouvelle vague d'intrusions dans la vie privée et de *marketing* invasif ? L'analyse des données nous aidera-t-elle à comprendre les communautés en ligne et les mouvements politiques ? Ou l'analyse sera-t-elle utilisée pour suivre les manifestants et réprimer la parole ? De grandes quantités de données transformeront-elles la manière dont nous étudions la communication et la culture humaines, ou réduiront-elles la palette des options de recherche et modifieront-elles le sens du terme "recherche" ? [17] ».

[17] « *Will large-scale search data help us create better tools, services, and public goods? Or will it usher in a new wave of privacy incursions and invasive marketing? Will data analytics help us understand online communities and political movements? Or will analytics be used to track protesters and suppress speech? Will large quantities of data transform how we study human communication and culture, or narrow the palette of research options and alter what 'research' means?* ».

Dans cet article, Boyd et Crawford décrivent le *big data* comme un phénomène socio-technique basé sur une triple interaction[18] :

> « (1) La technologie : maximiser la puissance de calcul et la précision algorithmique pour recueillir, analyser, lier et comparer des grands ensembles de données ;
> (2) L'analyse : s'appuyer sur de grands ensembles de données pour identifier des tendances afin de faire des réclamations économiques, sociales, techniques et légales ;
> (3) La mythologie : la croyance largement répandue que les grands jeux de donnent permettent de développer une forme d'intelligence et de connaissance supérieures pouvant produire des perceptions nouvelles auparavant impossibles, nimbées d'une aura de vérité, d'objectivité et de précision ».

Elles ajoutent que, selon cette triple configuration, le *big data* entraînerait un changement radical en termes épistémiques et éthiques, en particulier sur la recherche et la constitution du savoir, ce que nous avons précisé précédemment. La nécessité de sa compréhension s'avérait donc crucial, Dalton & Thatcher (2014) résumant ici les questionnements centraux de ces réflexions critiques :

> « Quelles conditions historiques amènent à la réalisation du *big data* et quelles sont-elles ?
> Qui contrôle le *big data*, sa production et ses analyses ?
> Quelles motivations et quels impératifs alimentent leur travail ?
> Qui sont les sujets du *big data* et quels savoirs produisent-ils ?
> Comment le *big data* est-il réellement appliqué dans la production des espaces, des lieux et des paysages ?
> Que devons-nous réaliser avec le *big data* et quels autres types de savoirs permettrait-il de produire ? [19] ».

Au-delà de son analyse globale doit s'ajouter une déconstruction de ses mythes. En effet, la quantification issue d'une mise en données du monde a été très tôt l'objet de critiques, en particulier vis-à-vis de la signification devant supposément découler des analyses de tendances. Apposer un sens sur des tendances sociétales, inexactes du fait des biais embarqués dans la conception et l'usage des logiciels de traitement des mégadonnées, est un leurre et ne peut refléter fidèlement la réalité :

[18] « *(1) Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets. (2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims. (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy.* ».

[19] « *What historical conditions lead to the realization of 'big data' such as they are? Who controls 'big data', their production and analysis? What motives and imperatives drive their work? Who are the subjects of 'big data' and what knowledges are they producing? How is 'big data' actually applied in the production of spaces, places and landscapes? What is to be done with 'big data' and what other kinds of knowledges could it help produce*? ».

> « Le travail de production de *big data*, ou plutôt des données brutes, est donc un travail de suppression de toute signification, afin que ces données brutes puissent être calculables et fonctionnent non plus comme des signes qui signifient quelque chose en rapport à ce qu'ils représentent, mais comme quelque chose qui se substitue à la réalité signifiante, la fasse disparaître. Se substitue ainsi à la réalité signifiante un ensemble de réseaux de données a-signifiantes qui fonctionnent comme des signaux, c'est-à-dire que bien qu'elles n'aient aucune signification, ou plutôt grâce à cela, elles deviennent calculables. En fait, c'est la définition du signal donnée par Umberto Eco : un signal, c'est un élément sans signification, qui ne signifie rien, mais, parce qu'il ne signifie rien justement, devient d'autant plus calculable » (Rouvroy & Stiegler, 2015).

Remettre l'action humaine au cœur de l'analyse est ainsi une visée de ces études :

> « Nous notons que les deux manières de parler des données adhèrent, et ont plutôt tendance à renforcer de façon commode, l'idée de « bon sens » selon laquelle les données proviennent de la réalité, la représentent ou la reflètent – qu'elles rendent compte de manière transparente de la nature du réel, plutôt que d'être quelque chose de fabriqué par les humains ou les machines [20] » (Atenas, Havemann, Rodés, & Podetti, 2023).

4.3.3.2. Les littératies critiques des données

Influencées, voire découlant de ces études, les littératies critiques poursuivent cette vision. A l'origine, la littératie critique (*critical literacy*) désigne « l'apprentissage de la lecture et de l'écriture comme partie intégrante du processus de prise de conscience de notre expérience en tant qu'elle est construite historiquement au sein de relations de pouvoir spécifiques [21] » (Anderson & Irvine, 1993, p. 82), s'intégrant dans la lignée des travaux de Paulo Freire. Elle est cette littératie qui remet en cause et la répartition des pouvoirs et « le *status quo* afin de découvrir des alternatives non seulement pour soi-même mais également pour le développement social »[22] (Pari & Shor, 1999, p. 1). Selon l'état de l'art de Lewison, Flint, & Sluys (2002), quatre dimensions la composent : déconstruire les lieux communs, adopter des points de vue multiples pour des enquêtes, se concentrer sur les problèmes sociopolitiques, prendre des mesures et promouvoir une justice sociale. L'objectif est d'intégrer ces éléments dans les pratiques éducatives et de se questionner sur le rôle du pouvoir sous-jacent dans les textes. Il s'agit plus globalement de maîtriser l'hétéroglossie, concept développé par l'historien

[20] « *We note that both ways of speaking about data buy into, and rather conveniently reinforce, the 'commonsense' idea that data comes from, and represents or reflects reality - that it transparently reports on the nature of the real, rather than being something made by human or machine* ».

[21] « *Critical literacy, then, is learning to read and write as part of the process of becoming conscious of one's experience as historically constructed within specific power relations.* ».

[22] « *Critical literacy thus challenges the status quo in an effort to discover alternative paths for self and social development.* ».

russe Mikhaïl Bakhtine (1895 - 1975) qui désigne « un autre discours dans une autre langue, qui sert à exprimer les intentions de l'auteur mais de façon réfractée » (Bakhtin et al., 1982), (Street, 2006).

Si la nécessité critique s'intègre dans les définitions de la *data literacy* à partir de Calzada-Prado & Marzal (2013), Il faudra attendre 2016 pour trouver une conceptualisation d'une littératie critique centrée sur les données sous la plume de Tygel & Kirsch (2016). Elle sera suivie de plusieurs autres, présentées en annexe 6. Les littératies (critiques) des données ont fait l'objet d'une synthèse par Sander (2023) que nous reprenons et traduisons dans la figure 4.1.

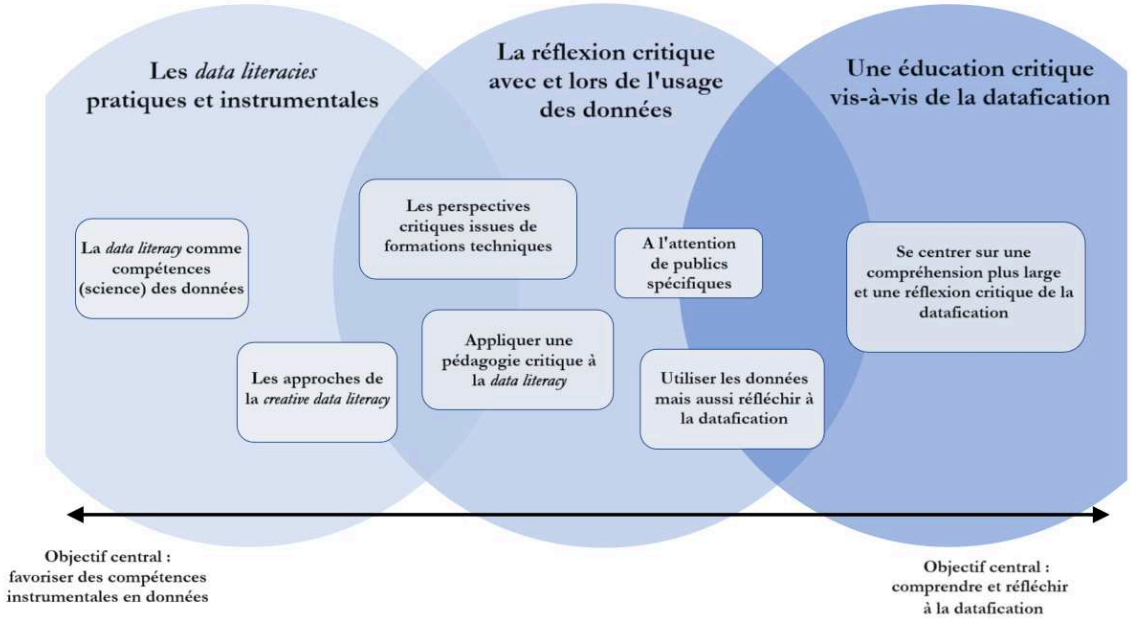

Figure 4.1.: Synthèse des littératies (critiques) des données - extraite et traduite de Sander (2023)

Trois grandes approches les caractérisent. La première approche, les *data literacies* pratiques et instrumentales, met en avant les compétences habituelles d'utilisation, nettoyage, analyse, interprétation et visualisation de jeux de données issus de la science des données. La seconde approche, la réflexion critique, s'applique seulement aux contenus médiatiques ou aux usages personnels des données. Elle n'aborde que peu la remise en cause de la datafication et s'attarde plutôt sur des points de vue individuels que collectifs. La troisième approche, la pédagogie critique vis-à-vis de la datafication, vise une compréhension et une réflexion plus large sur les incidences structurelles et systémiques de la datafication.

De manière constante, les littératies critiques des données ont pour objectif une remise en cause des *a priori* sur les données ainsi qu'une volonté d'encapacitation politique des individus, les publics ciblés différant selon les auteurs même si les « citoyens » restent malgré tout les plus cités. C'est ainsi toute une recherche et une remise en cause du pouvoir qui sert de socle aux initiatives. En un

sens, l'ensemble des réponses apportées par les littératies spécialisées contre la datafication et la dataveillance sont empreintes, voire se revendiquent, de cette vision.

4.3.4. La réponse systémique

L'analyse de la datafication dans ses aspects systémiques fait l'objet des travaux de plusieurs chercheurs. A commencer par Gray, Gerlitz, & Bounegru (2018) qui ont développé une littératie des infrastructures de données (*data infrastructure literacy*). Elle intègre les compétences classiques de lecture et de travail des jeux de données et ajoute la capacité d'intervenir au sein des infrastructures socio-techniques où les données sont créées et analysées. Plus qu'un simple transfert de compétences et d'extraction de valeur, pour ces auteurs il s'agirait d'une possibilité d'un engagement du public et d'expérimentations autour des infrastructures de la datafication.

> « ...faire de la place à la recherche collective, à l'expérimentation, à l'imagination et à l'intervention autour des données dans les programmes éducatifs et au-delà, y compris la manière dont les infrastructures de données peuvent être remises en question, contestées, remodelées et réutilisées pour s'aligner sur des intérêts et des publics autres que ceux initialement prévus [23] ».

Ina Sander a pour sa part proposé deux littératies spécialisées consacrées à cette thématique. La première est la littératie critique du *big data* (*critical big data literacy*) rapidement évoquée plus tôt. Inscrite dans les recherches issues des littératies critiques des médias et des recherches numériques critiques, elle doit amener au développement de la prise de conscience, de la compréhension et de la réflexion critique des individus sur les incidences du *big data*.

Poursuivant ses réflexions, Sander (2023) a ensuite conceptualisé une littératie critique de la datafication (*critical datafication literacy*) dans le but de développer un cadre pédagogique. Sander reprend et développe les points précédents qu'elle résume en trois objectifs. Tout d'abord, (1) promouvoir une compréhension systémique de la datafication par une prise en compte de ses processus, une compréhension de base des systèmes de données, afin de percevoir leur influence sur la vie des individus et sur le fonctionnement des sociétés. Puis, (2) d'encourager l'usage d'un esprit critique permettant de réfléchir et de concevoir différemment les technologies. D'atteindre en somme un *Mündigkeit*[24], soit pour l'autrice « la capacité d'affronter la société, la politique et l'économie modernes, de plus en

[23] « *...make space for collective inquiry, experimentation, imagination and intervention around data in educational programmes and beyond, including how data infrastructures can be challenged, contested, reshaped and repurposed to align with interests and publics other than those originally intended* ».

[24] Il s'agit d'un terme allemand signifiant « maturité » mais également « émancipation ». Source : https://dictionnaire.reverso.net/allemand-francais/M%C3%BCndigkeit

plus affectées et transformées par les technologies de données, de manière informée et autodéterminée » [25]. Enfin, (3) de permettre aux apprenants de devenir résilients (*resisting*) ou de s'émanciper vis-à-vis de la datafication. Sander note que si les stratégies individuelles de protection des données sont nécessaires, elles sont toutefois insuffisantes et appelle à aller au-delà des conseils de protections individuelles, à promouvoir et s'intégrer dans des actions collectives, en offrant par exemple aux citoyens les moyens de s'exprimer. De manière plus globale, l'objectif de cette littératie est de permettre d'imaginer d'autres futurs de données (*differents data futures*) et de se donner les moyens pour les accomplir.

4.3.5. L'activisme des données

Parmi les points évoqués dans les littératies critiques des données se trouve également l'activisme des données défini par Milan & Velden (2016) comme « un outil heuristique et polysémique employé pour réfléchir politiquement au *big data* selon la perspective des utilisateurs et des citoyens, en analysant en particulier l'évolution contemporaine de l'activisme vis-à-vis de la datafication [26] ». Cet activisme se situe pour ces auteurs aux frontières de l'activisme écologique, de la remise en cause de la datafication et de ses conséquences socio-politiques, et doit permettre « de réfléchir sur l'ubiquité des communications numériques et sur les dynamiques de médiation ayant cours dans la société des plateformes » ainsi que sur la « nature de l'activisme contemporain et de l'évolution de ses configurations médiatiques et de protestation [27] ». Il s'inscrit dans une logique politique nécessitant de prendre part aux initiatives liées aux données et de résister aux stratégies de collecte des données, en particulier dans le cas des données personnelles.

Ces dernières, considérées invariablement comme un nouvel or noir, une commodité centrale et une forme de capital (Artyushina, 2020), sont au cœur des enjeux commerciaux et politiques. De par la dénonciation de pratiques non éthiques, comme dans le cas d'Edward Snowden, se sont renforcées les pratiques de « sousveillance ». Ce terme, inventé en 1998 par Steve Mann, est basé sur le « réflexionisme », une philosophie et des procédures ayant pour objectif d'imiter/refléter et de lutter contre les organisations bureaucratiques en se réappropriant les technologies de contrôle (Mann, 1998). Plusieurs sous-tactiques l'ont intégré par la suite comme la « surveillance sombre » (*dark surveillance*) ayant pour but de résister et de remettre en question les méthodes de surveillances spécifiques aux populations noires (Browne, 2015), ou encore la « sousveillance numérique » (*digital sousveillance*).

[25] « *the ability to grapple with modern society, politics and economy that are increasingly affected and transformed by data technologies in an informed and self-determined manner* ».

[26] « *A heuristic, polysemic tool to think politically about big data from the perspective of users and citizens, analysing in particular the contemporary evolution of activism vis-à-vis datafication* ».

[27] « *the ubiquity of digital communication and mediation dynamics in the platform society (e. g., Castells 2009; van Dijck 2013), as well as on the nature of contemporary activism and its evolving protest/media configurations (e. g., Bennett 2012; Bennett/Segerberg 2013; Milan 2015b)* ».

L'objectif global de la sousveillance est « d'enregistrer les enregistreurs » (*recording the recorders*), de produire un « panoptique inversé » (*inverse panopticon*) (Mann, Nolan, & Wellman, 2003).

Dans ce cadre, Pangrazio & Selwyn (2019) ont proposé les littératies des données personnelles (*personal data literacies*) avec pour objectif de compenser l'asymétrie de l'information, à savoir l'ignorance des méthodes de collecte de nos données personnelles, des objectifs qu'elles servent et des manières dont elles sont utilisées (Brunton & Nissenbaum, 2015). Les données, étant construites socialement et étant sujettes à réinterprétation et à réutilisation, doivent être maîtrisées par les citoyens, ce qui, selon ces auteurs, n'est pas suffisamment abordé par la *data literacy*.

En particulier la capacité de déployer des tactiques de résistance au détournement des données personnelles pour les réemployer et répondre à des objectifs plus personnels et sociaux. Reconnaître les données personnelles est pour ces auteurs une première étape [28]. Quatre autres sont intégrées dans leur cadre conceptuel : comprendre les données (comprendre l'origine et la circulation des données personnelles), développer une approche réflexive sur les données (reconnaître les incidences des données sur nous-même et sur les autres), savoir utiliser les données (gérer et employer au mieux ces données) et développer des tactiques (créer différemment des données personnelles, à travers des tactiques de résistance ou d'opacité par exemple).

Hillman (2022) a pour sa part développé la littératie de la confidentialité des données (*data privacy literacy*). Centrée sur le domaine scolaire, elle étudie les effets sociotechniques des *edtechs*, ces technologies ayant pour but de faciliter l'enseignement et l'apprentissage comme *Kartable*, sur les écoles. Hillman appelle à leur régulation, la déconstruction de leurs discours, et plus largement à encapaciter les enfants pour leur permettre de résister à l'opacité du monde de l'entreprise. Il s'agit ici de recontextualiser l'École, d'en rappeler la « réalité » (*reality*) afin de comprendre que toute action entreprise en son sein est sujette à une perte de contrôle des données personnelles.

[28] Pour Pangrazio & Selwyn (2019) sont comprises comme données personnelles toute information permettant d'identifier un individu. Elles sont de trois types : les données transmises aux appareils et aux systèmes, les données extraites des utilisateurs par les appareils et systèmes, les traitements réalisés sur les données par les appareils et les systèmes.

4.3.6. Contrer les discriminations et révéler les invisibilisations

La question de la discrimination[29] et de l'invisibilisation[30] des problématiques liées aux expériences de vie de groupes marginalisés est un point de plus en plus soulevé dans la littérature (bien qu'existant depuis des décennies). Évoquer ces thématiques nécessite d'en faire de même pour les études intersectionnelles qui tendent à faire converger dans leurs analyses plusieurs de ces thématiques (la vision intersectionnelle étant elle-même revendiquée à plusieurs reprises dans la littérature).

L'intersectionnalité (*intersectionality*) est un terme conçu par la juriste américaine Kimberlé Crenshaw dans deux articles phares : *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics* (Crenshaw, 1989) et *Mapping the Margins : Intersectionality, Identity politics, and Violence Against Women of color* (Crenshaw, 1991), tous deux traduits en français dans l'ouvrage *Intersectionnalité* (Crenshaw & Delanoe, 2023). Nous nous baserons ainsi sur la traduction de Delanoe pour nos prochaines citations.

Prenant sa source dans les « études sur les femmes noires » (*Black Women's studies*) et dans « l'histoire intellectuelle des Noires » (pp. 10-11), l'intersectionnalité visait « à nommer l'expérience des femmes noires et à améliorer leur qualité de leur vie comme de celle de toutes les personnes exclues, marginalisées ou punies par les lois qui auraient dû les protéger » (p. 12), de faire valoir comment « race, genre et classe affectent le vécu des femmes noires aux États-Unis, en particulier sur le plan professionnel et légal » (p. 10). L'intersectionnalité vise à penser la discrimination autrement que par un prisme factoriel singulier (p. 13) car son expérience dépasse le racisme et le sexisme : ne pas prendre cela en considération fait abstraction de réalités conjointes vécues ici par les femmes noires. Sont cités en particulier les procès et les grilles de lectures des magistrats vis-à-vis des cas traités, majoritairement en défaveur des femmes noires dont l'expérience de vie est invisibilisée.

Dépassant par la suite le cadre du féminisme des femmes noires (*black feminism*), le concept s'est étendu à un ensemble varié de thématiques. Pour Jaunait & Chauvin (2013), l'intersectionnalité regroupait ainsi plusieurs fonctions :

[29] A titre illustratif, nous pouvons citer l'ouvrage *Weapons of Math destruction* d' O'Neil (2017) sur les inégalités inscrites dans les logiques algorithmiques. Bridges (2021) cite également les inégalités intégrées dans les distributeurs de savons, dans les reconnaissances faciales sur les couleurs de peaux foncées, dans le sexisme à l'œuvre dans les logiques de recrutements, dans les évaluations pour la contraction de crédits. S'y ajoute aussi la non-reconnaissance d'un genre (Keyes, 2018).

[30] D'Ignazio (2024) nomme cela les « données manquantes » (*missing data*), c'est-à-dire « les données qui sont négligées et ne font pas l'objet de priorisation, de collecte, de conservation et de publication par les institutions, et ce malgré les demandes politiques selon lesquelles ces données *devraient* être collectées et rendues disponibles », notre traduction de « *data that are neglected to be prioritized, collected, maintained and published by institutions, despite political demands that such data* should be *collected and made available* » (chapitre 2, paragr. 13).

> « [elle] est d'abord une théorie critique au sens où elle permet de formuler des intérêts normatifs spécifiques, ceux de minorités situées à l'intersection des grands axes de structuration des inégalités sociales et dont les intérêts ne sont pas représentés par des mouvements sociaux. Ce sont aux États-Unis des femmes noires qui ont contesté le monopole de représentation des mouvements auxquels elles cherchaient à participer, en dénonçant la façon dont certain.e.s représentant.e.s étaient considéré.e.s comme plus légitimes – en particulier les hommes dans le mouvement de libération noir et les femmes blanches de la classe moyenne dans le mouvement féministe. Comme critique, l'intersectionnalité pose ainsi la question de l'exclusion de certains groupes dans l'espace des mouvements sociaux comme dans celui du droit. Il s'agit autant de dénoncer des politiques de représentation excluant certaines minorités, que de souligner la façon dont les catégories juridiques résultant du combat contre les discriminations échouent à rendre compte de celles dont sont victimes des catégories de population situées à l'intersection de multiples situations minoritaires. ».

Corrélée aux études intersectionnelles, luttant contre l'« oppression des données » (*data oppression*) et empreinte de l'activisme des données, s'est constituée la « justice des données » (*data justice*). Selon Dencik, Hintz, & Cable (2019) et Dencik & Sánchez-Monedero (2022), cette dernière renvoie aux inégalités et discriminations à l'œuvre dans les logiques automatiques et algorithmiques. Elle est une réponse aux conséquences sociétales des technologies mais également des services basés sur les données. Elle a notamment pour but de favoriser une prise de conscience sur les pratiques de données dont les incidences se répercutent au sein de communautés spécifiques que nous allons à présent citer.

4.3.6.1. Le féminisme des données

Le féminisme des données (*data feminism*) a été développé par D'Ignazio & Klein (2020). Ces deux autrices ont souhaité ici réunir différents aspects du féminisme[31] dans un féminisme des données

[31] Le féminisme est défini comme le « courant de pensée et mouvement politique, social et culturel en faveur de l'égalité entre les femmes et les hommes » selon le dictionnaire Larousse, et « féministe » renvoie aux « femmes qui luttent pour d'autres femmes et s'identifient à toutes celles qui subissent l'oppression patriarcale ou la domination masculine » (Bourcier, 2012). Il est à noter que pour d'Ignazio et Klein, le féminisme des données concerne les femmes cisgenres, c'est-à-dire dont le sexe biologique correspond à l'identité sexuelle psychique, mais également les femmes transgenres, où le sexe biologique ne correspond pas à l'identité sexuelle psychique. Nous ajoutons cette précision du fait que certains courants féministes rejettent la présence des femmes transgenres dans leurs rangs et leurs réflexions (ces courants sont labellisés par leurs adversaires comme des groupes *TERF* (*Trans Exclusionary Radical Feminist*)) (Pearce, Erikainen, & Vincent, 2020).

« présentant une manière de penser les données, leur analyse et leur présentation qui nous renseigne sur les traditions issues de l'activisme féministe ainsi qu'à la pensée critique féministe [32] ».

Il s'agit d'une nouvelle manière de penser les usages et limites des données enrichie par « l'expérience directe, par un engagement dans l'action et par une pensée féministe intersectionnelle » (p. 8)[33]. L'objectif recherché est d'une part de comprendre comment les normes employées dans le traitement des données des sciences des données renforcent les inégalités déjà existantes, d'autre part d'employer la science des données pour défier et modifier la répartition du pouvoir (pp. 8-9).

Ceci passe par un examen de la manière dont opère le pouvoir, d'en « démasquer l'universalisme » (*unmask universalism*, expression employée par Davis (2008)), c'est-à-dire révéler la non-objectivité et la non-neutralité des choses, ces dernières étant marquées, influencées par les corps détenteurs du pouvoir (D'Ignazio & Bhargava, 2020), pour mieux le combattre. En ce sens, le féminisme des données doit remonter à la source des jeux de données pour caractériser et défaire les biais embarqués. Selon la vision de la philosophe américaine Donna Haraway, citée à plusieurs reprises dans leur ouvrage, tout savoir est situé socialement et politiquement, ce qu'elle nomme le « savoir situé » (*situated knowledge*) [34] (Haraway, 1988). Il en est de même pour les données qui ne sont pas neutres, encore moins quand elles sont traitées pour des sujets sensibles. Le versant technique ne doit ainsi pas être omis : les corps humains ne sont pas, dans cette vision, séparables des corps non-humains (ex : les animaux mais également les machines).

Le féminisme des données doit ainsi permettre de (1) promouvoir d'autres formes de savoirs qui façonnent la vie des personnes comme les émotions, (2) repenser les binarités et hiérarchies qui perpétuent les oppressions, (3) embrasser le pluralisme, (4) considérer les données dans leur contexte et rappeler ainsi qu'elles ne sont pas neutres mais issues de relations sociales inégales, et enfin (5) visibiliser le fait que le travail de la science des données naît de l'association de nombreuses personnes (D'Ignazio & Klein, 2020, pp. 17-18) . C'est en somme, rappeler que les personnes précèdent les données, que les données ne peuvent exister sans action humaine.

Parmi les sujets étudiés, l'on note les écarts salariaux, les traitements médiatiques ou encore les féminicides [35] qui font l'objet central de l'ouvrage *Counting feminicide : Data feminism in action* (D'Ignazio, 2024).

---

[32] « *...demonstrating a way of thinking about data, their analysis, and their display, that is informed by this tradition of feminist activism as well as the legacy of feminist critical thought* ».

[33] « *A way of thinking about data, both their uses and their limits, that is informed by direct experience, by a commitment to action, and by intersectional feminist thought*. »

[34] La réflexion d'Haraway s'inscrit dans les *Science studies* qui étudient les problématiques d'objectivité de la science. Haraway a ainsi appelé à une reconfiguration de la science, qui ne peut être à ses yeux objective du fait qu'elle est dominée les hommes et le patriarcat : elle est un domaine de cœrcition (Haraway, Allard, Gardey, & Magnan, 2007, p. 109).

[35] Selon Labrecque (2016), le terme désigne à l'origine « le meurtre misogyne de femmes par des hommes » et était nommé « fémicide ». Il va évoluer pour devenir « féminicide » selon une traduction littérale du terme *feminicidio* proposé

4.3.6.2. Le colonialisme des données

Nous avons évoqué précédemment une forme de « colonialisme des données » présentée par Couldry & Mejias (2019a) réduisant la vie humaine en données selon une logique d'exploitation capitaliste.

Cette expression a été reprise et réinterprétée par Ricaurte (2019). Pour cet auteur, le colonialisme des données est un régime centré sur les données [36] qui sert la « colonialité du pouvoir » (*coloniality of power*), à savoir la domination matérielle et symbolique de l'Europe sur les autres peuples (Quijano, 2007). A la différence du colonialisme ou de la colonisation, ce terme souligne « la dimension du rapport colonial "au-delà de sa période historique" » (Le Petitcorps & Desille, 2020). Apposé aux données, il imposerait des manières d'être, de penser, de ressentir qui amèneraient au rejet des humains de l'ordre social ainsi qu'au rejet des possibilités de mondes et d'épistémologies alternatives.

En ce sens, il reproduirait les inégalités de pouvoir exprimées en termes de violence telles que la « colonisation numérique » (*digital colonization*), les violences algorithmiques, les violences de genre, les discriminations de classe ou encore le racisme. Ricaurte appelle ainsi à une redéfinition de ce cadre pour inclure une compréhension du cadre systémique de la colonialité des données et développer contre ce dernier des résistances sous la forme de « désobéissances épistémiques » (*epistemic disobedience*)[37]. Il propose ainsi une approche multidimensionnelle intégrant les études décoloniales, intersectionnelles et féministes qui apporterait de nouvelles réflexions et de nouveaux moyens de défense. Mumford (2022) s'inscrira dans cette vision en rappelant que le colonialisme des données développé par Couldry & Mejias (2019a) ne fait que survoler la décolonialité, n'évoquant pas les travaux sur la colonialité présentés précédemment.

---

par Ríos (2010). Cette dernière précise que le féminicide recouvre trois dimensions : le manque de respect des droits humains des femmes, l'impunité dont bénéficient les meurtriers et l'irresponsabilité des autorités et surtout de l'État.

[36] Ce régime des données (*data-centric regime*) incorpore le capitalisme de surveillance, le capitalisme des plateformes et le capitalisme algorithmique.

[37] La désobéissance épistémique a été conceptualisée par le sémiologue argentin Walter Mignolo. Elle désigne le fait qu'une colonialité a toujours lieu dans les sociétés actuelles et influence les formes de savoir et de cognition, ce qu'il nomme la « matrice coloniale du pouvoir » (*colonial matrix of power*). Colonialité de la « modernité » qui présente la civilisation occidentale comme une civilisation de progrès, lui-même constitué de manière linéaire (Mignolo, 2013). La remise en cause de cette colonialité passe par une désobéissance épistémique, à savoir de décentraliser le savoir, de casser le lien (*delinking*) avec le savoir occidental persuadé de détenir une objectivité épistémique universelle, et de l'insérer dans de nouvelles temporalités et de nouveaux espaces de réflexion (comme par exemple ne plus prendre la Grèce Antique comme origine des réflexions) (Mignolo, 2011). C'est développer une pensée décoloniale dissociée de la « modernité » et basée sur une autre généalogie issue de la pensée d'Aimé Césaire (1913 - 2008), de Frantz Fannon (1925 - 1961) ou encore d'Ottobah Cugoano (1757 - 1791).

4.3.6.3. Les données des populations noires

Nous traduisons ici le terme de *black data* par « données sur les populations noires » pour mieux souligner leur objectif et marquer la différence avec les « données sombres » (*dark data*), souvent traduites justement par « données noires », qui désignent les données difficiles à trouver par les organisations car elles sont mal indexées, dépourvues de métadonnées, non collectées, mal archivées ou du fait d'un manque de compétences pour les traiter (Heidorn, 2008; Ionos, 2021; Schembera & Durán, 2020; Splunk, 2019).

Il est issu des *black data studies*, les études centrées sur les données employées pour décrire les vies des personnes noires. Elles se centrent à la fois sur les discriminations subies et sur l'invisibilisation de leur quotidien (Davis, Harper, & Christian, 2018).

Concernant la qualification même des données, deux auteurs ont proposé une description du concept de *black data*. Pour McGlotten (2016), selon une vision inscrite dans les *black queer studies*, ces études basées sur une vision intersectionnelle où les problématiques spécifiques aux populations à la fois noires et *queer* se recoupent :

> « J'utilise les données noires pour réfléchir à quelques manières historiques et contemporaines où les personnes *queer* et noires, comme d'autres personnes d'ascendance africaine et les personnes de couleur en général, sont saluées par le *big data*, dans lequel les *technes* de race et de racisme réduisent nos vies à de simples chiffres : nous apparaissons comme des matières premières, des flux de revenus, des écarts statistiques ou des vecteurs de risque. [38] » (p. 263).

Son objectif est de prendre en compte « de manière plus complète et objective » les vies des personnes noires et *queer*, et en ce sens de relier la notion de *black data* aux concepts « d'effacement, d'opacité et de cryptage ».

Pour Taylor (2024), les *Black datas* doivent « mettre en avant, défier, et destabiliser les discours sur les données qui, comme dans le cas des autres technologies, ont des constructions racialisées et des implications raciales explicites qui sont aseptisées ou négligées »[39]. Pour cet auteur, cette notion de *Black data* doit prendre en considération « comment les *technés* de la race et du racisme imprègnent les données de manière à refléter et à maintenir l'héritage de l'oppression raciale »[40]. Les *technés* se

[38] « *I use black data to think through some of the historical and contemporary ways black queer people, like other people of African descent and people of color more broadly, are hailed by big data, in which technes of race and racism reduce our lives to mere numbers: we appear as commodities, revenue streams, statistical deviations, or vectors of risk* ».

[39] « *I use Black Data to highlight, challenge, and disrupt normative discourses about data, which - as with other technologies - have racialized constructions and explicit racial implications that get sanitized or overlook* ».

[40] « *...how technés of race and racism imbue and imbibe data in ways that reflect and sustain legacies of racial oppression* ».

réfèrent ici aux pratiques qui « animent et promeuvent la race et le racisme, construites à partir d'une conceptualisation de la race comme une technologie sociale et le racisme comme une technique sociétale »[41].

L'auteur intègre son analyse dans le contexte de la datafication des étudiants et questionne le pouvoir politique des « pratiques archivistiques » (*archival pratice*), à son sens incapables de comprendre et de représenter fidèlement les expériences de vie des étudiants noirs. Ceci, du fait que la collecte, l'indexation, le rangement, la récupération et l'usage des données ne font pas l'objet, pour les institutions et les acteurs qui en ont la charge, d'un véritable questionnement et les confinent dans une vision raciale incomplète.

Il faut à son sens résister aux logiques bio-centrées (*biocentric logics*) qui considèrent les corps noirs comme de simples matériaux dont on dissocie les pensées, les sentiments et toutes expériences vécues au-delà du contexte universitaire, en remettant au centre des analyses l'identité noire. De rappeler qu'un étudiant est avant tout une personne noire avec ce que cela comporte de caractéristiques sociales. Connaître ces caractéristiques doit ainsi permettre de prendre en compte les spécificités des vies noires pour mieux faciliter le quotidien des étudiants noirs :

> « L'incapacité des structures, en charge du traitement des données au sein de l'enseignement supérieur américain, à comprendre la vie des Noirs a des conséquences, tout particulièrement parce que les données sont employées pour orienter et façonner les avenirs des Noirs. Comme McKittrick (2020) l'explique, la manière de vivre des Noirs n'est pas intégrée aux systèmes de grandes connaissances (de données), et la *dysselection*[42] est liée à la façon dont la *Blackness*[43] est positionnée dans les algorithmes et autres agencements de données comme irrationnelle et impossible à choisir. Les

[41]« *...the practices that animate and enact race and racism, building from the conceptualization of race as a social technology and racism as a societal technique.* ».

[42]Le terme *dysselection*, dont aucune traduction n'existe à notre connaissance en français, a été conceptualisé par la philosophe, romancière et essayiste jamaïcaine Sylvia Wynter à partir d'une analogie « méta-freudienne » (*meta-freudian*) et « méta-darwinienne » (*meta-Darwinianly*) où l'humanité est perçue comme une espèce hybride (Wynter & McKittrick, 2015). Dans ce contexte, Wynter (2003) prend le contrepied de Charles Darwin (*dysselection* s'opposant à *selection*) en précisant que les non sélectionnés (*dysselected*), en l'occurrence les populations non blanches, sont exclus des paradigmes ontologiques et donc exclus de l'humanité (*lack of normal human nature*) (Wynter, 1989). La publication de *L'origine des espèces* en 1859 fait basculer la définition de l'*Homme* qui passe d'un être rationnel conditionné par *Dieu* à un être biologiquement déterminé. Du fait du rôle prépondérant du biologique, la *Race* prend alors la place de *Dieu* comme déterminant de la réalité. Plus précisément, la *Race* fonctionne « au niveau de la réalité empirique comme incarnation de l'Autre dysgénique » (*level of empirical reality as the embodiment of the dysgenic Other*), faisant des populations non blanches des « autres », conceptualisés uniquement par ce critère et laissant donc la place à des théories et politiques racistes.

[43]La *blackness* renvoie aux caractéristiques spécifiques liées aux vies noires. Si le terme de « négritude » d'Aimé Césaire aurait ici pû être employé, il est historiquement et géographiquement situé et nous ne saurions dire s'il est adapté au contexte actuel. En l'absence d'une réponse définitive de la littérature, nous laisserons le terme non traduit.

> scènes d'assujettissement qui émergent pour les étudiants noirs, devenus des sujets de données, comprennent des discussions précaires sur les étudiants noirs souhaitables pour un recrutement sur la base de leurs données, des données noires à la construction idéalisée, qui conduisent à certaines visions imaginées et idéales des étudiants noirs et des décisions concernant l'intégration de rajectoires d'inscription des Noirs pour maximiser leur valeur et minimiser leurs risques [...] Les pratiques institutionnelles en matière de données effacent et excluent l'humanité des étudiants noirs - ignorant l'être, la pensée et les sentiments des étudiants noirs pour se concentrer ce qu'ils font et où. Certains aspects de la vie universitaire des étudiants noirs, non capturés par les pratiques normatives en matière de données, deviennent une sorte d'excès de données noires insondable et insaisissable au sein des logiques de l'enseignement supérieur [44] ».

4.3.6.4. Les données des populations indigènes

En continuité avec la colonialité des données, les travaux sur les données des populations indigènes (*indigenous data*) ont pour but d'échapper à la supposée objectivité des méthodologies occidentales. Méthodologies qui invisibilisent les vies indigènes, amenant à un « évitement quantitatif » (*quantative avoidance*) (Walter & Suina, 2019) où ces données ne sont pas recueillies. Par ailleurs, les données indigènes actuelles sont produites au sein de méthodes de recherches qui les décontextualisent et simplifient les problématiques liées aux vies indigènes [45]. Pour permettre aux populations indigènes de se baser sur des données indigènes fiables et permettre leur autodétermination, il serait nécessaire pour ces auteurs de recourir à des méthodologies indigènes. Ces dernières basent leurs recherches sur les points de vue, les perspectives, les valeurs et les expériences de vie des populations indigènes. Elles adoptent pour ce faire l'intersubjectivité selon deux modes :

> « l'intersubjectivité au sein des peuples et des manières d'être et de faire de ces peuples ; intégrant la culture traditionnelle et la culture actuelle, les croyances et les systèmes, les pratiques, l'identité et les manières de comprendre le monde et leur propre place, en

[44] « *The incapacity of US higher education data assemblages to capture Black life has implications, especially given their use to direct and shape Black futures. As McKittrick (2020) articulates Black livingness is outside the systems of big knowledge (data), and dysselection has to do with how Blackness is positionned within algorithms and other data assemblages as irrational and unchosable. The scenes of subjection that emerge for Black students, turned data subjects, include precarious discussions of which Black students are desirable for recruitment based on their data, idealized construction of Black data that lead to certain imagined-ideal visions of Black students, and decisions about Black enrollment trajectories to maximise their value and minimize their risk [...] Institutional data practices erase and preclude Black students' humanity - ignoring Black students being, thinking, and feeling to focus on where and what they do. Aspects of Black students' collegiate lives, not captured by normative data practices, become a kind of Black data excess unfathomable and uncapturable within the logics of higher education*».

[45] Maggie Walter nomme ces données les *5D*, correspondant à la disparité (*disparity*), la privation (*deprivation*), le désavantage (*disadvantage*), le dysfonctionnement (*dysfunction*) et la différence (*difference*) (Walter, 2016).

> tant que peuple, au sein de celui-ci ; et l'intersubjectivité en tant que personnes marginalisées, colonisées et dépossédées dont la vie quotidienne est encadrée et directement touchée par les relations et les interactions historiques et actuelles avec l'État-nation colonisateur. [46] ».

Les auteurs ajoutent par ailleurs trois recommandations : (1) développer des compétences techniques au sein des communautés liées au développement de sondages et de la collecte, de l'analyse et de la communication des données pour comprendre et produire des données adaptées, (2) renforcer la compréhension des méthodes et méthodologies de recherche au sein des partenaires tribaux en leur laissant déterminer quelles sont les données pertinentes à récolter et analyser, (3) défendre les méthodologies de recherches indigènes ainsi que la souveraineté des données indigènes pour garantir un contrôle des données.

Ce dernier point rejoint la thèse défendue dans l'ouvrage *Indigenous data sovereignty : Toward an agenda* (Kukutai & Taylor, 2016) où la souveraineté des données est définie comme une gestion de l'information en accord avec les lois, les pratiques et coutumes de l'État-nation où elle réside [47], devant permettre de « décoloniser » les données indigènes. Cette décolonisation intègre la possibilité de choisir ce qui doit être collecté dans les données et qui peut y avoir accès, de refléter les intérêts des peuples natifs (*native people*) et d'éviter d'intégrer des données qui leur seraient néfastes (Snipp, 2016, p. 52). C'est donc une réappropriation complète du cycle de vie des données selon des méthodologies et des visions dissociées mais qui ne sont pour autant pas en opposition frontale avec les logiques occidentales majoritairement à l'œuvre [48].

4.3.6.5. Les littératies raciales des données

Actuellement, deux littératies des données abordent frontalement les questions raciales présentées précédemment.

La première est la littératie raciale des données (*racial data literacy*), développée par Philip, Olivares-Pasillas, & Rocha (2016). Elle traite des questions raciales mais sous l'angle spécifique de la datavizualisation perçue comme un « artefact » pouvant reproduire les inégalités sociales et les hiérarchies sociétales au sein des salles de classe. Face à l'invisibilisation de la race (*colorblind*) dans les ana-

[46] « *intersubjectivity within peoplehood and the ways of being and doing of those peoples; inclusive of traditional and ongoing culture, belief and systems, practices, identity and ways of understanding the world and their own place, as a people, within it: and intersubjectivity as colonized, dispossessed marginalized peoples whose everyday life is framed through and directly impacted by their historical and ongoing relationship and interactions with the colonizing nation state* ».

[47] « *data sovereignty means managing information in a way that is consistent with the laws, practices and customs of the nation-state in which it is locate* » (p. 39).

[48] Il faut rappeler ici que le contexte de ces recherches est nord-américain et ne sont ainsi qu'un reflet de cet écosystème.

lyses, les auteurs proposent un remaniement des micro-contestations raciales et idéologiques (*racial-ideological micro-contestations*), ce concept développé et défini par le psychiatre Chester Pierce (1927-2016) comme « les échanges subtils, stupéfiants, souvent automatiques et non verbaux qui sont des "réprimandes" des [N]oirs par les agresseurs » [49]. Ces échanges peuvent prendre la forme d'insultes, d'invalidation des ressentis, de rejets, et ce à un niveau personnel ou systémique (Sue et al., 2007). La micro-agression s'oppose ici aux macro-agressions qui sont des démonstrations brutales, extrêmes, de ségrégation raciale (ex : lynchages).

Dans le cadre de cette littératie, il s'agit de reconnaître la datavisualisation comme le lieu possible de micro-agressions si elle est mal employée en cours. En ce sens, les auteurs appellent et à une formation des étudiants et des enseignants pour qu'ils deviennent conscients des biais raciaux présents dans les jeux de données et à un examen collectif des postulats idéologiques. La *data literacy* sans contextualisation politique ne peut amener pour ces auteurs qu'à une mauvaise pédagogie, amenant à toujours plus de mauvaises interprétations du quotidien des personnes victimes de racisme [50]. La littératie raciale des données est à leur sens une « véritable » littératie des données, à même de répondre aux défis d'une société en perpétuel changement.

La seconde est la vision de Johnson, Rydal Shapiro, DiSalvo, Rothschild, & DiSalvo (2021) qui applique la « théorie critique de la race » (*critical race theory*) [51] à ses analyses. Selon cette théorie, le racisme est considéré comme ordinaire (et non occasionnel) et la race est une construction sociale qui amplifie les expériences et les pensées des personnes marginalisées. Ainsi, cette littératie s'inscrit dans une approche intersectionnelle prenant en compte la race, la classe et le genre afin de déterminer les biais qui caractérisent les études et les pratiques sur les données. A la suite de l'analyse de *workshops*, où diverses problématiques appelant à des réponses par l'usage de données par un ensemble d'acteurs, les auteurs ont listé plusieurs problématiques.

Tout d'abord, les tensions entre le contexte de travail, les littératies critiques des données et l'agentivité épistémique (*epistemic agency*) où les données sont acceptées comme telles et non remises en question. Puis, la banalité (*ordinariness*) du racisme au point que ce dernier n'est même pas évoqué dans les réflexions lors des travaux sur les données. Enfin, ce qui est personnel est également communautaire et intersectionnel, au sens que tout choix résulte de logiques multiples et socialement entremê-

[49] « *subtle, stunning, often automatic, and nonverbal exchanges which are 'put downs' of [B]lacks by offenders* ».

[50] Bien que l'étude des auteurs est centrée sur les personnes noires, ils précisent néanmoins que leur littératie peut s'appliquer à tous types de populations où le critère de la race est pris en compte de manière défavorable. Le terme « racisé » pourrait ici s'appliquer à cette catégorie de personnes.

[51] Selon Delgado, Stefancic, & Harris (2012), la « théorie critique de la race » a été développée au cours des années 1970 au moment où une régression des acquis issus du mouvement pour les droits civiques aux États-Unis ainsi que de nouvelles formes de racisme ont commencé à émerger. Elle est décrite comme un « ensemble d'activistes et d'universitaires engagés dans l'étude et la transformation de la relation qui unit race, racisme et pouvoir » (*a collection of activists and scholars engaged in studying and transforming the relationship among race, racism, and power*, chapitre 1).

lées. Ce dernier point amène les auteurs à repenser les littératies critiques des données au-delà du simple point de vue individuel pour adopter une vision plus systémique.

Ils proposent par ailleurs des pistes pédagogiques : (1) la prise en compte du contexte de travail des données selon une vision critique, de comprendre qu'il existe des visions dominantes qui influent sur les méthodologies et cadres de penser qu'il s'agit de reconnaître et de décontruire, (2) l'ajout de contenus contextuels détaillant certaines réalités, comme les méthodes de surveillance de groupes racisés spécifiques, pour percevoir différemment les données, (3) de prendre la race en considération dans les analyses et (4) de savoir faire le lien entre les pratiques quotidiennes et professionnelles sur les données, de comprendre qu'elles ont une influence réciproque et ainsi permettre aux apprenants de mieux les employer.

4.3.6.6. Les données *queer*

Le terme *queer* signifie à l'origine « étrange, bizarre », désignant toute personne dont le comportement s'éloigne des normes socialement acceptables.

Il a été employé par la suite comme insulte à l'encontre des personnes déviant de la norme sexuelle (Eribon, 2003). Il a été redéfini par Eve Kosofsky Sedgwick comme « le maillage ouvert de possibilités, de lacunes, de chevauchements, de dissonances et de résonances, de laps et d'excès de signification, quand les éléments constitutifs du genre de quiconque, de la sexualité de quiconque, ne sont pas faits (ou ne peuvent pas être faits) pour signifier de manière monolithique » (Sedgwick, 1994, p. 7), traduction de Zielinska (2022).

Concernant les données *queer*, une proposition a été faite par plusieurs auteurs comme (Suen et al., 2020), (Duncan, 2021) ou encore (Ruberg & Ruelos, 2020). Est considérée comme telle toute recherche sur les conséquences socioculturelles des données (au sens des réalités qu'elles embarquent, qui les prédéfinissent) pour déterminer la corrélation entre les expériences de vie des personnes *LGBTQIA+*[52] et leur transcription au cœur des standards des données. Il s'agit d'une vision qui s'attelle à soulever l'objectivité et la pertinence des pratiques de traitement sur les données.

La question d'une déconstruction des visions hétéronormatives apposées aux données est recherchée afin de résoudre les problématiques de représentativité des jeux de données, en particulier les données démographiques recueillies dans le cadre de recensements, perçues comme des formes de savoirs partiels qui créent des vérités présentées comme objectives et dont le calcul est présenté comme une action neutre, impartiale (Browne, 2016).

[52]Lesbien, gay, bisexuel, transgenre, intersexe, asexuel. Le + renvoie à d'autres identités et orientations sexuelles.

> « Nous affirmons que les normes qui dominent les données démographiques sont insuffisantes pour prendre en compte la complexité qui caractérise nombre de vies lesbiennes, gays, bisexuelles, transgenres et queers [53] » (Ruberg & Ruelos, 2020).

L'objectif recherché n'est pas un rejet mais un décentrage des régimes de normalité, tout particulièrement les standards hétérosexuels et cisnormés, pour englober l'ensemble du spectre des identités (Duncan, 2021).

> « La critique queer est une critique, non de la normativité, mais de la normalisation c'est-à-dire de la tendance à juger des comportements sexuels, les identités et les désirs, à partir d'un seul ensemble de normes hégémoniques et dominantes issu d'une seule matrice » (Niedergang, 2023, p. 31).

La difficulté de conceptualisation des données *queer* repose sur la fluidité inhérente au concept de *queer* :

> « Si, comme le soutient la pensée *queer*, les sujets et les subjectivités sont fluides, instables et en perpétuel devenir, comment pouvons-nous recueillir des « données » sur ces sujets ténus et éphémères en utilisant les méthodes standards de collecte de données telles que les entretiens ou les questionnaires ? quelles significations pouvons-nous tirer et quel usage pouvons-nous faire de telles données quand elles ne sont que momentanément fixées et certaines ? [54] » (Nash, 2016, p. 2).

Ainsi la « queerisation » (*queering*) des données reste encore une interrogation, ce que Duncan (2021) soulignait, précisant que le choix du terme *queer* n'est pas sans conséquences, étant nécessairement limité pour représenter les vies *LGBTQIA+* [55].

Paul Guyan (2022) présentait dans son ouvrage *Queer data* une première approche méthodologique réflexive sur la collecte et le traitement des données, en cela proche d'une littératie critique des données, qu'il résumait dans un ensemble de questionnements centraux :

---

[53] « *We argue that dominant norms of demographic data are insufficient for accounting for the complexities that characterize many lesbian, gay, bisexual, transgender, and queer* ».

[54] « *If, as queer thinking argues, subjects and subjectivities are fluid, unstable and perpetually becoming, how can we gather 'data' from those tenuous and fleeting subjects using the standard methods of data collection such as interviews or questionnaires? what meanings can we draw from, and what use can we make of, such data when it is only momentarily fixed and certain?* ».

[55] A titre d'exemple, nous pouvons citer l'article de Scandurra, Mezza, Bochicchio, Valerio, & Vitelli (2019) qui cherchait à déterminer les niveaux de santé des personnes non-binaires et *genderqueer*, deux catégories qui échappent à la binarité de genre homme/femme et qui complexifient les approches méthodologiques médicales permettant de prendre en compte les réalités vécues par ces personnes ; les données recueillies, non « queerisées », s'avérant ici peu adaptées pour décrire leurs cas.

- Quel est votre but final ?
- Vos méthodes présentent-elles une restitution authentique des vies LGBTQ ?
- Qui prend les décisions sur les données qui ont un effet sur les personnes LGBTQ ?
- Est-ce que votre projet créé plus de bien que de mal ? Et pour qui ?
- Avez-vous (vraiment) besoin de plus de données ?
- Améliorez-vous la vie des personnes LGBTQ et examinez-vous d'un œil critique l'invisibilisation des caractéristiques majoritaires ?
- Vos méthodes de travail sont-elles ouvertes, accessibles et transparentes ?
- Les méthodes et systèmes de données néfastes peuvent-ils être réformés ?

4.4. Conclusion du chapitre

L'ensemble de ces littératies présentées démontrent une faiblesse sous-jacente des approches développées par la *data literacy* dans la compréhension et l'utilisation de ces données, tendant ici à montrer que la *data literacy* est trop généraliste. Revenant souvent dans la littérature, en particulier pour parler de la gouvernance des données, l'expression *one size does not fit all* désignant l'inadéquation d'une solution unique appliquée à l'ensemble des cas (Weber et al., 2009) semble ici adaptée pour souligner cet état de fait.

La conception généraliste de la *data literacy* n'est toutefois pas rejetée : elle constitue le centre névralgique sur lequel se constituent les littératies spécialisées. Ces dernières démontrent la capacité de la *data literacy* et plus globalement des données à se décliner et s'adapter pour répondre aux différents enjeux présentés. Elles forment ainsi une constellation de littératies qui entrent en résonance selon les aspects visés.

A la suite de cette synthèse, les travaux de Bhargava et al. (2015) prennent ici une toute autre ampleur. En 2015, ces auteurs avaient conceptualisé la *data literacy* comme « une littératie dans l'ère des données » (*literacy in the age of data*), définie comme « le désir et la capacité de prendre part de manière constructive à la société à travers et en lien avec les données »[56], devant nécessairement être adaptable aux contextes et problématiques émergentes et à venir :

> « Faire la promotion de la littératie dans l'âge des données ne doit pas seulement reposer sur les nouvelles technologies ou les nouveaux supports, mais implique de donner aux individus les moyens de naviguer dans leurs écosystèmes et sociétés actuels d'une manière qui soit significative et efficace pour eux. À l'ère des données, les nouvelles don-

[56] « *The desire and ability to constructively engage in society through and about data* ».

> nées et technologies continueront de remettre en question et de façonner nos capacités individuelles et collectives à apprendre, communiquer et prendre des décisions [57] ».

Le cas particulier des mutations de la *data literacy* opérées en réaction à la *datafication* et à la *dataveillance* permettent de rappeler un autre point essentiel de la littératie, à savoir la représentation du monde. Les biais contenus dans les données, au-delà de potentiellement fausser les résultats issus de leur analyse, peuvent engendrer des discriminations ou des invisibilisations de populations spécifiques ; les effets pouvant se recouper selon des logiques intersectionnelles. La prise de conscience et le décentrage des logiques à l'œuvre dans les productions de données deviennent dès lors un point critique central qui traverse les réflexions et offre une vision précise de ce fameux « pouvoir » des données qui influence les décisions et *in extenso* les vies humaines.

De manière plus globale, les appropriations de la *data literacy* présentées permettent de souligner que celles-ci ne se limitent pas à une discipline ou une thématique spécifique. Du fait de leur pervasivité, les données sont devenues l'affaire de l'ensemble des recherches et amèneront nécessairement à la conception prochaine de nouvelles littératies. Nous n'oublions néanmoins pas que la *data literacy* reste dépendante des discours : la prévalence d'une vision spécifique influencera nécessairement les appropriations comme nous avons pu le voir au cours du chapitre précédent dans le cas de l'intelligence artificielle et des algorithmes.

Dans le prochain chapitre, nous allons nous intéresser au cas français et tenter de déterminer si les représentations et appropriations de la *data literacy* diffèrent de celles analysées, nous permettant ou non de parler de *data literacy* à la française.

[57] « *Promoting literacy in the age of data should not solely be based in new technologies or mediums, but involve empowering people to navigate their current ecosystems and societies in ways that are meaningful and effective for them. In the age of data, new data and technologies will continue to challenge and shape our individual and collective capacities to learn, communicate and make decisions*. ».

Partie III.

Des visions françaises ?

Chapitre 5.

Le contexte français de la *data literacy*

> « *Le cyberespace. Une hallucination consensuelle ressentie au quotidien, dans le monde, par des milliards de techniciens autorisés, par des enfants y découvrant des concepts mathématiques… Une représentation graphique des données extraites des mémoires de tous les ordinateurs du système humain. Une impensable complexité. Des traits lumineux alignés dans le non espace de l'esprit, des amas et des constellations de données. Tels les éclairages d'une ville qui s'éloigne* »
>
> William Gibson, *Neuromancien*, 1984

5.1. Introduction

Nous avons pu précédemment décrire le concept de *data literacy* et souligner ses appropriations et évolutions multiples. Dans ce chapitre et le suivant, nous allons nous atteler à déterminer si nous pouvons évoquer une vision française de la *data literacy* et si oui, sous quelle(s) forme(s). Nous débuterons par une présentation des discours ayant accompagné la nécessité d'une *data literacy*, appelée le plus souvent en France *culture de la donnée*. Puis, nous présenterons les initiatives et les acteurs de l'écosystème français. Et enfin, nous détaillerons la conceptualisation de la *data literacy* selon un prisme universitaire et pédagogique.

5.2. La nécessité de la *data literacy* dans les discours

Concomitamment aux discours que nous allons présenter, des réglementations se sont développées pour encadrer le travail et l'accès aux données dont nous avons énuméré les plus représentatives en annexe 6.

5.2.1. Les discours institutionnels européens

Resituer le contexte européen est nécessaire avant d'aborder le cas français : la France étant intégrée au sein de l'Union Européenne et appliquant ses directives, elle est directement affectée par les décisions prises dans cette instance. La nécessité d'une maîtrise des données a été maintes fois soulignée par l'Union Européenne, dont les politiques se centraient à l'origine sur les données ouvertes. La *data literacy* a été de plus en plus sollicitée par la Commission Européenne vers la fin des années 2010, notamment dans le rapport *The Future of Government 2030+* :

> « De futures littératies sont nécessaires pour permettre aux citoyens de participer à la prise de décision anticipative, en reconnaissant le contexte d'incertitude et de complexité et en renforçant la résilience individuelle et sociétale pour travailler en collaboration afin d'y remédier. La maîtrise de la cyberlittératie et de la littératie des données sera importante pour que chacun puisse mieux comprendre le potentiel et les limites des plateformes numériques, leurs modèles économiques sous-jacents et leur gouvernance, afin d'éviter que la société et le gouvernement ne soient manipulés. La pensée critique doit être encouragée, à travers le système éducatif et au-delà sur le lieu de travail et dans la société civile, ce qui comprend la compréhension des médias numériques mais aussi d'autres aspects de la vie des gens. La littératie des politiques est également très importante, tant pour le présent que pour l'avenir : l'éducation à la citoyenneté existe déjà dans certains pays de l'UE, mais pas partout [1] » (Vesnic-Alujevic, Stoermer, Rudkin, Scapolo, & Kimbell, 2019, p. 84).

[1] « *Futures literacies are needed to enable citizens to participate in anticipatory decision making, recognising the context of uncertainty and complexity and building up individual and societal resilience to work collaboratively to address these. Cyber and data literacy will be important for everyone to understand better the potential and limitations of digital platforms, their underlying business models and their governance, in order to prevent society and government from being manipulated. Critical thinking should be nurtured, through the education system and beyond in the workplace and civil society, including understanding digital media but also other aspects of people's lives. Policy literacy is also very important, both for the present and for the future: citizenship education already exists in some EU countries but not everywhere.* ».

Un an plus tard, lors de la pandémie de Covid-19, cette volonté d'une maîtrise des données a été réaffirmée. La Commission Européenne (2020b) déclarait alors que les données étaient nécessaires pour faire progresser le savoir scientifique :

> « La pandémie de COVID-19 actuelle a montré l'importance que revêtent désormais les ressources numériques pour nos économies et comment les réseaux et la connectivité, les données, l'IA et le calcul haute performance ainsi que les compétences numériques élémentaires et avancées soutiennent nos économies et nos sociétés en permettant la poursuite du travail, le suivi de la propagation du virus et l'accélération des activités de recherche sur les traitements et les vaccins. » (p. 2).

Corollairement, la Commission a publié la même année un rapport sur la *Stratégie européenne sur la donnée* où était soulignée la faible appropriation de la *data literacy*[2] et le fait que ce manque affecterait « la capacité de l'UE à relever les défis de l'économie et de la société fondées sur les données » (pp. 10-11). Cette situation devait ainsi être compensée par l'apport d'un « parcours de renforcement des compétences » dans le cadre d'une « éducation générale aux données » par la formation aux « compétences numériques de base » (p. 26) (Commission Européenne, 2020a).

Plusieurs initiatives ont ainsi émergé comme le *DigComp* présenté en chapitre 1 qui intègre la *data literacy* dans sa seconde version ; le *Digital Innovation Lab (iLab)*[3] ayant pour but « développer des méthodologies, des modèles et des outils innovants, durables et réactifs pour améliorer l'élaboration des politiques et la prestation de services [4] » et qui a présenté en 2023 six *iTalks* sur la *data literacy*, notamment sur la question de son usage au sein du service public ; ou encore le programme *Data literacy in upper primary education (DALI4US)*[5] en 2024, pour une fin prévue en 2026, ayant pour objectif de « doter les enseignants du primaire de la confiance, des connaissances, des compétences et des attitudes nécessaires pour enseigner efficacement la littératie des données »[6].

[2] Il est important de noter que la version anglaise parle de *data literacy* tandis que sa traduction française parle « d'éducation numérique » et parfois d' « éducation générale aux données ».

[3] https://joinup.ec.europa.eu/collection/european-commission-digital-innovation-framework/solution/digital-innovation-lab-ilab

[4] « *developing innovative, sustainable and responsive methodologies, models and tools for improved policy-making and service delivery.* »

[5] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/projects-details/43353764/101132912/ERASMUS2027

[6] « *aims to equip primary school teachers with the necessary confidence, knowledge, skills, and attitudes to teach data literacy effectively*».

5.2.2. Les discours institutionnels français

En France, l'appui sur les données précède la nécessité d'une culture des données. Cela s'est d'abord exprimé par le truchement de l'*open data* au cours des années 2010, faits qui sont détaillés dans les travaux de (Chignard & Marchandise, 2012) et de Samuel Goëta, plus précisément dans sa thèse (Goëta, 2016) et dans son ouvrage *les données de la démocratie* (Goëta, 2024).

Au niveau de l'État, le rapport du représentant du numérique et ex-administrateur général des données Henri Verdier présentait en 2015 la nécessité d'une organisation globale des données « produites ou détenues par l'État » afin d'en assurer « la qualité, la fraîcheur, l'interopérabilité, la disponibilité dans des formats techniques en facilitant l'utilisation rapide et la meilleure circulation possible afin que chaque agent public de l'État comme des collectivités locales bénéficie des informations nécessaires à l'exercice de ses missions, et ce, dans le respect des secrets légaux qui protègent d'importantes libertés fondamentales et les intérêts fondamentaux de la nation » (Verdier, 2015a). La gouvernance des données était ici mise en avant et intégrait une logique ouverte basée sur le partage des données. Ce rapport sera suivi d'un second en 2018 intitulé *La donnée comme infrastructure essentielle* qui réalisera un état des lieux de la politique de la donnée et des progrès accomplis sur cette période (Verdier, 2018).

Mais le document où émerge véritablement la « culture de la donnée » est le rapport Bothorel (2020). Publié par le député Eric Bothorel sous le gouvernement de Jean Castex, il avait pour ambition des « réformes ambitieuses, notamment pour participer aux transformations en cours au niveau européen, mais aussi un grand nombre de mesures très raisonnables, qui n'appellent aucun "grand soir" et sont des actions réalisables sous un an, pouvant avoir des effets retentissants pour l'avenir. Si le gouvernement n'anime pas cette politique, la France manquera une occasion majeure de renforcer tout à la fois la confiance dans l'action publique, l'efficacité des politiques publiques, et la connaissance et l'innovation dans l'ensemble de l'économie, à un moment où la crise sanitaire et économique et le résultat du Grand Débat National en ont pourtant révélé le besoin impérieux » (p. 6). Il soulignait la nécessité grandissante d'une acculturation aux données des acteurs publics pour soutenir l'ouverture des données et faire progresser les politiques publiques :

> « Quatrièmement, la fonction publique a besoin d'une culture de la donnée et du code. Ce besoin fait partie de ceux qui ont été le plus mis en avant par la consultation publique. L'enjeu est d'agir sur le stock, et non seulement par les recrutements : trop de fonctionnaires, hauts placés et aux postes de commandement, éprouvent aujourd'hui de la peur à l'égard de l'ouverture des données, le plus souvent par ignorance. L'ouverture ne pourra progresser en attendant que des générations plus averties et sensibilisées aux enjeux parviennent jusqu'à ces postes. En général, les agents

> n'ont aujourd'hui que peu d'incitations à se former et à contribuer à la transformation de leur service par la donnée » (p.12).

Ce rapport sera suivi d'un certain nombre de feuilles de route au sein des ministères impulsées le 27 avril 2021 par le Premier Ministre Jean Castex lors de la publication de la circulaire 6264/SG *Politique publique de la donnée, des algorithmes et des codes sources* à l'attention des ministres, des ministres délégués, des secrétaires d'État et des Préfets de région dans laquelle est précisée que « la politique de la donnée doit constituer une politique stratégique de l'État dans ses relations avec tous ses partenaires, notamment dans les collectivités territoriales et les acteurs privés » (p. 2). Incitant à une « meilleure circulation de la donnée, des algorithmes et des codes, dans des formats ouverts et exploitables par des tiers » (p. 2) et à une synergie entre les services chargé du numérique des ministères, le Premier Ministre appelait à la création d'un administrateur ministériel des données, chargé d'élaborer la stratégie du ministère dans ce domaine, de coordonner les parties prenantes et d'être le point de contact des utilisateurs des données et des applications numériques relevant de leur périmètre (Castex, 2021).

Nous avons listé et synthétisé en annexe 7 l'ensemble des feuilles de route et les objectifs visés en termes de « culture de la donnée »[7]. Dans leur ensemble, les feuilles de route considèrent « la » donnée comme un actif « porteur de valeur » dont l'ouverture, le partage et l'exploitation sont nécessaires pour soutenir les décisions, anticiper les actions et mieux rendre compte des activités de l'État. La « culture de la donnée » est ici le moyen d'acculturer les agents des ministères à la « révolution des données », obtenue par des actions de sensibilisation (ex : des ateliers sur les caractéristiques des données et leur pilotage) et de formation, souvent techniques. Le tout afin de permettre une maîtrise assurant un partage des données entre acteurs. Cela s'est notamment concrétisé par la nomination d'administrateurs ministériels des données (AMD) dans tous les ministères le 15 mai 2021 en charge, avec une première session d'un « Conseil interministériel des Administrateurs de la Donnée (CIAD) » réalisée le 24 juin 2021 (Lucchesi, 2021).

En septembre 2023 est sorti le *Rapport de la mission Data et territoires*[8] qui devait répondre à la question suivante : « comment faciliter et améliorer l'exploitation des données par les collectivités territoriales ? » (Hennion, Altounian, Monthubet, Chignard, & Picavet, 2023). Dans ce cadre, la « culture de la donnée » sert de « premier élément » et de « condition préalable à une généralisation de l'utilisation des données par les collectivités » (p. 7). En effet, « la donnée a bien contribué, de manière substantielle, à la modernisation et à la transformation de l'action publique tant au niveau de l'État que des collectivités » (p. 21). De ce fait, « le développement de la culture de la donnée doit

[7] Les feuilles de route sont par ailleurs consultables sur ce lien : https://www.data.gouv.fr/fr/datasets/feuilles-de-route-ministerielles-sur-la-politique-de-la-donnee-des-algorithmes-et-des-codes-sources/

[8] Ce rapport a été commandité par le Ministre de la Transformation et de la Fonction publiques à Christine Hennion (conseillère municipale de Courbevoie), Magali Altounian (adjoint au maire de Nice et conseillère régionale Région Sud) et Bertrand Monthubert (président du Conseil National de l'Information Géolocalisée, d'Ekitia et d'OPenIG) avec l'appui de Simon Chignard (expert indépendant) et d'Olivier Picavet (Etalab).

permettre d'acquérir une meilleure compréhension des enjeux de la représentation du réel par des données, de l'importance des choix qui sont opérés qui ne sont pas seulement techniques » (p. 52). Il s'agit donc d'une volonté de maîtrise des données, mise en lumière par les initiatives d'*open data* au cours des années 2010, étendue depuis 2020 à l'ensemble des données. La « culture de la donnée » est ici l'objet de sensibilisations et de formations à l'attention des acteurs publics pour conduire la politique des données de l'État ; les données étant considérées comme le socle de la « souveraineté politique et numérique » de la France (voir la figure 5.1).



Figure 5.1.: Tweet du Ministre de l'Économie, des Finances et de la Souveraineté industrielle et numérique du 12 septembre 2022

5.2.3. Les discours des entreprises françaises

Les entreprises françaises véhiculent les discours internationaux abordés dans le chapitre 1 traduits par la nécessité d'une maîtrise des données. L'étude 2023 [9] de l'Observatoire de la « maturité data »[10], basée sur les retours de 288 répondants dont 72% de dirigeants et de cadres de PME, soulignait que « 24%, [soit un] 1 quart des entreprises se déclarent en capacité d'exploiter correctement leurs données » (slide 2). Si la culture des données est un sujet qui « se démocratise », 48% des décideurs « ont un niveau faible de connaissance des sujets data » et 55% des collaborateurs « ont un niveau faible de sensibilisation aux sujets data » (slide 16).

En tant que nécessité, la *data literacy*, traduite invariablement en « culture des données » (Talend, 2021), « culture data » (Datagalaxy, 2021), « culture data-driven » (Apogea, 2021), « datalphabétisation » (Fiaux, 2020) ou encore « data lettrisme » (Krim, 2022), apparaît au début des années 2010 dans la littérature professionnelle française.

[9] https://www.datasulting.com/observatoire-data/resultats-enquete-2023/

[10] Créé en 2020 par l'entreprise *Datasulting* (https://www.gpomag.fr/web/etudes/observatoire-de-la-maturite-data-des-entreprises), l'Observatoire a pour objectif d'aider à comprendre comment sont valorisées les données d'entreprises à

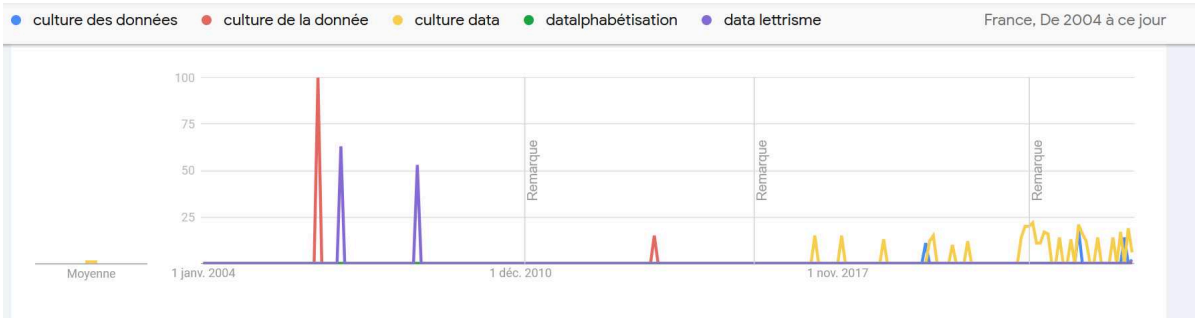

Figure 5.2.: Requêtes françaises obtenues à partir des termes « culture des données », « culture de la donnée », « culture data », « datalphabétisation » et « data lettrisme » dans Google Trends sur une période s'étalant de 2004 à 2024 - export réalisé le 24 avril 2024

En confrontant ces termes au sein des requêtes françaises sur *Google Trends* (voir la figure 5.2), nous avons pu noter que « culture de la / des données » précède le terme « culture data » aujourd'hui le plus recherché depuis 2015. Le terme « culture data-driven », si l'on aurait pû penser qu'il serait employé de façon équivalente, n'apparaît qu'une fois en janvier 2008 (100 requêtes). Les autres termes sont quasiment inexistants dans les requêtes. Etrangement, le terme « data literacy » n'apparaît quant à lui qu'à de rares reprises en avril 2005 et novembre 2006.

La *data literacy* est en revanche citée à la suite d'enquêtes sur le niveau de maîtrise des données des individus. Celle réalisée par l'entreprise Qlik en 2022 rappelait le rôle central de la *data literacy*, perçue comme la « compétence la plus demandée » pour 2030 [11], notamment du fait que nombre d'employés ne sont toujours pas « datalphabétisés » (12% en 2017 (Beky, 2017), 24% en 2018 (Qlik, 2018) ; 7% des employés français seulement se sentaient pleinement *data literate* en 2022 (Qlik, 2022a)) et que cette absence de compétences s'accompagnait d'une perte notable de capital pour les entreprises (9,8 milliards en 2020 (Les échos executive, 2020)). Avec l'engouement pour l'intelligence artificielle et les algorithmes, il a été rappelé que la « culture data » restait un élément central si l'on voulait explorer ces autres domaines (Chapron, 2024), notamment pour assurer une bonne qualité des données et ainsi permettre à l'IA et aux algorithmes de fonctionner correctement (Gutierrez, 2024).

Alimentant les logiques *data-driven* des entreprises, les données, majoritairement appelées *data*, doivent ainsi accompagner les décisions opérationnelles et stratégiques ainsi que renforcer leur *business model*. L'acculturation des employés est perçue comme un levier puissant pour entreprendre l'ensemble des opérations de traitement et de partage des données. Sans « culture data », la plupart

travers 5 volets que sont (1) le potentiel data, (2) la stratégie data, (3) la culture data, (4) l'organisation et la gouvernance, (5) les compétences data (https://observatoire-data.fr/).

[11] Cette étude a été citée au chapitre 1.

des projets en données échouent [12] : un constat qui a été à la source d'un remaniement des politiques de formation. Dans ce contexte, la « culture data » doit être nécessairement disséminée :

> « La transformation digitale est donc avant tout culturelle : elle est certes "réussie" lorsque les équipes se sont familiarisées avec de nouveaux outils, mais surtout lorsqu'elles comprennent les enjeux et l'intérêt de la data. Il en résulte une meilleure exploitation de la data, une plus grande efficacité dans le travail et des effets de synergies dans toute l'organisation. Une culture data-driven se construit. Elle doit être cohérente, comprise et appréhendée pour faire pleinement partie du quotidien des équipes. Les entreprises peuvent par exemple inscrire la data dans leurs routines - lors de plénières, d'entretiens annuels, et d'autres événements internes - ou encore impliquer les collaborateurs dans la génération et la compilation de la data pour les rendre acteurs des processus. Lorsque leur quotidien est animé par la data, les collaborateurs se l'approprient plus facilement » (Adreit, 2020).

Il s'agit donc d'instaurer une culture commune sur les données pour que l'ensemble des salariés soient formés et parlent un même langage. Or une difficulté constante se retrouve dans l'application de cette « culture data », notamment en termes de vulgarisation des contenus et d'adhésion des salariés. Selon certains discours, des postes spécifiques basés sur la conduite de l'acculturation, voire « l'évangélisation », des salariés existent.

Au-delà des traditionnels *chief data officer* (Data Galaxy, 2019) et des *data architect* (APEC, 2020), ont été proposés des *data literacy manager* (Chapron, 2021) ou encore des *data bakers* (Data Galaxy, 2021). Les premiers sont des experts techniques disposant de compétences communicationnelles fortes permettant de fédérer les acteurs et de leur faire accepter les formations tandis que les seconds est un terme inventé par l'entreprise *DataGalaxy* pour désigner l'ensemble des acteurs de la donnée (sont cités les *data scientists*, *chief data officers*, responsable BI, *data architects*, délégué à la protection des données (DPO) ou encore les *data stewards*) suffisamment spécialisés pour transmettre leurs connaissances. C'est en soi plus une fonction annexe d'un poste déjà existant qu'un poste à part entière.

Plusieurs propositions d'application de la « culture data » ont ainsi fleuri au cours des dernières décennies. Sont notamment citées des actions de sensibilisation (Bellec, 2022) et de formation (A5admin, 2019) des salariés, de souligner l'importance des données dans les actions des entreprises, et ce dans les gestes les plus quotidiens. C'est avant tout un ensemble d'actions de communication ayant pour but d'obtenir l'adhésion des acteurs et ainsi prévenir les freins potentiels au changement

[12] L'étude Harvard Business de 2021 estimait que 92% des entreprises ne parvenaient pas à devenir *data-driven* du fait d'un manque de culture data : https://hbr.org/resources/pdfs/comm/aws/ReassessingData-DrivenBusinessTransformation.pdf

(GPO Mag, 2023). Elle se bâtit sur le temps long et n'est pas une « fin en soi » (*Les bonnes pratiques pour développer une culture data dans votre entreprise*, 2024) mais un levier d'action pour garantir la pérennité des projets data. Ces propositions émanent d'entreprises françaises dépositaires d'un certain capital leur permettant d'entreprendre des actions d'acculturation (par le recrutement de personnel qualifié par exemple) mais également d'entreprises spécialisées spécifiquement sur les formations data qui ont émergé entre-temps [13] sur lesquelles nous allons revenir en seconde partie.

Dans un contexte concurrentiel, la « data » est un enjeu important, tant d'un point de vue financier, éthique et politique. Elle nourrie non seulement les décisions mais également les domaines connexes de l'IA, des algorithmes et des codes : elle est à nouveau ce système à part entière qui nécessite une vision englobante pour en assurer le parfait fonctionnement. Ainsi, la « culture data » accompagne et soutient les entreprises à affermir leur politique des données afin de devenir pleinement *data-driven*.

Il existe néanmoins une ambivalence de la « culture data » au sens de sa généalogie : si aujourd'hui les acteurs liés aux entreprises se revendiquent fréquemment de la *data literacy*, cela n'a pas été toujours été le cas ou du moins cela n'est pas visible, laissant entrevoir une autre généalogie (potentiellement hybride), voire une conceptualisation *ex-nihilo* basée sur la nécessité de répondre au besoin de développement d'une dite culture sur les données.

Par ailleurs, la question de la place des données au sein des entreprises a évolué. Nous avons pu le voir dans les précédents chapitres que si la maîtrise des données est loin d'être un objectif récent, la place accordée à cette maîtrise n'est toutefois plus la même. Longtemps placées sous contrôle des départements informatiques / SI, les données sont aujourd'hui de plus en plus traitées au sein de départements « data » ou « données » détachés des départements informatiques / SI. En effet, la création de « pôles / départements data », pilotés par un *chief data officer*, est de plus en plus sollicitée par les entreprises[14] pour adresser leurs objectifs et aiguiller leurs décisions.

Souvent intégrés dans une vision tirée de la *data science* qui implique le recrutement d'experts techniques dans une majorité de postes, les *data scientist* et les *data analyst* notamment, ces pôles sont fortement liés aux exigences des départements *marketing* qui s'appuient sur les analyses de données. A ce propos, la question de la *BI* (*business intelligence*) est un point constamment évoqué : la prédominance technique incarnée par les outils de type *dashboard* tendent à recentraliser les données

[13] Leur nombre exact ne peut pas être déterminé du fait que certaines combinent leurs activités « data » avec des activités « numériques » (telles que référencées sur https://www.francenum.gouv.fr/activateurs).
[14] Voir à ce propos (Datasulting, 2022), (Upward Data, n.d.) et (Morning Data Chat, 2023).

et leur gestion[15] sous un mode similaire à la base de données sous la *data administration* des années 1960/1970.

Les pôles data, s'ils ont une identité et une structuration propres, sont donc à l'intersection des enjeux de plusieurs départements connexes (informatique et marketing) et intégrés dans la stratégie *business* globale de leur organisation. La « culture data » sert en ce sens de liant en traversant l'ensemble des pôles et subit un réagencement global de par les influences culturelles propres à chaque département.

Enfin, ces pôles data sont touchés par la convergence évoquée aux chapitre 1 et 3 : l'on commence à voir apparaître des pôles « data & IA »[16] qui combinent les expertises des deux domaines. Il est important de noter que cette convergence ne concerne pas uniquement les entreprises mais bien d'autres types d'organisations[17].

5.2.4. Conclusion de la première partie

L'inflation des données et la nécessité de leur maîtrise à un niveau systémique, impulsées par le *big data*, a été réaffirmée depuis la pandémie de Covid-19. Si les organisations partagent une même volonté de faire des données une ressource centrale pour la conduite de leurs projets, elles éprouvent néanmoins de grandes difficultés pour acculturer les individus qui les composent. Le « phénomène des données » a entraîné une inflation de nouveaux pans d'expertise[18] et de nouveaux secteurs et services au sein des organisations. Or, si l'aspect technique a été traité, se pose encore la problématique de compréhension et d'adhésion de cette nouvelle configuration. C'est pour y répondre qu'un certain nombre d'initiatives ont vu le jour, portées par une diversité d'acteurs que nous allons maintenant présenter.

[15] Nous pouvons ici prendre pour exemple la gouvernance des données en entreprise qui est souvent traitée sous l'angle des logiciels et de la technique en général et ce sur l'ensemble du cycle de vie des données : *Informatica* et *Talend* pour le *Master Data Management*, *Tableau* et *Power BI* pour la *datavisualisation* et le *reporting*, *DataGalaxy* pour le catalogue des données, etc. Nous précisons néanmoins que la vision purement technique seule est de moins en moins sollicitée, à la place combinée avec des aspects culturels (la construction d'une culture organisationnelle commune et sa diffusion).

[16] Voir à ce propos (Benoit, 2022) et (Chanel, 2024).

[17] Voir par exemple (Adnouest, 2024).

[18] Dans sa note de recherche, Philippe Nieuwbourg listait en 2020 pas moins de 70 métiers liées aux données dont les contours ne sont pas encore réellement arrêtés aujourd'hui (Nieuwbourg, 2020).

5.3. Les acteurs et initiatives autour de la *data literacy*

5.3.1. Les acteurs entrepreneuriaux

Nous ne pouvons ici pas lister l'ensemble des acteurs entrepreneuriaux qui traitent spécifiquement de la formation aux données du fait de leur grand nombre. Nous pouvons néanmoins distinguer deux grandes catégories.

En premier lieu, les entreprises qui réalisent des formations data au sein de leurs équipes mais également auprès de publics externes, généralement leurs clients. Nous parlerons ici en particulier des *data schools* qui ont un rôle de plus en plus prépondérant. Les *data schools*, ou « écoles des données », sont des organismes de formation aux données, mais également à l'intelligence artificielle et aux algorithmes. Le premier exemple de ce type est le projet *School of data* de l'*Open Knowledge Foundation* que nous avons cité en chapitre 1 et qui avait pour but « de donner plus de pouvoir à la société civile en enseignant les compétences nécessaires pour réutiliser des données »[19]. En France, a existé l'*École des Données* qui proposait des cours ainsi que des hackhatons[20].

Il existe par ailleurs, sans aucun lien avec l'*Open Knowledge Foundation*, divers modèles d'école de données, souvent hybrides (Auffray, 2022). La *data school* de *La Poste* a été particulièrement mise en avant. Créée en 2022, elle a accueilli sa première promotion en 2023 selon des modalités spécifiques de formation :

> « Au travers un cursus qualifiant d'une durée de 3 à 12 mois, l'agence souhaite accompagner la montée en compétences dans quatre professions en plein essor, à savoir data product owner, data analyst, data engineer et data scientist. L'objectif est de former jusqu'à 250 profils par an de tous types aux métiers des data, de l'IA et développement informatique dans les 3 ans » (Arène, 2023).

Nous pouvons aussi citer l'*Ecole de la Data et de l'IA* de *Business & Decision*, un groupe d'*Orange* qui propose depuis 2018 des formations en *Business Intelligence, Intelligence Artificielle, Data Science, Data Engineering, Data Analyse* et d'autres sujets autour du domaine de la Data, du Digital et de l'IA [21]. Ou encore la *Mazars Data School* créée en 2023 pour « permettre à chaque collaborateur Mazars d'être formé et acculturé aux enjeux de la Data à travers des cursus adaptés, évolutifs et certifiants sur les solutions et technologies du programme » (Mazars, 2023).

[19]https://fr.okfn.org/ecole-des-donnees/
[20]https://ecoledesdonnees.org/index.html
[21]https://ecoledeladataetia.com/

En second lieu, les entreprises / *start-ups* entièrement dédiées à la formation data [22]. L'acculturation des publics cibles (employés, chef d'entreprises, etc.) soutient la définition et le pilotage de la stratégie data des entreprises, déployée selon des modalités de sensibilisations et de formations propres aux contextes locaux comme dans le cas du projet *Data & Quartiers* (2019 - 2022) lancé par *RésO Villes*, soutenu par l'*ANCT* et accompagné par la société *Civitéo* dirigée par Jacques Priol [23] qui visait l'expérimentation de nouveaux outils et de nouveaux usages des données sur le territoire ainsi qu'à doter les acteurs des quartiers « d'une culture des données utile à leurs missions » (Résovilles, 2022).

Le rapport récapitulatif de ce projet intègre dans son étape 1 la sensibilisation explicitée ainsi : « L'étape de sensibilisation est primordiale. Au démarrage de chaque projet, nous avons pris le temps d'acculturer les partenaires aux enjeux de la gestion des données. On parle aussi de data littératie et de data mindset » (Résovilles & Civitéo, 2022, p. 17).

Au-delà des personnes morales, des entrepreneurs individuels ont œuvré à la définition de la culture des données depuis près d'une décennie. C'est le cas de Simon Chignard. Il s'agit d'un consultant ayant eu une longue expérience au sein d'*Etalab* (2014-2021), ayant suivi l'élaboration de la loi pour une République Numérique de 2016 et ayant collaboré avec un certain nombre d'organismes sur des projets « données » comme la *FING*, *Open Data France* ainsi que des collectivités territoriales. Son expertise s'est centrée sur le développement des politiques d'*open data*, synthétisées dans un ouvrage éponyme (Chignard & Marchandise, 2012) et rappelées dans (Goëta, 2016). Il évoquait ainsi en 2012 la nécessité d'un développement d'une culture de la donnée pour soutenir l'appropriation de l'*open data* :

> « 3 – Développer une culture de la donnée. Le troisième type d'animation concerne la diffusion à un public encore plus large d'une culture de la donnée. C'est l'une des conditions (*nécessaire mais pas suffisante*) de l'appropriation de l'open data par tous. De telles actions peuvent aborder notamment la distinction entre une donnée et une information, mais peuvent aussi encourager à porter un regard critique sur les données, les conditions de leur « collecte » et de leur utilisation. A qui et à quoi sert cette donnée à l'origine ? Pourquoi a-t-elle été ouverte et dans quel but ? Que nous apprend-t-elle ? Qu'est-ce qu'elle ne nous dit pas ? [24] ».

Il participera par la suite au programme infolab de la *FING* dans lequel il contribuera à la conceptualisation de la culture des données que nous décrirons par la suite. Plus récemment en 2023, il précisait

[22] Nous citons ici à titre d'exemple, *Civiteo*(https://civiteo.fr/qui-sommes-nous/), *Data inceptio* (https://data-inceptio.com/en-savoir-plus-sur-le-campus-data-inceptio/), *Data pour tous* (https://datapourtous.fr/),*Datalogy* (https://datalogy-agency.com/), *Dataveyes* (https://dataveyes.com/fr), _ou encore *Role9* (https://www.role9.fr/).

[23] Jacques Priol est par ailleurs le président de *Data Publica* que nous allons présenter par la suite.

[24] https://donneesouvertes.info/2012/04/06/animer-lopen-data/

que la mise en place d'une culture de la donnée « passe aussi par faire prendre conscience que tout le monde peut être producteurs de données. Dans les ateliers de sensibilisation, on commence plutôt par la donnée collaborative, on construit ensemble une base de données à partir d'un sujet qui sensibilise les participants. C'est souvent en produisant soi-même de la donnée qu'on en comprend l'intérêt et les limites. » (Bornstein, 2023).

5.3.2. Les acteurs institutionnels

Coordonnant la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée depuis 2011, *Etalab*, et la *DINUM* qui l'englobe depuis 2019, ont soutenu en 2016 la création de *PIX*. Dans le cadre du dispositif des startups d'État proposé par la *DINUM*, *PIX* est un groupement d'intérêt public proposant une certification des compétences numériques [25]. Basé sur le *Digcomp* européen, son cadre de compétences numériques remplace le brevet informatique et internet (*B2i*) et le certificat informatique et internet (*C2i*) instaurés respectivement en 2000 et 2002. Il est aujourd'hui généralisé dans l'enseignement scolaire où son obtention est désormais obligatoire pour les élèves de troisième, de terminale (*CAP*, professionnelle, générale et technologique), et de *STS2* et *CPGE2* [26]. Il comporte cinq domaines divisés en seize compétences et cinq niveaux de maîtrise. Les cinq domaines sont « Informations et données », « Communication et collaboration », « Création de contenu », « Protection et sécurité », « Environnement numérique ». Les données et les informations sont ici liées dans une même catégorie qui intègre trois grandes compétences : « mener une recherche et une veille d'information », « gérer des données » et « traiter des données ».

[25]https://pix.fr/notre-histoire

[26]https://www.ac-paris.fr/developpement-et-certification-des-competences-numeriques-des-eleves-avec-pix-121798

Tableau 5.1.: Les compétences en données du PIX - tirées de Pix.fr, partie « compétences »

Compétences	**Descriptions**	**Thématiques associées**
Mener une recherche et une veille d'information	Mener une recherche et une veille d'information pour répondre à un besoin d'information et se tenir au courant de l'actualité d'un sujet tout en étant en mesure de vérifier les sources et la fiabilité de l'information (avec un moteur de recherche, au sein d'un réseau social, par abonnement à des flux ou des lettres d'information, ou tout autre moyen).	Web et navigation ; Moteur de recherche et requête ; Veille d'information, flux et curation ; Évaluation de l'information ; Source et citation ; Gouvernance d'internet et ouverture du web ; Abondance de l'information, filtrage et personnalisation ; Recul critique face à l'information et aux médias ; Droit d'auteur.
Gérer des données	Stocker et organiser des données pour les retrouver, les conserver et en faciliter l'accès et la gestion (avec un gestionnaire de fichiers, un espace de stockage en ligne, des tags, des classeurs, des bases de données, un système d'information, etc.).	Dossier et fichier ; Stockage et compression ; Transfert et synchronisation ; Recherche et méta-données ; Indexation sémantique et libellé (tag) ; Structuration des données ; Système d'information ; Localisation des données et droit applicable ; Modèles et stratégies économiques ; Sécurité du système d'information.

Compétences	Descriptions	Thématiques associées
Traiter des données	Appliquer des traitements à des données pour les analyser et les interpréter (avec un tableur, un programme, un logiciel de traitement d'enquête, une requête calcul dans une base de données, etc.).	Données quantitatives, type et format de données ; Calcul, traitement statistique et représentation graphique ; Flux de données ; Collecte et exploitation de données massives ; Pensée algorithmique et informatique ; Vie privée et confidentialité ; Interopérabilité.

Nous pouvons noter ici que les données sont traitées d'un point de vue technique et culturel, prenant en compte non seulement les données mais également les statistiques, les algorithmes, l'économie ou encore l'informatique, rappelant ici l'enchevêtrement de ces éléments.

5.3.3. Les acteurs associatifs

5.3.3.1. Les associations et coopératives

La FING

Conçu en 2000 par Daniel Kaplan et Jacques-François Marchandise comme « un projet de R&D collectif, d'une durée de vie limitée à 5 ans, centré sur les services, les applications et les usages de l'Internet "nouvelle génération" », l'association pour la Fondation d'un Internet Nouvelle Génération (*FING*) a perduré 22 ans, mettant l'accent sur les « enjeux sociétaux et transformatifs du numérique »[27]. Plusieurs actions ont été entreprises vis-à-vis des données comme l'*Open Data Garage* en 2011, l'*Open Data Week* de 2012 à 2014 ou encore le *Self Data Territorial* de 2018 à 2019 [28]. Concernant spécifiquement la culture des données, nous citerons deux initiatives centrales : la campagne Infolab et la *Data Literacy Conference*.

[27]https://fing.org/

[28]https://web.archive.org/web/20211101000000*/https://fing.org/toutes-les-actions.html

Développée en 2013 par Charles Népote, responsable du programme Infolabs, et Amandine Brugière, directrice de programme, la Campagne Infolab « entendait développer une culture de la donnée au service des entreprises et des acteurs du territoire. En lien avec un grand nombre de partenaires dont trois collectivités territoriales pilotes, il s'est agit d'inventer et d'expérimenter, auprès de publics ciblés, des dispositifs de médiation entre les besoins et les projets des entreprises et des acteurs des territoires et l'univers de données [29] ». Permettre le développement d'une culture et un état d'esprit des données était un objectif clé à atteindre et un certain nombre de travaux sur la conceptualisation de la culture des données ont été réalisés en ce sens. Nous pouvons ici citer ceux de Sarah Labelle [30], Armelle Gilliard [31], Simon Chignard et Charles Népote, résumés dans le premier cahier de la FING nommé « culture des données » :

> « La culture des données croise culture médiatique et culture informatique et mathématique. Elle repose comme toute culture sur le partage de connaissances et de pratiques. Il s'agit d'une part de connaître le mode de fonctionnement des médias informatisés qui, entre autres, collectent, communiquent, traitent des données [...] Et d'autre part, cela requiert d'être en mesure de mobiliser quelques bases informatiques et statistiques : par exemple, ouvrir un fichier CSV, appliquer un traitement mathématique à quelques données. Il ne s'agit pas de devenir un expert des données ou un data scientist » (Labelle et al., 2017, p. 4).

Deux objectifs centraux étaient visés dans la conceptualisation et le partage de cette culture :

> « Le développement de la culture des données est essentiel pour deux raisons. Tout d'abord, il est essentiel de construire une relation critique face aux médias et industries qui ont constitué les données comme éléments centraux du traitement algorithmique au cœur de leur modèle économique et éditorial. Ensuite, dans le champ professionnel, la culture des données permet de renouveler les manières de prendre des décisions et débattre, d'interagir entre services et avec l'extérieur de l'organisation » (Labelle et al., 2017, p. 4).

En lien avec le chapitre 1, la *data literacy*, ici en tant que culture des données, doit permettre d'atteindre un « état d'esprit de la donnée » que l'on retrouve sous la plume de Charles Népote :

[29] https://fing.org/toutes-les-actions/infolab.html

[30] Sarah Labelle est professeure des universités en Sciences de l'information et de la communication. Nous précisons ici cela car nous présenterons ses travaux en troisième partie de ce chapitre.

[31] Armelle Gilliard a également réalisé un cahier sur la gouvernance des données et tenait un site qui n'existe plus aujourd'hui, www.lareinemerlin.org, qui centralisait ses apports. A nouveau nous remarquons ici que la gouvernance des données a été traitée en même temps que la culture des données.

> « Ce qui est mal documenté, c'est comment ces entreprises ont acquis ce qu'on devrait appeler un "état d'esprit données". Cet état d'esprit repose sur le fait que les données aident à piloter l'activité, à mesurer, à avoir un retour sur tout ce qui est réalisé dans l'entreprise, permettant de réajuster en permanence la conduite de l'entreprise, dans tous les domaines : la production, le service commercial comme les RH » (Népote, 2016).

En lien direct avec le projet précédent, la *Data Literacy Conference* (2016 - 2018) avait pour objectif de délivrer « les clés d'une "culture de la donnée" enfin accessible aux non-spécialistes ». Elle réunissait des professionnels d'entreprises ainsi que des acteurs universitaires comme Catherine d'Ignazio plusieurs fois citée au cours de cette thèse.



Figure 5.3.: Logo de la data literacy conference - extrait du tweet du 12/09/2017 du compte @DataLitConf

Dès sa première édition tenue en 2016, la volonté de diffusion de la *data literacy* était posée :

> « L'idée montante d'une "littératie" des données consiste à permettre à tout décideur, innovateur, chercheur ou citoyen d'en comprendre les enjeux, d'en discuter les sources et les usages, et d'en tirer parti dans sa propre activité. Comment partager le pouvoir des données ? Comment développer un état d'esprit "données" dans les entreprises, les

> institutions publiques, les associations ? Pour répondre à ces questions, la Fing réunit pour la première fois les meilleurs experts mondiaux de la data literacy, mais aussi les témoignages d'acteurs de terrain, d'entreprises, d'administrations et de communautés agissantes [32] ».

Deux autres éditions suivront, amenant à la publication de près de plusieurs dizaines d'articles. La *Data literacy conference* s'achèvera sur sa troisième édition en 2018.

Open Data France

Open Data France, fondée en 2013 à Toulouse par un certain nombre de communes [33] et de partenaires, est une association poursuivant trois engagements : (1) Soutenir les collectivités dans leur démarche d'ouverture des données, (2) Développer une culture et une (ré)appropriation des données, (3) Renforcer la valorisation des données en participant à l'élaboration de cadres favorables à leurs usages [34].

Open Data France a été à l'origine de nombreuses initiatives vis-à-vis de l'ouverture des données mais également en culture des données. Parmi elles, Culture-D [35] (D désignant les données mais également le digital) a été conçu en 2021, piloté par la cheffe de projet Sarah Medjek. Quatre grands objectifs sont visés : (1) fédérer un écosystème d'acteurs de la donnée, (2) mener des actions d'acculturation à la donnée à destination des acteurs publics ou tout acteur de la médiation numérique, (3) développer l'agentivité des différents publics (esprit critique, capacité d'agir, de décider et de transformer), (4) élaborer des outils et des parcours de formation d'intérêt général [36].

Pour les atteindre, trois démarches sont poursuivies : une veille et une recension des outils et initiatives d'acculturation aux données, une élaboration basée sur l'identification des besoins et coconstructions des parcours d'acculturation aux données selon les publics visés, et enfin la diffusion de livrables et d'outils aux publics cibles [37]. Par souci de transparence, il est important de noter que nous avons fait partie du comité scientifique de ce projet de 2022 à 2023 et que nous avons participé à sa journée de lancement [38].

---

[32] https://web.archive.org/web/20160707145241/https://www.dataliteracyconference.net/

[33] La ville de Balma, la ville de Bordeaux, la Communauté Urbaine de Bordeaux, la commune de Brocas, la ville de Digne les Bains, le Conseil Général de la Gironde, le Conseil Général de Loire Atlantique, le Grand Lyon, la ville de Montpellier, la ville de Nantes, Nantes Métropole, Métropole Nice Cote d'Azur, la Région Provence Alpes Cote d'Azur, la ville de Rennes, Rennes Métropole, le Conseil Général de Saône et Loire, la ville de Toulouse et Toulouse Métropole.

[34] https://opendatafrance.fr/lassociation/

[35] Il s'agit à l'origine d'un programme élaboré en 2021 par la *FING* et *Open Data France*. Ceci s'explique du fait que les acteurs de la *FING* en charge du projet ont intégré *Open Data France* après la dissolution de celle-ci en 2022.

[36] Voir https://opendatafrance.fr/projets/culture-d/ et https://opendatafrance.gitbook.io/cultured

[37] https://opendatafrance.gitbook.io/culture-d-parcours-dacculturation-aux-donnees

[38] https://opendatafrance.fr/culture-d-ay-levenement-en-synthese/

Datactivist

Co-fondée en 2016 par Samuel Goëta et Joël Gombin à Aix-en-Provence (Libre à lire, 2024), *Datactivist* est une « société coopérative et participative qui se donne pour mission d'ouvrir les données et de les rendre utiles et utilisées [39] ». L'objectif poursuivi par le *pure player* est de placer les usagers au cœur des politiques d'*open data*. Elle s'inscrit dans trois valeurs : l'ouverture, l'émancipation, la coopération et l'épanouissement. Nous nous attardons ici plus particulièrement sur le volet « émancipation » qui est accompagné de cette affirmation :

> « Nous sommes conscients que donner accès aux données ne suffit pas. Défenseurs de la data literacy, nous inscrivant dans l'histoire de l'éducation populaire, nous soutenons que donner du sens aux données est une compétence essentielle qui ne doit plus être réservée à une frange de la population. Les données faisant tenir la réalité selon l'expression d'Alain Desrosières, nous voulons permettre à chacun de s'approprier les données sans naïveté, mais sans pessimisme [40] ».

L'intégration et la diffusion de la *data literacy* sont réalisées à travers la production de nombreux contenus et ateliers [41].

Le Cercle de la Donnée

Créé le 20 septembre 2018 à Paris, *Le Cercle de la Donnée* se présente comme « un think tank indépendant réunissant des professionnels de la donnée [...] s'intéressant aux usages de la donnée, au-delà de la seule dimension technologique [42] », afin de « faire émerger et transmettre les bonnes pratiques, approfondir les multiples expertises, briser les classiques cloisonnements professionnels et ainsi contribuer à la structuration d'une filière de la donnée, basée sur l'excellence, l'éthique et l'interdisciplinarité [43] ». Il poursuit en ce sens quatre objectifs : (1) Susciter le partage d'expériences, de connaissances et de points de vue, faire émerger de nouvelles pratiques et expertises, (2) Contribuer à la structuration d'une filière de professionnels de la donnée, (3) Coopérer notamment avec les filières d'enseignement aux fins de sensibilisation et de diffusion des travaux de l'association, (4) Sensibiliser les pouvoirs publics et décideurs sur l'importance de la donnée comme base de développement d'une économie numérique fiable et pérenne. Plusieurs productions ont été réalisées : des études (notamment sur l'intelligence artificielle), des prises de paroles notamment à travers la « Tribune B-Smart » [44].

---

[39] https://datactivist.coop/fr/
[40] https://datactivist.coop/fr/a-propos/
[41] https://datactivist.coop/fr/ressources/
[42] https://www.lecercledeladonnee.org/
[43] https://www.lecercledeladonnee.org/a-propos/
[44] https://www.lecercledeladonnee.org/category/publications/tribune-b-smart/

Data Publica

Le cas particulier de *Data Publica* doit ici être présenté. Il s'agit d'une start-up fondée en 2020 spécialisée dans les données entreprises, l'*open data*, le *big data* et la dataviz. Elle est issue de la collaboration de l'entreprise *Civiteo*, de *Datactivist*, d'*Innopublica* (cabinet de conseil) et de *Parme Avocats* (cabinet d'avocats) (Data Publica, 2019), toutes partageant « une vision et des valeurs communes sur les principes et les méthodes d'exploitation des données par les acteurs publics ou leurs partenaires privé » (Berthier, 2020).

Depuis 2020, *Data Publica* publie chaque année un rapport issu de *l'Observatoire Data Publica*[45], qui observe « les pratiques nouvelles de gestion publique des données », à savoir « l'émergence de « services publics locaux de la donnée », chartes éthiques, formes innovantes de gouvernance et de management de la donnée, prototypes de datascience et usages inédits d'algorithmes, recours à l'intelligence artificielle, etc. [46] ».

5.3.3.2. Les Infolabs

Centralisant les initiatives d'acculturation aux données, la campagne infolab a permis la fédération de structures infolab décrites comme des espaces collaboratifs dédiés « à la compréhension, la manipulation et l'exploration de données [47] ». Elles ne sont pas de nouvelles structures en tant que telles mais des structures existantes adhérant à une même philosophie, recensée dans une « charte infolab ». Celle-ci définit les infolabs selon quatre valeurs et trois missions [48]. Nous les avons listé dans le tableau 5.2.

[45] Il s'agit d'une association à but non lucratif fondée en 2020 par *Data Publica*.
[46] https://observatoire.data-publica.eu/
[47] https://web.archive.org/web/20160828102613/http://infolabs.io/
[48] https://web.archive.org/web/20161023043154/http://infolabs.io/content/charte-des-infolabs

Tableau 5.2.: Valeurs et missions de la charte des infolabs

Valeurs	Missions
1. L'Infolab est un espace d'accueil ouvert. Il recherche la diversité et la complémentarité des publics, des talents, et des envies. 2. L'Infolab est un espace accessible. Il n'est pas nécessaire d'avoir fait de longues études pour comprendre et y partager des choses. 3. L'Infolab privilégie le collectif et la collaboration. Le croisement est à la source des idées les plus neuves et les plus enrichissantes. 4. L'Infolab, si possible, est un lieu d'accueil continu. A tout moment chacun peut participer aux animations en cours et réaliser ses projets.	Découvrir. L'Infolab vous fait découvrir toute la richesse du monde des données numériques : chercher, voir, découvrir, manipuler, collecter, produire des données. Accompagner. L'Infolab vous guide et apporte un cadre à vos apprentissages. Il héberge, facilite voire propose des formations, rencontres, conférences, groupes de travail, etc. Produire. L'Infolab soutient, fait émerger, accompagne des réalisations et des projets fondés ou transformés par les données.

Sont ainsi intégrées des associations, des organismes d'État (comme *Etalab*) ou encore des universités. A l'orée de la première édition de la *Data Literacy Conference* en 2016, Charles Népote déclarait à propos des infolabs :

> « l'Infolab peut être une structure très perturbatrice ou une initiative sans lendemain. Chez Amazon, Facebook ou Google, il n'y en a jamais eu, car la culture de la donnée les imprègne. L'enjeu des Infolabs est de réussir à distribuer un "état d'esprit" comme la qualité ou le service client, c'est-à-dire de disparaître. Les évangélisés sont plus importants que l'évangéliste. » (Usine Digitale, 2016).

Une première recension des adhérents avait été réalisée en 2015 mais n'a plus été mise à jour et ne peut être retrouvée que sur *Internet Archive*. Nous allons ici citer spécifiquement les associations toujours en activité en 2024.

Le Tubà

Porté par l'association *Lyon Urban Data*, *le Tubà* réunit depuis 2011 des acteurs publics et privés « qui ont eu la volonté de créer une dynamique autour de la ville intelligente [49] ». Il se présente comme un « tiers-lieu des données, du numérique responsable et des transitions urbaines du territoire grand lyonnais, facilitant les rencontres et échanges entre entrepreneurs, grands groupes, collectivités, éta-

[49] https://www.tuba-lyon.com/tuba/origines/

blissements de l'enseignement supérieur et de la recherche, étudiants et citoyens intéressés par les enjeux numériques, données et urbains [50] ».

La Turbine

Fondée en à Grenoble, *la Turbine* centre son activité autour « du numérique et des données » par la co-conception de services numériques, la mobilisation de données numériques et la mise en place de tiers-lieu et d'innovation ouverte. Dans le cas des données sont recherchées : la valorisation des données, la conception de nouveaux services, l'accompagnement de stratégies data d'acteurs publics et privés, la sensibilisation aux enjeux des données [51].

Le Donut

Créé en 2016 à Marseille, *le Donut* « entend faire le pont entre les données, les collectivités territoriales, les professionnels de tous horizons et les citoyens usagers d'un territoire. A travers notamment des ateliers de médiation et des prestations évènementielles ou de formation sur mesure [52] ». Parmi ses champs d'action sont cités l'acculturation aux données en faveur des communs, la promotion de l'*open data* et la poursuite d'objectifs de développement durable. En partenariat avec *Datactivist*, ils ont notamment élaboré l'atelier *Data Mythes* dont les objectifs étaient de « casser (ou valider) des mythes sur le vélo à Marseille, de découvrir une méthode générique pour trouver et utiliser des données ouvertes mais aussi de pointer les limites des données disponibles par rapport aux questions exprimées par les participant(e)s » (Le Donut, 2018).

Fréquence Ecoles

En synergie avec les infolabs et *Datactivist* se rajoute *Fréquence Ecoles*. L'association lyonnaise, née en 1991 pour « sensibiliser les jeunes à une société de l'information et de la communication en plein essor [53] », centre ses activités sur la compréhension et le développement des usages du numérique. Sa présidente, Dorie Bruyas, est également depuis 2022 la présidente de la *Mednum* [54] dont *Fréquence Ecoles* est une sociétaire.

[50] https://www.tuba-lyon.com/

[51] https://turbine.coop/

[52] https://www.ledonut-marseille.com/notre-approche

[53] https://www.frequence-ecoles.org/mais-qui-sommes-nous

[54] La *Mednum* est une société coopérative d'intérêt collectif ayant pour objectif depuis 2017 de regrouper les acteurs de la médiation numérique. Parmi ses actions, on recense l'observatoire de l'inclusion numérique, réalisé avec le soutien de la Banque des territoires, qui décrit les structures de l'inclusion numérique selon les acteurs (https://lamednum.coop/actions/observatoire-inclusion-numerique-22/). Sur la centaine de sociétaires qui la compose, sont incluses un certain nombre d'initiatives déjà présentées, à savoir *Pix*, *la FING*, *Datactivist*, *Open Data France* et *La Turbine*.

Centré sur les usages des données, nous citerons ici le programme *datalireladata*, une plateforme d'acculturation aux données et aux algorithmes adressée au jeune public [55]. L'objectif est de lui permettre de (1) comprendre comment les données que nous partageons impactent nos pratiques informationnelles, (2) adopter une attitude critique face aux données que nous consommons et que nous produisons et (3) savoir comment utiliser les données pour devenir acteur·trice du monde numérique [56]. Le tout en rappelant le lien central avec l'Éducation aux Médias et à l'Information (*EMI*) [57]. Pour ce faire, des parcours d'activités selon l'âge des élèves [58] et un ensemble de documentations sont proposés.

En coordination avec *Datactivist* et à destination d'un public adolescent, elle a également produit *l'échappée verte*, un « jeu d'évasion pédagogique et ludique autour de la question des données, pour mieux les comprendre et les analyser. » dont les objectifs pédagogiques incluaient (1) la compréhension des données et du métier de *data scientist*, (2) apprendre comment lire, produire, analyser et interpréter les données, (3) comprendre ce qu'est la visualisation de données et apprendre à « datavisualiser », (4) Comprendre le rôle des data et des données personnelles dans l'économie numérique, (5) Connaître les enjeux de la collecte de données et développer un regard critique sur leur utilisation [59].

Dans leur ensemble, ces associations sont des médiateurs entre les citoyens et les institutions, maillant le territoire par des actions de sensibilisation et de formation aux données et au numérique.

5.3.4. Les acteurs universitaires et associés

5.3.4.1. Les formations universitaires

Une liste exhaustive des formations universitaires en *data literacy* ne peut être ici proposée du fait des changements fréquents de celles-ci, de leur intitulé ou descriptif qui n'indiquent aucune filiation avec le concept, ou encore du périmètre accordé aux données. Par conséquent, nous citerons celles qui ont explicitement revendiqué le lien avec la *data literacy* ou du moins qui soulignent une correspondance des pratiques avec ce concept. Nous avons déjà pu évoquer à la fin du chapitre 3 le master

[55] https://datalireladata.com/

[56] https://datalireladata.com/eduquer-aux-donnees

[57] Si des formations aux médias existaient déjà dès les années 1960, l'*EMI* n'a été officialisée qu'en tant que telle au sein de l'Éducation Nationale qu'avec la parution de l'article 53 de la loi d'orientation et de programmation pour la refondation de l'école de la République de 2013, puis de la circulaire de rentrée 2016 et diffusé au sein du cadre de référence de compétences numériques (*CRCN*) en 2019. Récemment, lors d'une audition au Sénat le 08 novembre 2023, le premier Ministre Gabriel Attal a proposé une dilution de l'*EMI* dans les programmes d'Éducation Morale et Civique (*EMC*).

[58] https://datalireladata.com/contexte-scolaire

[59] https://www.frequence-ecoles.org/lechappee-verte-escape-game-donnees

Médas du *CNAM*. Il a existé par ailleurs la licence *MIND* (Médiations de l'information numérique et des Données) de l'IUT de Bordeaux-Montaigne [60] ainsi que la licence *Data et Information* de l'IUT de Tours [61]. L'ensemble de ces formations centraient leur pédagogie sur la médiation aux données sur la base d'une vision documentaire que nous avions rappelée dans (Verdi & Le Deuff, 2020). Nous les retrouvons aujourd'hui au sein de la formation *B.U.T* Information-Communication, parcours information numérique dans les organisations [62].

5.3.4.2. Les projets de recherche

Plusieurs projets de recherche en culture des données dans l'éducation ont été entrepris à partir de 2016. Intégré dans une logique d'*open data* au sein du projet *Data-cultures* (Lehmans, 2017), le projet *PEPS-IDEX* « Médiation et valorisation des données pour l'éducation » (2016-2017) (Lehmans & Liquète, 2021) avait pour but d'identifier les acteurs, analyser l'accessibilité et les usages des données en action ainsi que médiatiser et mutualiser les ressources pédagogiques à partir d'entretiens auprès de plusieurs profils d'acteurs (collectivités territoriales, associations, enseignants élèves, etc.), par la collecte et l'analyse de scénarii pédagogiques ou encore la conception d'un prototype de plateforme collaborative (Capelle, Lehmans, & Liquete, 2017).

Il a été suivi par le projet *ODEFI* (Ouverture des Données pour l'Éducation, la Formation et l'Innovation) en 2020 inscrit dans le *GTNum DEFI* (Données pour l'Éducation, la Formation et l'Innovation) [63] à partir de 2022 ayant pour intérêt l'analyse du déploiement d'une littératie des données dans l'éducation. L'axe 2 du projet *ODEFI* a ainsi été repris dans le *GTNum DEFI*. Intitulé « Représentations et usages des données : médiation et médiatisation des savoirs avec les données, notamment ouvertes », il visait « à établir un diagnostic sur la façon dont les acteurs de l'éducation, qu'ils soient initiés ou non, envisagent et comprennent la question des données avec les élèves, à travers les perceptions de leurs propres compétences et systèmes de valeurs, leurs expériences, les pratiques scolaires et non scolaires, les représentations et les imaginaires » (Capelle, 2022).

[60] http://web-filieres.iut.u-bordeaux-montaigne.fr/lpmind/

[61] https://lpdatamanagement.univ-tours.fr/

[62] https://www.onisep.fr/ressources/univers-formation/formations/post-bac/but-information-communication-parcours-information-numerique-dans-les-organisations

[63] Les *GtNum*, pour Groupe Thématique Numérique, sont des projets de recherche sur les thématiques de l'éducation et du numérique soutenus par la Direction du Numérique pour l'Éducation sous contrôle du Ministère de l'Éducation Nationale et de la Jeunesse (voir : https://tube.reseau-canope.fr/w/4pKry4xa2D9PCycJhdSxWE).

5.3.4.3. Les organismes locaux

Au sein des universités, l'acculturation aux données est ancienne et a été entreprise selon des modalités spécifiques aux établissements : soit de manière externe avec les *URFIST*[64], soit de manière interne avec des structures spécifiques (souvent liées aux formations aux données de la recherche). Dans ce dernier cas, elles prennent la forme de « guichet unique des données », une structure qui centralise l'ensemble des demandes des acteurs universitaires et qui propose un accompagnement de ces derniers sur ces thématiques.

Nous pouvons ici citer le guichet unique de Nantes qui propose un accompagnement à la gestion des données de la recherche (Vigne, 2023), le service d'accompagnement aux données de la recherche de l'Université de Lorraine *ADOC Lorraine* qui propose des formations continues et des ateliers thématiques [65], ou encore le CEntre de formation et de soutien aux Données de la REcherche (*CEDRE*) de l'Université d'Aix-Marseille qui se base sur une triple ambition : « soutenir ou donner les moyens d'agir en matière de et sur les données de la recherche, offrir une offre de formation sur les bonnes pratiques autour des données, ainsi que renforcer les liens interdisciplinaires sont les principales missions de ce nouveau programme. » [66].

Il est à noter que ces initiatives émergent souvent sur le terreau des ateliers de la donnée. Ces derniers, impulsés par l'État, se positionnent « comme le point d'entrée en proximité des équipes de recherche sur toute nature de besoin relatif à la donnée [67] » et ont pour but d'accompagner les chercheurs dans les usages des données. La première campagne de certification de ces ateliers a débuté en 2021 et se poursuit encore aujourd'hui en 2024 (19 ont été recensés fin 2023 (Recherche Data Gouv, 2024)).

5.3.5. Conclusion de la seconde partie

Il existe donc en France, depuis plusieurs décennies, un grand nombre d'initiatives sur la *data literacy* soutenues par une pluralité d'acteurs pour répondre au besoin de maîtrise des données. Ces initiatives sont souvent intégrées dans des projets originellement portés sur la compréhension des écosystèmes numériques, soulignant à nouveau la passation effectuée entre le numérique et les données dans les discours (bien que dans les faits, les données sont essentiellement perçues comme numé-

[64] Fondées en 1982 et constituées en réseau depuis 2010, les Unités Régionales de Formation à l'Information Scientifique et Technique (*URFIST*) ont pour objectif « la recherche et la formation des usagers universitaires sur les outils, les pratiques et les évolutions de l'information scientifique. » (voir : https://sygefor.reseau-urfist.fr/#/).

[65] https://scienceouverte.univ-lorraine.fr/donnees-de-la-recherche-ul/donnees-de-la-recherche/

[66] https://www.univ-amu.fr/fr/public/cedre

[67] https://www.ouvrirlascience.fr/wp-content/uploads/2021/10/2021.10.11_AMI_Ateliers-de-la-donne%CC%81e.pdf

riques et donc les initiatives sur celles-ci cohabitent, voire sont intégrées dans des initiatives liées au numérique).

Les acteurs ont tissé entre eux des liens de partenariats et l'on retrouve ainsi nombre d'entre eux au sein des organismes et initiatives précités, ce que nous présentons ci-dessous dans la figure 5.4, une cartographie qui liste les interactions entre acteurs entre 2015 et 2024. Il est à noter que sur cette période temporelle, de nombreux changements ont eu lieu comme la dissolution de structures comme la *FING*, l'apparition et la disparition de certains acteurs, etc. C'est donc une vision spécifique, simplifiée, sur les interactions, l'exhaustivité des acteurs constituant déjà le socle de cette seconde partie de chapitre.

La *FING* a été un élément central, fédérateur des réflexions sur la *data literacy* en France. Nombre d'acteurs ont plus ou moins directement fait partie de la structure et participé à ses initiatives : on les retrouve aujourd'hui encore dans d'autres structures liées à cette même thématique. Ce qu'il convient de noter c'est la taille relativement restreinte de cet écosystème mais dont l'influence est perceptible.

Un autre point à souligner ici est la séparation des écosystèmes des acteurs du fait d'interactions non visibles dans les contenus publics consultés. L'absence d'interaction au sein de projets communs explique l'absence de la majorité des acteurs institutionnels et entrepreneuriaux sur cette cartographie. Nous précisons ici qu'une absence de liens dans les contenus publics consultés ne traduit pas nécessairement une absence effective de liens[68]. Cela permet néanmoins de souligner une certaine étanchéité des écosystèmes et ce à l'intérieur même de ces derniers au sens qu'une initiative n'est pas nécessairement connue d'autres acteurs appartenant au même écosystème.

5.4. Les conceptualisations universitaires et pédagogiques de la *data literacy*

A présent que les nécessités évoquées dans les discours et que les acteurs et initiatives autour des données ont été présentées, nous allons aborder les conceptualisations de la *data literacy* selon un prisme universitaire et pédagogique majoritairement situé en Sciences de l'information et de la communication. Avant toute chose, nous précisons ici que nous intégrons les travaux des chercheurs publiés en-dehors des formats traditionnels de la recherche (ex : articles scientifiques) en raison de leur importance définitionnelle.

[68]Nous avions réalisé en 2019 une expérience similaire de représentation cartographique basée sur l'étude de liens hypertextes entre contenus web et étions arrivés à la même interrogation (Verdi & Desfriches Doria, 2019).

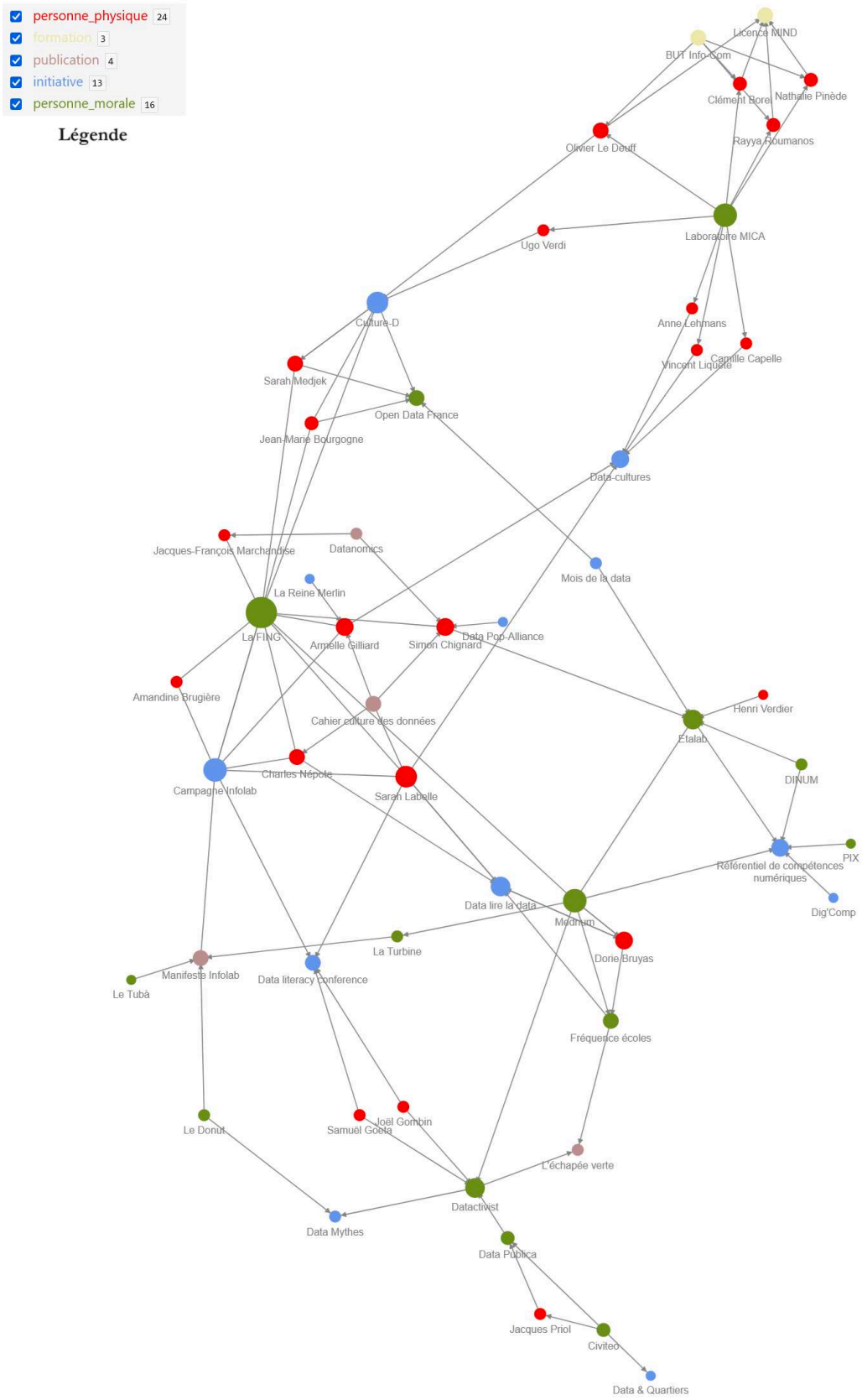

Figure 5.4.: Ecosystème simplifié des acteurs français ayant travaillé sur la *data literacy* sur une période comprise entre 2015-2024 - constituée à partir des sources publiques précitées et cartographiée avec le logiciel Cosma

Plusieurs termes ont été employés pour retranscrire la *data literacy* : culture des données (Labelle et al., 2017; Lehmans, 2017), culture de la donnée (Schöpfel, 2018), data littératie (Arruabarrena et al., 2019), littératie des données (Bodin, 2020), médiation des données (Pinède, 2020) et éducation aux données (Drot-Delange & Tort, 2022). Nous allons voir que l'affiliation à la *data literacy* n'est pas aussi directe que l'on pourrait le penser :

- Sarah Labelle déploie ses réflexions sur la « culture des données » en prenant appui à la fois sur la *data literacy* (et plus précisément (Bhargava et al., 2015)) mais également sur les réflexions (françaises) comme celle de Charles Népote, amenant ici à une généalogie hybride ;

- Anne Lehmans développe elle la « culture des données » en la liant fortement aux caractéristiques de l'*open data*, rappelant en cela notre réflexion amenée dans le chapitre 1 où nous précisions qu'une influence des caractéristiques de l'*open data* se retrouvait dans les définitions de la *data literacy*. Elle cite les travaux de Sarah Labelle et plus globalement ceux de la *FING*, et intègre (Bhargava et al., 2015) comme référence internationale ;

- Joachim Schöpfel lie sa vision de la « culture de la donnée » aux études sur les données de la recherche et le *data management*. La référence internationale citée sur la *data literacy* est ici celle de (Koltay, 2017) qui se centre sur les chercheurs et les *data librarians*.

- La « data-littératie » de Béa Arruabarrena, Gérald Kembellec et Ghislaine Chartron est la plus inspirée du contexte international, citant ainsi (Calzada-Prado & Marzal, 2013), (Hunt, 2005), (Koltay, 2016), (Mandinach & Gummer, 2013), (Matthews, 2016), (Qin & D'Ignazio, 2010) et (Wolff, Moore, Zdrahal, Hlosta, & Kuzilek, 2016). Ils ajoutent également des littératies spécialisées des données comme (Schneider, 2013) et (Carlson et al., 2011).

- La « littératie des données » de Franck Bodin et « l'éducation aux données » de Béatrice Drot-Delange se centrent sur une vision pédagogique. Dans le premier cas, les sources ne sont pas citées mais nous pouvons déceler les caractéristiques classiques de la *data literacy* qui sont associées à celles de l'*open data*, à l'usage de l'esprit critique et de la créativité. Dans le deuxième cas, l'affiliation est hybride, à la fois basée sur le contexte international ((Ridsdale et al., 2015), (Frank & Walker, 2016) et (Crusoe, 2016) pour la *data literacy*) et français ((Bodin, 2018) et (Lehmans, 2017)).

- La médiation aux données de Nathalie Pinède est la vision qui se différencie le plus des autres : la *data literacy* n'est pas citée et propose à la place une vision singulière où la médiation est basée sur un double socle technique et culturel permettant de traiter et dialoguer avec les données.

Cette première analyse souligne le fait que la vision française de la *data literacy* est complètement dépendante des acteurs et ne découle pas nécessairement des visions internationales, proposant à la place soit des visions hybrides, soit totalement indépendantes comme dans le cas de la médiation aux données. L'objectif ici sera de les présenter et de vérifier si des points de convergence existent.

5.4.1. Les définitions

En premier lieu, attardons-nous sur les définitions listées dans le tableau 5.3 :

Tableau 5.3.: Définitions des *data literacies* françaises

Auteurs	**Définitions**
Lehmans (2017)	« La culture de la donnée est composée de savoirs, de valeurs et de compétences statistiques, informatiques, médiatiques, et informationnelles qui donnent les moyens aux professionnels de comprendre le sens de leur activité de collecte et de mise en circulation des données, de la situer dans son contexte social d'usage, et de coordonner les gestes et les procédures professionnels. Cette culture est nécessairement propre à chaque organisation qui construit ses pratiques. »
Labelle et al. (2017)	« La culture des données est composée de savoirs et savoir-faire issus des statistiques et de l'informatique. La culture des données suppose une compréhension des médias et de la circulation de l'information. »
Schöpfel (2018)	« La maîtrise de la façon dont les données sont produites puis exploitées. »
Arruabarrena et al. (2019)	« Cette littératie intervient dans la compréhension des données pour guider le choix des analyses et de leur modalité de restitution. Cela signifie qu'il faut non seulement comprendre et analyser les données dans leur contexte informationnel de production et d'usages, mais aussi être capable d'en restituer les résultats et de communiquer sur plusieurs supports (présentations, rapport, visualisation. »

Auteurs	Définitions
Bodin (2020)	« Une certaine aptitude à comprendre et à utiliser les données au quotidien dans la vie courante, à la maison, à l'École, au travail et plus largement dans la vie citoyenne, en vue d'atteindre des buts personnels ou collectifs, d'étendre ses connaissances et ses capacités. En d'autres termes et à l'instar des littératies numériques, il s'agit de nos aptitudes à agir, avec les données . »

Une première lecture de ces définitions souligne d'une part le fait que la *data literacy* est un ensemble de connaissances, de compétences et d'attitudes, d'autre part qu'elle repose sur un socle à la fois technique et culturel permettant d'appréhender une multitude d'environnements enchevêtrés et d'agir au quotidien sur ces derniers. Culture informationnelle, culture informatique, culture numérique, culture statistique et culture médiatique : les connaissances développées au sein de ces domaines connexes sont pour ces auteurs nécessaires à l'acquisition d'une bonne maîtrise des données.

5.4.2. Les connaissances

Nous l'avons précisé, la *data literacy* intègre un certain nombre de connaissances issues de domaines d'études différents, de cultures différentes. Au sein de la *Data Littératie* développée par Arruabarrena et al. (2019) est intégrée une dimension critique basée sur l'usage des statistiques (faisant de la *statistical literacy* une sous-catégorie de la *data literacy*) et sur la *datavisualisation*. Un point intéressant à noter est le distinguo opéré entre « data littératie » et « culture des données » : cette dernière, en tant que posture et état d'esprit, renvoie plutôt à la notion de *data culture* que nous développerons par la suite [69].

Pour l'ensemble des auteurs cités, en support de cette posture et de cet état d'esprit critiques, comprendre ce que sont les données et les enjeux sociétaux qui leur sont attachés est une nécessité citoyenne. Ce qui est rappelé par Franck Bodin et Olivier Banus, respectivement directeur de l'*Atelier Canopé* 93 et directeur de l'*Atelier Canopé* 84 [70] : « Il y a un enjeu citoyen à maîtriser ces langages car ils sont omniprésents dans la société. Ils nous aident à penser le monde, à l'appréhender, à le construire, de fait à exercer notre citoyenneté. Développer des compétences de lecture et d'écriture dans ce domaine, c'est aussi exercer son esprit critique en interrogeant la fabrique des données, les conditions de leur écriture et les utilisations qui en sont faites. [... cette approche convoque...] une

[69] Le terme n'est pas cité par les auteurs, c'est une filiation *a posteriori* de notre part.

[70] Les Ateliers Canopé sont des organismes de formation des personnels enseignants, en particulier dans l'appropriation des outils et environnements numériques. https://www.reseau-canope.fr/qui-sommes-nous.html

pédagogie du faire, permettant d'apprendre en vivant une expérience créative et collaborative, ancrée, située dans le réel » (Bodin, 2020).

La question de la nature des données elles-mêmes est un point souvent abordé. Ces « obtenues » (Arruabarrena et al., 2019), vues comme essentiellement numériques, servent de « matière première de langages écrits et visuels » (Bodin, 2020) et contribuent à l'économie de la connaissance par une valorisation de l'information (Lehmans, 2017). En ce sens, *data literacy*, *media literacy*, *information literacy*, *computer literacy* et *digital literacy* forment un groupe de littéracies en interaction constante, renvoyant à l'aspect métalittéracique abordé en chapitre 3. La compréhension critique du *big data*, au sens notamment d'une remise en question des analyses de tendances issues de l'analyse des méga-données, mais également de l'*open data* et des *linked open data* est centrale aux réflexions data littéraciques. C'est une prise en compte et un décryptage d'un écosystème informationnel grandement numérisé qui a des conséquences notables sur le quotidien des individus.

5.4.3. Les compétences

Pour asseoir les connaissances précitées doivent être développées en parallèle des compétences couvrant l'ensemble du cycle de vie des données (voir le tableau 5.4).

Tableau 5.4.: Compétences associées aux *data literacies* françaises

Auteur	**Compétences citées**
Labelle et al. (2017)	« Il s'agit d'une part de connaître le mode de fonctionnement des médias informatisés qui, entre autres, collectent, communiquent, traitent des données. Regarder et configurer les paramètres de nos applications, comprendre la manière dont est produite une datavisualisation (représentation graphique de données). Et d'autre part, cela requiert d'être en mesure de mobiliser quelques bases informatiques et statistiques : par exemple, ouvrir un fichier CSV, appliquer un traitement mathématique à quelques données. »
Lehmans (2017)	« Il s'agit donc de comprendre à quoi sert une donnée et comment chacun peut agir sur ou faire avec les données, sans se limiter aux manipulations techniques des données. »

Auteur	**Compétences citées**
Arruabarrena et al. (2019)	(1) Collecter et préparer des données en vue d'un traitement informatisé, (2) Analyse des données pour produire des informations, des indicateurs utiles aux organisations ou à des individus, (3) Pilotage de la qualité des données avant et après le traitement des données au regard des valeurs sociétales, (4) Gestion d'un projet « data » dans une organisation, (5) Conception des services de données pour transformer des activités sociétales.
Bodin (2020)	(1) se constituer et mettre à jour une culture personnelle des données, (2) lire et écrire avec les données, (3) produire volontairement et décrire des données, (4) collaborer avec autrui dans le paysage des données, (5) être créatif en utilisant des données.
Pinède (2020)	« Il s'agit tout d'abord de permettre un "habillage" des données en travaillant par exemple la dimension visualisation afin de permettre leur valorisation et de faciliter leur compréhension dans une perspective orientée "design d'information". Mais cela peut être également, dans une approche plus pédagogique, l'accompagnement des publics pour une acculturation à des jeux de données, tels que proposés par les collectivités locales. Les médiations documentaires s'articulent avec les médiations de données, elles représentent ce cadre d'inscription nécessaire et fondamental à la production d'un environnement structuré, renseigné et adapté pour les documents, qu'ils soient numériques ou pas. Les médiations audiovisuelles constituent une déclinaison particulière autour de l'image et des vidéos, à la fois dans une logique de contextualisation et de construction du sens. Enfin les médiations collaboratives permettent de déployer des stratégies de médiations numériques à travers différents dispositifs du web, dans un croisement pluri-médias (Twitter, Facebook, etc.) favorisant des interactions diversifiées et réactives entre différents acteurs. »

Auteur	Compétences citées
Drot-Delange & Tort (2022)	« La modélisation des connaissances et compétences a montré la diversité des domaines qu'il conviendrait de couvrir : informatique, mais aussi informationnel et médiatique [...] L'usage des données nécessite également d'autres compétences, notamment informationnelles : savoir trouver les données pour résoudre un problème, savoir qu'elles existent, s'interroger sur les sources de ces données, sur leur sélection, sur les données absentes ou manquantes, etc. Ces compétences concernent la phase amont du travail sur les données, celle de leur acquisition, en portant un regard critique sur la proposition faite par l'institution qui les met à disposition et les outils qu'elle préconise. Il y a aussi des compétences nécessaires pour analyser et interpréter des données relevant d'une activité ou d'un champ spécifique. »

Ces compétences reposent sur un socle technique et culturel (au sens d'une culture générale doublée d'une posture critique offrant un recul sur les pratiques engagées) qui embrasse l'ensemble du cycle de vie des données (production, acquisition, traitement, analyse, représentation, partage, archivage, destruction). Parmi les compétences communes revient à plusieurs reprises la datavisualisation en tant qu'« habillage » (Pinède, 2020) qui permet de contrôler le « narratif » des données. De plus, ces compétences impliquent un engagement volontaire, une attitude portée vers le traitement des données : cela implique donc, au-delà ce qui a déjà été évoqué, des actes de médiation entre acteurs pour favoriser un ensemble d'attitudes au sein de cette culture commune.

5.4.4. L'aspect culturel

Dans une chaîne de rétroaction, les interactions des attitudes et des connaissances des acteurs font émerger des cultures singulières. Selon Lehmans (2017), une culture des données est un « assemblage de répertoires collectifs de représentation, d'action et de justification situé dans un contexte spécifique et caractérisé par une sensibilité, un univers de sens et une rationalité construits en relation avec des ensembles de données. Ces répertoires contribuent à façonner les données, leur statut et leurs effets sur le monde. Ils reposent sur des valeurs, des normes (explicites ou implicites), des littératies, des pratiques, des affects et des dispositifs techniques. Ces éléments forment un système possédant une cohérence interne mais non uniforme, traversé par des lignes d'ambivalence, de ten-

sion et de contradiction. Ils s'enchevêtrent avec d'autres cultures de données, toutes animées par des dynamiques à la fois locales et globales ».

Il s'agit donc d'une vision plurielle et dynamique où « des » cultures situées sont produites et peuvent se confronter, renvoyant ici à la vision d' Olson (1998) indiquée en chapitre 2 pour qui il existe non pas une mais plusieurs cultures de l'écrit « aux conséquences et implications différentes les uns des autres ». Ceci est rappelé par Labelle et al. (2017) qui précisent qu'il « n'est pas possible de parler de LA culture des données. Au contraire, en fonction du groupe social, de l'organisation, des spécificités pourront être identifiées par rapport au rôle attribué aux données et aux conditions de leur usage. Dans une approche large, les cultures des données constituent un environnement. Ces cultures nous entourent en permanence. Et nos pratiques nous permettent de nous inscrire à l'intérieur de ces cultures » (p. 6). Par ailleurs, Casemajor Loustau (2023) notait que les cultures des données n'était pas « des corpus de répertoires stables, mais plutôt des processus actifs de cultivation [71], travaillés par une multiplicité d'acteurs hétérogènes, incluant les publics ».

Elle ajoutera par ailleurs que :

> « La notion de culture de données réfère à un assemblage de répertoires collectifs de représentation, d'action et de justification situé dans un contexte spécifique et caractérisé par une sensibilité, un univers de sens et une rationalité construits en relation avec des ensembles de données. Ces répertoires contribuent à façonner les données, leur statut et leurs effets sur le monde. Ils reposent sur des valeurs, des normes (explicites ou implicites), des littératies, des pratiques, des affects et des dispositifs techniques. » (Casemajor Loustau, 2024)

Sens, rationalité, représentation : une multitude de cultures des données, contextuellement situées à la fois au niveau des individus mais également au niveau des organisations elles-mêmes, évoluent et s'enchevêtrent dans le temps, produisant ainsi nombre de représentations de la réalité (*weltanschauung*) parfois contraires, amenant à des frictions [72] symboliques et pratiques (ce qui est parfois

[71] Le terme de « cultivation des données » désigne dans ce cadre le fait que « diverses formes de données sont créées, nettoyées, ordonnées, récoltées et entrecroisées – par des acteurs multiples et distribués mais connectés, notamment des entreprises, des gouvernements, des développeurs, des annonceurs et des utilisateurs » (Casemajor Loustau, 2023).

[72] Le terme de « friction des données » (*data friction*) a été inventé par Edwards (2010) et a été repris par Bates (2018). Il désigne les problématiques liées aux facteurs socio-matériels qui, combinées, tendent à réduire et restreindre la production, le déplacement et l'utilisation des données. Cela est particulièrement le cas lorsqu'est imposée l'ouverture des données qui fait émerger des enjeux de pouvoir incarnés par une non-volonté des acteurs de partager leurs données. Ces enjeux de pouvoir liés aux données tendent d'une part à remettre en cause la supposée fluidité des données, rappelant qu'au contraire le partage des données ne va pas de soi, et d'autre part à souligner le caractère central des interactions entre acteurs au sein des organisations.

regroupés sous le terme de silotage pour désigner un écosystème de connaissances et de pratiques auto-référencées et imperméables aux apports extérieurs).

En ce sens, les cultures des données relèvent de processus de négociation (Lehmans, 2017) et l'on parle de maturité des organisations lorsqu'une vision commune, un sens partagé, désiloté, émerge. Cette connaissance organisationnelle doit être partagée par tous et rappelle le caractère universel de la *data literacy* qui ne doit pas être limitée à un cercle de spécialistes. En effet, la *data literacy*, au contraire de la *data science*, ne nécessite pas d'être développée à un niveau expert. Ce point de vue partagé par les auteurs rejoint ici les discours internationaux présentés dans les chapitres précédents (Labelle et al., 2017; Lehmans, 2017).

5.4.5. Mettre en place la *data literacy*

Se pose désormais la question de l'application pratique de la *data literacy*. Plusieurs pistes ont été abordées. Tout d'abord en termes d'objectifs où l'accessibilité et la visibilité des données ainsi que des pratiques associées est un enjeu central :

> « La première est que les données soient visibles et lisibles grâce à des bases de données permettant de les rechercher aisément ; la seconde est qu'elles soient suffisamment accessibles, pour être utilisées par des acteurs qui n'ont pas participé à leur processus de fabrication. La troisième dimension, encore peu ou pas envisagée, est le développement d'une littératie des données incluant la dimension visuelle, permettant d'acquérir une culture analytique et critique des éléments de traitement et de visualisation des données. Cette acculturation aux données, dans le cadre de la formation professionnelle notamment, comprend la sensibilisation, l'exploration de ressources, la connaissance de la législation et des contraintes des collectivités, la mise en œuvre de cartographies ouvertes et d'animations territoriales, notamment. » (Lehmans & Liquète, 2021).

Nous l'avons vu, la *data literacy* n'étant pas réservée aux spécialistes, elle doit pouvoir être diffusée à l'ensemble des acteurs. Pour ce faire, un certain nombre d'initiatives sont déployées : des explorations de ressources de données lors d'hackhatons par exemple, des sensibilisations et de vulgarisations sur l'importance des données au sein d'une organisation, des formations professionnelles spécifiques ou encore des mises à disposition de ressources.

Dans les travaux de Labelle et al. (2017) exposés dans la figure 5.5, trois phases doivent être déployées pour permettre le développement et la diffusion d'une culture de la donnée au sein d'une organisation, adaptées aux profils visés. La première phase, « l'amorçage », consiste à sensibiliser les acteurs aux risques et opportunités amenés par les données. La seconde phase, « la mobilisation », prend

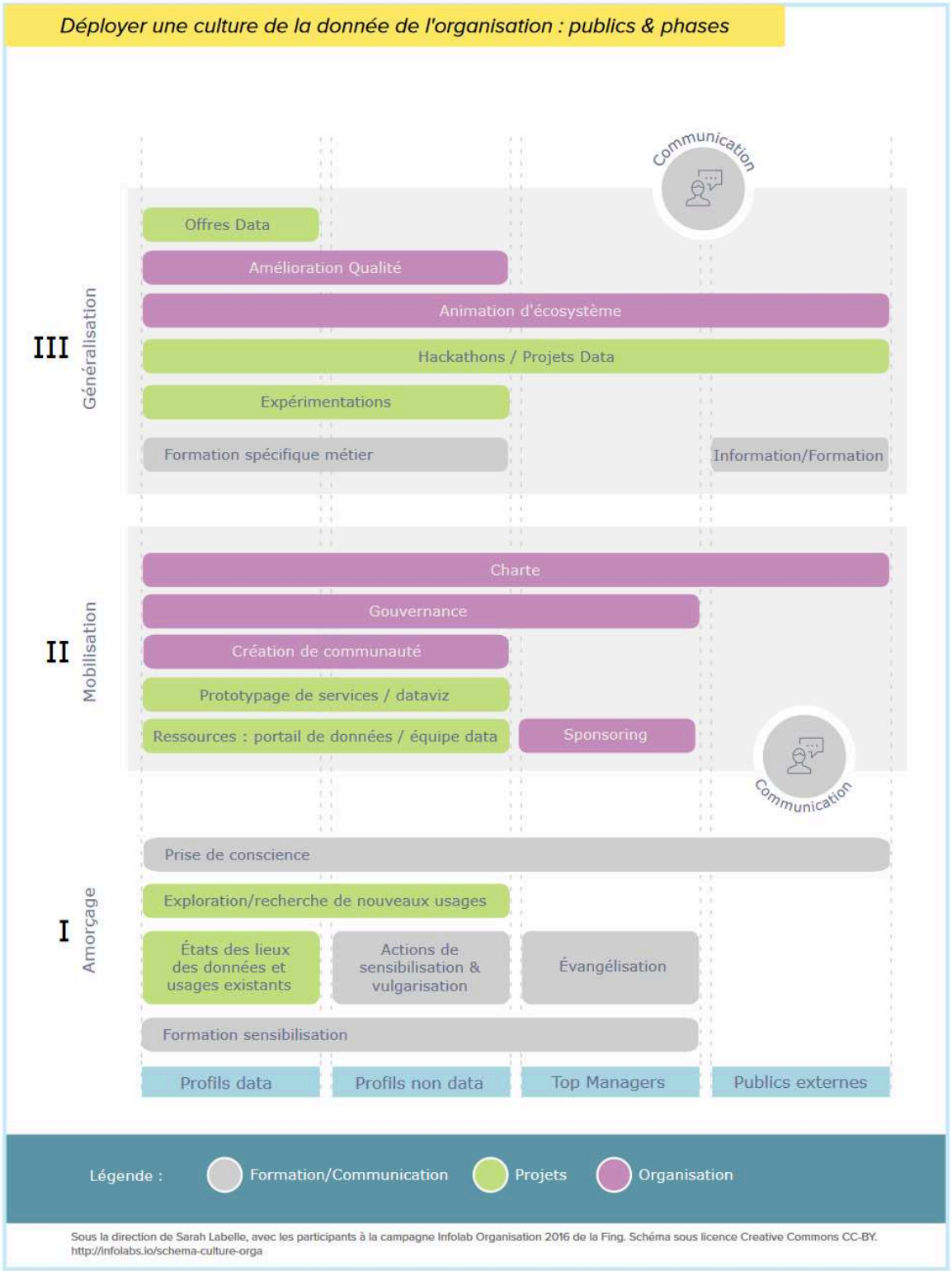

Figure 5.5.: Déployer une culture de la donnée de l'organisation : publics & phases - extrait de (Labelle et al., 2017, p. 8)

pour socle les connaissances consolidées des acteurs ainsi que leur volonté d'investissement pour développer des actions concrètes. La troisième phase, « la généralisation », vise à créer, à tous les niveaux de l'organisation, et aussi en direction des publics externes, « le réflexe de la donnée-ressource » où les données sont comprises comme des ressources importantes et employées « tant à des fins de performance opérationnelle qu'à des fins d'innovation » (pp. 8-9).

Nous retrouvons ici la distinction entre la sensibilisation et la formation, déjà visible dans les discours entrepreneuriaux, qui souligne la progressivité des démarches d'acculturation, un terme très souvent employé pour englober les initiatives liées à la culture des données. Courbot (2000) précisait qu'il demeurait « lié à toute une thématique sujette à controverse. Son utilisation amène, le plus souvent, à aborder des notions aussi polémiques que celles de la race, de l'ethnie, du rapport entre société dominante/société dominée, de la colonisation » et que son socle, la culture, posait problème du fait qu'il n'en existait (et n'en existe) pas de définition univoque ; le terme acculturation pouvant être donc compris de façon multiple.

Dans le cas de la *data literacy*, Casemajor Loustau (2024) indiquait qu'il induisait une conception « linéaire, mécanisme et unilatérale du changement culturel » où l'assimilation de la « vision adéquate » véhiculée par l'organisation suffisait à effacer les freins au développement. Cette approche qui ne permettrait pas par ailleurs « de comprendre les tensions qui se nouent dans les débats portant sur les principes supérieurs qui devraient guider le travail avec les données », pouvant même « contribuer à reproduire des répertoires en place qu'à engendrer des décalages transformatifs ».

Ainsi, l'acculturation des publics doit être interrogée, contextualisée selon les types d'organisation, de publics et des liens qui les unissent, d'objectifs visés, de temporalité et de budget dédiés. C'est une réflexion pédagogique globale qui est ici nécessaire pour s'assurer de la pertinence et de l'efficacité d'une culture des données.

5.4.6. La problématique des traductions

La traduction du terme *literacy* est une difficulté constante[73] qui a amené à de nombreuses réponses. Nous avions pu recenser « alphabétisation » (Goody, 1979), littérisme[74] et lettrisme[75], ou lettrure (Souchier, 2012). Le terme d'alphabétisation a souvent été employé comme en témoigne l'ouvrage central de Goody (1979) ou encore l'*UNESCO* qui emploie ce terme [76] depuis ses premières campagnes menées dès les années 1950 (Chartier & Rockwell, 2013). Selon plusieurs dictionnaires de

[73]Jaffré (2004) en parle comme d'un concept à « géométrie variable ».
[74]https://www.education.gouv.fr/bo/2005/37/CTNX0508668X.htm
[75]https://www.education.gouv.fr/bo/17/Hebdo6/CTNR1636064K.htm
[76]https://www.unesco.org/fr/literacy

langues françaises, l'alphabétisation peut être comprise comme l'enseignement de la lecture et de l'écriture à un groupe social déterminé, en l'occurrence les personnes analphabètes[77] [78]. Or, pour plusieurs auteurs, le concept d'alphabétisation est une forme réductrice de la littératie.

Dans la note de l'éditeur de la traduction française de l'ouvrage d'Olson, il est indiqué que jusqu'aux années 2000, l'adjectif *literate* était fréquemment traduit en français par « alphabétisé » avant d'être remplacé par « lettré », du fait « que la littératie, la culture de l'écrit, est un concept qui s'applique à toutes les sociétés disposant d'une écriture, que celle-ci soit alphabétique ou non ». (Olson, 2017). Pour Kara & Privat (2006), « littératie » permet de combler un vide sémantique car « lettrisme », « illétrisme », « lettré », « alphabétisé », « analphabétisé » ne leur donnent pas satisfaction. En effet, le terme de littératie, toujours selon ces auteurs, recouvre plusieurs avantages : il permet de bien intégrer la question de l'alphabétisation ainsi que l'entrée dans la culture de l'écrit, ce qui permet d'éviter l'écueil de comparer deux termes présentés comme antonymes, à savoir ceux de lettré et d'illettré.

En France, le terme de culture est fréquemment employé pour traduire *literacy*. Or, cette traduction n'est pas anodine et soulève de nombreuses difficultés comme le rappelle Lehmans (2018) qui considère ce terme comme « l'un des plus complexes et des plus discutés dans les sciences humaines et sociales » (p. 9), comme cela était notamment rappelé au sein de l'étude de Kroeber et Kluckhohn (1952) qui listait 161 sens différents du terme de culture. Ce choix est néanmoins revendiqué, d'une part du fait que le terme de littératie est beaucoup moins parlant que celui de culture pour un public français, d'autre part pour son caractère englobant : « On choisit ici de parler de culture et pas de compétences ou de connaissances, parce que c'est véritablement un ensemble de représentations, un univers de sens, une capacité à se projeter, un sentiment d'appartenance qui sont en jeu, et pas seulement la capacité à manipuler des outils informatiques » (Lehmans, 2017). Pour mieux appréhender cette difficulté, nous allons rappeler les caractéristiques générales de « la » culture et les confronter à celles de « la » littératie.

[77]https://www.larousse.fr/dictionnaires/francais/alphab%C3%A9tisation/2513
[78]https://dictionnaire.lerobert.com/definition/alphabetisation

5.4.6.1. Quelques définitions

Une définition[79] souvent citée est celle de l'anthropologue britannique Edward Burnett Tylor (1871) : « ce tout complexe qui englobe les connaissances, les croyances, l'art, la morale, la loi, la tradition et toutes autres dispositions et habitudes acquises par l'homme en tant que membre d'une société »[80]. En France, la culture a été définie dans le dictionnaire de l'Académie comme « le soin qu'on prend des arts et de l'esprit » (1762) puis comme « se dit figurément de l'application qu'on met à perfectionner les sciences, les arts, à développer les facultés de l'esprit » (Lefebvre, Tonnelat, Mauss, Niceforo, & Weber, 1930). Dans son introduction à la sociologie générale, Rocher la désignait comme un ensemble de « manières de penser, de sentir et d'agir plus ou moins formalisées qui, étant apprises et partagées par une pluralité de personnes, servent, d'une manière à la fois objective et symbolique, à constituer ces personnes en une collectivité particulière et distincte » (Rocher, 2012).

En 2004, l'*UNESCO* précisait que dans « son sens le plus large, la culture peut aujourd'hui être considérée comme l'ensemble des traits distinctifs, spirituels et matériels, intellectuels et affectifs, qui caractérisent une société ou un groupe social. Elle englobe, outre les arts et les lettres, les modes de vie, les droits fondamentaux de l'être humain, les systèmes de valeurs, les traditions et les croyances. » (UNESCO, 2004). La culture possède donc un caractère englobant dont la formalisation des attitudes détermine les manières acceptables de se comporter au sein de la société (Goodenough, 1981). La multiplicité des traits distinctifs a permis un apport varié de caractéristiques possibles adossées à la culture comme l'Etat (*Staat*) chez Novalis, l'art pour Goethe ou encore la tolérance pour Schiller (Lefebvre et al., 1930), rappelant que les cultures sont multiples, contextuellement situées et évolutives.

[79] Plusieurs synthèses ont été proposées, au-delà d'une recension de définitions. Forquin (1996) divisait la culture selon cinq sens : (1) philosophique (la culture s'oppose à la nature ; elle se découvre et se transmet, elle n'est pas un donné), (2) traditionnel et individuel (éventail de connaissances et de compétences), (3) objectif et descriptif (l'ensemble des traits caractéristiques d'une société ou d'un groupe), (4) patrimonial, différentialiste et identitaire (connaissances et compétences propres à une communauté humaine qui engendre une distinction des cultures), (5) universaliste et unitaire (dépassement des particularismes). Kambutcher, cité Amat & Maigné (2022), décelait pour sa part un sens classique (idéal de formation des capacités humaines, accompli au moyen d'un détour par des œuvres et des institutions), un sens sociologique ou ethnologique (ensemble de pratiques et de représentations propres à une population) et un sens ontologique ou transcendantal (domaine d'objets, de processus ou de légalités à distinguer (sur un mode problématique) de la nature) (p. 7). Pour Geuss (1996), la culture peut également être perçue soit comme un phénomène collectif (dans la lignée d'Hegel et Weber), soit comme quelque chose d'individuel et de progressif (dans la lignée d'Humboldt qui emploie pour ce faire *Bildung*). En Sciences de l'Information et de la Communication, Liquete (2014) listait pour sa part deux acceptations de la culture : « celle qui met l'accent sur des savoirs de référence, des contenus d'autorité, renvoyant à l'idée d'une culture générale, classique, discriminant des auteurs, des littératures et des gisements et accentuant de facto des éléments de connaissance », et « celle qui renvoie à des manières de vivre, d'appréhender le monde, de considérer l'autre » (p. 10).

[80] « *…is that complex whole which includes knowledge, belief, art, law, morals, custom, and any other capabilities and habits acquired by man as a member of society* ».

5.4.6.2. Culture et nature

Pour Horkheimer (2021), la notion de culture est apparentée au concept du « formé » (*das Geformte*). Elle se dissocie selon deux concepts allemands : la *Bildung* et la *Kultur*. *Bildung* serait la mise en forme d'une nature primitive informe. *Kultur* proviendrait de *colere*, prend soin, en lien avec l'agriculture, et plus généralement aux produits des hommes « qui sont là "comme les fleurs des champs", aux œuvres d'art, aux livres, aux systèmes religieux ou philosophiques révélateurs des particularités d'un peuple » (Elias & Kamnitzer, 2003). En se basant sur les écrits de Goethe, Tonnelat (1930) résumera la notion en tant que « façons de vivre et de penser collectives qu'un peuple s'est formées durant un long contact avec la nature, au cours des luttes qu'il a dû mener contre elle. Ce sont aussi les transformations qu'il a fait subir à la nature qui l'entoure. Ces façons collectives de penser et de sentir, voire d'agir, ou encore de s'organiser dans un pays déterminé, sous un certain ciel, sont d'autre part le fondement nécessaire de toute œuvre d'art. » (p. 70).

Si la *Kultur* est traditionnellement opposée à la *Natur*, comme nous pouvons le noter chez Freud (2010) qui précisait : « le mot "civilisation"[81] désigne la somme des réalisations et des institutions par lesquelles notre vie s'écarte de celle de nos ancêtres animaux, et qui servent deux buts : protéger l'homme contre la nature et réguler ses rapports avec ses semblables » (p. 83), cette dichotomie a fait l'objet d'une remise en cause. A titre d'exemple, dans *Par-delà nature et culture*, Descola (2015) appelait à dépasser ce clivage et arguait que la nature « ne dévoile pas son essence grâce aux efforts cumulés d'une cohorte de grands esprits et d'artisans ingénieux, mais qu'elle est peu à peu construite comme un dispositif ontologique d'un genre particulier servant d'assise à la cosmogenèse des Modernes » (p. 130).

Pour Berque (1999), nature et culture ne sont pas antinomiques[82] car la nature est toute sauf naturelle : « la nature est forcément traduite en termes propres à une culture ; elle est intégrée au monde que l'homme est capable de concevoir, de percevoir et d'aménager. Dans ce cas, la nature ne se définit pas sans ou contre l'homme mais par l'homme » (p. 52). Selon cette vision, la nature est le fait d'une médiance[83], à savoir « le sens ou l'idiosyncrasie d'un certain milieu, c'est-à-dire la relation d'une so-

[81] Le terme original est *Kultur*.

[82] « Le dualisme moderne, et ses dérivés métabasistes, n'a pu en l'affaire que réduire l'humain aux mécanismes de la nature (c'est ce qu'on appelle le *scientisme*), ou bien faire de la culture – notamment dans les vues régnantes à propos du sémiotique – un aérostat détaché de tout enracinement terrestre ; ce qui n'est qu'une énième expression de la dichotomie cartésienne *res extensa/res cogitans*, et revient à opter soit pour la Terre, soit pour le Monde. Comme si l'horizon ne liait pas l'un à l'autre ! » (Berque, 2002).

[83] Selon les propres mots de son auteur : « Le néologisme *médiance* rend ici le japonais *fûdosei* □□□, concept créé par Watsuji à partir du terme *fûdo* □□, lequel s'écrit « vent (□) – terre (□) », et qu'après moult réflexions je m'étais résolu à traduire par « milieu » dans le sens traditionnel que la géographie humaine a donné à ce terme [...] en japonais toutefois, le sens ordinaire de *fûdo* est celui de « milieu local ». La plupart des lecteurs japonais de *Fûdo* entendent donc le concept de *fûdosei* au sens de « milieu-localité », i.e. « singularité (locale, régionale, nationale) ». En ce sens, *fûdosei* pourrait être rendu par *contréité* (ce qui donnerait en allemand *Gegendheit*, en pensant à la *Gegend* heideggérienne) ; mais je préfère

ciété à son environnement » (Berque, 2009), du fait que « que l'être humain se compose pour moitié d'un corps individuel et pour moitié d'un milieu éco-techno-symbolique ; l'ensemble des milieux formant l'écoumène : la relation humaine à l'étendue terrestre » (Berque, 2002).

5.4.6.3. Culture et civilisation

L'interaction culture / civilisation est également importante. La civilisation a une acceptation large selon Elias & Kamnitzer (2003), pouvant se référer à « un degré de l'évolution technique, aux règles de savoir-vivre, au développement de la connaissance scientifique, aux idées et usages religieux » (p.11). Pour Lefebvre et al. (1930), la civilisation possède deux versants. L'un désignant l'« ensemble des caractères que présente aux regards d'un observateur la vie collective d'un groupement humain : vie matérielle ; vie intellectuelle, vie morale, vie politique et - par quoi remplacer cette expression vicieuse ? - vie sociale » (p. 12) et étant d'ordre collectif. L'autre renvoyant à la notion de progrès, « des défaillances, des grandeurs ou des faiblesses de la civilisation » (p.12). Or, la notion de civilisation est très différente selon la généalogie considérée : pour les Anglais et les Français, elle est assez similaire et regrouperait pour Elias & Kamnitzer (2003) ce qui fait la fierté d'une nation, les progrès de l'Occident et de l'humanité en général. Elle désignerait à la fois les réalisations et les comportements de fierté liés à celles-ci et engloberait tout un ensemble de faits (économiques, moraux, politiques, religieux, sociaux et techniques).

Pour les Allemands, c'est la *Kultur* qui renvoie à la fierté liée à ses propres créations et à sa propre nature. Au contraire de la civilisation à l'anglaise et à la française, elle délimite le périmètre de ses caractéristiques, limitées aux données intellectuelles, artistiques et religieuses (p. 13) comme nous pouvons à nouveau le voir chez Freud (2010) intégrait pour sa part « La beauté, l'ordre et la propreté occupent à l'évidence une place particulière parmi les exigences de la civilisation » (p. 89) et ajoutait « le trait qui nous semble le mieux caractériser la civilisation réside dans le respect et la pratique des activités psychiques supérieures, des réalisations intellectuelles, scientifiques et artistiques, et dans le rôle majeur accordé aux idées dans la vie des hommes » (p. 90).

Le terme *Zivilisation* malgré la ressemblance avec celui de civilisation contient une vision péjorative vis-à-vis des cultures française et anglaise, perçues comme superficielles et corrompues (Mann, Lilla, & Morris, 2021). Dumont (1987) résumait ainsi que « la "civilisation", chose des Occidentaux, s'oppose à la Kultur, qui caractérise l'Allemagne. C'est un thème rebattu, devenu général à l'époque, comme le slogan majeur par lequel l'Allemagne répond à la propagande ennemie ». *Kultur* et *Zivilisation* s'opposent donc : « le terme de *Kultur*, dans l'Allemagne du 18e siècle, sert à caractériser la distinction et la finesse des manières ou des façons de vivre et de penser collectives qu'un peuple

m'en tenir à *médiance*, qui rend fidèlement la définition que Watsuji donne lui-même de *fûdosei*, et qui – comme je le montrerai plus loin - est d'une bien plus grande portée ontologique » (Berque, 2014)

s'est formé en rapport avec son milieu, alors que celui de *Zivilisation* renvoie à un sens moral se rapportant à la maîtrise que l'homme acquiert sur lui-même et à l'organisation correspondante de ses rapports avec ses semblables » (Bert, 2010). Meyran (2009) précisera que cette opposition entre les termes *culture / Kultur / Zivilisation / civilisation* est d'ordre politique, résultant d'une lutte pour la domination culturelle entre la France et l'Allemagne au XIXe siècle et ce jusqu'au début du XXe siècle. Rocher (1970) résumait ainsi que la civilisation pouvait être aujourd'hui employée (avec réticence néanmoins de la part des sociologues et des anthropologues) pour d'une part désigner un ensemble de cultures particulières ayant entre elles des affinités ou des origines commune, et d'autre part pour désigner les sociétés présentant un stade avancé de développement (ex : progrès scientifique et technique).

Vis-à-vis de la littératie, nous pouvons entrevoir une connexion avec la civilisation au travers de la thèse aujourd'hui désuète de la *literacy thesis*, parfois nommée *literacy hypothesis*, qui systématisait à l'échelle civilisationnelle le fait que « l'invention, l'adoption et l'application d'un nouveau mode et d'une nouvelle technologie de la communication, à savoir l'écriture, a modifié les pratiques sociales de la société ainsi que les processus cognitifs des personnes impliquées » (Olson & Lejosne, 2006).

5.4.6.4. Sens et représentation

Découlant de ce qui a été précédemment présenté, la culture est ainsi la réalité « telle qu'elle est chargée de sens » en raison de son « rapport à la valeur » (Rickert & Orth, 1997). En effet, la culture « n'existe que si elle soulève la question du sens de la vie et de la recherche de valeurs » (Ellul & Porquet, 2020); la valeur étant définie par Dewey, Truc, & Bidet (2011) comme une qualité associée à un évènement, une situation, un objet ou une personne : elle est une composante de toute expérience immédiate impliquant l'appréciation ou non d'une situation.

Par ailleurs, la culture est un « comportement symbolique », à savoir « des perceptions et modèles structurés du monde, des croyances à son égard qui se jouent en termes symboliques » (Sherzer, 2012). L'ensemble de ces symboliques, de ces constructions de sens et de valeurs, façonnent des représentations de la réalité (*weltanschauung*) et représentent un « patrimoine commun de représentations en quoi se reconnaissent les sujets qui revendiquent la même appartenance sociale et symbolique » (Lamizet, 1992, p. 44).

Rocher (1970) rapprochait cela à une fonction spécifique de la culture qu'est « l'univers psychique » et qui permet de forger les imaginaires des individus. L'imaginaire étant cet « ensemble des productions d'une fonction mentale appelée imagination, à la fois de l'ordre de la reproduction, par le pouvoir qu'elle a de faire revivre des perceptions déjà éprouvées, et créatrice dans sa capacité à former des images selon des combinaisons inédites ». (Giust-Desprairies, 2009). Plus qu'un produit de

l'imagination, il peut être conceptualisé « comme médiation entre le monde intérieur et le monde extérieur, ce qui donne une tout autre signification aux symboles, signes et allégories qu'il mobilise. » (Berdoulay, Castro, & Gomès, 2005) et donne ainsi vie aux représentations, conséquences de ce que Castoriadis nomme « l'imagination radicale ».

Nous l'avons vu dans le chapitre 2, la littératie véhicule également des symboliques, des sens et tout un ensemble de représentations de la réalité qui diffèrent selon les cultures de l'écrit étudiées. La littératie en ce sens produit les mêmes effets que la culture et se base sur des mécanismes d'action similaires.

5.4.6.5. Transmission et technique

Cet ensemble de représentations, d'imaginaire et de sens se transmet « par apport successifs, par adaptation lentes, mûries et intégrées, de générations en générations» (Ellul, 1987), démontrant le caractère dynamique et évolutif de la culture. Morin (2011) voyait justement en elle « l'ensemble des pratiques et des idées qui parviennent à se maintenir dans le temps ou à se diffuser dans l'espace grâce à la transmission : les traditions ». Il évoque à ce propos que la culture se propage selon deux modes : la diffusion (la distribution d'idées dans l'espace et le temps) et la transmission (la passation d'un individu à un autre, nécessitant donc une interaction) [84]. Dans la lignée d'auteurs comme Robert Brunton et Arthur Morice Hocart (1883 - 1939), Morin rejette la vision, alors partagée par des auteurs comme George Herbert Mead (1863 - 1931) ou Emile Durkheim (1858 - 1917), d'une prétendue homogénéité de la culture, parfaitement transmise à l'entièreté des individus d'une société[85]. C'est ce que notait Goody qui précisait que chaque culture possédait « un noyau de doute, sa propre autocritique susceptible de déboucher sur des formes de comportements opposées » (Goody, 2006, p. 279).

Le processus de transmission est présent également dans la littératie (Olson & Torrance, 2009) et lie cette dernière aux caractéristiques contenues dans l'*Erziehung* et la *Bildung*. Pour Sander (2019), *Erziehung* « correspond à l'influence intentionnelle et pédagogique exercée sur le jeune public (comme un savoir, des normes, des valeurs ou des compétences précises) » alors que *Bildung* « vise l'indépendance du soutien pédagogique et est liée à des aspects tels que la maturité, le développement de la personnalité, la capacité de jugement et la compréhension réflexive des expériences du monde. Alors qu'Erziehung se concentre généralement sur des idées concrètes sur ce qui doit être réalisé chez ses destinataires, Bildung ne vise pas des objectifs définis dans un sens utilitaire,

[84] Les traditions ont un statut particulier pour Morin : elles doivent leur distribution grâce à leur transmission. Dans ce cas de figure seulement, distribution et transmission sont interchangeables.

[85] Parlant en particulier des sociétés « primitives », Mead (1940) précisait que la transmission culturelle était « inévitable et parfaitement efficace », traduction de « *the inevitability and complete effectiveness of the transmission of culture* », ce que Morin (2011) rejette.

même si cela peut être utile au développement personnel ainsi qu'à la réussite professionnelle »[86]. Il ajoutera que si l'*Erziehung* s'arrête à un moment donné, la *Bildung* continue tout au long de la vie, se rapportant en cela à la « formation tout au long de la vie ».

Cette transmission n'est possible que grâce à la technique (Stiegler, 1998). En effet, la culture est de fait technique pour Ellul (1987) car conçue par l'homme « dans et à partir de son univers matériel, social, quotidien et concret », au même titre que la littératie notamment en raison du fait qu'elle est une « technologie de l'intellect ». Simondon (2012) ne disait pas autre chose :

> « l'opposition dressée entre la culture et technique, entre l'homme et la machine, est fausse et sans fondement; elle ne recouvre qu'ignorance ou ressentiment » (p. 9).

5.4.6.6. Culture et littératie

Nous avons pu constater que culture et littératie sont deux concepts aux caractéristiques proches mais que des différences notables les séparent. Tout d'abord d'un point de vue définitionnel où les termes *literacy* et *culture* sont séparés dans la langue anglais, ce que précise Lehmans (2018) où la culture pour les anglo-saxons se situerait généralement dans le sens anthropologique de « pratique partagée », tandis que *literacy* désignerait « ce qui se construit de façon dynamique dans un processus d'apprentissage, pas nécessairement sous une forme scolaire classique (une classe, un enseignant, une discipline, une évaluation) mais le plus souvent dans un cadre académique ». Elle ajoute que la fluidité du terme de culture en France expliquerait « une partie de malentendus entre littératie et culture que l'on voudrait synonymes mais qui désignent plutôt, respectivement, un processus et un état » (Lehmans, 2018, p. 30).

Puis, en termes de périmètre : si la littératie reflète et transmet la culture, elle ne l'englobe pas entièrement (Olson, 2017) : « il s'agit d'un concept qui embrasse les ressources culturelles d'une tradition lettrée - ce qui inclue le(s) système(s) d'écriture(s) de cette tradition - et l'ensemble des compétences nécessaires pour exploiter cette ressource culturelle [87] » (Brockmeier & Olson, 2002). La littératie, bien que de nature culturelle, ne représente en fait qu'un périmètre spécifique de « la » culture.

---

[86] « *Erziehung describes intentional, pedagogical influence on young people (for example, by imparting knowledge, norms, values, or abilities). Bildung, in contrast, aims at the independence of pedagogical support and is connected to aspects such as maturity, the development of personality, the ability to make judgements, and reflective understanding of experiences of the world. Whereas Erziehung is usually focused on concrete ideas of what should be achieved in its recipients, Bildung is not aimed at defined goals in a utilitarian sense, even though it may be useful for personal development as well as for professional success* ».

[87] « *It is a concept that embraces the cultural resources of a literate tradition – including the writing system(s) of this tradition – and the ensemble of the abilities necessary to exploit these cultural resource*».

L'article *What is a Culture of literacy* de Brockmeier & Olson (2002), au-delà de rappeler dans leur titre la séparation des deux concepts, notent que l'emploi du terme culture est problématique :

> « Cependant, ce que signifie exactement le terme "culture" ou "culturel" dans ces contextes reste obscur. Même s'il y a eu une diffusion inflationniste du vocabulaire culturel dans les sciences humaines, cette diffusion s'est faite au détriment de la précision. Dans de nombreux contextes, le terme "culture" pourrait facilement être remplacé par des termes tels que "société", "histoire", "représentations sociales" ou "politique", sans perte ni gain de sens. Le caractère insaisissable du sens de la "culture" n'est pas propre aux discussions sur l'écriture et la littératie, ni un phénomène récent. [...] Souvent, le terme "culture" est utilisé pour définir une perspective à travers ses oppositions à certaines perspectives étiquetées existantes (par exemple "cognitives", "nativistes", etc.), plutôt que de servir d'heuristique intellectuelle générale pour une nouvelle compréhension de questions complexes [88] ».

Littératie et culture ne sont donc pas équivalents. Le terme de culture présente l'avantage d'être plus englobant et plus compréhensible pour un public français. Néanmoins, le terme « culture des données » pose problème, au-delà des points suscités, car le terme de « data culture » existe également dans la littérature.[89]

5.4.6.7. La *data culture*

Nous avions rapidement évoqué la *data culture*[90] dans le chapitre 1. Selon Falk (2021) :

[88] « *However, what exactly the term "culture" or "cultural" means in these contexts remains obscure. Although there has been an inflationary spread of the cultural vocabulary in the human sciences, this spread has been at the expense of precision. In many contexts, the term "culture" could easily be replaced by such terms as "society," "history," "social representations," or "politics" without any loss or gain in meaning. The elusiveness of the meaning of "culture" is not unique to discussions on writing and literacy, nor is it a recent phenomena. In their famous survey, carried out half a century ago, Kroeber and Kluckhohn (1952) discussed 161 different meanings of "culture" in the human sciences. This may be an underestimate given the rise and academic institutionalization of new disciplines such as "Cultural Studies," "Communication," and "Media Studies." [...] Often 'culture' is used as a term to define a perspective through its opposition to some existing labeled perspectives (e.g. 'cognitive,' 'nativist,' etc.), rather than serving as a general intellectual heuristic for a new understanding of complex issues." (Crawford and Valsiner 1999: 262).* »

[89] Nous rappelons ici que cette difficulté de traduction s'est posée dans d'autres pays. En allemand, la *data literacy* est traduite par *datenkompetenz* et fait de la *datakultur* une sous-catégorie de la première (Schüller, 2021). En italien et espagnol, les termes *Alfabetizzazione ai dati* (Raffaghelli, 2017) et *Alfabetización en datos* (Martín-González & Iglesias-Rodríguez, 2022) intègrent pour leurs auteurs une part culturelle.

[90] Pour éviter toute confusion avec le terme « culture des données » qui est une traduction souvent employée pour la *data literacy*, nous laisserons le terme non traduit.

> « Les *data cultures* ont émergé au sein du gouvernement moderne, à la fois aux niveaux national et local. Ces cultures sont construites par les conditions socio-matérielles et les pratiques des outils numériques et des services qui ont généré et accumulé des données d'utilisateurs et des métadonnées à travers une multitude d'interactions numériques tout au long du dispositif numérique du secteur public. [91] » (Falk, 2021, p. 38).

Plusieurs tentatives de définitions ont existé : nous les avons recensées en annexe 8. Elles démontrent une instabilité du concept comme l'a été, et peut parfois encore l'être, la *data literacy*. Comme le rappellent Kremser & Brunauer (2019) ainsi que Oliver, Cranefield, Lilley, & Lewellen (2023), ce terme ne rencontre pas de consensus et fait donc l'objet d'appropriations multiples qui concourent à des définitions et conceptions multiples du concept. En effet, la *data culture* évoque soit une même vision sur les données partagée dans une organisation, soit un ensemble de pratiques de données établies au sein d'une organisation sous formes de comportements et de normes, soit un état d'esprit acceptant le changement impulsé par les nouvelles pratiques des données, ou une clé de compréhension et de d'exploration de la complexité des données. Il s'agit d'une distinction entre un aspect individuel et un aspect collectif [92].

Ce terme, s'il n'est pas un équivalent de la *data literacy*, reste fortement lié à cette dernière : la *data culture* est en effet souvent présentée comme sous-catégorie de la *data literacy*, en particulier dans les cadres de compétences où elle représente un ensemble de connaissances générales sur les données à l'instar de Yang & Li (2023) qui englobent sous ce terme de « savoir reconnaître l'importance des données et de soutenir un environnement qui génère un usage critique des données pour l'apprentissage, la recherche et la prise de décision [93] ».

Sternkopf & Mueller (2018) reprennent cette conception de la *data culture* comme sous-catégorie de la *data literacy* qu'ils développent au sein d'un modèle de maturité [94]. Ce modèle est composé de quatre niveaux d'évolution successifs qui se déploient ainsi dans le cas de la *data culture* : un niveau 1 « incertitude » où le terme de « donnée » n'est pas compris et provoque des insécurités ; un niveau 2 « éveil » où la donnée est perçue comme un concept intéressant et les bénéfices qu'on peut en tirer sont appréciés, les doutes se portant sur les cas pratiques et ce qui doit être attendus des données ; un

[91] « *data-cultures have emerged within the modern-day government, both on national and local levels. These cultures are constituted by the socio-material conditions and practices of digital tools and services that generated and accumulated user- and meta-data through a multitude of digital interactions throughout the public sector's digital dispositif*. ».

[92] Nous retrouvons cela dans la distinction entre « culture de l'information » et « culture informationnelle » telle qu'évoquée dans (Le Deuff, 2009) et (Liquete, 2014).

[93] « *Recognizes the importance of data and supports an environment that fosters critical use of data for learning, research, and decision-making* » (p. 13).

[94] Un modèle de maturité sert d'outil de diagnostic (Kohlegger, Maier, & Thalmann, 2009) ayant été testé dans un contexte spécifique (Wilkinson, 2014) et comprenant deux perspectives potentielles : un cycle de vie ou un potentiel de performance (McBride, 2010). Il intègre des niveaux de maturité qui représentent une ligne d'évolution anticipée ou désirée (Becker, Knackstedt, & Pöppelbuß, 2009).

niveau 3 « certitude » où la donnée est perçue comme une source de progrès et comme soutien aux activités existantes et à venir, et où les directives managériales soutiennent les initiatives en matières de données ; un niveau 4 « maîtrise des données » où les barrières psychologiques des données ont été levées et où les politiques managériales sont en faveur des initiatives de données. Plus qu'un socle de base de connaissances sur les données, c'est également ici un état d'esprit ouvert au changement [95], à l'acceptation des données et des nouvelles pratiques institutionnelles impulsées par celles-ci.

Pour Bates (2017), cette *data culture* influence la manière dont les données sont gérées sur l'ensemble de leur cycle de vie et façonne les « propriétés matérielles » des données (*material properties of data*), à savoir leur persistence, leur durabilité, leur spatialité, leur taille, leur mobilité, etc. De plus, elle influence le développement des infrastructures sur lesquelles sont assises les données. Il s'agit donc d'une manière de voir et d'agir sur le monde à travers les données (ce que nous pourrions nommer une *daten-weltanschauung*), de souligner et développer les conditions matérielles de production et de gestion des données.

En particulier les pratiques de données (*data practices*) qui recouvrent l'ensemble des usages des données (Thessen & Patterson, 2011). En effet, selon Aragona & Zindato (2016), la *data culture* intègre deux éléments : les changements méthodologiques et organisationnels qui constituent la production des données dans une période temporelle déterminée et la quantité de données sociales existantes dans cette période.

Selon la vision de Radermacher (2017) fortement imprégnée des statistiques, la *data culture* doit permettre :

- de comprendre la chaîne de processus de génération de connaissances à partir de données dans les conditions actuelles avec ses sous-processus et leurs exigences et tâches caractéristiques ;

  - de développer une structure adéquate de programmes éducatifs adaptés aux différentes étapes du processus de cette chaîne de valeur ainsi qu'aux acteurs censés être compétents à une étape et à un niveau respectifs ;

  - de transmettre non seulement les connaissances et les compétences, mais également les valeurs et les attitudes requises au niveau du processus respectif dans cette formation à la maîtrise des données ;

[95] Ce qui a été nommé « enculturation » par Herskovits et que nous avons cité au chapitre 3 pour parler de la gouvernance des données.

- de réviser les principes existants en matière d'éthique, de bonne gouvernance et de qualité des statistiques afin qu'ils puissent également être pertinents et orienter les *data scientists*.

- de proposer des formes faciles à comprendre mais solides pour la communication de faits et de chiffres ;

- d'impliquer activement les citoyens dans la conception et la production de statistiques ; la co-création et la coproduction peuvent contribuer à combler de manière constructive le fossé entre les experts techniquement qualifiés en production de statistiques et le grand public ;

- de traiter scientifiquement les relations entre statistique et État, entre données et société dans une sociologie de la quantification et de la politique des nombres. [96]

Basés sur ces observations, Poirier & Costelloe-Kuehn (2019) envisagent la *data culture* selon une approche heuristique à plusieurs niveaux d'analyse : méta (la façon dont les forces caractérisées par les échelles ici présentées sont évoquées, interroge le discours dominant et les contre-récits guidant la manière dont une communauté valorise le partage de données), macro (les structures financières et juridiques qui soutiennent le travail de partage de données des communautés et des organisations), méso (les réseaux et les organisations), micro (les coutumes et les pratiques, à la fois des pratiques de données et des pratiques de recherche), techno (la disponibilité, l'accessibilité et la pertinence des technologies et des standards de données soutenant les pratiques partagées de données), de données (l'architecture et la configuration des données) et nano (les croyances des chercheurs apportées au partage des pratiques de données).

Ainsi, la *data culture* est donc cette culture générale des données intégrée à la *data literacy* qui génère une culture organisationnelle impliquant une vision mais également des normes et des pratiques spécifiques sur les données. La *data culture* est également comprise ici comme un engagement, une attitude favorable au changement impulsée par les nouvelles pratiques de données. C'est donc le

[96] « *To understand the process chain of generating knowledge from data under today's conditions with its sub-processes and their characteristic requirements and tasks. To develop an adequate structure of educational programmes that is geared to the different process stages in this value chain as well as actors, supposed to be competent at a respective stage and level. To impart not only the knowledge and skills but also the values and attitudes that are required at the respective process level in this data literacy training. To review the existing principles for ethics, good governance and quality in statistics so that they can also be relevant and guiding for data scientists. To offer easy-to-understand, yet solid forms for the communication of facts and figures. To actively involve citizens in the design and production of statistics; co-creation and co-production can help to constructively bridge the gap between technically skilled experts in statistics production and the lay public. To scientifically process the interrelationship between statistics and the state, between data and society in a sociology of the quantification and of the politics of numbers* ».

lieu des représentations ainsi que des interactions avec le monde grâce et par les données (voir la figure 5.6).

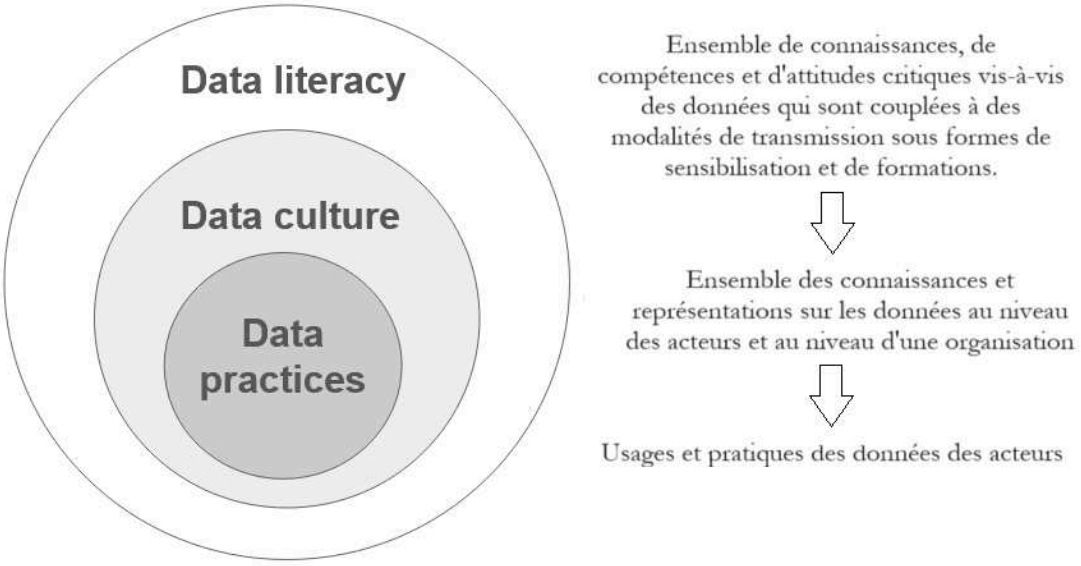

Figure 5.6.: Data literacy et data culture

En point supplémentaire, au même titre que la *data literacy*, une vision critique (au sens de littératie critique) a été développée pour la *data culture* : Raffaghelli & Sangrà (2023) ont ainsi proposé une *fair data culture* renvoyant à la possibilité de rendre visible les éléments constitutifs d'une data culture pour toutes les personnes concernées, de créer des espaces qui leur seront dédiés pour prendre part et transformer la *data culture* existante.

5.4.7. Conclusion de la troisième partie

La *data literacy* connaît donc une transposition française, en particulier par le truchement des Sciences de l'Information et de la Communication depuis les années 2010. Les caractéristiques développées contiennent certains éléments arrivés plus tard dans la conceptualisation anglo-saxonne, en particulier la posture critique, ce qui souligne à notre sens l'imprégnation des réflexions antérieures sur l'*information literacy* qui a suivi une conceptualisation et des développements très similaires et qui, nous l'avons vu, entretient une relation symbiotique avec la *data literacy*.

Ce qui est renforcé par le fait que nombre des chercheurs ont précédemment, voire continuent de travailler sur l'*information literacy*, notamment dans le cadre de l'*Éducation aux Médias et à l'Information*. La vision pédagogique est d'ailleurs assez importante lorsque la littératie est abordée et cela se ressent particulièrement dans son étude en France. En effet, si les compétences visées sont

très similaires à celles recherchées dans la *data literacy* anglo-saxonne, la vraie ligne de démarcation française se situe dans ses approches. Médiation, éducation, culture : les prismes d'analyses pédagogiques et culturelles permettent de sortir d'une vision purement fonctionnelle et technique (bien qu'indispensable) simplement basée sur la production de curricula et de cadres de compétence pour élargir le périmètre d'étude. Néanmoins, la didactique de la *data literacy* est loin d'être résolue et est toujours un enjeu actuel.

Il en est de même pour la traduction de la *data literacy*. Nous avions à l'origine proposé littératie des données pour parler de la *data literacy* et de culture des données si une vision spécifiquement française émergeait (Verdi, 2023b). Une interrogation également abordée par Casemajor Loustau qui distingue la littératie des données et la culture des données :

> « Alors que la notion de littératie des données est davantage centrée sur les apprentissages engagés au niveau individuel, la notion de culture de données est plus large et plus englobante : par culture, on entend ici un ensemble de répertoires de sensibilité et de signification collective partagés au sein d'un groupe. C'est à l'intérieur de ce cadre culturel plus large qu'on peut situer la notion de public de données : les manières d'imaginer les publics comme entités collectives au sein d'environnements numériques, les manières de les constituer en pratique (via la collecte de données, la médiation algorithmique et les mesures d'audiences), les manières dont ils s'assemblent effectivement à travers les environnements numériques, et les manières de se confronter à leur agentivité – toutes dépendent en bonne partie de schémas d'action et d'interprétation d'ordre culturel. » (Casemajor Loustau, 2023).

Sur la base des éléments apportés au sein de ce chapitre, nous serions tentés de parler respectivement de littératie des données pour la *data literacy* et de culture des données pour la *data culture*. Or, nous avons pu noter que parler de littératie ne va pas de soi chez un public français au contraire de la culture, et mélanger ainsi ces deux traductions compliquerait plus qu'autre chose les représentations des acteurs. Par ailleurs, la *data culture* se rapproche de ce que nous nommons la culture générale, un terme tout aussi compliqué à appréhender, dont le périmètre est également l'affaire de visions contradictoires. En outre, la culture générale est souvent perçue comme une connaissance reçue où l'individu a un rôle passif, ce que Freire nommait la « pédagogie bancaire » (Freire, 1968), et évacue les compétences pratiques. Sur cette base imparfaite et dans un souci de clarification, nous proposons néanmoins « culture des données » pour *data literacy*, « culture personnelle des données » pour parler de la *data culture* à un niveau individuel et « culture institutionnelle des données » pour parler de la vision partagée au sein d'une institution.

5.5. Conclusion du chapitre

Un constat peut être ici dressé concernant les généalogies conceptuelles : celles-ci sont propres à chaque acteur et l'on ne peut ainsi pas parler d'une généalogie unique. Si des emprunts à la *data literacy* anglo-saxonne sont manifestes, ce n'est pas le cas de tous les acteurs et pour nombre d'entre eux décéler les sources inspiratrices est difficile, en particulier dans le cas des entreprises où la « culture data » semble propre à chaque entreprise.

Dans ce dernier cadre, si la « culture data » se retrouve fréquemment dans les discours des entreprises, laissant penser qu'il existe une vision commune, nous avons pu également déceler d'autres dénominations comme « datalphabétisation » et « data lettrisme » mais celles-ci ne dépassent que peu le périmètre d'un article ou d'une publication et une véritable conceptualisation manque. Et il en est de même pour les autres types d'acteurs comme les acteurs universitaires qui proposent également des visions différentes, à défaut d'être concurrentes. Les écosystèmes des acteurs n'ont donc pas de visions complètement homogènes en leur propre sein.

En revanche, ces ensembles d'appropriations ne présentent pas de différences fondamentales en termes de compétences, notamment celles basées sur le cycle de vie des données qui sont ici assez similaires. La vision critique au sens d'une recontextualisation des données est par ailleurs commune à l'ensemble des discours ; la manipulation des données sans prise en compte des contextes socio-économiques étant ici une entrave aux « bonnes » pratiques des données.

Là où les visions diffèrent, c'est d'une part sur les prismes d'analyse et de transmission, d'autre part sur les objectifs poursuivis.

Les prismes d'analyse peuvent être déduis directement des intitulés. Médiation, éducation, (data)alphétisation, lettrisme, culture : de ces concepts semblent découler des structures de pensée et des applications spécifiques de la *data literacy* totalement différentes.

Dans le cas des acteurs entrepreneuriaux et institutionnels, il est important de souligner que la gouvernance des données n'est jamais absente des discours et est toujours liée à la *data literacy* qui permet son élaboration et sa pérennité, et ce depuis les années 2010 : c'est donc une logique de gestion parcimonieuse et efficace des données qui dicte les recherches sur la *data literacy*, seule supposément à même de transformer les pratiques professionnelles pour répondre à cette nouvelle nécessité organisationnelle entièrement *data-driven*.

Dans les cas des objectifs poursuivis, les projets traversent le prisme des thématiques présentées : ils portent soit sur la réussite de projets économiques et sur le soutien à la prise de décisions dans le cas des entreprises, soit sur la bonne gestion des services publics dans le cas des acteurs institutionnels

ou encore sur l'éducation citoyenne dans le cas des projets de recherche ; les associations et coopératives sont plus éclectiques et tendent à faire converger plusieurs de ces objectifs. A nouveau, une homogénéité ne s'établit pas ici.

Pour dépasser le cadre des publications, le chapitre suivant présentera les résultats de notre enquête auprès d'un ensemble d'acteurs français représentatifs ayant travaillé sur la conceptualisation et/où l'application de la *data literacy*. Nous interrogerons l'ensemble des représentations et approfondirons des points peu abordés, peu explicités ou non résolus dans les travaux présentés ci-dessus.

Chapitre 6.

Un point de vue français de la *data literacy* ?

« *Data! data! data! I can't make bricks without clay*»

Arthur Conan Doyle, *les Hêtres Rouges*, 1892

6.1. Introduction

Le chapitre précédent a été l'occasion de faire un point sur les discours tenus par différents acteurs français sur la *data literacy*. L'objectif de ce chapitre est d'approfondir les points précédemment abordés au moyen d'une enquête qualitative afin de vérifier d'une part si une évolution a eu lieu dans la pensée de certains acteurs s'étant déjà exprimé sur cette thématique et d'autre part de souligner les spécificités de leur appropriation du concept.

6.2. Première partie : terrain et méthodologie

Nous inscrivons notre démarche dans une recherche qualitative et plus précisément dans une enquête qualitative de terrain, définie par Paillé & Mucchielli (2021a) comme « la recherche qui implique un contact personnel avec les sujets de la recherche, principalement par des entretiens et par l'observation des pratiques dans les milieux mêmes où évoluent les acteurs » (paragr.4). Notre objectif étant de solliciter les représentations des personnes interrogées à travers leurs discours, il est nécessaire de faire émerger « la perception qu'un acteur élabore d'une situation donnée » car elle

« constitue pour lui la réalité de cette situation » (Bertaux, 2016, p. 31) et de reconstituer le « monde collectif dont chacun des acteurs n'a qu'un petit bout, bien qu'il participe à la construction totale collective » (Paillé & Mucchielli, 2021c, paragr. 7).

6.2.1. L'entretien semi-directif

Pour ce faire, nous avons opté pour la réalisation d'entretiens semi-directifs. L'entretien semi-directif « se veut une forme d'interaction proche de la conversation, grâce à l'adaptation continue des interrogations et interventions du chercheur à l'échange en cours » (Nossik, 2011, paragr. 6). En effet, à la différence d'un entretien directif, il permet à l'enquêteur de délimiter « la forme et l'ordre de ses questions aux réactions de l'enquêté au fil de l'interaction » (Nossik, 2011, paragr. 6) et de positionner l'enquêté « dans un rôle d'informateur/-trice, de détenteur/-trice d'un savoir (commun, non scientifique) précieux sur le thème d'intérêt de l'enquêteur-trice » (Pin, 2023). Il permet ainsi d'explorer des hypothèses « sans qu'elles soient toutes définitives » (Sauvayre, 2013, Chapitre 1, paragr. 28).

6.2.2. L'échantillon d'analyse

6.2.2.1. La catégorisation des personnes interrogées

Nous avons procédé ensuite à une catégorisation des personnes interrogées pour faciliter leur regroupement et les analyses *a posteriori*. Dans le cas présent, la catégorisation des publics interrogés selon leur milieu professionnel de rattachement doit permettre de refléter des écosystèmes supposément séparés, ayant leurs propres logiques et imaginaires ; Sfez (2002) précisait en effet qu'« il y aurait ainsi une collection d'imaginaires particuliers propres à des métiers, professions ou vocations » (p. 33-34). La catégorisation est en ce sens « l'attribution même de la signification » (Paillé & Mucchielli, 2021b, paragr. 13) désignant directement le phénomène étudié et permettant de construire les échantillons où seront appliquées les analyses :

> « C'est en fonction de ce phénomène de *variété des positions* – et donc des points de vue des acteurs qui les occupent – que l'on construira progressivement un échantillon, en y incluant les différentes catégories d'agents/acteurs, y compris des subdivisions apparues en cours d'enquête (ouvriers syndiqués et non syndiqués, militants et non-militants). Aucune catégorie d'acteurs ne détient à elle seule toute la connaissance objective ; mais la vision de chacune contient sa part de vérité ; et c'est par leur mise en rapport critique, par le chercheur, que passe notamment le travail de construction d'une représentation

> mentale plus objective, d'un meilleur modèle de l'objet social étudié. » (Bertaux, 2016, p. 31).

Notre choix s'est porté sur quatre catégories socio-professionnelles que nous avons nommées « universitaire », « entrepreneuriale », « institutionnelle » et « associative ». Celles-ci reflètent celles des acteurs rencontrés lors de notre état de l'art du chapitre 5.

- Est dite « universitaire » toute personne travaillant pour un établissement de l'enseignement supérieur (ex : université, IUT, école de commerce, etc.) et ce quel que soit son statut (ex : les maîtres de conférences, les professeurs des universités, les professeurs associés, etc.) ;

- Est dite « entrepreneuriale » toute personne exerçant son activité dans une structure privée comme des start-ups, des entreprises, des industries, etc. ;

- Est dite « institutionnelle » toute personne affiliée à une structure de service public comme les métropoles, les villes, les départements, les ministères, etc. ;

- Est dite « associative » toute personne intégrée à une association mais également à une coopérative (ex : des *think tanks*).

6.2.2.2. La représentativité des personnes interrogées

La représentativité des acteurs non seulement vis-à-vis de la thématique questionnée, mais également au sein de leur catégorie est un questionnement récurrent dans ce type d'enquête. Pour Padieu (1974), la représentativité « n'est pas une propriété individuelle. C'est une propriété de l'échantillon dans son ensemble » :

> « la représentativité est de nature synthétique. Un député, un délégué, un ambassadeur, le président d'une association représente une collectivité. Cela ne veut pas dire qu'il n'exprime pas ou pas forcément, sa propre opinion; il s'efforce de rendre compte des opinions, des désirs, des intérêts de ses mandants, dont il fait une synthèse. Le caractère synthétique, nécessaire pour conférer la propriété de représentativité, exige alors de prendre en considération la totalité de l'échantillon. Les unités qui composent l'échantillon ne sont pas, n'ont pas à être représentatives. L'échantillon dans son ensemble peut l'être » (pp. 65-66).

Du fait que le nombre de personnes travaillant, non pas sur les données, mais sur les stratégies de formation aux données est relativement limité, nous les avons filtrées en fonction de leur expérience

significative vis-à-vis de ces stratégies. Cette expérience a été premièrement déduite de la lecture de toute la littérature disponible : descriptif de projets réalisés, productions scientifiques publiées, rapports rédigés, contenus web affichés, etc. Cela a été complété par la suite par les retours des premières personnes interrogées qui nous ont aiguillé vers d'autres personnes ressources.

Il est évident que toute catégorisation est subjective : « créer une catégorie, c'est, déjà, mettre en marche l'articulation du sens des représentations, des vécus et des événements consignés. » (Paillé & Mucchielli, 2012, Chapitre 12, paragr. 12). Son périmètre, censé traduire une certaine homogénéité, révélant par ailleurs des pratiques hétérogènes et des opinions souvent discordantes entre les individus catégorisés.

En outre, les postes occupés au sein d'une même catégorie présentent également des différences notables, notamment dans la catégorie « entrepreneuriale » qui intègre un(e) data journaliste, des chef(fe)s d'entreprise, des consultant(e)s, etc. Comme le précisait Durkheim (2020) : « Il n'est pas d'uniformité sociale qui ne permette toute une échelle de gradations individuelles, il n'est pas de fait collectif qui s'impose de manière uniforme à tous les individus. » (p. 15). Nous soutenons néanmoins que la catégorisation reste nécessaire et que l'analyse permettra de déterminer si une homogénéité de pensées et de pratiques est présente.

Un dernier point a été pris en compte, celui de savoir comment catégoriser des acteurs ayant eu une expérience significative dans plusieurs milieux. Deux solutions ont été envisagées :

- soit ne pas catégoriser l'acteur et le présenter comme un « électron libre »,
- soit l'intégrer dans une catégorie représentative de son parcours.

C'est cette dernière solution qui a été retenue.

6.2.3. Les étapes suivies

L'enquête de terrain a duré deux ans (2021 – 2023). Elle a suivi une progression fluctuante et parfois rétroactive selon les apports des personnes interrogées. Elle est intervenue après une première année de thèse dédiée à la réalisation d'un état de l'art scientifique afin de définir les caractéristiques de la *data literacy* et de déterminer les matériaux manquants pour procéder à une analyse plus approfondie de la thématique. Nous listons ci-dessous les cinq étapes suivies pendant cette enquête :

1. Le *listing* des acteurs potentiellement pertinents : pour ce faire, tous les contenus papiers et numériques ont été consultés. Une lecture combinée par une consultation de la presse spécialisée et de la littérature scientifique [avant septembre 2021] ;

2. la démarche de contact : une fois les acteurs identifiés, le mail a été le moyen principal pour initier une première prise de contact. 92 personnes ont ainsi été contactées à deux, voire trois reprises [septembre 2021 - décembre 2022] ;

3. la réalisation des entretiens [octobre 2021 - juillet 2023] : après 28 heures d'entretien, nous avons fini par ressentir un « phénomène de saturation » au sens de (Bertaux, 1980, p. 205), à savoir « le phénomène par lequel, passé un certain nombre d'entretiens (biographiques ou non, d'ailleurs), le chercheur ou l'équipe a l'impression de ne plus rien apprendre de nouveau, du moins en ce qui concerne l'objet sociologique de l'enquête » et avons cessé nos démarches d'enquête en 2023 ;

4. la transcription des entretiens : plusieurs logiciels ont été employés avant d'abandonner leur usage en raison de trop grandes erreurs de transcription dues notamment à l'usage d'un vocabulaire beaucoup trop spécifique [septembre 2022 - mars 2023] ;

5. l'analyse des entretiens [avril 2023 - septembre 2023].

32 personnes ont été finalement interrogées réparties en fonction de la chronologie des entretiens. Nous comptabilisons 9 « universitaires », 9 « entrepreneurs », 7 « associatifs » et 7 « institutionnels ».

Tableau 6.1.: Les personnes interrogées entre 2021 et 2023

Typologie d'acteur	**Date de l'entretien**
Entrepreneur 1	19 octobre 2021
Universitaire 1	28 octobre 2021
Institutionnel 1	20 décembre 2021
Universitaire 2	06 janvier 2022
Entrepreneur 2	09 février 2022
Entrepreneur 3	10 février 2022
Associatif 1	04 mars 2022
Associatif 2	21 mars 2022
Institutionnel 2	24 mars 2022
Associatif 3	31 mars 2022
Universitaire 3	01 avril 2022
Associatif 4	13 avril 2022
Universitaire 4	20 avril 2022
Institutionnel 3	06 mai 2022
Entrepreneur 4	11 mai 2022
Entrepreneur 5	18 mai 2022

Typologie d'acteur	Date de l'entretien
Institutionnel 4	20 mai 2022
Universitaire 5	31 mai 2022
Entrepreneur 6	22 juin 2022
Entrepreneur 7	24 juin 2022
Entrepreneur 8	27 juin 2022
Universitaire 6	01 juillet 2022
Universitaire 7	11 juillet 2022
Universitaire 8	13 juillet 2022
Entrepreneur 9	05 octobre 2022
Associatif 5	24 octobre 2022
Universitaire 9	25 novembre 2022
Institutionnel 5	28 novembre 2022
Institutionnel 6	06 décembre 2022
Associatif 6	07 décembre 2022
Institutionnel 7	10 février 2023
Associatif 7	26 juillet 2023

Nous avons recontacté l'ensemble des personnes interrogées début 2024 : celles-ci témoignaient de l'évolution de leurs pensées et de leurs représentations ; certaines se distanciant de leurs propos à la relecture des transcriptions. Ceci démontre l'extrême rapidité d'appropriation des données par les acteurs et le fort impact en conséquence sur l'évolution des discours en seulement quelques années (deux ou trois ans selon les acteurs interrogés). Les résultats des entretiens sont donc représentatifs de la période 2021 - 2023.

6.2.4. Les modalités de l'entretien

6.2.4.1. Le guide d'entretien

Un guide d'entretien a été conçu (présent en annexe 10). Ce dernier est un « premier travail de traduction des hypothèses de recherche en indicateurs concrets et de reformulation des questions de recherche (pour soi) en questions d'enquête (pour les interviewés) » (Blanchet & Gotman, 2012). Il était composé d'une quinzaine de sous-thématiques et était programmé pour que les entretiens durent entre 40mn et 1h selon la longueur des réponses. Il servait de cadre global mais il est souvent arrivé d'en sortir précisément pour approfondir certaines réponses, nécessitant alors de déployer une nouvelle batterie de questions et rallongeant de fait la durée parfois longue de l'entretien. Les points

abordés dans le guide suivaient un fil rouge à travers trois grands aspects : l'aspect conceptuel, l'aspect des enjeux et l'aspect applicatif. Il s'agissait de recueillir un maximum d'informations en laissant suffisamment de marge de manœuvre à la personne interrogée pour ne pas limiter son propos, ce qui aurait pu être le cas en intégrant des questions trop spécifiques ou situées à un niveau de granularité trop important.

6.2.4.2. Le déroulement des entretiens

La personne interrogée était au fait du but de l'entretien et de ses modalités mais n'avait en revanche pas accès aux questions pour éviter d'influencer ses réponses. De rares personnes ont reçu certaines questions en amont dans le but de s'assurer de leur pertinence dans ce domaine ; plusieurs d'entre elles ont ainsi précisé ne pas se sentir compétentes et ont donc été retirées de notre liste d'enquêté(e)s. Les entretiens ont tous été réalisés en visio-conférence du fait de la répartition des acteurs sur l'ensemble du territoire français. Les vidéos et fichiers audio ont été enregistrés et déposés dans un serveur sécurisé pour permettre leurs transcriptions, accessibles uniquement par nous-même.

6.2.4.3. L'anonymisation des personnes interrogées

Dès le départ s'est posée la question de l'anonymisation des résultats. De par l'instauration du *RGPD* et le respect d'une éthique de recherche, un formulaire de consentement a été transféré aux participant(e)s en amont des entretiens. Il récapitulait l'ensemble du projet de thèse, les modalités de l'entretien et les contacts. L'ensemble des transcriptions ont ensuite été anonymisées et pseudonymisées. Celles-ci, sur autorisation des personnes interrogées, ont été intégrées dans les annexes 11 à 14. Toutes ne sont pas présentes du fait du refus de certaines personnes interrogées qui ne souhaitaient pas voir l'intégralité de leur apport présenté.

6.3. Deuxième partie : synthèse des résultats

Pour débuter cette synthèse, nous nous sommes intéressés en premier lieu au vocabulaire employé pour parler de la *data literacy* (listé dans la figure 6.1).

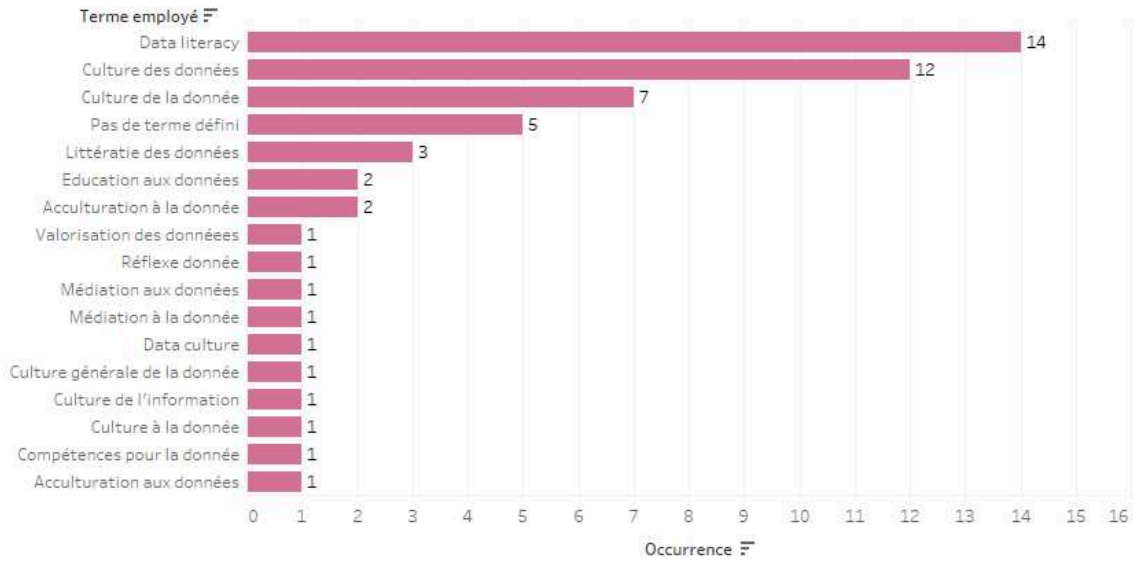

Figure 6.1.: Occurrence des termes employés par les acteurs

6.3.1. Les termes employés

Nous avons pu mettre en évidence la présence de 16 termes, soit plus que ce que nous avions repéré dans la littérature. « Data literacy » est le plus employé (n=14), suivi de près par « culture des données » (n=12) déployé selon plusieurs modalités : « culture de la donnée », « culture générale de la donnée », « culture à la donnée » et « data culture ». S'ensuivent d'autres termes qui connaissent également des variations : « médiation aux données » et « médiation à la donnée », « acculturation aux données » et « acculturation à la donnée ». Et enfin, quatre termes sans variations : « compétences pour la donnée », « éducation aux données », « littératie des données » et « réflexe donnée ». « Culture de l'information » est employé quant à lui dans un contexte plus englobant. Tout ceci rejoint notre précédente analyse dans le chapitre 5 qui présentait déjà une myriade de traductions françaises. Nous avons corrélé ces usages selon les typologies de publics (voir la figure 6.2) :

La plupart des acteurs utilisent plusieurs termes à la fois et ces choix sont circonstanciels. Dans ce cadre, « data literacy » et « culture des données » sont employés par la totalité des typologies de publics interrogés, présentant ici un certain consensus. L'emploi des autres termes semble, cette fois-ci, relever de logiques plus individuelles que typologiques. « Culture data » n'apparaît étrangement pas ici, remplacé par les termes suscités.

Le tableau 6.2 intègre les grands concepts mobilisés et fusionne les variances sous une dénomination commune (ex : « acculturation » contient les termes « acculturation à la donnée » et « acculturation

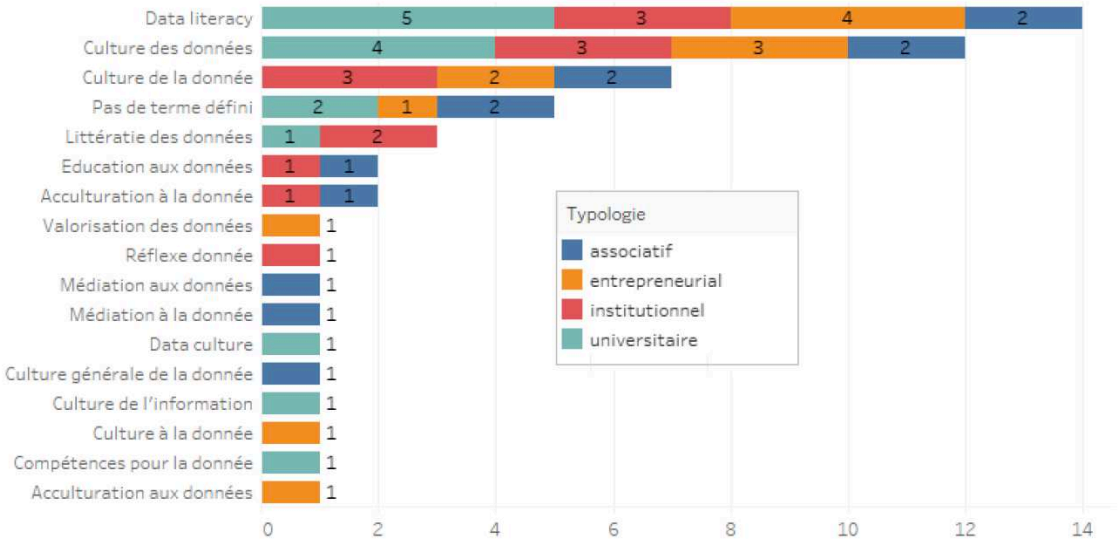

Figure 6.2.: Occurrence des termes selon la typologie des acteurs

aux données ») afin de déceler une potentielle tendance d'appropriation. Les termes les plus employés sont en gras pour plus de visibilité :

Tableau 6.2.: Concepts mobilisés selon les typologies d'acteurs

Terme/Public	Associatif	Entrepreneurial	Institutionnel	Universitaire	Total
Acculturation	1	1	1	0	3
Compétences	0	0	0	1	1
Culture	**5**	**6**	**6**	**5**	**23**
Data literacy	2	4	3	5	14
Éducation	1	0	0	0	1
Littératie	0	0	2	1	3
Médiation	2	0	0	0	2
Réflexe	0	0	1	0	1
Valorisation	0	1	0	0	1
Absence de terme	2	1	0	2	5
Total	13	13	13	15	54

Le milieu « associatif » est le plus enclin à l'usage d'une multitude de concepts différents en raison de la diversité des actions entreprises au sein de leurs structures comme nous avions pu le constater dans le chapitre précédent. Néanmoins, nous remarquons une mobilisation assez forte du concept de culture (n=22). En comparaison avec la figure précédente, le terme « data literacy », originellement

le plus cité, passe au second plan. Les raisons à cela seront évoquées par la suite. Il faut ici retenir que les concepts sollicités relèvent tous d'une vision et d'une visée particulières :

- « Acculturation » englobe l'ensemble des initiatives de formation des publics aux usages des données ;
- « Compétences » met l'accent sur l'aspect fonctionnel des formations ;
- « Culture » renvoie à la notion de culture générale et intègre tout aussi bien des connaissances (ex : savoir à quoi sert une donnée) que des compétences techniques (ex : savoir nettoyer des données) ;
- « Éducation » renvoie d'une part à la pédagogie faite auprès des publics et d'autre part à la concrétisation de ces formations au sein de l'Éducation Nationale ;
- « Littératie » intègre un aspect anthropologique de mise en contexte de pratiques de lecture et d'écriture ;
- « Médiation » souligne la nécessité de former des personnels intermédiaires pouvant comprendre les spécificités techniques et étant capables de les vulgariser ;
- « Réflexe » traduit un « état d'esprit », un automatisme à atteindre lorsque confronté à un traitement de données ;
- « Valorisation » désigne à la fois une compétence technique (ex : utilisation des métadonnées pour améliorer l'accès aux contenus) et une compétence communicationnelle (ex : faire la promotion d'un portail *open data*).

Plusieurs acteurs délaissent l'usage d'une dénomination en raison (1) de l'existence d'un terme plus généraliste (ex : le numérique) ou (2) parce que les techniques d'acculturation aux usages des données ne nécessitent pas, pour eux, d'être englobées sous une appellation spécifique.

6.3.1.1. La littératie et la problématique de l'anglais

Le terme *data literacy* pose des difficultés de compréhension, en premier lieu en raison de l'anglais. Chez les acteurs de type « institutionnel », l'anglais est proscrit au sein des administrations publiques, nécessitant donc une solution alternative. Pour d'autres acteurs, l'existence d'équivalents français permet d'évacuer cette question :

> « Ce n'est pas un terme que je vais employer pour une raison très simple : c'est qu'aujourd'hui mes principales actions professionnelles sont avec des petites collectivités et des agents de petites collectivités, ou même des DGS, voire des élus. On reste dans tous les cas dans un cadre de petites collectivités. Donc tout ce qui va être de l'ordre des anglicismes… déjà open data, ce n'est pas simple. On m'a déjà demandé : "data c'est l'acronyme de quoi ?". Donc on part de très loin. » [Entrepreneur 6].

L'anglais peut même, à certains égards, être synonyme de méfiance, voire de rejet :

> « Ce n'est pas un terme qu'on emploie tant que ça, sauf dans certains cercles, dans certaines réunions ou dans certaines interventions. Ce n'est pas un terme qu'on emploie pour se définir. Ce n'est pas un terme qu'on emploie non plus pour expliquer ce qu'il y a derrière parce qu'il me semble que c'est un terme qui reste assez peu utilisé : je ne dirais pas élitiste mais assez peu utilisé, assez peu connu ; lorsqu'on parle au grand public, c'est certain. Mais même lorsqu'on parle à d'autres personnes, surtout des entreprises ou des organisations publiques (et d'autant plus que lorsqu'on parle à des organisations publiques), il peut y avoir en plus une méfiance quant à l'usage de l'anglais… ce n'est pas systématique bien sûr mais cela peut arriver. » [Associatif 7].

> « Le terme data fait peur ! L'utilisation de la data fait technique, *literacy* est un terme anglais, donc oui entre techniciens on peut parler *data literacy* mais quand on parle à des collègues ou quand on parle à des élus, moi je parle de culture de la donnée. On utilise trop de mots anglais dans le numérique ce que l'on reproche souvent, et l'enjeu est de rendre accessible le sujet et non d'utiliser du jargon technique. » [Institutionnel 7].

A l'inverse, celui-ci peut être encouragé en raison de son imprégnation dans les pratiques quotidiennes et de la facilitation qu'il apporte à la compréhension du concept :

> « Alors dans notre milieu de data journalistes, on est très « frenglish » en permanence parce qu'on est dans un milieu qui baigne énormément dans le monde anglo-saxon que ce soit pour les outils, pour les inspirations, parce que c'est vrai que le data journalisme est plutôt né aux États-Unis et Angleterre donc c'est vrai que je continue de dire *data literacy* et j'ai bien du mal à sortir de ce terme. Après plus généralement quand j'en parle autour de moi je vais plus dire la culture des données. » [Entrepreneur 2].

> « *Data literacy*, *mindset*, j'emploie ces mots anglo-saxons parce qu'en fait ils contribuent, surtout parmi les gens qui ont des pré-formatage audit/conseil anglo-saxon, à leur fixer les idées dans la langue de Shakespeare. » [Entrepreneur 1].

Ainsi, l'emploi de l'anglais est très circonstanciel : on note une certaine homogénéité au sein du public « institutionnel », en particulier les acteurs officiant dans des administrations publiques qui rejettent l'anglais par nécessité, tandis que l'inverse se présente pour un public « entrepreneurial » où l'anglais est couramment employé. Les publics « universitaire » et « associatif » interrogés sont plus partagés sur ce point : on en vient ici à des choix du terme *data literacy* plus dictés par des logiques individuelles que déterminés par l'organisation de rattachement. Si une compréhension commune émerge

chez les acteurs à l'évocation du terme *data literacy*, nous pouvons souligner toutefois la difficulté de sa réception chez leurs destinataires.

Pour compenser cette difficulté et ainsi proposer un terme plus adéquat pour un public francophone, plusieurs traductions sont proposées. Ces dernières embarquent de nouvelles réalités parfois éloignées de la vision originelle de la *data literacy*, et opèrent un transfert de concepts, la culture étant le concept le plus usité, préféré à celui de littératie. Celle-ci est considérée par plusieurs acteurs comme un concept plus généraliste et englobant dont le caractère peu restrictif facilite l'intégration d'un panel plus large d'enjeux comme le précise par exemple [Entrepreneur 7] : « la culture des données a une dimension plus collective, organisationnelle et donc qui met des enjeux différents. Je dirais que la culture des données intègre mieux cela ». La question de l'équivalence entre culture des/de la/aux/à la donnée(s) et *data literacy* est quelque fois questionnée, interrogeant la pertinence d'une telle traduction. Par ailleurs, le lien avec l'acculturation est sous-jacent aux visions déployées ; la culture et la médiation impliquant *de facto* cette notion pour les personnes interrogées.

> « C'est pour ça qu'on a employé le terme de culture : la culture c'est à la fois une compréhension on va dire globale, conceptuelle, une entrée intellectuelle sur un domaine. Et c'est aussi une pratique : parce que la culture elle se vit et elle se fait et elle se concrétise, donc c'est pour ça qu'on a parlé de culture. De comprendre, quand on parle de données, de quoi il s'agit, à quoi on fait référence, d'avoir en fait un peu tout le bagage définitionnel derrière. De savoir un petit peu comment ça impacte la vie de chacun : c'est pour ça qu'on parlait aussi de culture, en fait ça nous définit dans notre être à la société et de notre rapport à l'autre, c'est pour ça qu'on parlait de culture. Et avec la culture, l'idée de pouvoir créer et d'apporter les contenus nouveaux. [...] Quand on parle de culture des données, on parle aussi de repères qui permettent d'englober les champs de la donnée sans être enfermé dans l'un ou dans l'autre, d'avoir la vue globale générale. » [Institutionnel 1].

> « Pour moi, la culture des données c'est plus large. C'est aussi moins prescriptif, voire moins restrictif, c'est-à-dire que c'est un peu la culture qui peut se constituer de différentes manières par le faire, par les discussions... *data literacy*, je l'associe plus à des démarches éducatives ou pédagogiques qu'on peut développer dans le cadre plutôt institutionnel. [...] Je parlais plutôt de la culture des données dans mes travaux mais les deux termes me conviennent : je n'ai rien contre *data literacy*. Pour moi, *data literacy*, ça implique une référence plus importante à la question de la pédagogie. Avec la culture des données, on n'est pas obligé de passer par l'éducation en fait. Alors que la *data literacy*, ça suppose une démarche éducative et prescriptive quelque part. Pour moi, il y a moins de liberté dans la *data literacy* que dans la culture des données. » [Universitaire 4].

Le lien entre culture et littératie est évoqué par plusieurs acteurs dont les démarches s'inscrivent autour de stratégies pédagogiques. Si les deux termes sont dissociés, le choix d'une dénomination adéquate semble en revanche difficile pour transcrire au mieux la *data literacy* comme le rappelle [Universitaire 2] : « j'emploie plutôt le terme de littératie des données en français. Mais j'utilise aussi le terme de culture des données. Alors, effectivement, parfois on emploie l'un pour l'autre. On ne sait pas lequel choisir ». Comme nous avons pu le constater au cours du chapitre 5, la compréhension de la littératie n'est pas une chose aisée pour un public francophone et nous retrouvons cela au sein des entretiens effectués.

La « culture » reste donc une première porte d'entrée vulgarisatrice permettant d'intégrer les publics et de faciliter leur compréhension des thématiques des données. La « data culture » est mentionnée comme équivalente à la « culture des données » par [Universitaire 6] mais ne renvoie pas à la notion anglo-saxonne présentée en chapitre 5. Il s'agit ici d'une vision singulière seulement rencontrée uniquement chez cette personne :

> « Je revendique plutôt, en tout cas par rapport à mon parcours, la *data culture*, c'est-à-dire la culture des données. Mais on n'est pas sur le même registre : c'est-à-dire qu'on est sur quelque chose de plus générique qui est quelque chose justement de l'ordre de l'anthropologie politique ou les modalités de communication autour notamment de la *data literacy*. Mais c'est pas par synonyme : on n'est pas sur le même positionnement, on n'est pas sur les mêmes attentes. » [Universitaire 6].

6.3.1.2. Singulier, pluriel : l'essentialisme de(s) donnée(s) à l'épreuve

Par ailleurs, la présence d'une essentialisation n'est pas abordée par les acteurs qui l'utilisent, employant invariablement « la » ou « les » données. A l'exception d'un acteur qui précise : « il y a cette idée de dire "culture de la donnée" cela ne va pas : on ne dit pas la donnée. Il y a tellement de types de données différentes, cela essentialise... ". Et à la fin de cela, on passe à "culture des données" avec cette idée de pouvoir prendre en considération plus largement ce champ qui n'est à l'époque pas nouveau mais qui en tout cas prend une importance nouvelle à la fois dans les politiques publiques et aussi dans les interrogations en lien avec la place de l'informatique à l'école ; l'apprentissage du code notamment » [Universitaire 3]. La *data literacy* traiterait l'ensemble des données et apporterait un traitement spécifique pour chacune d'entre elles. Nous approfondirons ce point dans la partie « Pourquoi former ? » où nous interrogerons les acteurs sur les typologies de données au cœur des enjeux actuels.

Nous avons pu ainsi constater que plusieurs critères entrent ainsi en jeu dans le choix d'un terme :

1. La connaissance de celui-ci et de ce qu'il peut recouvrir non seulement pour l'acteur mais également pour le public destinataire ;

2. L'usage (imposé ou non) d'une langue spécifique ;

3. La familiarité des usages avec un terme fermement enraciné.

Nous notons par ailleurs que, malgré le fait que nous avons affaire à des personnes lettrées (au sens de *literate*), des problèmes de compréhension sont perceptibles (en particulier vis-à-vis de la littératie) et sont redoublés du fait des interactions quotidiennes avec d'un côté l'environnement de travail (ce qui inclue un encadrement des modalités d'expression) et d'un autre côté les personnes destinataires des discours, nécessitant un réagencement constant du vocabulaire employé.

6.3.2. Les caractéristiques de la *data literacy*

6.3.2.1. Caractériser la *data literacy*

En passant en revue l'ensemble des définitions proposées, nous constatons l'emploi d'un certain nombre de concepts ou de pratiques : accès, appétence, aptitude, capacité, compétence, compréhension, connaissance, formation, habileté, réflexe, science et sensibilisation (intégrés dans la figure 6.3).

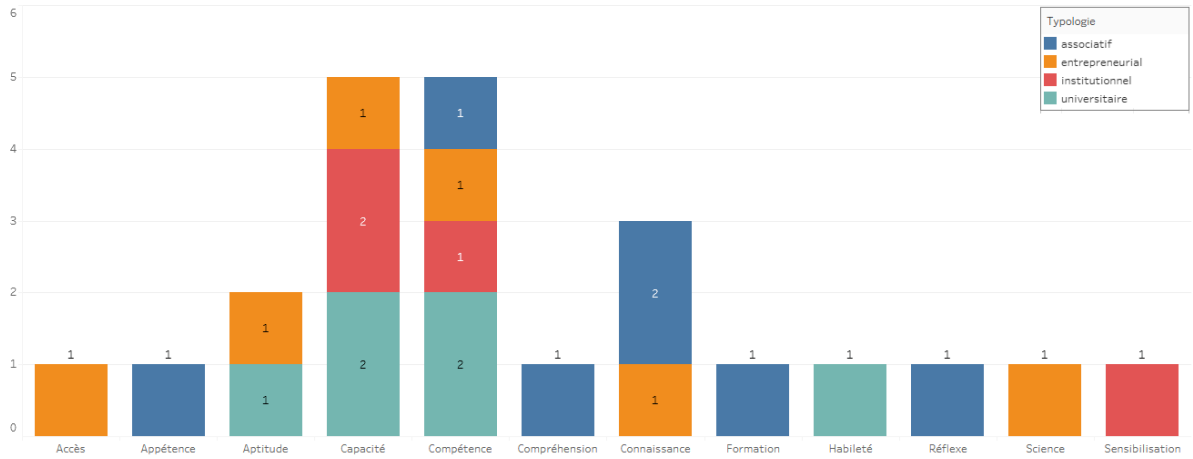

Figure 6.3.: Termes cités selon les typologies d'acteurs

Dans leur ensemble, les termes évoqués ne sont pas définis. La « compétence » et la « capacité » sont citées 5 fois chacune ; « compétence » étant le terme employé par toutes les typologies de publics. Les deux termes sont parfois employés invariablement et renvoient à une encapacitation des

publics formés, qu'il s'agisse de connaissances générales sur les enjeux des données ou d'un début de maîtrise technique comme le cite [Associatif 4] : « Il y a trois axes : comprendre, maîtriser, agir. Il y a comprendre où sont les données, ce qu'elles nous font au quotidien. Puis, être un peu plus capable de les utiliser ou de les mettre à distance. Enfin, savoir même les utiliser pour pouvoir finalement être force de proposition ». La finalité de la *data literacy* est de servir le citoyen, de faciliter son quotidien ; l'aspect professionnel n'intervient qu'au second plan.

> « Pour les élus, c'est de les sensibiliser pour les faire se rendre compte des données dont ils disposent et de la manière dont ils pourraient les utiliser pour améliorer les politiques publiques. Et par rapport au citoyen, c'est plus une acculturation au sens de lui permettre de se rendre compte aussi de comment peuvent être utilisées ses données personnelles à des fins d'intérêt général et de ne pas tout de suite penser à l'exploitation commerciale qu'on peut en faire. C'est donc essayer de montrer aussi les côtés positifs de partager ses données » [Institutionnel 4].

> « Faire différentes activités du quotidien, peu importe, en partant des données et en étant capable de les exploiter, de les interpréter et puis de s'en servir pour agir, pour faire des choses. Et la culture des données, je la vois comme quelque chose de plus large qui ferait partie d'une culture numérique au sens très large dans laquelle il y aurait à la fois des connaissances, des compétences, des savoirs mais aussi des pratiques qui sont à la fois personnelles mais aussi partagées avec d'autres individus. Et donc, la culture des données, cela serait tout ce qu'on est capable d'aller chercher, d'aller faire avec des données. C'est pour moi plus large, c'est une culture en termes de pratiques au quotidien » [Universitaire 2].

En lien direct avec les précédents points, une question générale a été posée, à savoir quelle pouvait être la « nature » de la *data literacy*. La nature, à défaut de terme plus pertinent, étant à comprendre ici comme l'ensemble des caractéristiques intrinsèques de la *data literacy* qui construisent ses représentations et permettent de l'identifier. Pour faciliter la compréhension de cette question, qui se posait en récapitulatif des échanges, nous avons proposé plusieurs exemples : une culture, un état d'esprit, un ensemble de techniques et de pratiques. Nous avons ajouté que cela pouvait recouvrir une partie ou l'ensemble des exemples, voire être totalement autre chose. Nous intégrons la possibilité que ces exemples aient influencé les visions des personnes interrogées et prendront donc un certain recul vis-à-vis des réponses apportées. Nous répertorions dans la figure 6.4 l'ensemble des dimensions et caractéristiques associées :

Les propositions faites lors des entretiens arrivent en tête du nombre de citations : la « culture » est l'aspect le plus sollicité (n=14), suivi de près par « l'ensemble de pratiques et de techniques » (n=12) et par « l'état d'esprit » (n=7).

Nature proposée	Occurrence
Culture	14
Ensemble de pratiques et de techniques	12
Etat d'esprit	7
Ensemble de compétences	3
Stratégie	2
Actions d'éducation	2
Environnement	1
Ensemble de méthodes	1
Ensemble de formations	1
Actions de sensibilisation	1

Figure 6.4.: Nature(s) de la data literacy

Chacune de ces dimensions ne sont pas nécessairement séparées entre elles ou ne résument pas à elles seules la *data literacy* : elles peuvent s'exprimer différemment selon les niveaux de formation déployés et selon les publics auxquelles elles s'adressent. Dans le cas du public institutionnel, la *data literacy* est d'abord une culture générale qui peut être étendue par la suite à des actes techniques. [Universitaire 7] distingue d'une part « le grand » public pour qui la *data literacy* se matérialisera sous une forme de culture, « d'intégration du raisonnement et de lecture des livrables de la data et de recours à la data pour s'orienter dans nos activités », et d'autre part les professionnels qui cherchent à monter en compétences sur le traitement des données.

De plus, les dimensions peuvent s'intégrer les unes dans les autres : la culture peut s'incarner dans des aspects pratiques et techniques (et donc ne pas être simplement incantatoire), l'état d'esprit peut découler de la stratégie et inversement, etc. Il n'y a aucun consensus sur le périmètre couvert, ni sur les interconnexions des termes ; les visions variant très sensiblement d'un individu à l'autre.

Un autre argument (n=2) soutient explicitement que la *data literacy* ne peut en aucun cas se réduire à des compétences techniques :

> « Je dirais que c'est une éducation civique à la donnée. A la donnée qui nous entoure, à la donnée que l'on produit en tant qu'être humain dans nos interactions sociales, dans nos interactions numériques, dans nos interactions avec la ville, avec notre environnement [...] et je dirais que c'est un peu tout ça à la fois parce qu'on ne peut pas se contenter de la data uniquement dans son aspect technique. L'aspect technique ce sont des outils, je peux les maîtriser : finalement, c'est une compétence technique. Mais si je me contente de ça, je ne comprends pas l'aspect politique, tout simplement. » [Entrepreneur 2].

En outre, le numérique est souvent couplé avec la *data literacy* perçue alors comme une littératie du numérique, intégrée à la « culture numérique » :

> « Il y a toute cette dimension aussi quand on parle de l'acculturation à la donnée celle du modèle économique [...] pour moi le numérique ce n'est pas technique, c'est politique ! Et ce sont les valeurs. Parce que derrière chaque outil qu'on utilise, c'est Jacques François Marchandise qui disait cela mais on embarque quelque chose, on embarque des valeurs, on embarque un modèle économique, on embarque plein de choses, et là, tout ce qu'on va mettre en place avec le département, c'est cela : la culture du numérique » [Institutionnel 7].

> « Et je pense que la culture des données, je le vois comme quelque chose de plus large qui ferait partie d'une culture numérique au sens très large dans laquelle il y aurait à la fois des connaissances, des compétences, des savoirs mais aussi des pratiques qui sont à la fois personnelles mais aussi partagées avec d'autres individus » [Universitaire 2].

Ainsi, en tant que culture (générale), la *data literacy* sensibilise les individus aux logiques et enjeux des données. Elle permet en ce sens le développement d'une posture critique vis-à-vis des réalités embarquées dans les données, ces faits bruts qu'il s'agit d'analyser pour extraire une richesse informationnelle. Une analyse enrichie par un ensemble de compétences techniques nécessaires pour un traitement approfondi des données.

6.3.2.2. Les visions sur les données

Comme précisé au début de ce chapitre, l'un de nos objectifs est de faire émerger les représentations des personnes interrogées autour des données pour ensuite déterminer si cela a une influence sur leurs pratiques. Pendant les entretiens, nous avons donc demandé aux acteurs de donner leur définition des données.

En premier lieu, les données sont perçues comme des matériaux bruts qu'il est nécessaire de retravailler pour en extraire la richesse. Des « enregistrements du réel » qui captent les phénomènes et sont une représentation particulière du monde. Néanmoins, si elles sont brutes, elles ne sont toutefois pas neutres comme le souligne [Entrepreneur 2] : « Si on ne comprend pas que la donnée est un objet politique au sens où elle est faite par des humains, enregistrée par des humains, calculée par des humains, et donc biaisée par des humains en fait, on ne comprend pas que la data n'est pas neutre. A partir du moment où l'on a compris ça, on ne la voit plus comme cet objet qui est juste en face de nous avec notre thermomètre : tout d'un coup, elle prend une autre dimension, donc on ne peut pas se contenter de l'aspect juste technique ».

Le lien entre donnée et information est principalement tissé par le public « universitaire ». Dans ce contexte, la donnée n'est pas un synonyme d'information car en elle-même elle n'a pas de sens, elle est « hyper sèche » [Universitaire 6] et nécessite d'être au préalable bien structurée avant d'être agrégée à d'autres données et former ainsi de l'information. A l'inverse, d'autres acteurs associent donnée et information, en particulier dans le cas des données informatiques et/ou numériques :

> « Pour moi une donnée, c'est un ensemble structuré d'informations. Sa structuration est souvent positionnée dans un outil qu'on appelle le tableur. Et cette structuration en lignes et en colonnes d'une information, entre guillemets, « propre », c'est-à-dire : dans une cellule, j'ai une information, qu'elle soit textuelle ou chiffrée, qui me permet d'appliquer un certain nombre de formules de calcul et de sortir de l'information une information autre, une information synthétique, de cet ensemble de lignes et de colonnes. Donc pour moi c'est basiquement ça : une information structurée et utilisable immédiatement pour produire une nouvelle couche d'informations synthétiques » [Entrepreneur 2].

> « Ça dépend si c'est une donnée informatique ou pas. Une donnée, quand on parle data en général, on utilise le terme anglais parce qu'une donnée c'est une information. Donc soit on parle effectivement du terme donnée = information publique et dans ce cas-là une information peut avoir tout type de forme, tout type de diffusion, de producteurs, etc. Après, quand on parle data, avec le prisme un peu « culture de la donnée », pour moi on est plus sur une donnée informatique au sens de données informatiquement réutilisables, donc avec un format classique, CSV, Json, geojson. Des formats réexploitables et assez ouverts » [Institutionnel 5].

Une donnée informatique et/ou numérique dont la spécificité est d'être un encodage d'information(s) compréhensible(s) seulement par les ordinateurs :

> « C'est une façon d'encoder l'information qui permet son traitement par des machines, par des algorithmes, par des ordinateurs, de façon programmatique, qui permettent des échanges d'informations entre des machines, mais qui n'est pas un format d'échange qui est fait pour des humains à la différence du langage ou de la photo qui est une façon d'encapsuler de l'information, de transmettre de l'information qui est plutôt fait pour des humains. » [Entrepreneur 9].

Le périmètre des données informatiques et/ou numériques peut par ailleurs dépasser celui des données et inclure l'intelligence artificielle et les algorithmes, soulignant une convergence des concepts.

> « Alors pour moi la donnée c'est à prendre en compte dans une acceptation plus large que la donnée de la base des données. En fait, la donnée c'est une matière brute dont la nature a beaucoup changé avec l'informatique [...] ce qui change par rapport à toute cette histoire de données, et de *data culture*, de *data literacy*, etc., c'est que la donnée, elle est devenue numérique, donc extrêmement fluide et toute l'ubiquité dont on entend parler par ailleurs, eh bien c'est la donnée qui reflète ça parce qu'au bout du compte, comme ça peut transiter très vite, tout le monde peut en avoir des usages dont certains sont complètement innovants. Et il faut arriver à considérer cela comme une matière première dans laquelle vont se loger d'autres réalités » [Entrepreneur 1].

Une distinction est ainsi faite entre la donnée en tant que concept et la donnée en tant que matériau nativement numérique. Mais également entre la donnée et l'information : ce point est ici variable, les deux termes pouvant recouvrir une même réalité. Dans l'ensemble des cas, nous retrouvons l'idée d'une captation de fait qui nous donne une certaine mesure de la réalité : la *data literacy* offrirait donc une vision du monde spécifique selon les données analysées.

6.3.2.3. Les interconnexions conceptuelles

Au-delà de ses caractéristiques définitionnelles et fonctionnelles, la *data literacy* est connectée à plusieurs concepts couvrant les pratiques de données : la gouvernance des données, la science des données, l'éthique des données et l'esprit critique vis-à-vis des données.

6.3.2.3.1. La gouvernance des données

La gouvernance des données est tour à tour désignée comme un « ordonnancement », une « organisation » ou une « politique » qui encadre la production et le partage de données de qualité. Elle est vue comme un vecteur de « confiance » dans les référentiels alimentant les données et ne doit donc pas être négligée. La *data literacy* joue un rôle de support, voire de catalyseur de la mise en place de la gouvernance grâce à une plus grande connaissance des données, de leurs traitements et de leurs enjeux permettant ainsi d'alimenter les réflexions. Comme « chantier », « brique », voire comme « déclinaison opérationnelle » de la gouvernance, elle s'inscrit dans la réflexion stratégique de long terme en étant pleinement intégrée à ses impératifs. A l'inverse, pour d'autres auteurs, elle n'est pas une pierre angulaire, elle est au contraire une surcouche, une « méta-enveloppe » qui intègre toutes les facettes des traitements de données. Néanmoins, quelle que soit son niveau d'intégration, la *data literacy* est perçue comme indispensable pour consolider la gouvernance: elle est un « élément complémentaire et facilitateur ». Elle est ce « socle commun », ce « minimum vital » culturel dont la transversalité permet de traiter l'ensemble des problématiques des usages des données.

6.3.2.3.2. La science des données

La science des données est perçue comme une science, une boite à outils (ce terme revient à plusieurs reprises) ou encore un éventail de techniques expertes employées pour traiter les données. Les réponses apportées sont ici plus tranchées que pour la gouvernance. Pour une majorité d'acteurs, la *data literacy* est un concept à part, décorrélée de la science des données et ce pour plusieurs raisons. La première porte sur l'expertise : les personnes formées à la science des données sont des experts. La seconde souligne les objectifs de massification des connaissances de la *data literacy*, ce qui n'est pas le cas pour la science des données. La troisième, plus spécifique, fait de la *data literacy* une littératie médiatrice et surplombante, permettant d'avoir une vision de l'écosystème des données. De fait, elle ne se trouve pas aux mêmes échelles que celles de la science des données ou de la gouvernance des données pour [Universitaire 4].

6.3.2.3.3. L'éthique des données

L'éthique des données est indissociable de la *data literacy* pour l'ensemble des personnes interrogées. C'est un code de conduite instauré pour évacuer les problématiques d'usages en proposant des données de qualité, un service fiable engageant la confiance des usagers :

> « De mon point de vue, l'éthique des données désignerait le code de conduite que doivent se fixer les professionnels qui produisent ou exploitent un jeu de données. Elle dépend donc de la technologie dont on parle, du contenu et de la vocation du jeu de données. 1 - la prise en compte et la formulation des biais que comporte une analyse de données : il y a là derrière cela de l'honnêteté intellectuelle (éviter de 'faire dire' aux données ce qui nous arrange, présenter toutes les informations aux destinataires pour qu'ils puissent juger des résultats) et de la vulgarisation. 2 - l'attention portée aux données personnelles lorsqu'elles sont en présence, et notamment à la transparence et à l'usage de ces données pour des finalités légitimes : ici, le simple respect du Règlement européen est suffisamment normatif et va même sans doute au-delà de ce que l'éthique intuitive attendrait. 3 - Protéger les données d'une fuite ou d'un mésusage : dès qu'elles sont confidentielles, sensibles, privées ou personnelles, les pratiques de sécurité font partie de l'éthique du jeu de données, autant qu'elles concourent au point précédent » [Entrepreneur 4].

Elle doit concerner l'ensemble des usages des données :

> « L'éthique des données doit être vue dans l'ensemble de la chaîne de traitement. Comment les données sont-elles acquises ? Quels croisement sont réalisés ? Quelles ques-

> tions sont posées dans les analyses et quelles décisions sont prises ? Toutes ces questions ont des réponses personnelles, individuelles, mais nous rappellent que la donnée est puissante. Les impacts seront positifs ou négatifs selon les intentions de leurs utilisateurs. La transparence imposée par le RGPD nous oblige à plus d'éthique car un traitement compris par les personnes concernées peut plus facilement être challengé. » [Entrepreneur 7].

Elle est dans certains cas un idéal à poursuivre qui dépasse les simples normes encadrant les pratiques :

> « On se rapproche beaucoup d'*Ekitia* [1] sur ce champs-là avec la Déclaration de Montréal. En fait, la question de l'éthique pour nous c'est la seule condition qui fera qui fera que nos habitants, nos citoyens, auront confiance. Ce n'est qu'à cette condition que l'on pourra devenir tiers de confiance. On l'a vu sur la petite opération que je décrivais tout à l'heure. L'éthique c'est le cadre qui permet de gérer le partage de la donnée. Si on ne l'a pas et si on ne la garantit pas, on ne fera rien. » [Institutionnel 7].

Le terme ainsi que son périmètre d'application peuvent toutefois être questionnés :

> « C'est un concept qui me parle mais que j'ai de mal à mettre en concret. Donc il y a un côté « protection des données à caractère personnel » ça c'est clair. Un côté « pas de décisions automatiques sur des individus directement ». Aussi, de la sécurité technique, de la cybersécurité. Ce n'est pas de l'éthique mais ça en fait partie : ça fait partie de l'intégrité, de la confidentialité. Il y a quelque chose qui pointe qui est la sobriété aussi mais qui va en contradiction avec le big data. Donc voilà, ça reste difficile à mettre en mots pour l'instant » [Institutionnel 6].

> « Je ne suis pas un grand fan du mot éthique. Parce que pour moi le mot éthique, il renvoie à la morale. Au final l'éthique est-ce qu'elle est générique, standardisée ? pour toutes les personnes, toutes les cultures et pour tous les modes de fonctionnement ? Je ne sais pas. Ça se trouve on a chacun notre éthique et notre mode de fonctionnement. Et on trouve que l'éthique de l'autre n'est pas la même et pas la bonne donc qu'est-ce que cela signifie ? qu'il faut standardiser et du coup la conformer pour créer un cadre. Ou qui dit qu'on n'a pas le droit d'aller-là et on fait ça. C'est un peu comme la morale, chacun a un peu la sienne. » [Universitaire 9].

[1] Ekitia est une « association de préfiguration d'un pôle d'économie de la donnée » : https://www.ekitia.fr/une-association-de-prefiguration/

6.3.2.3.4. L'esprit critique vis-à-vis des usages des données

Le dernier concept associé à la *data literacy* est l'esprit critique vis-à-vis des données. Les personnes interrogées à ce sujet évoquent plusieurs facettes que nous avons ici synthétisées dans le tableau 6.3:

Tableau 6.3.: Facettes de l'esprit critique vis-à-vis des données

Facettes	**Sous-facettes**
Avoir une capacité de recul vis-à-vis	du potentiel des données (peuvent-elles répondre aux demandes ? Ont-elles un potentiel néfaste ?)
	de l'information qui va découler des données
	des logiciels employés (qui les a créés ? Dans quel but ? Comment sont effectuées les tâches ? Quelles sont leurs limites ?)
	des limites et des biais des données
Connaître les besoins en données	quelles données sont-elles nécessaires pour répondre à ma problématique ?
	de quelles données disposons-nous ?
	comment obtenir les données manquantes ?
	les données récupérées / produites sont-elles pertinentes et/ou de qualité ?
Savoir recontextualiser les données	comment ont t-elles été collectées / produites / partagées ?
	qui les a collecté / produit / partagé ?
	où ont-elles été collectées / produites / partagées ?
	dans quel contexte ont-elles été collectées / produites / partagées ?
	pourquoi ont-elles été collectées / produites / partagées ? Quelle est l'intentionnalité ?

Il s'agit d'adopter un recul, voire une « posture » vis-à-vis des usages des données, de comprendre son cycle de vie et d'être conscient qu'un ensemble d'intervenants a pris part à leur élaboration, qu'une intentionnalité est à l'œuvre et est porteuse de certains biais qui impactent les usages. Il s'agit de démystifier les données, de leur ôter leur caractère de « vérité absolue ». C'est avoir conscience de leur potentiel et de leurs limites pour répondre à des problématiques spécifiques.

> « Forcément, une analyse de données, une exploitation de données, c'est la rencontre entre quelque d'objectif et quelque chose de subjectif. Il y a une subjectivité dans l'action d'interroger des données. On va faire un certain nombre de choix méthodologiques, de parti pris. Mais ça c'est propre à toutes les sciences. Une expérience scientifique elle se fait sur la base d'hypothèses, elle se fait sur la base de parti pris et elle conduit à un résultat qui pour autant est quand même scientifique. Mais l'esprit critique, c'est justement de se dire : cette expérience, il va falloir qu'elle soit reproductible. Il va falloir qu'on la réitère et c'est le fait qu'elle puisse être reproduite et réitérée et qu'elle puisse être éventuellement réfutée qui va faire qu'on est bien dans une discipline scientifique.» [Entrepreneur 4].

La notion fait toutefois l'objet de critiques. Soit dans sa dénomination qui est vue comme peu pertinente, soit dans son emploi actuel qui n'apporte pas de réflexions nouvelles ou encore sur le fait que les critiques sont trop centrées sur les données et non pas sur leurs usages. Pour [Associatif 3], le terme n'a tout simplement pas de sens :

> « Alors je ne sais pas ce que c'est être critique envers la donnée. J'ai l'impression que les données ont toujours été question à critiques : il faut questionner ce que sont les données. Mais je crois qu'au-delà des données elles-mêmes, aujourd'hui, peut-être que ce qu'on peut questionner ce sont les usages justement autour des données. Mais de toute manière, les données sont des objets à qui, en tant qu'utilisateurs, on donne un sens à travers les usages qu'on en fait. Donc finalement je ne sais pas à quel point c'est une critique envers les données ».

La place de la *data literacy* dans l'écosystème est donc variable : extérieure à la science des données dont elle ne serait au mieux qu'une première étape, support (intégré ou non) de la gouvernance des données, elle alterne entre une position située à l'intersection de ces deux concepts, un sous-domaine et une vision englobante. Un consensus n'est pas ici trouvé. En revanche, l'éthique et l'esprit critique, majoritairement sollicités, doivent être intégrés aux caractéristiques de la *data literacy* pour dépasser un simple statut technique et répondre à des nécessités démocratiques et citoyennes.

6.3.3. Une formation en *data literacy*

Nous venons ici de présenter les dimensions de la *data literacy*. Nous nous interrogerons maintenant sur ce que recouvre une formation en données et les moyens mis en œuvre pour son implémentation.

6.3.3.1. Pourquoi former ?

Cette volonté de formation est le fait de nombreux enjeux liés aux usages des données dont la prise de conscience générale découle d'une succession d'évènements marquants cités par les acteurs :

- Les discours mettant en avant des pratiques de données liés au *big data* et à l'*open data* ;
- La massification des données et la démocratisation de leur accès ;
- Les scandales autour des traitements des données : l'affaire *Schrems* (2013), les révélations d'Edward Snowden (2013) et l'affaire *Cambridge Analytica* (2018) ;
- La promulgation de nouvelles réglementations : la loi n°78-753 et la loi Informatique et Libertés (1978), la loi *Inspire* (2007), la loi pour une République Numérique (2016), le Règlement sur la protection des données (2018), le *Data Governance Act* (2023) et le *Data Act* (non publiés au moment de la rédaction de la thèse) ;
- La pandémie de Covid-19 et la médiatisation des exploitations de données (en particulier les visualisations graphiques des données) ;
- L'irruption des logiciels d'intelligences artificielles génératives comme *ChatGPT* ou *DALL-E* (ce point concerne les entretiens à partir d'octobre 2022; date de généralisation de ces objets dans les discours).

Il s'agit d'une (re)découverte constante non seulement du potentiel économique des données dont s'emparent rapidement les entreprises, mais également des enjeux démocratiques et citoyens vis-à-vis des connaissances des usages réalisés grâce aux données.

> « Alors je l'ai rencontré très tôt, bien avant l'open data avec les premiers systèmes de *business intelligence*. Je travaille depuis les années 1990. Dans les années 90, déjà 2000, j'étais sur des systèmes décisionnels, des processus où la donnée était un moyen d'optimisation, de performance. Par contre c'était très spécialisé et très fermé : on était à la fois sur des segments d'activité très spécifiques qui étaient bien souvent le financier, la compta, les RH. C'est le pilotage par le chiffre, le chiffre en lui-même. Avant il n'y avait pas de pilotage par le chiffre. Depuis dix, quinze ans, ça a envahit tout l'espace. On se dit aujourd'hui : "il faut trouver de la donnée". La culture de tout associer à de la donnée est vraiment très récente je trouve. Parce qu'il y a encore cinq à sept ans, on n'en était encore qu'à des processus de gestion. [...] Donc on a un inversement : c'est un renversement de tendance depuis l'open data. Et il y a aussi tout ce qu'il s'est passé

> à côté de l'open data, à savoir le big data et les GAFA. C'est la prise de conscience de la donnée, même si on ne sait pas trop encore ce que ça représente, ce que ça signifie, quels sont les enjeux. C'est présent et on en fait un sujet. »[Entrepreneur 8].

Un point important, souligné par [Universitaire 7], porte sur la teneur politique de la *data literacy* et du rôle des statistiques dans ce contexte :

> « On a vécu dans l'ère des discours et on continue à y vivre bien sûr, mais je pense que c'est une forme de retour à l'objectivation et à des données statistiques et factuelles qui quelque part contrebalancent des prises de positions parfois très idéologiques. Je pense qu'on a un balancier qui revient au centre et qui tient de l'idée qu'on peut de plus en plus avoir recours à la *data literacy* pour cadrer, pour gouverner, parce qu'on a de plus en plus de données. Et je pense qu'il faut les mettre en balance avec des raisonnements plus qualitatifs et politiques. Donc je pense que c'est ce mouvement global et massif du retour des données statistiques et factuelles qui marque la *data literacy* dans son ensemble. [...] Je pense que la politique publique sur tout ce qui est *open data*, tout ce qui relève des données de la recherche, tout ce qui est transparence, ce sont les moments forts qui ont fait qu'on peut dire aujourd'hui que c'est peut-être à partir de l'émergence de ces mouvements-là que la data est redevenue très importante. Ainsi que le problème de la souveraineté de la data. Je pense que ça date d'une petite dizaine d'années à peu près [...] je pense que la technologie fait que les sujets deviennent de plus en plus importants. La technologie est de plus en plus mûre pour traiter ces données, les stocker. De nouvelles architectures sont là. [...] Je pense que les forces qui ont consolidé la *data literacy* viennent du politique, viennent de l'économie, de la techno. Dans les sociétés numériques, le « capital matériel » entre guillemets est quand même maintenant aussi important pour avoir l'avantage concurrentiel. Les entreprises du numérique et leur importance aujourd'hui ont accompagné aussi ce poids de la data. Selon moi, cela ne naît pas de rien : cela naît de forces conjoncturelles fortes qu'il faut trouver au croisement de l'économie, du politique et de la maturité des technos.» [Universitaire 7].

Au-delà des points cités, l'aspect perceptible de cette nécessité se retrouve également dans les changements de pratiques des organisations. Des changements communicationnels révélateurs d'une évolution paradigmatique :

> « L'évolution du nom des directions ! Il y a quatre, cinq ans, on parlait de la direction du digital et [nom d'organisation] est un bon exemple pour ça. Il y a quatre, cinq ans quand ils ont débuté la transformation digitale, la direction s'appelait la direction du digital. C'est devenu il y a trois ans la direction de la data et du digital. Et maintenant c'est devenu, je ne sais plus dans quel ordre je t'avoue, mais je me souviens que l'ordre

était stratégique : data, digital et innovation. Il y a quatre, cinq ans c'est passé du digital à la data. ». [Entrepreneur 3].

« On a beaucoup de sollicitations là-dessus parce que c'est le terme même de donnée/data qui est au cœur de l'économie d'aujourd'hui, au cœur des réflexions sur le numérique aujourd'hui donc effectivement on est passé de la culture numérique à la culture des données. C'est un effet de la mode en quelque sorte mais aussi une évolution de nos sociétés puisqu'en même temps on constate qu'on exploite davantage des données qu'avant et on a des moyens beaucoup plus poussés aussi pour les exploiter ». [Universitaire 4].

C'est tout un ensemble de discours qui se succèdent et qui amènent l'ensemble des acteurs à faire de la donnée un enjeu central :

« Alors si on considère la data au sens de ses réceptions, il y a toujours un enjeu aujourd'hui. Un enjeu fort qui a été celui de la transparence, de la confiance, de l'*open data*. Ça c'est clair et c'est fort. C'est le contexte des politiques publiques : la data porte quand même de gros enjeux de transparence, de confiance et de lutte contre tout ce qui s'apparente à la fraude. Pour l'entreprise, la valeur c'est la performance. Il ne faut pas se leurrer, l'entreprise est toujours dans des logiques de gain de productivité et l'enjeu principal c'est la performance dans les entreprises, que ce soit la *BI*, que ce soit le *reporting* quotidien, pour développer de nouveaux marchés, de nouveaux segments, etc. Pour les scientifiques, pour les gens qui sont plus dans le savoir, c'est l'intégrité, c'est la qualité de la preuve, c'est la vie scientifique. On a donc trois registres un peu différents mais forcément encore une fois qui se greffent par rapport à ce qui anime profondément et prioritairement ces communautés-là, ces objectifs-là en tous cas. Donc après, du point de vue médiatique, je pense que pour nos sociétés, l'enjeu de la data c'est la performativité du message que tu transmets. Elle a quelque chose de fort dans la persuasion, dans la performativité du message, dans les arènes publiques, la sphère publique. En infocom c'est important ça, d'un côté média. » [Universitaire 7].

6.3.3.1.1. Les grands enjeux autour des données

Les données sont au cœur de grands enjeux auxquels la *data literacy* doit permettre de répondre. Ces enjeux cités ne sont pas corrélés avec les typologies de rattachement des personnes interrogées : il s'agit ici d'avis personnels mûris par l'expérience.

Nous avons décelé cinq grands types d'enjeux qui concernent tout aussi bien le niveau micro (l'individu) que le niveau macro (l'ensemble des structures soutenant les prises de décisions), à savoir les enjeux techniques, logistiques, éthiques, économiques et citoyens (voir le tableau 6.4).

Tableau 6.4.: Les types d'enjeux autour des données

Type d'enjeux	**Exemples cités**
Techniques	connaître son patrimoine de données ; Disposer de données de qualité, de références, fiables ; Produire des données ; Traiter et exploiter des données ; « Mixer » les données au sens d'une utilisation connexe de plusieurs jeux de données ; Interpréter les données ; Savoir réguler les flux de données ; Permettre l'interopérabilité des données en vue de leurs réutilisations ;
Logistiques	Intégrer une gouvernance des données ; Encourager la souveraineté des jeux de données ; Donner le pilotage de la donnée aux métiers traitant des données ;
Éthiques	Instaurer la transparence des données et des algorithmes ;
Économiques	Développer des modèles économiques basés sur les données ouvertes ; Briser le monopole des GAFAM sur les données ;
Citoyens	Permettre une meilleure connaissance de notre environnement pour prendre des décisions éclairées (ex : les programmes politiques) ; Mettre le citoyen au centre des stratégies de décisions ; Permettre au citoyen d'avoir le contrôle de ses données (dans son accès, sa production, son appropriation et ses réutilisations) ; Généraliser les formations aux données dans un souci démocratique ;

	Former le citoyen à l'ensemble des enjeux (numériques, économiques et politiques) des données ;

Le premier type d'enjeux recouvre l'ensemble des techniques nécessaires dans l'obtention et le traitement des données afin d'en tirer des avantages, en particulier la production de nouveaux services. Comme vu précédemment, l'ensemble du cycle de vie de la donnée est concerné : les publics doivent être formés pour traiter chaque étape dudit processus (la production, la collecte, l'accès, le traitement, les réutilisations, etc.).

Suivent les enjeux logistiques qui sont directement dépendants des enjeux techniques : ils doivent permettre l'implémentation d'une gouvernance permettant de soutenir de façon pérenne les pratiques raisonnées et respectueuses de traitement des données, normées par l'éthique des données.

Le troisième type d'enjeux centrés sur l'éthique encadre les enjeux techniques et logistiques et recouvre la transparence des données et des algorithmes, la frugalité des données et le respect des réglementations (en particulier le *RGPD*). Il s'agit donc d'une interaction entre des enjeux écologiques, juridiques et citoyens.

Dans le quatrième type d'enjeux, les enjeux économiques, une volonté forte se dégage des entretiens : celle de développer un modèle économique opposé à la « logique des GAFAM » qui doit permettre d'instaurer une certaine souveraineté et de développer une monétisation des données ouvertes. Sur ce dernier point, il faut noter que le *Data Governance Act*, qui permet justement l'instauration de redevances sur les réutilisations des données, n'existait pas encore au moment des entretiens (il a été mis en application en septembre 2023).

Les derniers enjeux concernent les citoyens. Il est question d'une encapacitation des citoyens, conscients des enjeux des données, capables de comprendre les flux des données et de les maîtriser afin de prendre des décisions éclairées.

L'ensemble de ces enjeux sont éminemment politiques, nécessitant des décisions fortes pour être résolus.

6.3.3.1.2. Les types de données au cœur des enjeux

Au cœur de l'ensemble de ces enjeux se trouvent un certain nombre de données qu'il s'agit de traiter en priorité (voir le tableau 6.5).

Tableau 6.5.: Les types de données cités

Type de donnée cité	Occurrence
Donnée à caractère personnel	13
Donnée publique	3
Donnée de santé	3
Donnée privée	2
Donnée de recherche	2
Donnée de géolocalisation / mobilité	2
Donnée urbaine	1
Donnée ouverte	1
Donnée métier	1
Donnée financière	1
Donnée environnementale	1
Donnée de vente immobilière	1
Donnée de consommation énergétique	1
Donnée d'imagerie spatiale	1
Donnée d'entreprise	1
Donnée bancaire	1

Les données à caractère personnel sont les données les plus citées (n=13). Cette préoccupation peut être expliquée en grande partie par l'instauration du *RGPD* qui encadre leurs usages. Suivent bien après les données publiques et les données de santé (respectivement n=3), les données privées, de recherche et de mobilité (respectivement n=2).

Selon la figure 6.5, les données à caractère personnel sont citées par la grande majorité des acteurs et ce quelle que soit la typologie à laquelle ils appartiennent. C'est un axe de réflexion et d'appréhension central qui recouvre l'ensemble des enjeux cités précédemment :

> « Eh bien, il y a cette histoire de données personnelles qui est très sensible dans le milieu de l'éducation parce que les profs ont réalisé assez récemment qu'ils sont entrain de mobiliser, voire de mettre en circulation les données personnelles, les élèves sont sensibilisés à leurs propres données personnelles, avec des outils comme Pronote par exemple. Je suis presque certain(e) que maintenant il y a très peu d'enseignants qui réalisent que Pronote ça a longtemps était un outil totalement privé à travers lequel on héberge des données personnelles d'élèves pas forcément en France, etc. Et c'est finalement assez récemment qu'on s'est dit qu'il fallait peut-être retrouver la maîtrise française, la souveraineté sur ces données personnelles des élèves. » [Universitaire 1].

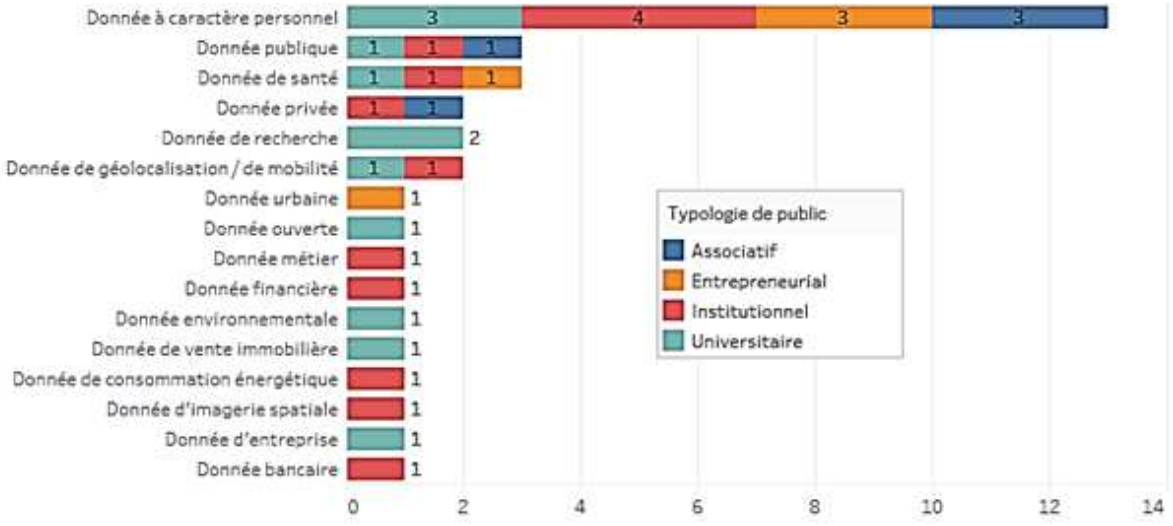

Figure 6.5.: Liste des données citées selon les typologies de publics

Les données publiques et les données de santé sont également très prisées par une multiplicité de publics en raison de la possibilité de leur appropriation par des entités tierces. Les dernières données évoquées quant à elles correspondent à des pratiques spécifiques liées aux contextes professionnels des personnes interrogées et ne font pas l'objet d'enjeux globaux.

6.3.3.1.3. Les réglementations sur les usages des données

Nous venons de présenter un certain nombre de données qui font l'objet de réglementations spécifiques. Nous supposons que le cadre réglementaire influe de ce fait sur leurs manipulations, et par extension sur la résolution des enjeux précédemment présentés, c'est pourquoi nous avons interrogé les participant(e)s à ce sujet (voir le tableau 6.6).

Tableau 6.6.: Les réglementations citées

Réglementations	**Occurrences**
RGPD (2018)	28
Loi pour une République numérique (2016)	12
loi n°78-753 (1978)	5
Loi Informatique et libertés (1978)	1
Data act (2023)	1

Le Règlement général pour la protection des données est la réglementation la plus citée (n=28). Sa réception est très variable selon les acteurs. Il est reconnu pour avoir permis une prise de conscience

des problématiques liées aux usages des données. En revanche, sa réception par les organisations est perçue très différemment. D'un point de vue positif, le *RGPD* est présenté comme un catalyseur de prise de conscience sur les enjeux des données. Pour [Institutionnel 3], il permet « d'impulser le changement et de mobiliser les troupes ». D'un point de vue négatif, il limite les pratiques et s'avère contre-productif vis-à-vis de ses objectifs. Selon [Institutionnel 1], il réduit tout à un aspect purement sécuritaire, « la seule entrée qui comptait ». Par ailleurs, [Institutionnel 6] note que le *RGPD* tend à ne proposer qu'une vision unique centrée autour des données à caractère personnel. Il est de fait vécu pour [Associatif 7], non pas comme « une acculturation désirée, désirable, souhaitable » mais subie et dont les effets attendus ne sont pas perceptibles. Un entre-deux est néanmoins présent chez certains acteurs qui trouve la réglementation nécessaire, utile, mais qui doit encore être améliorée, notamment dans la gestion des *cookies* :

> « Je travaille à mettre en conformité mon client sur le *RGPD* donc je le sens beaucoup. Mon avis là-dessus c'est qu'il y a une intention assez intéressante de poser un cadre de travail et de collaboration vis-à-vis de l'exploitation des données. Un avis qui est plus personnel c'est que parfois, à vouloir avoir une bonne intention de protection, en réalité on va dégrader l'expérience des législateurs, et là je prends un exemple un peu à part avec la gestion des cookies : je trouve que c'est infernal que dès que l'on va sur un site de devoir accepter ou refuser les cookies. On a des normes européennes qui imposent des pratiques, très bien, ce n'est pas pensé UIC, c'est pensé protection. Et c'est ce que je regrette beaucoup c'est que finalement on ne protège pas l'individu s'il accepte systématiquement pour être tranquille. On n'a pas éduqué finalement sur ce sujet. » [Entrepreneur 7].

La seconde réglementation la plus citée est la loi pour une République numérique (n=12), dite « loi Lemaire ». Elle est jugée selon les mêmes critères que le *RGPD*, à savoir sur son utilité, la portée de ses actions et sa réception par les organisations :

> « Pour moi, la Loi Lemaire aujourd'hui est dépassée. Alors c'était très beau dans son concept de tout ouvrir, mais après pour des questions de chronophage, de priorité, de sobriété, on ne peut pas tout ouvrir tout le temps et plutôt ouvrons quand on sait que cela peut être utile et que derrière il y a du service qui sera proposé ou de la transparence publique, cela se réfléchit, il faut une stratégie, on ne part pas bille en tête en sortant tous les jeux de données possibles.» [Institutionnel 7].

> « Sur la loi pour une République Numérique, je pense que le bilan aujourd'hui est beaucoup plus mitigé. Disons que cela a été un outil de prise de conscience, d'acculturation, de pédagogie. Je ne pense pas que d'un point de vue juridique on puisse considérer, en tout cas sur le sujet de l'ouverture des données, que cela a été un outil contraignant et

> opérationnel. En tout cas, ça n'a pas été perçu comme ça par les acteurs. En 2022 aujourd'hui, on a encore l'essentiel des organisations publiques pour qui oui l'*open data* c'est vaguement inscrit dans la loi mais bon à la fin des fins on fait bien comme on veut.. [...] Je pense que ce qu'on peut dire c'est que la loi pour une République Numérique illustre le fait, et je pense que c'est largement partagé, qu'il ne suffit pas d'écrire les choses dans la loi pour qu'elles se traduisent en réalité et donc soit on est dans des approches qui sont purement incitatives où finalement on compte sur l'alignement de certains intérêts ou certaines incitations pour que les choses se passent.» [Associatif 6].

Nous retrouvons à nouveau dans les discours l'aspect positif d'une prise de conscience impulsée par la réglementation, contrebalancé toutefois par une atteinte limitée des objectifs d'ouverture.

Un rappel de la loi n°78-753 (qui a instauré la *CADA*) et de la « loi Informatique et Libertés » (née conjointement avec la *CNIL*) est effectué pour souligner leur rôle fondateur dans l'accès et la gestion des données (administratives pour la loi n°78-753, personnelles pour la loi Informatique et Libertés). Le *Data Act* enfin est cité comme réglementation à suivre sans plus de précisions.

La pertinence de la *data literacy* trouve donc un écho dans la résolution de grands enjeux qui imprègnent la société dans chacune de ses strates et impactent le quotidien des individus. Une formation doit s'effectuer pour impulser une conscientisation de ces derniers afin de développer les attitudes et les compétences nécessaires à cette résolution.

6.3.3.2. Qui former ?

Si les problématiques liées aux données peuvent en théorie s'appliquer à l'ensemble des individus, une question demeure sur leur formation :

> « Est-ce que ça doit concerner tout le monde ? Oui. Est-ce que ça doit concerner tout le monde de la même manière ? Non. Je pense qu'il y a différents niveaux en termes de strates à toucher. Et dans ces strates, différents messages à faire passer. » [Institutionnel 3].

La formation de l'ensemble des publics est partagée par une majorité de personnes interrogées et ce quelle que soit leur typologie :

> « Ça concerne tout le monde mais c'est comme la culture générale : tout le monde en a besoin mais ça ne peut pas être équitablement réparti. Donc évidemment il y a des

> experts et des profanes mais il faut un socle commun, un minimum commun qui reste à définir mais qui est probablement insuffisant à l'heure actuelle. » [Associatif 5].

> « Oui, j'ai la conviction que ça doit concerner tout le monde. J'ai cette conviction et elle est partagée je pense par mon organisation. On dit régulièrement que cela ne doit pas être un sujet réservé à des codirs, des conex, aux élus, etc. : c'est vraiment un sujet qui concerne tout le monde et ça se comprend aisément parce qu'en fait les données sont souvent saisies, manipulées par des agents de terrain, des agents de guichet, etc. Et si à eux on ne leur explique pas l'importance du sujet : pourquoi ce qu'ils font est important, comment ensuite ça a des répercussions ailleurs, etc, on a aucune chance d'avoir des données de qualité, exploitables à la fin. Donc même d'un point de vue très utilitariste il y a un intérêt à ce que ça concerne tout le monde. » [Associatif 6].

Certains acteurs estiment cependant que la formation aux usages des données n'est pas nécessaire pour l'ensemble des individus : une distinction est en effet faite entre les citoyens, qui doivent tous avoir un socle minimal de connaissances vis-à-vis des données, et les salariés/employés qui n'ont pas tous besoin d'être formés ; tout va dépendre du poste qu'ils occupent. À cela s'ajoutent des logiques pragmatiques de financement, pointées par certains acteurs pour justifier ce point :

> « Après la question c'est : est-ce que cela doit être réalisé pour tout le monde ? Pas forcément pour tout le monde, en tout cas pas à la même échelle. Et là je suis vraiment dans les entreprises, je ne me place pas au niveau citoyen : je ne ne suis pas certain qu'un cariste ait besoin du même niveau d'information sur la donnée qu'un gestionnaire de paie qui saisit dans un applicatif et qui a un rôle très important dans la qualité de la donnée. Que le cariste sache ce que l'on fait de ces données là on est plus dans l'ordre de l'information versus, dans un autre cas, plutôt de la formation. Différentes échelles mais une action de *data literacy* qui peut être intéressante dans les deux cas. Après au niveau citoyen je pense qu'on doit tous avoir un socle de connaissances au niveau citoyen tout simplement parce que l'on a des données qui ont une valeur que l'on sous-estime radicalement, et aujourd'hui très peu de gens ont conscience de ce qui est fait de leurs données par les *GAFAM* au quotidien. » [Entrepreneur 7].

En ce sens, les personnes détentrices des décisions, comme les élus dans le cas des collectivités territoriales, doivent être sensibilisées pour que des actions concrètes aient lieu.

Une opinion est néanmoins complètement partagée, celle qui souligne le fait que les formations ne doivent pas être réservées à un public d'experts techniques, à des « élites » seules détentrices de ces savoirs. De même, l'objectif n'est pas de former des experts. Il s'agit de former les individus pour qu'ils

aient suffisamment de connaissances pour savoir utiliser les données et ainsi répondre aux problèmes du quotidien.

6.3.3.3. A quoi former ?

En reprenant l'ensemble des enjeux évoqués précédemment, nous pouvons résumer ici les éléments pivots d'une formation en *data literacy*.

6.3.3.3.1. Connaître les caractéristiques et enjeux des données

Le premier attendu d'une formation en *data literacy* est de posséder des connaissances générales sur ce que recouvrent les données, qu'il s'agisse de leurs caractéristiques, de leurs enjeux et des bénéfices qu'elles peuvent apporter :

> « En priorité une acculturation à la donnée, d'être capable d'avoir ce schéma de pensée, de dire comment fonctionne la donnée, c'est cette pensée de croisement de données pour extraire de l'information, c'est vraiment une démarche d'analyse quantitative qui permet de dire "ok j'ai une donnée qui donne une information. J'ai d'autres données qui donnent d'autres informations. Si je les mets ensemble, je produis de la valeur. Je produis quelque chose d'intéressant. » [Entrepreneur 7].

> « Alors moi je dirais : peut-être dans un premier temps comprendre ce qu'est une donnée parce qu'au final cela peut être tout et n'importe quoi. Et du coup, de bien de faire la distinction entre les données à caractère personnel mais aussi des données « autres » : des données géographiques, des données énergétiques, des données industrielles... tout ça, c'est de la donnée. Et ce ne sont pas les mêmes enjeux que cela soulève d'utiliser telle ou telle sorte de donnée. C'est d'abord de prendre conscience des différentes catégories de données qui existent, je pense. Et aussi de savoir, pour chaque catégorie, quels sont les enjeux que cela soulève. Quelles sont les limites à ne pas dépasser par rapport aux valeurs et à notre législation aussi. Et d'un autre côté, de montrer en quoi l'exploitation de cette donnée-là peut servir l'intérêt général » [Institutionnel 4].

6.3.3.3.2. Une certaine connaissance du cadre réglementaire

Comme le précise l'adage « nul n'est censé ignorer la loi », la connaissance du cadre réglementaire est nécessaire. Sa connaissance spécifie d'une part les moyens pour appliquer de « bonnes pratiques »

en termes d'usages des données et permet d'autre part de consolider les connaissances préalables sur ce que recouvrent les données :

> « Au minimum, je pense que tout le monde doit être au fait du cadre légal actuel. Qu'est-ce que c'est qu'une donnée personnelle, qu'est-ce qu'on a le droit de demander et de pas demander, et qu'est-ce qu'on a le droit de stocker, de ne pas stocker, et c'est déjà beaucoup ». [Associatif 5].

6.3.3.3.3. Le développement de compétences techniques

Nous avons vu que le *data literate* n'a pas vocation à devenir un expert technique. Toutefois, une certaine maîtrise technique peut être attendue et est un socle utile pour approfondir les analyses des données :

> « Et puis après ce sont des formations, des compétences techniques qui permettent d'utiliser la donnée, de vraiment agir sur la donnée, de la valoriser. Mais aussi se rendre compte que la donnée n'est pas sans conséquence : ça a quand même un impact sur la société de tout numériser et d'avoir ces usages. C'est important de relativiser, de ne pas tout miser là-dessus. De comprendre l'impact qu'a la donnée et son utilisation sur notre société, sur notre environnement. Et finalement de se poser la question de ces pratiques » [Institutionnel 4].

6.3.3.3.4. Le développement d'un questionnement critique et la quête de l'autonomie

Au-delà de développer les connaissances sur l'aspect réglementaire des données, sur leurs enjeux et de former à certaines compétences techniques, la *data literacy* doit insuffler de nouveaux schémas de pensée chez les individus afin qu'ils gagnent en autonomie :

> « Donc de pas être dans un sentiment et dans une relation de domination par rapport aux données et de se dire « ce n'est pas *Google* qui a la réponse, c'est bien moi qui a la réponse ». Oui, *Google* a plein de réponses mais je dois toujours être lucide quand j'ai affaire à une donnée. De savoir qui la produit, dans quelles intentions, de me poser les bonnes questions pour savoir quelles sont les sources, qu'est-ce qu'on essaye de me faire comprendre, de bien être au cœur de tout cela. Et pour moi la *data literacy* donne des clés, elle donne des outils pour pouvoir être autonome, pour pouvoir le faire. Mais il faut un accompagnement assez structuré pour que nos interlocuteurs puissent être autonomes. [...] Donc la culture des données, elle doit surtout aider les plus fragiles à pouvoir se défendre sur ces facettes-là. » [Associatif 1].

C'est développer une réflexion critique permettant d'identifier les « systèmes d'intention » incorporés dans les données, de relever les caractéristiques propres à toute technique.

> « Être capable justement de les comparer et d'être capable de repérer les systèmes d'intention. La genèse des données mises à disposition et finalement comment elles ont été agrégées, mise à disposition. » [Universitaire 6].

> « C'est de penser les données dans les espaces dans lesquels elles sont produites et donc de ne pas strictement réduire les données à des questions d'informatique. Je pense que c'est le deuxième aspect central. C'est d'être dans une réflexion qui permet d'inscrire les données en lien avec l'éducation aux médias et à l'information d'une part, et d'autre part de le raccrocher à tout un ensemble d'aspects techniques, la dimension, la place des technologies, leur rôle et leurs conditions de manipulation. En tout cas c'est ce mouvement entre *EMI* et la technicité, les gestes techniques qui sont liés à l'usage de ces deux dispositifs d'instruments liés aux données, au traitement de données. » [Universitaire 3].

6.3.3.4. Comment former ?

6.3.3.4.1. Parler de tout sauf des données

Le premier conseil partagé par les acteurs est la nécessité de « parler de tout sauf des données » : cette évacuation du sujet central d'une formation en *data literacy* vise à désamorcer le possible désintérêt pour ce type de formation découlant d'une incompréhension de ce que recouvrent les données. Pour servir cet objectif, il est nécessaire en premier lieu de rendre limpide la technicité des données, quitte à évacuer cet aspect au début des formations :

> « Pas de compétences techniques mais je dirais une appropriation de concepts techniques. Parce que quand on parle de la donnée on ne peut pas ne pas parler des concepts techniques. Et je pense qu'une bonne acculturation c'est une acculturation qui arrive à partager ces concepts techniques de manière super simple comme-ci vous les expliquiez à un enfant » [Entrepreneur 5].

> « J'essaye toujours d'aborder les questions sous un angle usages, services, et pas du tout technique [...] j'ai toujours facilité l'open data sans parler d'open data, en ne parlant que services et utilité.» [Institutionnel 7].

« Pour nous, du point de vue du pédagogue, c'était une culture et une mise en pratique, forcément. Le plus possible on se disait que tant qu'on parlait des données c'était du brouhaha. C'est le fait de pouvoir donner à voir, que les gens puissent prendre, puissent manipuler, puissent changer. On sortait le plus possible de l'ordinateur pour qu'on puisse avoir d'autres appréhensions, les plus sensorielles possibles, par rapport aux données afin que les gens puissent se les approprier.» [Associatif 1].

6.3.3.4.2. Etre « pratique »

Il faut être « pratique », c'est-à-dire ne pas venir avec des idées générales mais au contraire partir des pratiques quotidiennes des publics pour opérer une recontextualisation et ainsi faciliter l'imprégnation du bénéfice premier découlant des usages des données.

« Ça peut prendre plein de formes et ça agrège plein de choses. Tu as la partie vocabulaire. Tu as la partie « les cas d'usage ». Qu'est-ce que tu vas pouvoir faire de la donnée ? Le fait de comprendre tous les usages qu'il va y avoir derrière, ça fait partie de la culture. Connaître les outils que tu vas pouvoir utiliser par rapport à la donnée, ça fait aussi partie de la culture. [...] Mon objectif, c'est d'accompagner les gens. C'est de les amener à cette étape là où ils vont se dire « oui, c'est pour moi la donnée et oui je décide de m'y engager » quels que soient leurs métiers [...] Si tu n'as pas tous tes collaborateurs qui ont été acculturés à la donnée, tu ne peux pas passer à l'étape d'après. Ta gouvernance de la data, elle n'a de sens que si tes collaborateurs, ils savent de quoi tu parles. [...] Pour certaines populations, ça va être la gamification qui fait le côté défi, le jeu [...] Mais je pense que dans le cadre de l'entreprise, surtout la première étape c'est réussir à comprendre quel est, parmi tout le parcours de la donnée au sein de leur entreprise, à quel endroit ils sont en fait, quel maillon de la chaîne ils sont. Je pense que déjà si déjà tes collaborateurs, chacun a compris par rapport à ton job où est-ce qu'il est, je pense que c'est déjà énorme. [...] Tu vois, on est à l'étage de la sensibilisation, pas à l'étage de la maîtrise et de la formation qui là pour moi sera pris en charge par les RH et c'est complètement une autre histoire.» [Entrepreneur 3].

Il s'agit de déconstruire les préjugés pour pouvoir aborder par la suite des notions plus complexes :

« La technique pédagogique qu'on avait employée, c'était de poser la question que vous venez de poser et on leur disait par exemple : « ce matin vous êtes venus en formation, vous êtes venus en atelier, quelles sont les données dont vous vous êtes servis ? quelles sont les données que vous avez générées ? » ou « quelles sont les données qu'il y a dans cette salle ? Quelles sont les données qu'il y a sur vous ? ». On leur posait ce genre de

> questions et ils se mettaient à chercher sur les radiateurs, ils se mettaient à chercher dans leurs fringues ou sur les bouteilles d'eau ou ils nous parlaient de la température, la circulation, et ce genre de choses. En tout cas, avec un groupe de cinq ça marche déjà très bien. Avec un groupe de trente, il n'y a pas de soucis : ils arrivent en moins de dix minutes à nous trouver une cinquantaine de données. Et après on faisait de la déconstruction en disant : « ces données elles sont tiers, propriétaires, comment on agit dessus ? est-ce qu'on peut agir dessus ? comment on s'en sert ? dans quel système d'informations elles sont ? ». Donc, par exemple, si on me donnait la référence de la table, je disais : « elles sont dans le système d'informations du fabricant, dans le système d'informations du livreur, dans le système d'informations de celui qui nous hébergeait, dans les processus d'inventaire, dans les processus de comptabilité, etc. ». On expliquait que ces données-là étaient ressaisies et recaptées à plusieurs endroits et on essayait de mettre un point fort sur pourquoi on captait ces informations-là et quelles étaient les finalités. Par exemple, si je demandais quelle était la température de la pièce et que je suis technicien en réglage de températures ou en économie d'énergies ou en confort dans un hôtel de luxe pour mes participants, je ne vais pas capter la température au même endroit et je ne vais pas être vigilant à la même finesse de la température que je captais. En s'appuyant sur cet exemple, on essayait de leur faire toucher pour la première fois du doigt, parce que c'est une notion qui est difficile à comprendre, que la donnée elle est décidée, elle est construite : il y a une intention puissante qui est derrière. Parfois on voyait des *data scientists* qui prenaient des données de plusieurs endroits différents et qui faisaient une belle salade niçoise avec : mais il fallait bien regarder ce qu'on saisissait comme données pour pouvoir en sortir de l'intelligence et que l'intelligence elle était toujours du côté de l'humain. On insistait énormément là-dessus pour que nos interlocuteurs se sentent moins dépossédés par rapport aux données parce que souvent il y avait un complexe d'infériorité, d'informatique, de mathématiques et j'en passe. Plus la suprématie technocratique usuelle dans les discours. On essayait de casser un peu les pattes à quelques idées préconçues sur cette question de définition des données. » [Associatif 1].

6.3.3.4.3. Etre « ludique »

En lien avec le point précédent, en plus d'évacuer les difficultés de représentations et d'a priori des publics par un recentrage autour de pratiques « concrètes », une formation sur les données doit être faite de manière « ludique », c'est-à-dire de proposer des activités à l'opposé d'un cours théorique. Il faut en ce sens abandonner les lourds contenus et être succinct :

> « Le premier support qui avait été utilisé pour l'accompagnement du [nom d'organisation] c'était un powerpoint de 35 slides. Aujourd'hui, j'arrive avec 5

slides. Et je ne parle pas d'entrée des données : je parle d'abord de service numérique. » [Entrepreneur 6].

« Et très souvent le premier réflexe ça va être de partir dans un planning de production de contenus mais ce n'est pas une bonne idée parce qu'en fait je me suis vraiment rendu compte que ce qu'il faut faire avant tout ça, parce qu'en fait les gens qui produisent les contenus ce sont des gens qui sont déjà acculturés [...] Alors oui, la grande différence, la grande particularité de l'acculturation, c'est que tu vas toucher des gens qui n'en ont rien à foutre. Quand tu proposes à des gens de leur apprendre l'anglais, ils voient tout de suite à quoi ça va pouvoir leur servir et ils ont envie de partir en vacances en Angleterre donc ça tombe bien, et donc ils vont avoir une démarche d'action. Alors que quand tu vas voir des collaborateurs et tu vas leur dire « Coucou, je vais t'acculturer à la data », le gars il se dit « mais ça ne me concerne pas ! En quoi ça me concerne ? Attends, moi je suis caissière. Ce n'est pas ma vie, je ne connais même pas ce mot, c'est un truc d'ingénieur, c'est un truc de tecos, rien à voir avec moi ». Les gens ne se sentent pas du tout concernés. Et donc tu vas avoir tout un travail d'abord. Si tu leur fais directement des contenus où tu expliques à ta caissière c'est quoi un *data scientist*, mais elle n'en a rien à carrer ! [...] tu dois toucher des gens qui n'en ont rien à faire. Et tu dois d'abord réussir à aller les toucher : c'est la première étape, ce n'est pas toujours simple. Tu dois réussir à les sensibiliser. A les faire sortir de leur zone de confort. » [Entrepreneur 3].

Cette aspect ludique peut s'incarner dans l'usage de jeux permettant de fédérer les publics :

« Nous on aime bien faire des jeux sérieux sur ce genre de question. Nous-mêmes on fait de l'acculturation sur les données, sur l'urbanisme, sur plein de choses, et au final le format qui marche le mieux c'est le jeu en fait. [...] un format qui peut être sympa c'est le jeu de rôle où on distribue des cartes au hasard de différents rôles aux gens autour d'une table et ça permet de voir quels sont les désirs et les contraintes des uns et des autres, et comprendre où des abus peuvent avoir lieu. Et du coup d'avoir une meilleure prise de conscience pour que tout le monde soit au fait de ce qu'il se passe » [Associatif 5].

6.3.3.4.4. Former dès l'enfance

L'Éducation Nationale est un point sollicité à de nombreuses reprises au cours des entretiens : la formation aux données doit être enseignée dès le plus jeune âge dans un cursus commun. L'école étant un lieu que tous les citoyens fréquentent au cours de leur existence, les formations seraient donc généralisables à tous et les contenus plus facilement assimilables et intégrés dans l'imaginaire collectif.

Ces cours doivent recouvrir un champ étendu de thématiques pour comprendre ce « monde de la donnée » qui inclue non seulement les données mais également les outils numériques.

Des notions qui pourraient faire l'objet d'une discipline à part entière, soit être disséminés dans un grand nombre de cours. Dans les deux cas, une grande part de transversalité doit être présente :

> « Et la solution, ça serait le rapport Jules Ferry 3.0 : vous nous mettez une épreuve au bac où il y ait un mobilier des habiletés numériques qui soit, pourquoi pas, outillé, alimenté par le traitement des données » [Institutionnel 1].

> « Je pense que ça devrait être intégré à la culture générale, la culture de la *literacy*, de la donnée. Donc ça devrait être un élément de culture générale. Pour moi, on apprend à faire un certain nombre de choses à l'école, on fait un peu de stats, on fait un peu de maths, on en fait jusqu'au bac à la limite, et ça devrait en faire partie pour pouvoir être un citoyen, pour prendre ces décisions en connaissance de cause, il faut être capable d'analyser un discours et pour analyser un discours t'apprends à l'école les rudiments de la rhétorique, de comprendre ce qu'il se dit, la typologie des arguments, un peu de philo pour prendre un peu de hauteur et de recul sur ce qui se dit et sur ce qui se fait, la manière de le faire, sur l'homme dans la société également. Donc ça c'est une vision philosophique de la sociologie et de l'anthropologie. Il faudrait la même chose pour la data, pour prendre des décisions éclairées dans à peu près tous les aspects de sa vie, il faudrait être capable de travailler la data.» [Universitaire 8].

En corollaire s'ajoute une double nécessité, celle de former les enseignants à ces thématiques et celle d'englober un panel large d'acteurs pour intégrer les modalités de formation et de médiation.

> « Il va falloir former les profs. Pour donner un exemple : est arrivé en 2019 la compétence du programme Sciences numériques et techniques *SNT*. Le programme SNT, il y avait une compétence à l'intérieur qui est la compétence, donc obligatoire pour toutes les secondes de France, "contribuer de manière collaborative à Open Street Map". Cela arrive dans le programme, c'est une obligation. Tout enfant de seconde doit faire, doit répondre à cette demande. Sauf que pour ça il faut qu'en face il y ait des enseignants qui sachent de quoi on parle. Alors je ne sais pas comment c'est arrivé dans le programme. Nous on ne le sait pas. On est assez intrigués sur la question. Comment le comité des programmes de l'éducation nationale a intégré, validé cette compétence. Mais par contre ce que l'on voit c'est qu'il n'y a aucune formation qui accompagne et donc là on a limite du vandalisme à des moments par des élèves de seconde à qui on donne les clés d'un commun numérique autour de la donnée cartographique, non en-

cadrés, et ne comprenant rien. Là par exemple il y a beaucoup de choses à faire mais ça prend du temps » [Associatif 2].

« La question serait comment on met en place une culture des données auprès des élèves à l'école et comment on peut développer ça. Je pense, il y a l'école et les partenaires extérieurs, parce que les médiateurs, les associations, l'éducation populaire sont très actifs là-dessus. C'est aussi par les militants du web, de l'open source, des militants autour de toutes ces valeurs-là autour du web que cela peut aussi passer donc je pense que c'est déjà par une communauté d'acteurs assez engagés. Que ce soit les enseignants, les cadres de l'éducation qui participent aussi justement à favoriser la mise en place de ce type d'action là, d'éducation autour des données. Et puis je pense aussi au monde associatif, à tous les militants qui mettent en place des journées d'action, de médiation de la donnée avec des jeunes, donc je pense que c'est en multipliant les actions de formation, de médiation, qu'on peut contribuer à ça : diffuser une culture des données. » [Universitaire 2].

6.3.3.4.5. Les niveaux de formation

Comme nous l'évoquions dans le chapitre 1, la *data literacy* est conceptualisée selon une formation appliquée par degrés que nous retrouvons ici dans les discours. Trois niveaux sont citées par les acteurs : la sensibilisation, la formation culturelle et la formation technique.

Premier niveau : la sensibilisation

Il s'agit de susciter l'intérêt des individus aux thématiques des données en parlant un langage qu'ils comprennent. On se trouve ici en amont même d'une démarche de formation caractérisée par les deuxième et troisième niveaux.

« C'est tout ce qui se trouve avant la formation. L'acculturation, la culture de la donnée, c'est tout ce qui se trouve avant que les gens que tu vas acculturer rentrent dans l'étape de la formation où là ils font une démarche volontaire de leur part de...enfin " volontaire", parfois on leur impose de se former. Mais où ils sont obligés personnellement de se former, en disant "je vais me former à la donnée", à un sujet particulier autour de la donnée. Donc tout ce qui est culture, c'est tout ce qui est avant. C'est pour ça que ce n'est pas facile à définir parce qu'on est dans l'ordre de la sensibilisation.» [Entrepreneur 3].

Deuxième niveau : la formation culturelle

Il s'agit ici de développer l'ensemble des connaissances de base liés aux usages des données ; un « socle minimum de formation / socle obligatoire » par lequel les individus gagneraient en autonomie. Dans les discours, l'étape de sensibilisation se confond souvent avec l'acculturation ; sensibiliser aux enjeux des données et insuffler une culture générale sur leurs caractéristiques allant ici de pair.

> « Il y a peut-être deux niveaux. Il y a le côté culturel : "qu'est-ce que c'est ? Qu'est-ce que je peux en faire ?". Et un deuxième niveau qui serait plus : pour aller plus loin, ceux qui veulent vraiment développer les compétences numériques, et que ça fasse l'objet de formations spécifiques. » [Institutionnel 4].

Troisième niveau : la formation technique

La troisième étape intègre les compétences techniques et numériques : on dépasse ici le stade conceptuel pour traiter directement les données. Cette étape n'est pas réservée à tous, elle ne concerne que les personnes devant traiter les données dans le cadre professionnel ou bien les individus intéressés pour renforcer leurs compétences.

> « C'est-à-dire qu'il y a le premier niveau de la prise de conscience et de la compréhension. Pour moi, il y a le deuxième niveau de l'appropriation. C'est-à-dire qu'on est capable finalement d'avoir cette première base de compréhension de ce monde-là et de comment c'est collecté et à quoi ça sert. Et l'appropriation, c'est-à-dire qu'on est capable finalement de développer des compétences de manipulation, d'articulation, de jouer avec, etc. [...] Et puis pour moi, la troisième étape, c'est cette compétence et cette capacité d'aller plus loin. C'est-à-dire d'être dans une démarche où on peut utiliser des outils plus haut niveau, de manipulation, d'articulation, de jouer avec les données et là on passe rapidement sur de la programmation, du code et des solutions finalement moins propriétaires, plus *open source*» [Universitaire 9].

La *data literacy* s'applique donc par niveaux dont une première étape intervenant en amont des formations est centrée sur un éveil de l'intérêt des publics, lui-même facilité par des présentations ludiques et « pratiques » éloignées de toute vision théorique. Cette dernière pourra être introduite une fois l'intérêt suscité et une compréhension minimale des enjeux des données intégrée.

6.3.3.4.6. Exemple de projets

Un nombre volumineux de projets autour des données ont été réalisés au fil des années en France. Nous avons questionné les acteurs sur ceux qu'ils considéraient comme emblématiques et les avons intégrés dans le tableau ci-dessous.

Tableau 6.7.: Les projets représentatifs évoqués par les acteurs

Projet sollicité	**Porteur du projet**	**Nature du projet**
Agent 001	Freya Games, Savoir*devenir, Sicoval	Jeu sérieux
Belle Colline	Ekitia	Jeu sérieux
Culture D	Open Data France	Plateforme de ressources
Data lireladata	Fréquence écoles	Plateforme de ressources et de formations
Data literacy conference	FING	Conférence
Datajournos	Wedodata, Contexte et autres acteurs data journalistes	Plateforme de ressources
Dataland	Fréquence écoles	Activité pédagogique
Dataviz	CLEMI	Outil
Donnez, données moi	Tubà	Exposition
Doranum	Doranum	Plateforme de formations
Gare aux cookies	Tubà	Jeu sérieux
GTNUM DEFI	Universitaires	Projet ANR
Humanum	Humanum	Infrastructure de recherche
Initiatives du grand Poitiers	Métropole de Poitiers	Évènements
Jeu sérieux sur le RGPD	Associatif 1	Jeu sérieux
L'échappée verte	Fréquence écoles	Escape game
Master MEDAS	CNAM	Formation
Non précisé	Mydata	Bande-dessinée
Non précisé	Ekitia	Jeu sérieux
Non précisé	Infolab Rennes	Dataviz
Parcours de formation RH	Non précisé	Formation
Pas de dénomination	AirBnb	Stratégie globale
Prisme	Prisme	Chatbot
Référentiel en compétences numériques + contenus	PIX	Plateforme de formation et de contenus
Sanmanto	Datactivist	Escape game
The Uber Gates	Financial Times	Jeu sérieux
Vidéos de Métapolis	Métapolis	Contenu vidéo

Nous retrouvons certains projets présentés dans le chapitre 5. Ils s'incarnent dans une grande variété de contenus pouvant s'adresser à tous types de publics et de tous âges, ne limitant pas l'accès aux

ressources à une cible spécifique. En synthétisant et opérant une refonte des catégories de projets nous obtenons les résultats suivants (voir la figure 6.6) :

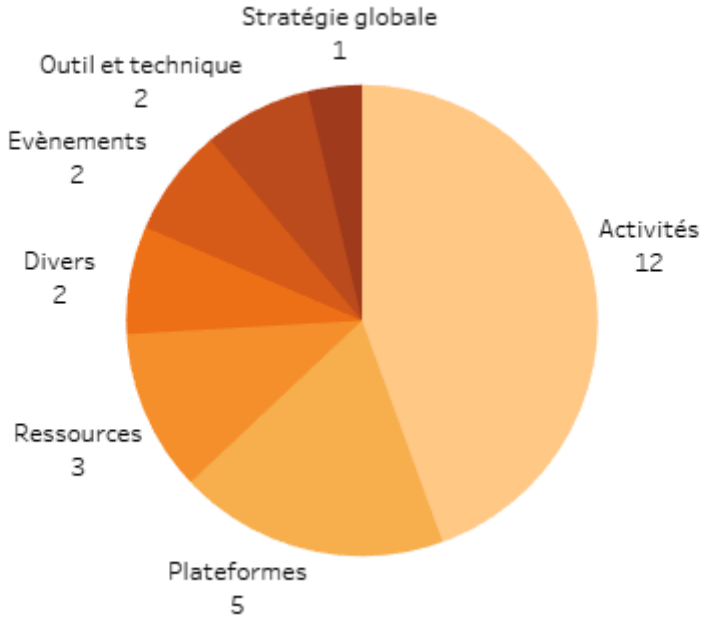

Figure 6.6.: Liste des grandes catégories de projets

Sont cités en majorité des activités (12 réponses) réparties selon trois formes : des activités de formation, des activités ludiques et des activités pédagogiques.

L'activité ludique est le mode de projet le plus sollicité par toutes les catégories d'acteurs (8 réponses). Il répond aux critères précédemment définis d'une formation en *data literacy*, à savoir de susciter l'intérêt en proposant des modalités engageantes permettant d'intégrer un panel large de publics en termes d'âge et de formation. Les enfants, les adolescents et même les adultes sont concernés. Les jeux sérieux sont ainsi le mode d'action le plus sollicité du fait qu'ils abordent un grand éventail de données, permettant de traiter leurs problématiques affiliées (ex : *open data, big data*, etc.). C'est en ce sens une porte d'entrée permettant de remplir le premier et le second niveau de la *data literacy* et pouvant être déployée au sein de l'ensemble des activités (formation et pédagogie).

Concernant les contenus, ils recouvrent un ensemble varié de formats et traitent les thématiques des données au sens large, de l'angle le plus général (ex : quelle protection des données personnelles ?) au plus spécifique (ex : data journalisme, réglementations, bonnes pratiques, etc.). On y trouve des rapports, des fiches synthétiques, des vidéos de vulgarisation ou encore des bandes-dessinées qui peuvent

faire l'objet d'exposition comme l'exposition « Données, donnez-moi des données » du *Tubà* [2]. Leur multiplicité est un atout défendu par les acteurs qui citent à ce propos l'utilité des plateformes centralisatrices comme celles liées aux projets *Culture-D, Datajournos, Data lire la data, Doranum* et *PIX*. Ces plateformes mettent en avant, et ce dès la page d'accueil, les contenus dont elles disposent et invitent les usagers à les parcourir.

Au-delà d'une mise à disposition de contenus, ces plateformes sont l'occasion d'échanges et de formations des utilisateurs. Elles servent de repères informationnels dans un écosystème de contenus sur les données très dense comme cela est visible sur la page d'accueil du site *Datajournos*[3].

Parmi les autres aspects évoqués, nous trouvons donc des formations spécifiques comme le master *MEDAS* du *CNAM* et une formation RH, des outils et méthode à employer (la dataviz revenant assez souvent comme méthode centrale), des initatives comme le *GTNUM Defi* et l'infrastructure de recherche *Humanum*.

Et enfin des évènements, en particulier la *Data literacy conference*, sont cités comme moyen utile de promotion de l'ensemble des contenus cités mais surtout comme lieux permettant la rencontre de publics très différents qui ne sont pas nécessairement des spécialistes. Selon les activités proposées, tous les niveaux de formation peuvent être atteints :

> « La Data literacy conference, c'était un évènement vraiment intéressant qui arrivait à capter des acteurs qui étaient justement à la fois des acteurs métiers, à la fois des acteurs plutôt experts, des acteurs publics, des acteurs privés, voire même plutôt des acteurs de terrain et je trouvais ça vraiment intéressant. Cela remonte à quelques années donc les questions se sont étoffées, complexifiées depuis mais je trouvais que c'était une bonne façon et avec des formats intéressants pour acculturer à un premier niveau à la fois sur les enjeux, voire avec des sessions un peu plus techniques. Ce n'est pas forcément l'exemple de projet qui a eu du succès mais pour le coup je trouve que ça a aussi suscité des vocations et je pense que pas mal d'acteurs qu'on y croisait il y a quelques années ont développé un intérêt, un attrait et une expertise pour la donnée depuis, donc ça c'était un chouette projet. » [Associatif 7].

La *data literacy* ne se limite ainsi pas à des formations conçues sur un mode descendant. Elle se décline en une multitude de contenus, institués comme point d'accès pour une grande variété de publics.

[2] https://www.tuba-lyon.com/projet/legende-urbaine-donnees-personnelles/
[3] https://datajournos.fr

6.4. Conclusion du chapitre

L'enquête de terrain nous a permis de préciser les contours de la *data literacy* tels que perçus par les acteurs français. Elle souligne tout d'abord une non stabilisation des traductions, chacune tournée vers un objectif particulier. Le terme de culture, comme évoqué dans le chapitre précédent, est un terme très utilisé car jugé plus efficace pour faciliter la compréhension des enjeux autour des données et pour englober un large ensemble de pratiques et de connaissances. La *data literacy* y est principalement perçue comme une culture (et ce sans que les acteurs emploient nécessairement la dénomination « culture » pour traduire la *data literacy*) et un ensemble de compétences techniques et intellectuelles devant permettre aux individus de s'approprier l'univers des données dans leur quotidien et de pouvoir agir en citoyens « éclairés ».

En effet, la *data literacy* n'est pas l'affaire de spécialistes mais de l'ensemble des individus qui, à la suite de campagnes de sensibilisations et de formations, deviennent des « lettrés » à même de percevoir et s'emparer des enjeux techniques, logistiques, éthiques, économiques et citoyens des données dont le *big data*, l'*open data*, le potentiel économique et démocratique des données ainsi que toutes les problématiques liées à leur mésusage sont les exemples les plus représentatifs. Il est à noter que si le terme de culture est évoqué, ce n'est pas le cas de celui de *data culture* au sens international, à l'exception d'une conceptualisation spécifique d'un acteur qui en fait l'équivalent de la « culture des données ».

La *data literacy* est donc une culture au sens d'un ensemble de pratiques, de connaissances et de représentations partagées. Elle est plus spécifiquement une culture intégrée au numérique où les données sont principalement perçues comme nativement numériques, soulignant à nouveau le fait que nous avons plus affaire à une *digital data literacy* qu'à une *data literacy*. Mais plus que cela, c'est toute la prise en compte de l'écosystème numérique qu'elle permet ici de questionner.

L'introduction des données dans les discours suit celles de l'informatique et du numérique plusieurs décennies plus tôt : la *data literacy* hérite donc de réflexions antérieures et s'inscrit dans un héritage conceptuel particulier qui brouille aujourd'hui les frontières des périmètres des données, des algorithmes et de l'intelligence artificielle, point que nous avions déjà aperçu dans la littérature anglo-saxonne. Et au-delà du numérique, on retrouve également dans les discours les interactions conceptuelles de la *data literacy* avec l'éthique des données, la gouvernance des données, la *data science* et l'esprit critique vis-à-vis des données sur des modes très similaires à ce qui est proposé dans la littérature, démontrant une imprégnation de ces concepts dans les représentations des acteurs.

Les données elles-mêmes ne sont pas perçues de la même manière selon les acteurs et il en est de même pour les types de données, bien que les données à caractère personnel sont les plus citées, rappelant ici l'effet durable du *RGPD* dans les pratiques professionnelles. Il doit être ici noté l'importance des

réglementations qui influencent les pratiques professionnelles des données : leur connaissance est ainsi considérée comme vitale pour les acteurs et doit donc être une partie intégrante des formations en *data literacy*. En comparant les caractéristiques proposées et le reste des réponses apportées, on note par ailleurs que la vision sur les données, qu'il s'agisse de l'interaction avec l'information ou du format (numérique ou non) n'a pas d'incidence sur les formations. Le seul élément récurrent réellement tangible est la richesse pouvant être potentiellement extraite des données (et qui justifie la mise en place desdites formations).

Vis-à-vis des caractéristiques de la *data literacy* elle-même, celles présentées sont en résonance avec la littérature anglo-saxonne et ce quel que soit la typologie d'acteur sollicitée (bien que tous ne s'en réclament pas explicitement) : il s'agit de traiter les données sur l'ensemble de leur cycle de vie en adoptant une posture critique et éthique.

Pour « acculturer » les acteurs, plusieurs niveaux ont été bien définis au cours de l'enquête : la sensibilisation d'une part, puis la formation culturelle et la formation technique d'autre part. Des règles générales doivent être pour cela respectées : partir des pratiques quotidiennes des acteurs pour être « parlant », évacuer les données au départ pour mieux les réintégrer par la suite, éviter les contenus de type *Powerpoint* et privilégier à la place des approches pratiques et ludiques où les acteurs sont directement impliqués dans les processus de traitement des données (les contenus plus théoriques viendront après). Si la technicité intervient plus tard dans le processus, elle ne doit pas être complètement évacuée néanmoins pour que les futurs lettrés aient une connaissance pratique de ce que recouvre les pratiques des données et ce sans que nécessité d'être un expert technique comme dans le cas de la science des données.

Nous nous sommes demandés si les milieux de rattachement avaient une influence sur les discours. Cette interrogation n'appelle pas une réponse tranchée sur l'ensemble des points abordés. Sur certains points, comme l'usage d'une dénomination, le type de données à traiter prioritairement ou encore les objectifs visés par ce traitement des données, une certaine homogénéité se dégage nettement, en particulier chez les acteurs de type « institutionnel » où l'anglais est proscrit. Sur d'autres, le milieu de rattachement n'a aucun poids : la préférence d'une pratique par rapport à une autre découle d'une vision singulière mûrit par l'expérience.

A lumière des chapitres 5 et 6, nous ne pouvons pas parler d'une conception homogène française, tirée d'une source unique, de la *data literacy* : si certaines dénominations sont communes, ce n'est pas le cas des sources mentionnées (et c'est clairement le cas pour le public « universitaire ») qui diffèrent d'une personne à l'autre. Nous avons affaire à des généalogies hybrides qui nourrissent des visions singulières. Il existe néanmoins des invariants, notamment dans les compétences et connaissances à développer, se rapprochant en cela fortement des visions anglo-saxonnes.

Les méthodes « d'acculturation » sont également similaires à celles rencontrées dans le monde anglo-saxon au sens d'un continuum, retirant ici la dichotomie « lettrée / non lettrée ». En revanche, nous maintenons notre propos énoncé dans le chapitre 5, à savoir que c'est le prisme d'appropriation qui marque la différence là où le contexte anglo-saxon est plus générique dans ses approches, très centré sur les compétences comme nous avons pu le voir dans le chapitre 1 dans les référentiels de compétences où la *data culture* et plus globalement les aspects culturels sont minoritaires, tout en rappelant toutefois que la littérature anglo-saxonne est bien plus riche que la littérature française en termes de thématiques sociétales abordées (ex : le féminisme des données).

Ainsi, si les problématiques pédagogiques sont bien présentes dans la littérature internationale, la question des médiations et des implications culturelles sont plus spécifiques aux réflexions françaises. Tout cela est néanmoins très susceptible d'évoluer comme nous l'ont témoigné les acteurs interrogés plusieurs années après cette enquête dont les avis ont changé sur de nombreux points.

Nous sommes donc en présence de *data literacies* françaises assez diverses bien que comportant quelques invariants (importance des données, volet critique, compétences et connaissances à enseigner). Et ceci reflète le paysage international qui est lui-même non homogène comme en témoignent les littératies spécialisées des données.

Conclusion générale

> « *I'm a qualitative researcher. I collect stories ; that's what I do. And maybe stories are data with a soul.* »
>
> Brené Brown, *The power of vulnerability*, 2010

L'intérêt porté aux données, dont l'aspect supposément « révolutionnaire » est souvent mis en avant dans les discours, est en réalité loin d'être récent : il s'ancre dans un passé plus lointain, émergeant dans les disciplines scientifiques avant de s'étendre à d'autres sphères économico-commerciales pour aujourd'hui concerner l'ensemble des individus. Une volonté de maîtrise centrée à l'origine autour d'une culture technique développée sous l'égide de la *data science* mais dont l'existence seule s'est accompagnée d'échecs successifs dans l'élaboration de projets centrés sur les données.

Le vocabulaire de la crise, au sens d'une tension survenue lors d'une rupture d'équilibre, accompagne chaque littératie et il en est de même pour la *data literacy* : les nombreux changements en cours amenés par la démocratisation et la généralisation des techniques et technologies numériques de production et de gestion des données ont influencé de nombreux pans des sociétés, dont l'émergence de nouveaux emplois et de « pôles data » sont les aspects les plus visibles, et ont été condensés, essentialisés au sein du vocable de « révolution des données », porteur à la fois des espérances et des craintes des organisations.

Cet essentialisme appliqué aux données (réduites en « la » donnée ou « la » data) et au mouvement qui les accompagne (ladite « révolution ») a obscurci une situation naturellement difficile à appréhender. La *data literacy*, premièrement conceptualisée en 2004, a été maintes fois réappropriée et employée comme une solution, si ce n'est « la » solution, pour remédier à cette crise liée aux manques de connaissances et de compétences des individus. Or, ces réappropriations ont contribué à diluer sa supposée unicité, véhiculant un kaléidoscope toujours plus étendu de ses représentations masquant les invariants, les méthodologies et les objectifs visés. Et finalement à perdre de vue ce qui est entendu par une maîtrise des données.

Notre recherche portait donc une ambition, celle de proposer un état de l'art des représentations du concept de la *data literacy* au sein des discours et d'apporter une vision claire de ses caractéristiques.

Nous avions en ce sens posé trois hypothèses en soutien à notre problématique générale qui était de **définir ce qu'est une littératie appliquée aux données.**

Réponse aux hypothèses de recherche

Nos trois hypothèses découlaient l'une de l'autre et proposaient ainsi de voir la *data literacy* comme (1) un objet frontière (2) influencé par les domaines de rattachement des porteurs des discours et (3) nous amenant à parler de *data literacies* et non de « la » *data literacy*. Nous allons ici les traiter non pas séparément mais ensemble du fait qu'elles ne peuvent être dissociées.

Nous l'avons précisé en introduction, l'objet frontière est une entité qui améliore la capacité d'une idée, d'une théorie ou d'une pratique à se traduire au-delà de frontières culturellement établies (Fox, 2011). Il « qualifie la manière dont les acteurs établissent et maintiennent une cohérence entre des mondes sociaux en interaction » (Trompette & Vinck, 2009, paragr. 10), qui négocient leurs différences et créent une mise en accord, sans consensus préalable, de leurs points de vue respectifs, facilitée par la « flexibilité interprétative » inhérente à l'objet frontière lui-même (Star, 2010).

A travers les six chapitres de cette thèse, nous avons pu déceler que quatre grands types d'acteurs [4], tout aussi bien au niveau anglo-saxon que français, promeuvent une multiplicité de représentations de la *data literacy*. Ces représentations, si diverses soient-elles, possèdent des invariants, à savoir (1) une volonté de maîtrise des données obtenue par le développement de connaissances et de compétences, (2) une mise en valeur des données perçues comme des atouts et soutenant les décisions des organisations mais également des individus dans leur quotidien et (3) un déploiement continu de formations (dont font partie les curricula universitaires), de référentiels, de cadres de compétences ou encore de certifications, permettant de développer et de mesurer les taux de littératie des individus.

Un quatrième invariant est plus récent et se base sur (4) un décentrage des questionnements liés au raisonnement scientifique, du traitement des données perçues comme des ressources uniquement associées à la recherche scientifique (à l'exception, dans une certaine mesure, de certaines littératies spécialisées plus versées dans ce domaine comme la *research data literacy* et la *science data literacy*) pour s'intégrer dans une vision numérique et une technicité informatique.

Selon cette vision, les données, majoritairement perçues comme numériques, nécessitent des compétences techniques spécifiques pour être administrées et c'est en ce sens que nous devrions parler non pas de *data literacy* mais de *digital data literacy*. En plus de ces représentations numériques des données et des compétences techniques associées, la *digital data literacy* porte en elle une interroga-

[4] Nous les avions nommés « associatif », « entrepreneurial », « institutionnel » et « universitaire ».

tion des écosystèmes numériques comme dans le cas des données, des intelligences artificielles et des algorithmes. Si le titre de *digital data literacy* peut sembler englobant, nous précisons que nous ne confondons pas pour autant *digital literacy* et *data literacy* : les deux littératies ne se déploient pas sur les mêmes plans et ne partagent pas les mêmes objectifs malgré les connexions précitées comme nous avons pu le constater dans le chapitre 3. La *digital data literacy* ne concerne qu'un pan spécifique de l'écosystème numérique et limite son prisme d'analyse à travers les données.

Cette apparente unicité de la *(digital) data literacy*, véhiculée par ses invariants, en particulier dans les premières étapes de sa conceptualisation, a occulté involontairement [5] une réalité plus nuancée. Dès son origine, elle a été conceptualisée au sein de deux domaines différents de recherche (les bibliothèques et les statistiques pour Schield, la pédagogie pour Love), nous donnant déjà ici un aperçu d'une part de la capacité de la thématique des données à se traduire au-delà d'un domaine spécifique de recherche, et au-delà du contexte même de la recherche quand celle-ci a été appropriée par d'autres types d'acteurs, notamment les entreprises, renvoyant à la plasticité interprétative inhérente à l'objet frontière. Ce qu'il convient ici de noter, c'est l'absence, plus ou moins prononcée, d'une homogénéité des visions au-delà des invariants suscités (comme dans le cas du choix des compétences à développer en priorité par exemple).

Les dénominations en sont un reflet et soulignent les difficultés encore actuelles sur la manière idoine de transmettre les connaissances et compétences sur les données, renvoyant aux problématiques rencontrées par la littératie elle-même que nous avons rappelées dans le chapitre 2.

Dénominations anglo-saxonnes	Dénominations françaises	Autres dénominations
Data literacy *Datacy* *Data acumen* *Data sense* *Data wise*	Acculturation à la / aux donnée(s) Compétences pour la donnée Culture data Culture à la / de la / des donnée(s) Culture de l'information Culture générale de la donnée Data culture Data lettrisme Data littératie Datalphabétisation Education à la / aux donnée(s) Littératie des données Médiation à la / des donnée(s) Réflexe donnée Valorisation des données	*Alfabetización en datos* (Espagne) *Alfabetizzazione ai dati* (Italie) *Datenkompetenz* (Allemagne) Littératie des données (Québec)

Figure 6.7.: Dénominations et traductions de la *data literacy* en France et à l'international

[5] Dès 2002, Tedesco rappelait que la *data literacy* n'était pas officiellement définie et les premières publications, notamment celles de Schield et Love en 2004, n'affirmaient nullement une vision unique ou stabilisée du concept.

Nous avons pu constater sur ce point que certaines dénominations sont employées par des acteurs aux domaines de rattachement spécifiques comme dans le cas de la « culture data » typiquement liée aux acteurs français de type « entrepreneurial ». Or, si des dénominations peuvent se retrouver fréquemment au sein de certains domaines de rattachement, les caractéristiques embarquées ne sont pas nécessairement les mêmes. En reprenant l'exemple de la « culture data », il n'est pas avéré que ce concept recouvre les mêmes réalités en termes de compétences, de connaissances ou d'objectifs suivis. Et il en est de même pour les conceptions anglo-saxonnes également multiples, différant même au sein d'une seule typologie d'acteurs (comme nous avons pu le voir notamment lors de notre enquête sur *Twitter/X* où ce sont en majorité des acteurs individuels et non des organisations qui promeuvent les visions sur la *data literacy*), et qui connaissent, à un degré moindre que dans le cas français, des difficultés de stabilisation des dénominations.

Si le terme *data literacy* est très majoritairement employé dans la littérature anglo-saxonne, nous avons pu voir apparaître des terminologies concurrentes comme *data acumen* et *data wise* qui nourrissent ici des réflexions différentes bien que ne débordant pas du cadre des visions utilitaires, voire « minimale » et « autonome », de la littératie. *Data sense* est une exception au sens que ce concept, selon la vision de Lupton (2017) présentée dans le chapitre 1, développe une approche charnelle où les sens humains deviennent les médiateurs des interactions hommes-données-numérique. Il ne peut être défini comme une littératie spécialisée du fait qu'il se décentre des visions littéraciques utilitaires pour asseoir une nouvelle généalogie. Mais cet exemple prouve que les univers anglo-saxons et français sont confrontés aux mêmes interrogations, notamment en termes de prismes de conceptualisation et de déploiement.

Nous avions pris soin de séparer ces deux univers pour tenter de déceler des différences fondamentales : il n'en est rien. Il existe évidemment des variations comme nous avons pu le voir dans le cas français où un plus grand nombre de prismes d'appropriation est employé (comme la culture, la médiation, l'éducation, etc.) et où se pose une grande problématique de dénomination [6] mais les questionnements et les disparités de contenus sont les mêmes. Bien que nous n'ayons pas développé cela dans la thèse, nous ajoutons ici qu'en-dehors du monde anglo-saxon, nous retrouvons cela dans d'autres pays comme ceux présentés précédemment (à savoir l'Allemagne, le Canada, l'Espagne et l'Italie) : ce n'est donc pas une spécificité française.

Ainsi, les domaines de rattachement jouent un rôle certain dans les visions des *data literacies* mais ce rôle n'est jamais total ou totalisant. Pour certains domaines de rattachement comme les entreprises et les institutions étatiques, l'emprise organisationnelle est importante, en particulier dans le choix des dénominations et des objectifs poursuivis. Pour d'autres, l'emprise est plus faible, laissant plus de place aux visions individuelles : c'est alors souvent l'expérience de l'acteur lui-même qui va influer sur

[6] Nous précisons « dénomination » et non « traduction » du fait que les visions françaises ne découlent pas toutes d'une transposition de la *data literacy*.

ses choix comme nous avions pu le déceler dans les chapitre 1 et 6. Ceci s'explique notamment par le caractère polysémique, polymorphe, des données qui laisse place à plus d'interprétations de la part des acteurs.

Nous avons pu le constater dans le cas des littératies spécialisées qui se déploient *ad infinitum*, s'hybridant tout aussi bien avec des domaines de recherche, qu'avec des thématiques sociétales ou des spécificités techniques. Ces visions démontrent à nouveau la capacité de la *data literacy* à se traduire au-delà de frontières culturellement établies et à embrasser un spectre toujours plus large de thématiques. C'est en ce sens que nous pouvons parler de *data literacies* et non de « la » *data literacy* : il n'existe pas de *data literacy* sous une forme supposément pure, univoque et universelle. Ce qui existe, ce sont des invariants data littéraciques sur lesquels se fondent une pluralité de visions parfois concurrentes, rendues possibles par la plasticité interprétative des données.

Qu'est-ce qu'une littératie des données ?

Une littératie des données est donc un ensemble d'invariants sur lesquels s'adossent un certain nombre de conceptions (utilitaires et/ou idéologiques). Avec les apports éthiques et critiques du milieu des années 2010 souvent inspirés par les travaux de Paulo Freire, nous pouvons dire qu'une littératie des données est aujourd'hui, avant toutes considérations utilitaires, une réflexion, une posture intellectuelle basée sur le désenchantement. Au même titre que la datalogie qui visait à « dissiper l'aspect mystique » entourant les ordinateurs, les *data literacies* dissipent l'aspect mystique des données et de leurs usages, déconstruisent les mythes tels que définis par Roland Barthes. Il ne s'agit pas de dire qu'à l'inverse les données ne porteraient aucun poids, qu'elles ne seraient investies d'aucun « pouvoir » car c'est précisément l'inverse qui a été montré, en particulier dans le cas des littératies spécialisées liées aux rejets des discriminations. Et encore plus lorsque l'on évoque les discours : du fait que toute communication est automatiquement adossée par le pouvoir (Douyère, 2016), l'on ne peut dissocier les logiques de pouvoir des enjeux des données portés par les discours. Tout discours véhicule des représentations de la réalité qui se cristallisent et s'incarnent, refaçonnant les quotidiens des individus. Ce que nous avons pu constater tout au long de cette thèse.

Le vocable de citoyen (*citizen* en anglais), évoqué à de nombreuses reprises dans les publications, marque ce basculement des littératies des données minimales et utilitaires vers des littératies des données maximales et idéologiques (ici compris doublement d'une part selon la vision de Brian Street présentée dans le chapitre 2 et d'autre part comme une littératie investie d'un pouvoir politique). Le citoyen est, selon le dictionnaire *Larousse*, une « personne jouissant, dans l'État dont il relève, des droits civils et politiques, et notamment du droit de vote (par opposition aux étrangers) [7] ».

[7] https://www.larousse.fr/dictionnaires/francais/citoyen/16241

Cette vision est loin d'être anodine : elle renvoie à la vie en société, au principe de responsabilité et souligne le lien politique sous-jacent aux pratiques citoyennes. En ce sens, les *data literacies* servent des objectifs politiques basés sur la résolution de problèmes du quotidien et embrassent logiquement une infinité de thématiques, ce que nous avons pu déceler avec l'existence des littératies spécialisées. La plasticité d'une littératie des données est permise et amplifiée par le caractère polysémique des données, à la fois statut, matériau et système.

Une littératie des données est donc cette encapacitation des personnes, des citoyens, tant au niveau des connaissances que des compétences, pour résoudre des problématiques du quotidien dont la portée dépasse le simple cadre micro (comme dans le cas des biais de profilages des personnes basés sur le déploiement d'intelligences artificielles entraînées sur des jeux de données limités). Elle répond en cela à des enjeux informationnels, rejoignant ici notre analyse posée en 2020 : « Pourtant les enjeux de la littératie des données reviennent finalement à des principes d'analyse informationnelle avant d'en penser la stratégie d'exploitation communicationnelle. » (Verdi & Le Deuff, 2020). Maîtriser les données, c'est avant tout se réapproprier des méthodes critiques de décryptage des réalités (qui bien qu'artificielles, construites, ont des effets bien palpables) qui pourront ensuite être influencées, réagencées, du fait de l'encapacitation politique et informationnelle des individus. En ce sens, les littératies des données sont en symbiose avec les littératies informationnelles, bien au-delà de la simple connexion conceptuelle entre donnée et information. Devenir pleinement un(e) lettré(e) des données, c'est selon nous devenir également un(e) lettré(e) informationnel (*information literate*).

Les perspectives de recherche

En lien avec les convergences conceptuelles et littéraciques évoquées dans le chapitre 3 et traduites d'une part dans les discours par cette donnée devenue une « donnée-système » englobant des réalités plus grandes, et d'autre part par le rapprochement de métiers et de pôles professionnels, en l'occurrence les données et l'intelligence artificielle, nous défendons ici une approche littéracique basée non pas sur les concepts analysés individuellement mais sur une vision systémique, sur l'appropriation par les *SIC* du système.

En effet, si les approches littéraciques démontrent une véritable recherche de la part de leurs auteurs et nous apportent des éléments précieux de réflexion, elles posent néanmoins de nombreux problèmes. Si les approches littéraciques permettent de questionner les réalités, elles conditionnent et limitent notre regard. Bien que la question du continuum a permis de sortir de l'écueil de la dichotomie « lettré - non lettré », nous constatons néanmoins la permanence des visions utilitaires basées sur l'évaluation des compétences.

De plus, elles s'attaquent aux concepts de façon isolée (les données pour la *data literacy*, les algorithmes pour l'*algorithm literacy*, l'intelligence artificielle dans l'*IA literacy*, etc.), proposant de fait des solutions isolées et tendant, involontairement dans la plupart des cas, à établir des hiérarchies entre ces derniers. Les littératies basées ces concepts vont ensuite entrer dans une querelle où une seule émergera et imposera temporairement sa vision sur les autres. Dans le cas de la *data literacy*, nous avons pu montrer que les querelles littéraciques ont été vives (en particulier avec la *statistical literacy*) et ne sont pas encore réglées : elles se poursuivent aujourd'hui avec l'arrivée des littératies des algorithmes et de l'IA et nous forcent à nous poser la question du concept dominant, reflétant en cela un asservissement des littératies aux discours dominants ; les littératies perdant de leur force de réflexion, d'analyse objective, pour se conformer, au gré des tendances, à des modèles standards.

Et par-delà les hiérarchies, un problème d'explicitation des liens entre des littératies différentes, et *in extenso* des concepts qu'elles portent, se pose, en particulier dans le cas des littératies spécialisées. Nous le répétons : les recherches sur chacune des thématiques précitées, en particulier les thématiques sociétales comme le féminisme des données, sont centrales mais tendent à se siloter du fait de leur spécialisation.

Tout ceci contribue à une logique d'entassement toujours plus conséquente de publications, une constellation de concepts et de littératies dont les liens et les frontières deviennent de moins en moins visibles, ce que constatait déjà Le Deuff (2012) :

> « Le constat d'une convergence médiatique liée au numérique succède à une période qui pourrait correspondre à une « querelle » des littératies. Les littératies informationnelles (information literacy) et médiatiques (media literacy) se sont donc disputées pendant plusieurs années une forme de leadership, chacune prétendant contenir l'autre. En parallèle, la littératie informatique (computer literacy) a connu également des évolutions du fait du numérique au point d'être souvent intégrée au projet d'une littératie du numérique (digital literacy). L'émergence de nouvelles littératies ont accru cette complexité et cette dispersion des forces. Les actuelles perspectives offertes par la translittératie (transliteracy) (Thomas, 2007) ou bien encore de la MIL (Media and Information Literacy) permettraient de les regrouper toutes dans un même projet à condition de réellement parvenir à de réelles avancées sans répéter des étapes précédemment déjà opérées. »

Ce que nous craignons ici dans la convergence actuelle est une situation limitée à de simples fusions de littératies, comme dans le cas de la *data and IA literacy* ou de l'*algorithm and data literacy project*, qui finiront par proposer à nouveau des modèles utilitaires qui évacueront la complexité systémique des concepts, diluant ces derniers et limitant grandement le développement d'approches didactiques offrant aux apprenants une compréhension concrète des savoirs transmis.

Bien qu'indispensable, limiter les *data literacies* à un versant purement technique s'est soldé, nous l'avons vu, par des échecs. Et si des aspects culturels ont été intégrés dans les enseignements et formations, ils ne suffisent pas pour avoir une vision complète. Nous avons conscience qu'il est impossible d'aborder tout un écosystème de manière fine au sein de chaque publication.

Ce que nous proposons ici ce n'est pas le développement d'une littératie systémique qui permettrait de tout englober : nous retomberions dans le piège de l'entassement littéracique. A la place, nous proposons une systémie des littératies en lien avec une systémie des concepts comme le « système-information » évoqué en chapitre 3 [8], pensée comme réflexion globale et interactive où, à la manière du système tel que défini dans le chapitre 3, est pris en considération le fait que les littératies ont une influence notables les unes vis-à-vis des autres et ne peuvent exister les unes sans les autres du fait des concepts qu'elles portent. Par-delà une simple vision théorique, cette systémie aurait l'avantage de proposer de nouvelles didactiques, plus complètes et plus précises qui permettraient d'éviter l'ajout au compte-goutte de nouveaux enseignements souvent dissociés les uns des autres [9].

En tant que base de nos futures réflexions sur la systémie littéracique, nous faisons ici un rappel de notre proposition de *data literacy* distribuée développée dans (Verdi & Le Deuff, 2020). Basée sur l'idée d'un retour aux idées de Paul Otlet, en particulier celle du réseau, elle avait pour but d'être :

> « Une condition pour envisager des réseaux ou milieux métastables, c'est-à-dire bénéficiant d'infrastructures communes partagées et entretenues collectivement, tout en garantissant une neutralité ainsi que des potentialités d'innovation. Cette manière de concevoir la littératie des données permettrait d'éviter les milieux dissociés que dénonçait Stiegler (Stiegler, 2006) qui ont le double défaut de vouloir penser et révéler le monde tout en dissimulant ses méthodes et présupposés. La constitution de la littératie des données distribuée, telle que nous avons cherché à la définir ici, permettrait aussi aux différentes communautés de pouvoir construire des savoirs et savoir-faire en partageant des valeurs communes avec d'autres communautés ».

[8] Ceci reflète par ailleurs un manque spécifique rencontré au sein des *SIC* : celui de la synthèse systématique, d'un état de l'art permanent des recherches permettant d'éviter un éparpillement ou un doublon des recherches. Nos publications traduisent cette volonté de synthèse systématique, bien que nécessairement imparfaite, contribuant à poser des repères et cette thèse s'inscrit pleinement dans cet objectif.

[9] Tout récemment, la Ministre de l'Éducation nationale et de la Jeunesse Nicole Belloubet, dans sa lettre de saisine du 14 mars 2024, a amené le sujet de l'intelligence artificielle comme enseignement supplémentaire : « les programmes contribueront, pour chaque discipline, à créer une culture de l'IA, au travers en particulier de l'utilisation raisonnée des possibilités offertes par les outils utilisant l'intelligence artificielle. Ils identifieront pour chaque année les cas d'usage où l'apport de l'IA représente une véritable plus-value et les notions à connaître. Cette harmonisation et cette révision de l'ensemble des autres programmes devront entrer en application à la rentrée 2026 » (source : https://www.education.gouv.fr/le-conseil-superieur-des-programmes-41570). Nous retrouvons ici le même type de discours que pour les données et l'informatique avant cela : les sujets changent, les discours et les enjeux véhiculés restent les mêmes.

C'est donc un type de système qui était ici sollicité, la distribution, révélant les liens entre les connaissances et neutralisant les opacités. C'est selon cette perspective systémique que nous développerons nos futurs travaux.

Bibliographie

A5admin. (2019, 15 juillet). La Data Literacy ou l'art de donner du sens à la donnée : le véritable enjeu de nos entreprises. *A5sys*. Repéré à https://www.a5sys.com/la-data-literacy-ou-lart-de-donner-du-sens-a-la-donnee-le-veritable-enjeu-de-nos-entreprises/

Abner, K. (2020). Data Literacy as Digital Humanities Literacy: Exploration of Threshold Concepts [Text]. *dh+lib*. Repéré à https://acrl.ala.org/dh/2020/06/22/data-literacy-as-digital-humanities-literacy-exploration-of-threshold-concepts/

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424-438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*, 3-9.

Adnouest. (2024, 14 mai). ADN Data & IA. *ADN Ouest*. Repéré à https://www.adnouest.org/communaute/adn-data

Adreit, H. (2020, 12 octobre). La culture « data driven » : La condition pour réussir sa transformation digitale. *Journal du net*. Repéré à https://www.journaldunet.com/big-data/1494515-la-culture-data-driven-une-condition-sine-qua-none-pour-reussir-sa-transformation-digitale/

Agesilaou, A., & Kyza, E. A. (2021). Empowering Students to be Data Literate: The Design and Implementation of a Learning Environment to Foster Critical Data Literacy. Repéré à https://repository.isls.org//handle/1/7503

Albrechtslund, A., & Lauritsen, P. (2013). Spaces of Everyday Surveillance: Unfolding an analytical concept of participation. *Geoforum*, *49*, 310-316. https://doi.org/10.1016/j.geoforum.2013.04.016

Albury, K., Burgess, J., Light, B., Race, K., & Wilken, R. (2017). Data cultures of mobile dating and hook-up apps: Emerging issues for critical social science research. *Big Data & Society*, *4*(2), 2053951717720950. https://doi.org/10.1177/2053951717720950

Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: a comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management*, *31*(2), 300-316. https://doi.org/10.1108/JEIM-01-2017-0007

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, *23*(5-6), 839-859. https://doi.org/10.1007/s00779-017-1104-3

Amat, M., & Maigné, C. (2022). *Philosophie de la culture*. Paris: Vrin. Repéré à https://www.decitre.fr/livres/philosophie-de-la-culture-9782711630264.html

American Library Association. (1989). *Presidential Committee on Information Literacy: Final Report*. Repéré à https://www.ala.org/acrl/publications/whitepapers/presidential

Amoore, L. (2020). *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. (S.l.): Duke University Press.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. Repéré à https://www.wired.com/2008/06/pb-theory/

Anderson, G. L., & Irvine, P. (1993). Informing critical literacy with ethnography. Dans *Critical literacy : politics, praxis, and the postmodern* (pp. 81-104). Etats-Unis: State University of New York.

André, F. (2014). Chapitre 5. Déluge des données de la recherche ? *Information et strategie*, 77-95. Repéré à https://www-cairn-info.proxybib-pp.cnam.fr/big-data--9782804189150-page-77.htm

Andrejevic, M. (2014). Big Data, Big Questions| The Big Data Divide. *International Journal of Communication*, *8*(0), 17. Repéré à https://ijoc.org/index.php/ijoc/article/view/2161

APEC. (2020). *Les métiers de la data*. APEC. Repéré à https://corporate.apec.fr/files/live/sites/corporate/files/Nos%20%c3%a9tudes/pdf/Les-metiers-de-la-data.pdf

Apogea. (2021, 26 avril). Instaurer une culture data driven : 5 principes. *Apogea*. Repéré à https://www.apogea.fr/5-principes-pour-instaurer-une-culture-data-driven-dans-votre-entreprise/

Aragona, B., & Zindato, D. (2016). Counting people in the data revolution era: challenges and opportunities for population censuses*. *International Review of Sociology*, *26*(3), 367-385. https://doi.org/10.1080/03906701.2016.1244926

Arène, V. (2023, 10 mars). La Poste ouvre la 1ere promotion de son école de la data et IA. *Le Monde Informatique*. Repéré à https://www.lemondeinformatique.fr/actualites/lire-la-poste-ouvre-la-1ere-promotion-de-son-ecole-de-la-data-et-ia-89793.html

Arnove, R. F., & Graff, H. J. (Éds). (1987). *National Literacy Campaigns: Historical and Comparative Perspectives*. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-0505-5

Arruabarrena, B., Kembellec, G., & Chartron, G. (2019). Data littératie & SHS : développer des compétences pour l'analyse des données. Dans *CODATA*. Marne la Vallée, France. Repéré à https://hal.archives-ouvertes.fr/hal-03154432

Arruda, S. A. G. (2020). Exploring Data Literacy: Concepts and Determinants for Data Skills Development. Dans *Exploring Data Literacy: Concepts and Determinants for Data Skills Development*. Universidade de Coimbra. Repéré à https://estudogeral.sib.uc.pt/handle/10316/94716

Artyushina, A. (2020). Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto. *Telematics and Informatics*, *55*, 101456. https://doi.org/10.1016/j.tele.2020.101456

Atchison, W. F., Conte, S. D., Hamblen, J. W., Hull, T. E., Keenan, T. A., Kehl, W. B., ... Young, D. M. (1968). Curriculum 68: Recommendations for academic programs in computer science: a report of the ACM curriculum committee on computer science. *Communications of the ACM*, *11*(3), 151-197. https://doi.org/10.1145/362929.362976

Atenas, J., Havemann, L., Rodés, V., & Podetti, M. (2023). Critical data literacy in praxis: An open education approach for academic development. *Edutec. Revista Electrónica de Tecnología Educativa*, (85), 49-67. https://doi.org/10.21556/edutec.2023.85.2851

Athanases, S. Z., Bennett, L. H., & Wahleithner, J. M. (2013). Fostering Data Literacy Through Preservice Teacher Inquiry in English Language Arts. *The Teacher Educator*, *48*(1), 8-28. https://doi.org/10.1080/08878730.2012.740151

Audureau, W. (2024, 20 avril). Tout comprendre à l'intelligence artificielle, cette technologie source de nombreux malentendus. *Le Monde.fr*. Repéré à https://www.lemonde.fr/les-decodeurs/article/2024/04/20/tout-comprendre-a-l-intelligence-artificielle-cette-technologie-source-de-nombreux-malentendus_6228954_4355770.html

Auffray, C. (2022, 15 décembre). Créer la symbiose Data/Métiers : un passage obligé. *Alliancy*. Repéré à https://www.alliancy.fr/symbiose-data-metiers

Auge, M. (2019). *Les Formes de l'oubli*. (S.l.): Payot.

Aurell, M. (2011). *Le Chevalier lettré : Savoir et conduite de l'aristocratie aux XIIe et XIIIe siècles*. Paris: Fayard.

Avenati, O., & Chardel, P.-A. (2016). *Datalogie : formes et imaginaires du numérique*. (S.l.): LOCO.

Aviram, A., & Eshet-Alkalai, Y. (2006). Towards a Theory of Digital Literacy: Three Scenarios for the Next Steps. *European Journal of Open, Distance and E-learning*, *9*(1). Repéré à https://old.eurodl.org/?p=archives&year=2006&halfyear=1&article=223

Bachimont, B. (2017). Le numérique comme milieu : enjeux épistémologiques et phénoménologiques. : Principes pour une science des données. *Interfaces numériques*, *4*(3), 402-402. https://doi.org/10.25965/interfaces-numeriques.386

Bachimont, B., & Verlaet, L. (2021). Traces, données et preuves en contexte numérique : quelles acceptions interdisciplinaires? *Revue Intelligibilité du Numérique*, (2). https://doi.org/10.34745/numerev_1712

Bakhtin, M. M., Holquist, M., & Emerson, C. (1982). *The Dialogic Imagination: Four Essays* (Reprint édition). Austin, Tex: University of Texas Press.

Balloni, A. J. (2018). A new peer review standard using a no-hierarchical dikw pyramid. *International Journal of Education*.

Bargagliotti, A., Binder, W., Blakesley, L., Eusufzai, Z., Fitzpatrick, B., Ford, M., ... Zachariah, T. (2020). Undergraduate Learning Outcomes for Achieving Data Acumen. *Journal of Statistics Education*, *28*(2), 197-211. https://doi.org/10.1080/10691898.2020.1776653

Barthes, R. (1957). *Mythologies*. Paris: POINTS.

Bates, J. (2017). Data cultures, power and the city. Dans *Data and the City*. (S.l.): Routledge.

Bates, J. (2018). The politics of data friction. *Journal of Documentation*, *74*(2), 412-429. https://doi.org/10.1108/JD-05-2017-0080

Bates, M. J. (2005). Information and knowledge: an evolutionary framework for information science. *Information Research*, *10*(4). Repéré à http://informationr.net/ir/10-4/paper239.html

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Business & Information Systems Engineering*, *1*(3), 213-222. https://doi.org/10.1007/s12599-009-0044-5

Beer, D. (2016). *Metric Power* (Palgrave Macmillan).

Beky, A. (2017, 16 octobre). Analyse de la données : les Français sont des «datanalphabètes». *Silicon*. Repéré à https://www.silicon.fr/litteratie-donnees-salaries-francais-187011.html

Bellec, A.-C. (2022, 20 juin). Diffuser une culture de la donnée au sein de son organisation. *Opendatasoft*. Repéré à https://www.opendatasoft.com/fr/blog/comment-diffuser-la-culture-de-la-donnee-dans-votre-organisation/

Benoit, M.-C. (2022). Docaposte annonce la création d'un pôle en data et intelligence artificielle. *ActuIA*. Repéré à https://www.actuia.com/actualite/docaposte-annonce-la-creation-dun-pole-en-data-et-intelligence-artificielle/

Benson, S. E. (2009). Understanding Literate Language: Developmental and Clinical Issues. *Contemporary Issues in Communication Science and Disorders*, *36*(Fall), 174-178. https://doi.org/10.1044/cicsd_36_F_174

Benzécri, J. P. (1976). Histoire et préhistoire de l'analyse des données. Partie I La préhistoire. *Les cahiers de l'analyse des données*, *1*(1), 9-32.

Ben-Zvi, D., & Garfield, J. (1999). *Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges* (p. 15). https://doi.org/10.1007/1-4020-2278-6_1

Berdoulay, V., Castro, I., & Gomès, P. C. D. C. (2005). L'espace public entre mythe, imaginaire et culture. *Cahiers de géographie du Québec*, *45*(126), 413-428. https://doi.org/10.7202/023001ar

Bergqvist, D., Troëng, T., Elfström, J., Hedberg, B., Ljungström, K.-G., Norgren, L., & Örtenwall, P. (1998). Auditing surgical outcome: ten years with the Swedish vascular registry-Swedvasc. *European Journal of Surgery*, *164*(S7), 3-32. https://doi.org/10.1080/11024159850191607

Berndt, M., Schmidt, F. M., Sailer, M., Fischer, F., Fischer, M. R., & Zottmann, J. M. (2021). Investigating statistical literacy and scientific reasoning & argumentation in medical-, social sciences-, and economics students. *Learning and Individual Differences*, *86*, 101963. https://doi.org/10.1016/j.lindif.2020.101963

Bernhard, P. (2003). *Déclaration de Prague : vers une société compétente dans l'usage de l'information*. Prague. Repéré à https://www.enssib.fr/bibliotheque-numerique/documents/1900-declaration-de-prague-vers-une-societe-competente-dans-l-usage-de-l-information.pdf

Berque, A. (1999). *Médiance de milieux en paysages*. (S.l.): Belin.

Berque, A. (2002). Indigènes au-delà de l'exotisme. *Diogène*, *200*(4), 46-57. https://doi.org/10.3917/dio.200.0046

Berque, A. (2009). Chapitre 5. Sens. Dans *Écoumène* (pp. 187-238). Paris: Belin. Repéré à https://www.cairn.info/ecoumene--9782701151779-p-187.htm

Berque, A. (2014). Chapitre III. La médiance. Dans *La mésologie : Pourquoi et pour quoi faire ?* (pp. 31-37). Nanterre: Presses universitaires de Paris Nanterre. https://doi.org/10.4000/books.pupo.10793

Berret, C., & Phillips, C. (2016). *Teaching data and computational journalism*. New York, NY, USA: Columbia Journalism School, Knight Foundation. Repéré à https://journalism.columbia.edu/system/files/content/teaching_data_and_computational_journalism.pdf

Berry, D. M. (2011). The Computational Turn: Thinking About the Digital Humanities.

Bert, J.-F. (2010). Éléments pour une histoire de la notion de civilisation. *Vingtieme Siecle. Revue d'histoire*, *n° 106*(2), 71-80. Repéré à https://www.cairn.info/revue-vingtieme-siecle-revue-d-histoire-2010-2-page-71.htm

Bertaux, D. (1980). L'approche Biographique: Sa Validité Méthodologique, Ses Potentialités. *Cahiers Internationaux de Sociologie*, *69*, 197-225. Repéré à https://www.jstor.org/stable/40689912

Bertaux, D. (2016). 1. La perspective ethnosociologique. Dans *Le récit de vie* (Vol. 4e éd., pp. 17-37). Paris: Armand Colin. Repéré à https://www.cairn.info/le-recit-de-vie--9782200601614-p-17.htm

Berthier, D. (2020). Data Publica, une alliance au service de la donnée publique. *Smart City Mag*. Repéré à https://www.smartcitymag.fr/article/528/data-publica-une-alliance-au-service-de-la-donnee-publique

Bertin, J. (1967). *Sémiologie graphique. Les diagrammes, les réseaux, les cartes*. Repéré à https://www.decitre.fr/livres/semiologie-graphique-9782713220272.html

Bertrand, Y. (2011). *Culture organisationnelle*. (S.l.): PUQ.

Beynon-Davies, P. (2016). Instituting facts: Data structures and institutional order. *Information and Organization*, *26*(1), 28-44. https://doi.org/10.1016/j.infoandorg.2016.04.001

Beynon-Davies, P. (2021). *Business Analysis and Design: Understanding Innovation in Organisation* (1st ed. 2021 edition). (S.l.): Palgrave Macmillan.

Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. Repéré à https://dspace.mit.edu/handle/1721.1/123471

Blanchet, A., & Gotman, A. (2012). *L'entretien. L'enquête et ses méthodes* (2e éd.). Paris: Armand Colin. Repéré à https://www.decitre.fr/livres/l-entretien-9782200248093.html

Blodgett, J. H., & Schultz, C. K. (1969). Herman hollerith: data processing pioneer. *American Documentation*, *20*(3), 221-226. https://doi.org/10.1002/asi.4630200307

Bocala, C., & Boudett, K. (2015). Teaching Educators Habits of Mind for Using Data Wisely. *Teachers College Record*, *117*(4), 1-20. Repéré à https://www.tcrecord.org/Content.asp?ContentId=17853

Bodin, F. (2018, 17 octobre). Apprendre avec les données numériques, les datasprints pédagogiques : le cas de Traces de Soldats. *Digital Praxis*. Repéré à https://medium.com/digital-praxis/apprendre-avec-les-donn%C3%A9es-num%C3%A9riques-avec-les-datasprints-p%C3%A9dagogiques-le-cas-de-traces-de-50475fbc6b8e

Bodin, F. (2020). La littératie des données comme enjeu d'éducation, d'action et de construction d'une citoyenneté.... *Medium*. Repéré à https://medium.com/@franckbodin/la-litt%C3%A9ratie-des-donn%C3%A9es-comme-enjeu-d%C3%A9ducation-d-action-et-de-construction-d-une-citoyennet%C3%A9-9cd48779470d

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 1857. https://doi.org/10.1073/pnas.1807180116

Bornstein, S. (2023, 12 février). Interview de Simon Chignard : « L'open data n'est pas l'outil de la transparence ». *42mag.fr*. Repéré à https://42mag.fr/2023/02/interview-de-simon-chignard-lopen-data-nest-pas-loutil-de-la-transparence/

Bostrom, N. (2017). *Superintelligence*. Malakoff: Dunod.

Bothorel, E. (2020). *Mission Bothorel : pour une politique publique de la donnée*. Paris: Assemble Nationale. Repéré à https://www.gouvernement.fr/remise-du-rapport-sur-la-politique-publique-de-la-donnee-des-algorithmes-et-des-codes-sources

Bouchard, R., & Kadi, L. (2012). Didactiques de l'écrit et nouvelles pratiques d'écriture. *Le Français dans le Monde, Recherches et applications*, (51), 168. Repéré à http://fipf.org/sites/fipf.org/files/n51_001-168_web.pdf

Bounegru, L., Chambers, L., & Gray, J. (2012). *The Data Journalism Handbook 1*. (S.l.): European Journalism Centre. Repéré à https://s3.eu-central-1.amazonaws.com/datajournalismcom/handbooks/The-Data-Journalism-Handbook-1.pdf

Bourcier, M.-H. (2012). *Comprendre le féminisme: Guide graphique*. (S.l.): Max Milo.

Bourdieu, P. (2014). *Langage et pouvoir symbolique*. Paris: Points.

Bowker, G. C. (2013). Data Flakes: An Afterword to « Raw Data » Is an Oxymoron. Dans *« Raw Data » Is an Oxymoron* (pp. 167-171). (S.l.): MIT Press. Repéré à https://ieeexplore.ieee.org/document/6462159

Boyd, danah, & Crawford, K. (2012). Critical questions for big data : Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878

Breiman, L. (2001). Statistical modeling: The two cultures. *undefined*. Repéré à https://www.semanticscholar.org/paper/Statistical-modeling%3A-The-two-cultures-Breiman/e5df6bc6da5653ad98e754b08f63326c2e52b372

Bressler, S. (2020). Fondements et enjeux de la data-philosophie. *Management & Datascience*, *4*(4). https://doi.org/10.36863/mds.a.13148

Bressler, S. (2023). *Data-Philosophie*. Paris: LA ROUTE.

Bridges, L. E. (2021). Digital failure: Unbecoming the « good » data subject through entropic, fugitive, and queer data. *Big Data & Society*, *8*(1), 205395172097788. https://doi.org/10.1177/2053951720977882

Bridle, J. (2018). Opinion: Data isn't the new oil — it's the new nuclear power. *ideas.ted.com*. Repéré à https://ideas.ted.com/opinion-data-isnt-the-new-oil-its-the-new-nuclear-power/

Brigham, C. C. (1935). *Examining Fellowship Applicants: A Report Made to the Social Science Research Council on the Method of Selecting Fellows for First-year Graduate Study*. (S.l.): Social science research council.

Brochier, E. (2016). Toute donnée est une perte d'information. Dans *La donnée n'est pas donnée : stratégie & big data* (pp. 41-52). Paris: Editions Kawa.

Brockmeier, J., & Olson, D. R. (2002). What is a culture of literacy ? Dans *Literacy, narrative and culture* (pp. 1-15). (S.l.): Routledge.

Brotcorne, P., & Valenduc, G. (2009). Les compétences numériques et les inégalités dans les usages d'internet. Comment réduire ces inégalités ? *Les Cahiers du numérique*, *5*(1), 45-68. Repéré à https://www.cairn.info/revue-les-cahiers-du-numerique-2009-1-page-45.htm

Brown, S. (2021). How to build data literacy in your company. *corporateinnovation.mit.edu*. Repéré à https://corporateinnovation.mit.edu/wp-content/uploads/2021/03/How-to-build-data-literacy-in-your-company-MIT-Sloan.pdf

Browne, K. (2016). Queer Quantification or Queer(y)ing Quantification. Dans (pp. 231-250). https://doi.org/10.4324/9781315603223-15

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. (S.l.): Duke University Press. https://doi.org/10.1215/9780822375302

Bruce, C. (1997). The Seven Faces of Information Literacy.

Brugger, P. (2001). From haunted brain to haunted science: A cognitive neuroscience view of paranormal and pseudoscientific thought. *Hauntings and Poltergeists: Multidisciplinary Perspectives*, 195-213.

Brunton, F., & Nissenbaum, H. (2015). *Obfuscation: A User's Guide for Privacy and Protest*. Cambridge, MA, USA: MIT Press.

Buckland, M. K. (1991). Information as thing. *Journal of the American Society for Information Science*, *42*(5), 351-360. https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C351::AID-ASI5%3E3.0.CO;2-3

Burstein, L. (1983). The Use of Existing Data Bases in Program Evaluation and School Improvement. *Educational Evaluation and Policy Analysis*, *6*(3), 307-318. https://doi.org/10.3102/01623737006003307

Calvi, T. (2018). Le marché de l'emploi français des Data Scientists, une infographie de Joblift. *ActuIA*. Repéré à https://www.actuia.com/actualite/le-marche-de-lemploi-francais-des-data-scientists-une-infographie-de-joblift/

Calzada-Prado, F. J., & Marzal, M. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*, *63*. https://doi.org/10.1515/libri-2013-0010

Cao, L. (2017). Data Science: A Comprehensive Overview. *ACM Computing Surveys*, *50*(3), 43:1-43:42. https://doi.org/10.1145/3076253

Capelle, C. (2022). ODEFI devient le GTNum DEFI. *Data culture*. Repéré à https://doi.org/10.58079/nha7

Capelle, C., Lehmans, A., & Liquete, V. (2017). De la visibilité à la médiation : l'open data et ses usages en éducation [Colloque international]. Colloque international communication présentée au « Big data et visibilité en ligne : un enjeu pluridisciplinaire de l'économie numérique », Fort de France, Martinique.

Cardon, D. (2015). *A quoi rêvent les algorithmes. Nos vies à l'heure*. Paris: Le Seuil.

Carey, M., Grainger, P., & Christie, M. (2018). Preparing preservice teachers to be data literate: a Queensland case study. *Asia-Pacific Journal of Teacher Education*, *46*(3), 267-278. https://doi.org/10.1080/1359866X.2017.1402860

Carlson, J., Fosmire, M., Miller, C., & Nelson, M. S. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Libraries Faculty and Staff Scholarship and Research*. Repéré à https://docs.lib.purdue.edu/lib_fsdocs/23

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. *Internet Policy Review*, *9*(2), 1-22. https://doi.org/10.14763/2020.2.1481

Casemajor Loustau, N. (2023). Cultures de données et publics de données : conceptualisation critique. Dans Millerand, F., Coutant, A., L.-. Toth, G., & Millette (Éds), *Datafication et publics de données. Penser la mise en donnée de la société*. (S.l.): Presses de l'Université de Montréal. Repéré à https://hal.science/hal-04254885

Casemajor Loustau, N. (2024). Cultures de données : regard transatlantique sur la mise en données des contenus et des publics. Dans *Culture en régime numérique. Questionner les pratiques, catégories et méthodes*. Repéré à https://hal.science/hal-04426291

Castex, J. (2021). Circulaire n°6264/SG du 27 avril 2021 relative à la politique publique de la donnée, des algorithmes et des codes sources. [Circulaire]. Repéré à https://www.legifrance.gouv.fr/circulaire/id/45162

Chanel, S. (2024, 22 janvier). Yann Legrand, directeur du pôle Data & IA d'Epsilon France. *Stratégies*. Repéré à https://www.strategies.fr/actualites/mouvements/LQ2800098C/yann-legrand-directeur-du-pole-data-ia-depsilon-france.html

Chapron, H. (2021, 1 février). Le Data Literacy manager : définition de fonction. *Cadre & Dirigeant Magazine*. Repéré à https://www.cadre-dirigeant-magazine.com/manager/le-data-literacy-manager-latout-cle-des-entreprises-pilotees-par-la-donnee/

Chapron, H. (2024, 20 février). Investir dans l'IA – oui, mais après un travail sur la donnée. *Journal du net*. Repéré à https://www.journaldunet.com/intelligence-artificielle/1528369-investir-dans-l-ia-oui-mais-apres-un-travail-sur-la-donnee/

Charaudeau, P. (2014). *Le discours politique. Les masques du pouvoir*. Repéré à https://www.decitre.fr/livres/le-discours-politique-9782359351309.html

Charles, V., & Gherman, T. (2013). Achieving Competitive Advantage Through Big Data. Strategic Implications.

Chartier, A.-M., & Rockwell, E. (2013). Histoire comparée des outils et débats sur la lecture des débutants : alphabétisation vs literacy. *Histoire de l'éducation*, (138), 5-16. https://doi.org/10.4000/histoire-education.2647

Chartron, G., & Broudoux, É. (2015). Enjeux géopolitiques des données, asymétries déterminantes. Dans *Big Data - Open Data : Quelles valeurs ? Quels enjeux ?* (pp. 65-83). Louvain-la-Neuve: De Boeck Supérieur. https://doi.org/10.3917/dbu.chron.2015.01.0065

Chauvigné, C., & Coulet, J.-C. (2010). L'approche par compétences : un nouveau paradigme pour la pédagogie universitaire ? *Revue française de pédagogie. Recherches en éducation*, (172), 15-28. https://doi.org/10.4000/rfp.2169

Chignard, S., & Benyayer, L.-D. (2015). *Datanomics Les nouveaux business models des données* (1er édition). Limoges: FYP EDITIONS.

Chignard, S., & Marchandise, J.-F. (2012). *L'Open data : Comprendre l'ouverture des données publiques* (1er édition). Paris: FYP éditions.

Chirio, M. (2007). Le pouvoir en un mot : les militaires brésiliens et la « révolution » du 31 mars 1964. *Nuevo Mundo Mundos Nuevos. Nouveaux mondes mondes nouveaux - Novo Mundo Mundos Novos - New world New worlds*. https://doi.org/10.4000/nuevomundo.3887

Cimatti, B. (2016). Definition, development, assessment of soft skills and their role for the quality of organizations and enterprises. *International Journal for Quality Research*, *10*, 97-130. https://doi.org/10.18421/IJQR10.01-05

Claes, A., & Philippette, T. (2020). Defining a critical data literacy for recommender systems: A media-grounded approach. *Journal of Media Literacy Education*, *12*(3), 17-29. https://doi.org/https://doi.org/10.23860/JMLE-2020-12-3-3

Clara. (2024, 29 février). La data literacy: définition, enjeux et comment la développer en entreprise. *Haskn*. Repéré à https://www.haskn.com/data-literacy/

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, *31*(5), 498-512. https://doi.org/10.1145/42411.42413

Clarke, R., & Greenleaf, G. (2017). Dataveillance Regulation: A Research Framework. *Journal of Law, Information and Science*, *25*, 104. Repéré à https://heinonline.org/HOL/Page?handle=hein.journals/jlinfos25&id=114&div=&collection=

Cleveland, W. S. (2001). Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review*, *69*(1), 21-26. https://doi.org/10.1111/j.1751-5823.2001.tb00477.x

Cohen, M. S. R. (1958). *Marcel Cohen. La Grande invention de l'écriture et son évolution*.

Cohen, R. (2006). BI Strategy : What's in a Name? Data Governance Roles, Responsibilities and Results Factors. *DM Review*, (8).

Cohen, S. (2021). 23. Ways of Doing Data Journalism. Dans *23. Ways of Doing Data Journalism* (pp. 157-161). (S.l.): Amsterdam University Press. https://doi.org/10.1515/9789048542079-024

Coiro, J., Knobel, M., Lankshear, C., & Leu, D. J. (2008). *Handbook of Research on New Literacies* (Reprint édition). New York: Routledge.

Collins, J., & Blot, R. (2003). *Literacy and Literacies: Texts, Power, and Identity*. New York: Cambridge University Press.

Comité national pilote d'éthique du numérique. (2022). *Pour une éthique du numérique*. Paris: PUF. Repéré à https://www.puf.com/pour-une-ethique-du-numerique

Commission Européenne. (2020a). *Communication de la commission au parlement européen, au conseil, au comité économique et social européen et au comité des régions : Une stratégie européenne*

pour les données. Commission Européenne. Repéré à https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A52020DC0066

Commission Européenne. (2020b). *Indice relatif à l'économie et à la société numériques (DESI) 2020*. Bruxelles, Belgique: Commission Européenne. Repéré à https://digital-strategy.ec.europa.eu/en/policies/desi-france

Conrad, N. J. (2018). Is Digital Literacy Changing the Way We Think? *LAnnee psychologique*, *Vol. 118*(4), 371-376. Repéré à https://www.cairn.info/revue-l-annee-psychologique-2018-4-page-371.htm

Conte, S. D., Hamblen, J. W., Kehl, W. B., Navarro, S. O., Rheinboldt, W. C., Young, D. M., & Atchinson, W. F. (1965). An undergraduate program in computer science—preliminary recommendations. *Communications of the ACM*, *8*(9), 543-552. https://doi.org/10.1145/365559.366069

Correia, A., & Água, P. (2021). *A holistic perspective on Data Governance*. https://doi.org/10.22495/cgsetpt12

Correll, M., & Garrison, L. A. (2024, 7 février). When the Body Became Data: Historical Data Cultures and Anatomical Illustration. arXiv. https://doi.org/10.48550/arXiv.2402.05014

Corti, L. (2005). Survey Data in Teaching Project (SDiT): Enhancing Critical Thinking and Data Literacy. *IASSIST Quarterly*, *28*(2), 39. https://doi.org/10.29173/iq796

Couldry, N., & Mejias, U. A. (2019a). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, *20*(4), 336-349. https://doi.org/10.1177/1527476418796632

Couldry, N., & Mejias, U. A. (2019b). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press.

Courbot, C. (2000). De l'acculturation aux processus d'acculturation, de l'anthropologie à l'histoire. Petite histoire d'un terme connoté. *Hypothèses*, *3*(1), 121-129. https://doi.org/10.3917/hyp.991.0121

Crawford, K. (2013). The Hidden Biases in Big Data. *Harvard Business Review*. Repéré à https://hbr.org/2013/04/the-hidden-biases-in-big-data

Crawford, K. (2022). *Contre-atlas de l'intelligence artificielle: Une cartographie politique, sociale et environnementale de l'IA*. Traduction par L. Bury, Paris: ZULMA.

Crenshaw, K. (1989). Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. *University of Chicago Legal Forum*, 139-167.

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, *43*(6), 1241-1299. https://doi.org/10.2307/1229039

Crenshaw, K., & Delanoe, E. (2023). *Intersectionnalité: Deux essais*. Paris: Payot.

Crowther, G. (1959). *The Crowther Report (1959) : A report of the Central Advisory Council for Education (England)* (Rapport No. Vol.1). Londres, Royaume-Uni: Her Majesty's Stationery Office. Repéré à https://education-uk.org/documents/crowther/crowther1959-1.html

Crusoe, D. (2016). Data Literacy defined pro populo: To read this article, please provide a little information. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3276

CSG. (2020). Developing civic data literacy. *Civic Switchboard Guide*. Repéré à https://civic-switchboard.gitbook.io/guide/library-roles/developing-civic-data-literacy

Cuny, D. (2015, 19 janvier). Données personnelles : le sujet polémique du futur projet de loi Numérique. *La Tribune*. Repéré à https://www.latribune.fr/technos-medias/20150119tribcacf5f6de/donnees-personnelles-le-sujet-polemique-du-futur-projet-de-loi-numerique.html

D'Ambrosio, U. (1999). Literacy, Matheracy, and Technocracy: A Trivium for Today. *Mathematical Thinking and Learning*, *1*, 131-153. https://doi.org/10.1207/s15327833mtl0102_3

D'Ignazio, C. (2017). Creative data literacy: Bridging the gap between the data-haves and data-have nots. *Information Design Journal*, *23*(1), 6-18. https://doi.org/10.1075/idj.23.1.03dig

D'Ignazio, C. (2024). *Counting Feminicide: Data Feminism in Action*. (S.l.): The MIT Press.

D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. *bloomberg*, 6.

D'Ignazio, C., & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3280

D'Ignazio, C., & Bhargava, R. (2020). Data visualization literacy: A feminist starting point. Dans M. Engebretsen & H. Kennedy (Éds), *Data Visualization in Society* (pp. 207-222). (S.l.): Amsterdam University Press. https://doi.org/10.2307/j.ctvzgb8c7.19

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. Cambridge, Massachusetts: The MIT Press.

Dalton, C., & Thatcher, J. (2014). What Does A Critical Data Studies Look Like, And Why Do We Care? Repéré à https://www.societyandspace.org/articles/what-does-a-critical-data-studies-look-like-and-why-do-we-care

Dangol, A., & Dasgupta, S. (2023). Constructionist approaches to critical data literacy: A review. Dans *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 112-123). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3585088.3589367

Dani, B.-Z., & Joan, G. (2004). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges. Dans D. Ben-Zvi & J. Garfield (Éds), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 3-15). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-2278-6_1

Data Galaxy. (2019, 18 octobre). Chief Data Officer et culture data, l'alliance indispensable ! - DataGalaxy. *Data Galaxy*. Repéré à https://www.datagalaxy.com/fr/blog/chief-data-officer-et-culture-data/

Data Galaxy. (2021, 24 mai). Glossaire des métiers de la Data. *Data Galaxy*. Repéré à https://www.datagalaxy.com/fr/ressources/glossaire-des-metiers-de-la-data/

Data Publica. (2019, 16 septembre). Création de Data Publica « vos données au service de l'intérêt général ». *DécryptaGéo, l'information géographique*. Repéré à https://decryptageo.fr/creation-de-data-publica-donnees-interet-general/

Data Quality Campaign. (2018). Administrator Data Literacy Fosters Student Success. Repéré à https://dataqualitycampaign.org/resource/administrator-data-literacy-fosters-student-success/

Data Revolution Group. (2014). *A world that counts : mobilising the data revolution for sustainable development*. Independent Expert Advisory Group on a Data Revolution for Sustainable Development. Repéré à https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

DataCamp. (2023). *The State of Data Literacy 2023*. DataCamp.

Datagalaxy. (2021). Pourquoi la culture data est essentielle dans l'entreprise ? Repéré à https://www.datagalaxy.com/fr/blog/importance-culture-data-entreprise/

Datasulting. (2022). Pourquoi créer un pôle data dans son entreprise ? – Datasulting. Repéré à https://www.datasulting.com/articles/pourquoi-donnees-pole-data/

Davenport, T. H., & Patil, D. J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*. Repéré à https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

David, J. (2015). Literacy-Litéracie-littératie : évolution et destinée d'un concept. *Le français aujourd'hui*, *190*(3), 9-22. https://doi.org/10.3917/lfa.190.0009

Davidson, E., Wessel, L., Winter, J. S., & Winter, S. (2023). Future directions for scholarship on data governance, digital innovation, and grand challenges. *Information and Organization*, *33*(1), 100454. https://doi.org/10.1016/j.infoandorg.2023.100454

Davis, C. H. F., Harper, S. R., & Christian, W. A. (2018). Black Data Matter: Connecting Education Research to the Movement for Black Lives. Dans *Envisioning Public Scholarship for Our Time*. (S.l.): Routledge.

Davis, K. (2008). Intersectionality as buzzword: A sociology of science perspective on what makes a feminist theory successful. *Feminist Theory*, *9*(1), 67-85. https://doi.org/10.1177/1464700108086364

Day, R. E. (2014). *Indexing It All: The Subject in the Age of Documentation, Information, and Data*. Cambridge, Massachusetts: (s.n.).

De Saussure, F. (1971). *Cours de linguistique générale*. Paris: Payot.

Deahl, E. (2014). *Better the Data You Know: Developing Youth Data Literacy in Schools and Informal Learning Environments* (Rapport No. ID 2445621). Rochester, NY: Social Science Research Network. https://doi.org/10.2139/ssrn.2445621

Deleuze, G., & Guattari, F. (2005). *Qu'est-ce que la philosophie ?* (Collection "Reprise" édition). Paris: Les Editions de Minuit.

Delgado, R., Stefancic, J., & Harris, A. (2012). *Critical Race Theory: An Introduction, Second Edition*. (S.l.): NYU Press. Repéré à https://www.jstor.org/stable/j.ctt9qg9h2

Delort, P. (2018). *Le Big Data*. (S.l.): Humensis.

Dencik, L., Hintz, A., & Cable, J. (2019). Towards data justice: Bridging anti-surveillance and social justice activism 1. Dans *Data Politics*. (S.l.): Routledge.

Dencik, L., & Sánchez-Monedero, J. (2022). Data justice. *Internet Policy Review*, *11*(1), 16. https://doi.org/10.1007/978-3-319-95960-3_300049

Derrida, J. (1967). *De la grammatologie* (Paris, Les Editions de Minuit). Paris: Editions de Minuit.

Descola, P. (2015). Chapitre III. Le grand partage. Dans *Par-delà nature et culture* (Vol. 2e éd., pp. 114-165). Paris: Gallimard. Repéré à https://www.cairn.info/par-dela-nature-et-culture--9782070465873-p-114.htm

Desrosières, A. (2000). L'histoire de la statistique comme genre : style d'écriture et usages sociaux. *Genèses*, *39*(2), 121-137. https://doi.org/10.3917/gen.039.0121

Desrosières, A. (2010). Introduction — Prendre appui sur des choses. Dans *La politique des grands nombres* (pp. 7-25). Paris: La Découverte. Repéré à https://www.cairn.info/la-politique-des-grands-nombres--9782707165046-p-7.htm

DeVito, M. A. (2021). Adaptive Folk Theorization as a Path to Algorithmic Literacy on Changing Platforms. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 339:1-339:38. https://doi.org/10.1145/3476080

Dewey, J., Truc, G., & Bidet, A. (2011). *La formation des valeurs*. Paris: Empêcheurs de penser rond.

Dhar, V. (2013). Data Science and Prediction. *Communications of the ACM*, pp. 64-73. Repéré à https://cacm.acm.org/magazines/2013/12/169933-data-science-and-prediction/fulltext

Diderot, D., & Alambert, J. L. R. d'. (1751). *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers* (Vol. 5). Paris: (s.n.). Repéré à https://www.zotero.org/groups/667230/avh-r/items/SZRM2GUN

Dijck, J. van. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197-208. https://doi.org/10.24908/ss.v12i2.4776

Dogruel, L. (2021). What is Algorithm Literacy? A Conceptualization and Challenges Regarding its Empirical Measurement. Dans M. Taddicken & C. Schumann (Éds), *Algorithms and Communication* (Vol. 9, pp. 67-93). Berlin: (s.n.). https://doi.org/https://doi.org/10.48541/dcr.v9.3

Dogruel, L., Masur, P., & Joeckel, S. (2021). Development and Validation of an Algorithm Literacy Scale for Internet Users. *Communication Methods and Measures*, *0*(0), 1-19. https://doi.org/10.1080/19312458.2021.1968361

Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, *26*(4), 745-766. https://doi.org/10.1080/10618600.2017.1384734

Douglas, A. S. (1966). IFIP-ICC Vocabulary of Information Processing. *The Computer Journal*, *9*(2), 136. https://doi.org/10.1093/comjnl/9.2.136

Douven, I., Elqayam, S., Gärdenfors, P., & Mirabile, P. (2021). Conceptual Spaces and the Strength of Similarity-based Arguments. *Cognition*. https://doi.org/10.1016/j.cognition.2021.104951

Douyère, D. (2016). De la recherche en communication : matérialité, critique et idéologie(s). *Questions de communication*, (29), 195-221. https://doi.org/10.4000/questionsdecommunication.10461

Dressen, A. (2021). From Digital Literacy to Data Literacy: How Much Digital Literacy Do We Need in the Art History Curriculum? *International Journal for Digital Art History*, (6), 5.02-5.08. https://doi.org/10.11588/dah.2021.6.81735

Droesbeke, J.-J., Vermandele, C., & Didier, E. (2018). *Histoire(s) de(s) données numériques*. (S.l.): EDP Sciences.

Drot-Delange, B., & Tort, F. (2022). Éducation aux données ou enseignement des données : quelles humanités numériques au lycée ? *Humanités numériques*, (5). https://doi.org/10.4000/revueh n.2908

Drucker, J. (2011). Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, *005*(1).

Drucker, P. F. (2005). Be Data Literate -- Know What to Know. *Wall Street Journal*. Repéré à https: //www.wsj.com/articles/SB113208395700897890

Dumont, L. (1987). Individualisme « apolitique » : la « Kultur » dans les Considérations de Thomas Mann. Dans *Sur l'individu* (pp. 38-53). Paris: Le Seuil. https://doi.org/10.3917/ls.veyne.1987.01 .0038

Duncan, L. (2021). *Queer Data: Medical Quantification and What Counts About Counting*. Ph.D. Ann Arbor, United States. Repéré à https://www.proquest.com/docview/2555593344/abstract /6314572AA4F648CDPQ/1

Dupont, P., & Grandaty, M. (2012). Littératie à l'école maternelle : les écrits et oraux réflexifs. *Lettrure*, (Lettrure 2), 50. Repéré à https://univ-tlse2.hal.science/hal-01196494

Durkheim, É. (2020). *Textes 1. Éléments d'une théorie sociale*. (S.l.): Minuit.

Dykes, B. (2017). Why Companies Must Close The Data Literacy Divide. *Forbes*. Repéré à https: //www.forbes.com/sites/brentdykes/2017/03/09/why-companies-must-close-the-data-literacy-divide/

Earl, L., & Katz, S. (2002). Leading Schools in a Data-Rich World. https://doi.org/10.1007/978-94-010-0375-9_34

Educational Weekly. (1884). The Problem of Illiteracy. *Educational Weekly*, *3*(22), 9-9. Repéré à https: //www.jstor.org/stable/44009068

Edwards, P. N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA, USA: MIT Press.

Eggers, W. D. (2013). Data as the new currency. *Deloitte Insights*. Repéré à https://www2.deloitte. com/us/en/insights/deloitte-review/issue-13/data-as-the-new-currency.html

Elias, N., & Kamnitzer, P. (2003). *La civilisation des moeurs*. Paris: Pocket.

Elisa Raffaghelli, J. (2020). Is Data Literacy a Catalyst of Social Justice? A Response from Nine Data Literacy Initiatives in Higher Education. *Education Sciences*, *10*(9), 233. https://doi.org/10.3390/ educsci10090233

Ellul, J. (1987). Peut-Il Exister Une «Culture Technicienne»? *Revue Internationale de Philosophie*, *41*(161 (2)), 216-233. Repéré à https://www.jstor.org/stable/23946736

Ellul, J., & Porquet, J.-L. (2012). *Le système technicien*. (S.l.): Cherche midi.

Ellul, J., & Porquet, J.-L. (2020). *Le bluff technologique*. Repéré à https://www.fayard.fr/pluriel/le-bluff-technologique-9782818502273

Engelsen, K. S., & Smith, K. (2014). Assessment Literacy. Dans C. Wyatt-Smith, V. Klenowski, & P. Colbert (Éds), *Designing Assessment for Quality Learning* (pp. 91-107). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-5902-2_6

Eribon, D. (2003). *Dictionnaire des cultures gays et lesbiennes*. Paris: Larousse.

Euclide, & Peyrard, F. (François). (1814). *Les oeuvres d'Euclide*. (S.l.): Paris, Chez M. Patris. Repéré à http://archive.org/details/lesuvresdeuclide03eucl

Eurostat. (2023). EU digital skills divide: cities outpace rural areas - Products Eurostat News - Eurostat. Repéré à https://ec.europa.eu/eurostat/en/web/products-eurostat-news/w/ddn-20230320-2

Falk, P. (2021). Towards a Public Sector Data Culture: Data as an Individual and Communal Resource in Progressing Democracy. Dans G. Concilio, P. Pucci, L. Raes, & G. Mareels (Éds), *The Data Shake: Opportunities and Obstacles for Urban Policy Making* (pp. 35-45). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-63693-7_3

Farjat, G. (2004). *Pour un droit économique* (1er éd.). Paris: Presses Universitaires de France. Repéré à https://www.lgdj.fr/pour-un-droit-economique-9782130545071.html

Federal Data Strategy. (2020). *FDS Framework : Mission, Principles, Practices, and Actions*. Washington, USA: Federal Data Strategy. Repéré à https://strategy.data.gov/assets/docs/2020-federal-data-strategy-framework.pdf

Février, J. (1995). *Histoire de l'écriture* (Reprod. en fac-sim édition). Paris: Payot.

Fiaux, C. (2020). La datalphabétisation devient un enjeu majeur au sein des entreprises. *https://www.daf-mag.fr/*. Repéré à https://www.daf-mag.fr/Thematique/data-methodologie-1238/Breves/datalphabetisation-devient-enjeu-majeur-sein-entreprises-350798.htm#&utm_source=social_share&utm_medium=share_button&utm_campaign=share_button

Floridi, L. (2009). Philosophical Conceptions of Information. Dans G. Sommaruga (Éd.), *Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information* (pp. 13-53). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-00659-3_2

Floridi, L. (2011). *The Philosophy of Information*. Oxford ; New York: OUP Oxford.

Floridi, L. (Éd.). (2015). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. (S.l.): Springer International Publishing. https://doi.org/10.1007/978-3-319-04093-6

Floridi, L., Cath, C., & Taddeo, M. (2019). Digital Ethics: Its Nature and Scope. Dans (pp. 9-17). https://doi.org/10.1007/978-3-030-17152-0_2

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160360. https://doi.org/10.1098/rsta.2016.0360

Fœssel, M. (2015). Un monde de données. *Libération*. Repéré à https://www.liberation.fr/chroniques/2015/05/01/un-monde-de-donnees_1282052/

Fontichiaro, K., & Oehrli, J. A. (2016). Why Data Literacy Matters. *Knowledge Quest*, *44*(5), 21-27. Repéré à https://eric.ed.gov/?id=EJ1099487

Fontichiaro, K., Oehrli, J. A., & Lennex, A. (2017). *Creating Data Literate Students*. (S.l.): Maize Books/Michigan Publishing.

Forquin, J.-C. (1996). *Ecole et culture, 2e édition. Le point de vue des sociologues britanniques* (2e édition). Sainte-Foy: De Boeck.

Forrester Consulting. (2021). *The Great Data Literacy Gap: Demand For Data Skills Exceeds Supply Need For Data Skills Are On The Rise, Can Academia Accelerate Learning To Meet Them?* Repéré à https://www.tableau.com/sites/default/files/2021-06/Tableau_Data_Literacy_Report.pdf

Fotopoulou, A. (2020). Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility. *Information, Communication & Society*, *0*(0), 1-18. https://doi.org/10.1080/1369118X.2020.1716041

Foucault, M. (1994). *Dits et Ecrits, 1954-1988. Tome III. 1976-1979*. (S.l.): Gallimard. Repéré à https://www.amazon.fr/Dits-Ecrits-1954-1988-III-1976-1979/dp/2070739880

Fourcade, M., & Healy, K. (2016). Seeing like a market. *Socio-Economic Review*, *15*(1), 9-29. https://doi.org/10.1093/ser/mww033

Fox, N. J. (2011). Boundary Objects, Social Meanings and the Success of New Technologies. *Sociology*, *45*(1), 70-85. https://doi.org/10.1177/0038038510387196

Frank, M., & Walker, J. (2016). Some Key Challenges for Data Literacy. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3288

Frank, M., Walker, J., Attard, J., & Tygel, A. (2016). Data Literacy - What is it and how can we make it happen? *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3274

Freire, P. (1968). *La Pédagogie des opprimés* (1er édition). Traduction par I. Pereira, E. Dupau, & M. Kerhoas, (S.l.): Agone.

Freud, S. (2010). *Le Malaise dans la civilisation*. Traduction par A. Weill, Paris: Payot & Rivages.

Frické, M. (2009). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, *35*(2), 131-142. https://doi.org/10.1177/0165551508094050

Frøkjær, E. (2014). Fagfolk på nobelprisniveau: Dårlig plan at fusionere datalogi og matematik. *University Post*. Repéré à https://uniavisen.dk/fagfolk-paa-nobelprisniveau-daarlig-plan-at-fusionere-datalogi-og-matematik/

Fu, X., Wojak, A., Neagu, D., Ridley, M., & Travis, K. (2011). Data governance in predictive toxicology: A review. *Journal of Cheminformatics*, *3*(1), 24. https://doi.org/10.1186/1758-2946-3-24

Gaffiot, F. (1934). litteratus. Dans *Dictionnaire latin-français* (p. 917). Hachette. Repéré à https://www.lexilogos.com/latin/gaffiot.php?q=litteratus

Gal, I. (2002). Adults' Statistical Literacy: Meanings, Components, Responsibilities, 38.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. (S.l.): The MIT Press. https://doi.org/10.7551/mitpress/2076.001.0001

Gärdenfors, P. (2001). Les espaces conceptuels. *Intellectica*, *32*(1), 185-205. https://doi.org/10.3406/intel.2001.1617

Gebre, E. (2022). Conceptions and perspectives of data literacy in secondary education. *British Journal of Educational Technology*, *53*(5), 1080-1095. https://doi.org/10.1111/bjet.13246

Gee, J. P. (1989). Literacy, Discourse, and Linguistics: Introduction. *Journal of Education*, *171*(1), 5-17. https://doi.org/10.1177/002205748917100101

Gemignani, Z., Gemignani, C., Galentino, R., & Schuermann, P. (2014). *Data Fluency: Empowering Your Organization with Effective Data Communication*. (S.l.): John Wiley & Sons.

GEN_SCAN. (2023). Rentrée 2023 : les derniers chiffres de l'emploi et de la formation au numérique en France. *La Grande Ecole du Numérique*. Repéré à https://www.grandeecolenumerique.fr/actualites/rentree-2023-les-derniers-chiffres-de-lemploi-et-de-la-formation-au-numerique-en-france

Gerlitz, C., & Rieder, B. (2014). Techniques of intersection -metrics as connecters and separators in twitter research. *AoIR Selected Papers of Internet Research*. Repéré à https://spir.aoir.org/ojs/index.php/spir/article/view/8890

Geuss, R. (1996). Kultur, Bildung, Geist. *History and Theory*, *35*(2), 151-164. https://doi.org/10.2307/2505359

Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G., & Dey, S. (2023). A systematic literature review of data literacy education. *Journal of Business & Finance Librarianship*, *28*(2), 112-127. https://doi.org/10.1080/08963568.2023.2171552

Gibson, P., & Mourad, T. (2018). The growing importance of data literacy in life science education. *American journal of botany*, *105*(12). https://doi.org/10.1002/ajb2.1195

Gille, B. (1980). *Les sources statistiques de l'histoire de France : Des enquêtes du XVIIe siècle à 1870*. (S.l.): Librairie Droz.

Gillenson, M. L. (1985). Trends in Data Administration. *MIS Quarterly*, *9*(4), 317. https://doi.org/10.2307/249232

Gilster, P. (1998). *Digital Literacy* (New e. édition). New York, NY: John Wiley & Sons.

Gitelman, L. (Éd.). (2013). *« Raw Data » Is an Oxymoron*. Cambridge, MA, USA: MIT Press.

Giust-Desprairies, F. (2009). 2. Représentation, idéologie, imaginaire. Dans *L'imaginaire collectif* (pp. 43-73). Toulouse: Érès. Repéré à https://www.cairn.info/l-imaginaire-collectif--9782749211329-p-43.htm

Goëta, S. (2016). *Instaurer des données, instaurer des publics : une enquête sociologique dans les coulisses de l'open data*. These de doctorat. Paris, ENST. Repéré à https://www.theses.fr/2016ENST0045

Goëta, S. (2024). *Les données de la démocratie : open data, pouvoirs et contre-pouvoirs.* (S.l.): C&F Editions. Repéré à https://www.decitre.fr/livres/les-donnees-de-la-democratie-9782376620716.html

Goodenough, W. H. (1981). *Culture, Language and Society* (2nd Revised edition). Menlo Park, Calif: Benjamin-Cummings Publishing Co.,Subs. of Addison Wesley Longman,US.

Goody, J. (1979). *La raison graphique : la domestication de la pensée sauvage*. Paris: Les Editions de Minuit.

Goody, J. (2006). *La peur des représentations: L'ambivalence à l'égard des images, du théâtre, de la fiction, des reliques et de la sexualité*. Paris: La Découverte.

Goody, J., & Lejosne, J.-C. (2006). La technologie de l'intellect. *Pratiques*, *131*(1), 7-30. https://doi.org/10.3406/prati.2006.2114

Goody, J., & Watt, I. (1963). The Consequences of Literacy. *Comparative Studies in Society and History*, *5*(3), 304-345. https://doi.org/10.1017/S0010417500001730

Gould, R. (2017). Data literacy is statistical literacy, 4.

GPO Mag. (2023). Conseils pour bâtir une culture data dans les entreprises. Repéré à https://www.gpomag.fr/web/tribunes-libres/data-batir-une-culture

Graber, M., Megard Mutezintare, C.-L., & Gakuba, T.-O. (2010). Chapitre IV. Processus d'acculturation. Dans *Les étudiants d'Afrique subsaharienne : Représentations et discours des acteurs des Hautes écoles de la santé et du social sur les processus et les conditions d'apprentissage* (pp. 51-64). Genève: Éditions ies. https://doi.org/10.4000/books.ies.155

Graff, H. J. (1979). *The literacy myth: Literacy and social structure in the nineteenth-century city* (1st edition). New York: Academic Press.

Graff, H. J. (2010). The literacy myth at thirty. *Journal of Social History*, *43*(3), 635-661. Repéré à http://www.jstor.com/stable/20685428

Gran, A.-B., Booth, P., & Bucher, T. (2021). To be or not to be algorithm aware: a question of a new digital divide? *Information, Communication & Society*, *24*(12), 1779-1796. https://doi.org/10.1080/1369118X.2020.1736124

Granger, G.-G. (1955). *Méthodologie économique*. (S.l.): Presses Universitaires de France.

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2), 205395171878631 6. https://doi.org/10.1177/2053951718786316

Green, T. (2008). The Partnering of Church and School in Nineteenth-Century Sweden. *Journal of Church and State*, *50*(2), 331-349. Repéré à https://www.jstor.org/stable/23921524

Grundmann, H. (1958). Litteratus - illitteratus. *Archiv für Kulturgeschichte*, *40*(jg), 1-65. https://doi.org/10.7788/akg-1958-jg02

Guillaud, H. (2012). *Un Monde de données*. S.l: PublieNet.

Guler, G. (2019). *Data literacy from theory to reality: How does it look?*

Gutierrez, D. (2024, 14 mars). Heard on the Street – 3/14/2024. *insideBIGDATA*. Repéré à https://insidebigdata.com/2024/03/14/heard-on-the-street-3-14-2024/

Guyan, K. (2022). *Queer Data*. (S.l.): Bloomsbury Publishing PLC. Repéré à https://www.waterstones.com/book/queer-data/kevin-guyan/9781350230729

Hacking, I. (2002). *Historical Ontology* (Revised edition). Cambridge, Mass.: Harvard University Press.

Hamelink, C. (1976). An Alternative to News. *Journal of Communication*, *26*(4), 120-123. https://doi.org/10.1111/j.1460-2466.1976.tb01947.x

Hand, D. (2018). Aspects of Data Ethics in a Changing World: Where Are We Now? *Big Data*. https://doi.org/10.1089/big.2018.0083

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies*, *14*(3), 575-599. https://doi.org/10.2307/3178066

Haraway, D., Allard, L., Gardey, D., & Magnan, N. (2007). *Manifeste cyborg et autres essais : Sciences - Fictions - Féminismes*. Paris: Editions Exils.

Harfoush, F. A. (2021). Data Ethics: Facts and Fiction. Dans F. A. Harfoush (Éd.), *From Big Data to Intelligent Data: An Applied Perspective* (pp. 87-99). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-76990-1_7

Hargittai, E. (2002). Second-Level Digital Divide: Differences in People's Online Skills. *First Monday*. https://doi.org/10.5210/fm.v7i4.942

Harris, J. (2012). Data Is Useless Without the Skills to Analyze It. *Harvard Business Review*. Repéré à https://hbr.org/2012/09/data-is-useless-without-the-skills

Hartnett, R. J. (1989). *Project Management Software: Proper Selection for Use Within Air Force Systems Command*. AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF SYSTEMS AND LOGISTICS. Repéré à https://apps.dtic.mil/sti/citations/ADA215575

Hautea, S., Dasgupta, S., & Hill, B. M. (2017). Youth Perspectives on Critical Data Literacies. Dans *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 919-930). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3025453.3025823

Havelock, E. A. (1963). *Preface to Plato*. (S.l.): Harvard University Press. https://doi.org/10.2307/j.ctv1smjvnt

Hayashi, C. (1998). What is Data Science ? Fundamental Concepts and a Heuristic Example. Dans C. Hayashi, K. Yajima, H.-H. Bock, N. Ohsumi, Y. Tanaka, & Y. Baba (Éds), *Data Science, Classification, and Related Methods* (pp. 40-51). Tokyo: Springer Japan. https://doi.org/10.1007/978-4-431-65950-1_3

Hayashi, C., Yajima, K., Bock, H.-H., Ohsumi, N., Tanaka, Y., & Baba, Y. (Éds). (1998). *Data Science, Classification, and Related Methods: Proceedings of the Fifth Conference of the International Federation of Classification Societies (IFCS-96), Kobe, Japan, March 27–30, 1996*. Tokyo: Springer Japan. https://doi.org/10.1007/978-4-431-65950-1

Head, A. J., Fister, B., & MacMillan, M. (2020). *Information Literacy in the Age of Algorithms: Student Experiences with News and Information, and the Need for Change*. Project Information Literacy. Repéré à https://eric.ed.gov/?id=ED605109

Heidorn, P. (2008). Shedding Light on the Dark Data in the Long Tail of Science. *Library Trends*, *57*, 280-299. https://doi.org/10.1353/lib.0.0036

Heilinger, J.-C. (2022). The Ethics of AI Ethics. A Constructive Critique. *Philosophy & Technology*, *35*(3), 61. https://doi.org/10.1007/s13347-022-00557-9

Hennion, C., Altounian, M., Monthubet, B., Chignard, S., & Picavet, O. (2023). *Rapport de la mission Data et territoires*. Ministère de la Transformation et de la Fonction Publiques. Repéré à https://www.transformation.gouv.fr/files/ressource/Rapport_Mission_Data_Territoires.pdf

Hepp, A., Jarke, J., & Kramp, L. (2022). New Perspectives in Critical Data Studies: The Ambivalences of Data Power—An Introduction. Dans A. Hepp, J. Jarke, & L. Kramp (Éds), *New Perspectives in Critical Data Studies: The Ambivalences of Data Power* (pp. 1-23). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0_1

Heravi, B. (2018). 3Ws of Data Journalism Education: What, where and who? *Journalism Practice*. https://doi.org/10.1080/17512786.2018.1463167

Herskovits, M. J. (1952). *Les Bases de l'anthropologie culturelle*. Paris: (s.n.). Repéré à https://unesdoc.unesco.org/ark:/48223/pf0000244386

Herzog, D. (2015). *Data Literacy: A User's Guide*. (S.l.): SAGE Publications.

Higounet, C. (2003). *Que sais-je : l'écriture* (11e édition). Paris: Presses Universitaires de France - PUF.

Hillerich, R. L. (1976). Toward an Assessable Definition of Literacy. *The English Journal*, *65*(2), 50. https://doi.org/10.2307/814811

Hillman, V. (2022). Data Privacy Literacy as a Subversive Instrument to Datafication, 22.

HLPP 2015. (2013). *New Global Partnership : eradicate poverty and transform economies through sustainable development.The Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda*. New York, NY, USA: United Nation Publications. Repéré à https://www.un.org/sg/sites/www.un.org.sg/files/files/HLP_P2015_Report.pdf

Høg, E. (2017). GIER: A Danish computer from 1961 with a role in the modern revolution of astronomy. *Proceedings of the International Astronomical Union*, *12*(S330), 92-93. https://doi.org/10.1017/S1743921317005336

Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. *Proceedings of the National Academy of Sciences*, *112*(36), 11199-11204. https://doi.org/10.1073/pnas.1505329112

Hora, M., Bouwma-Gearhart, J., & Park, H. (2017). Data driven decision-making in the era of accountability: Fostering faculty data cultures for learning. *The Review of Higher Education*, *40*, 391-426. https://doi.org/10.1353/rhe.2017.0013

Horkheimer, M. (2021). Le concept de culture (1952). Dans *Philosophie de la culture* (pp. 201-214). Paris: Vrin.

Huawei. (2021). Intelligent_World_2030. *huawei*. Repéré à https://www.huawei.com/fr/news/fr/2021/intelligent_world_2030

Hughes, O. (2022). Employers are desperate for data scientists as demand booms. *ZDNET*. Repéré à https://www.zdnet.com/article/employers-are-desperate-for-data-scientists-as-demand-booms/

Hügi, J., & Schneider, R. (2014). Linked Open Data Literacy for Librarians. Dans (Vol. 492, pp. 109-117). https://doi.org/10.1007/978-3-319-14136-7_12

Hunt, K. (2005). The Challenges of Integrating Data Literacy into the Curriculum in an Undergraduate Institution. *IASSIST Quarterly*, *28*(2), 12. https://doi.org/10.29173/iq791

Ibekwe, F. (2017). Vers la datafication de la société ? Dans *Transition digitale, handicaps et travail social* (LEH Edition, pp. 31-49). (S.l.): Vincent Meyer.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2), 2053951716674238. https://doi.org/10.1177/2053951716674238

International Symposium for literacy. (1975). Déclaration de Persepolis. Repéré à https://en.unesco.org/sites/default/files/persepolis-declaration-literacy-1975-en.pdf

Ionos. (2021). Dark Data. *IONOS Digital Guide*. Repéré à https://www.ionos.fr/digitalguide/web-marketing/vendre-sur-internet/dark-data/

Irgens, G. A., Simon, K., Wise, A., Philip, T., Olivares, M. C., Van Wart, S., … Kahn, J. (2020). Data Literacies and Social Justice: Exploring Critical Data Literacies through Sociocultural Perspectives. Repéré à https://repository.isls.org//handle/1/6665

Iwasiński, Ł. (2020). Theoretical Bases of Critical Data Studies. *Zagadnienia Informacji Naukowej – Studia Informacyjne*. Repéré à https://www.academia.edu/44647685/Theoretical_Bases_of_Critical_Data_Studies

Jacobs, E. (2024, 14 avril). The rise of the chief AI officer. *The Financial Times*. Repéré à https://www.ft.com/content/4c5d823e-004b-4f00-9536-9fea93330453

Jaeger, P. T., Bertot, J. C., Thompson, K. M., Katz, S. M., & DeCoster, E. J. (2012). The Intersection of Public Policy and Public Access: Digital Divides, Digital Literacy, Digital Inclusion, and Public Libraries. *Public Library Quarterly*, *31*(1), 1-20. https://doi.org/10.1080/01616846.2012.654728

Jaffré, J.-P. (2004). La litéracie : histoire d'un mot, effets d'une notion.

Janicaud, D. (1985). *La Puissance du rationnel*. Paris: Gallimard.

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Jaton, F. (2019). « Pardonnez cette platitude » : de l'intérêt des ethnographies de laboratoire pour l'étude des processus algorithmiques. *Zilsel*, *5*(1), 315-339. https://doi.org/10.3917/zil.005.0315

Jaunait, A., & Chauvin, S. (2013). Intersectionnalité. Dans *Dictionnaire. Genre et science politique* (pp. 286-297). Paris: Presses de Sciences Po. https://doi.org/10.3917/scpo.achi.2013.01.0286

Jeanneret, Y. (2017). *Y-a-t-il (vraiment) des technologies de l'information ? : Nouvelle édition revue et corrigée*. Villeneuve d'Ascq: Presses universitaires du Septentrion. https://doi.org/10.4000/books.septentrion.13894

Jeanneret, Y. (2019). L'urbanité, c'est là et ça va donner. *Questions de communication*, (36), 25-42. https://doi.org/10.4000/questionsdecommunication.20807

Jeanneret, Y., & Souchier, E. (2001). Présentation. *Communication & Langages*, *128*(1), 33-33. https://doi.org/10.3406/colan.2001.3072

Jeanneret, Y., & Souchier, E. (2002). La communication médiatisée est-elle un « usage » ? *Communication & Langages*, *132*(1), 5-27. https://doi.org/10.3406/colan.2002.3135

Jennex, M. (2017). Big Data, the Internet of Things, and the Revised Knowledge Pyramid. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *48*, 69-79. https://doi.org/10.1145/3158421.3158427

Jennex, Murray., & Bartczak, Summer. (2013). A Revised Knowledge Pyramid. *International Journal of Knowledge Management (IJKM)*, *9*(3), 19-30. https://doi.org/10.4018/ijkm.2013070102

Johansson, E. (1987). Literacy Campaigns in Sweden. Dans R. F. Arnove & H. J. Graff (Éds), *National Literacy Campaigns: Historical and Comparative Perspectives* (pp. 65-98). Boston, MA: Springer US. https://doi.org/10.1007/978-1-4899-0505-5_4

Johnson, B., Rydal Shapiro, B., DiSalvo, B., Rothschild, A., & DiSalvo, C. (2021). Exploring Approaches to Data Literacy Through a Critical Race Theory Perspective. Dans *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1-15). New York, NY, USA: Association for Computing Machinery. Repéré à https://doi.org/10.1145/3411764.3445141

Johnson, C. A. (2012). *The Information Diet: A Case for Conscious Comsumption* (1st edition). Sebastopol, CA: O'Reilly Media.

Jordan Morrow: Why everyone should be data literate | Jordan Morrow | TEDxBoise | TED Talk [12mn32]. (2019). TED. Repéré à https://www.ted.com/talks/jordan_morrow_why_everyone_should_be_data_literate/transcript

Jung, G. (1996). *Problèmes de l'âme moderne*. Paris: Buchet-Chastel. Repéré à https://www.decitre.fr/livres/problemes-de-l-ame-moderne-9782702013823.html

Kaestle, C. F. (1985). The History of Literacy and the History of Readers. *Review of Research in Education*, *12*, 11-53. https://doi.org/10.2307/1167145

Kant, I. (2001). *Critique de la raison pure*. Traduction par A. Renaut, (S.l.): Garnier-Flammarion.

Kara, M., & Privat, J.-M. (2006). La littératie. Autour de Jack Goody. *Pratiques*, (131/132).

Kayser-Bril, N. (2012). Become Data Literate in 3 Simple Steps. Dans *The Data Journalism Handbook 1* (pp. 126-128). (S.l.): European Journalism Centre.

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *2*(9), 100314. https://doi.org/10.1016/j.patter.2021.100314

Keller, S. A., & Shipp, S. (2021). Data Acumen in Action. *Notices of the American Mathematical Society*, *68*(09), 1. https://doi.org/10.1090/noti2353

Keller, S. A., Shipp, S. S., Schroeder, A. D., & Korkmaz, G. (2020). Doing Data Science: A Framework and Case Study. *Harvard Data Science Review*, *2*(1). https://doi.org/10.1162/99608f92.2d83f7f5

Keyes, O. (2018). The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 88:1-88:22. https://doi.org/10.1145/3274357

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, *53*(1), 148-152. https://doi.org/10.1145/1629175.1629210

Kiemer, K., Gröschner, A., Kunter, M., & Seidel, T. (2018). Instructional and motivational classroom discourse and their relationship with teacher autonomy and competence support—findings from teacher professional development. *European Journal of Psychology of Education*, *33*(2), 377-402. https://doi.org/10.1007/s10212-016-0324-7

Kilpatrick, J. (2001). Understanding mathematical literacy: The contribution of research. *Educational Studies in Mathematics*, *47*(1), 101-116. https://doi.org/10.1023/A:1017973827514

Kitchin, R. (2014a). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1), 2053951714528481. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (1st edition). Los Angeles, California: SAGE Publications Ltd.

Kitchin, R., & Lauriault, T. (2014, 30 juillet). Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. [SSRN Scholarly Paper], Rochester, NY. Repéré à https://papers.ssrn.com/abstract=2474112

Kiyindou, A. (2007). De la fracture numérique à la fracture cognitive : pour une nouvelle approche de la société de l'information. Dans. Communication présentée au Thémat'IC 2007 « La maîtrise de l'information par les adultes : enjeux et méthodes », Strasbourg. Repéré à https://www.enssib.fr/bibliotheque-numerique/notices/1948-de-la-fracture-numerique-a-la-fracture-cognitive-pour-une-nouvelle-approche-de-la-societe-de-l-information

Kleckner, A. (2020). Improving Healthcare Data Literacy. Repéré à https://www.healthcatalyst.com/insights/improving-healthcare-data-literacy/

Klein, J. (2012). *A Commentary on Plato's Meno*. (S.l.): The University of North Carolina Press.

Klock-Fontanille, I. (2014). Penser l'écriture : corps, supports et pratiques. *Communication & langages*, *182*(4), 29-43. https://doi.org/10.3917/comla.182.0029

Klock-Fontanille, I. (2016). Repenser l'écriturePour une grammatologie intégrationnelle. *Actes Sémiotiques*, (119). https://doi.org/10.25965/as.5623

Knaflič, L. (2015). Psychological aspects of literacy. *Libellarium: journal for the research of writing, books, and cultural heritage institutions*, *7*. https://doi.org/10.15291/libellarium.v7i1.195

Kohlegger, M., Maier, R., & Thalmann, S. (2009). Understanding Maturity Models. Results of a Structured Content Analysis. Dans. Communication présentée au I-KNOW '09 and I-SEMANTICS '09. Repéré à https://www.researchgate.net/publication/215312013_Understanding_Maturity_Models_Results_of_a_Structured_Content_Analysis

Koltay, T. (2015). Data literacy: in search of a name and identity. *Journal of Documentation*, *71*(2), 401-415. https://doi.org/10.1108/JD-02-2014-0026

Koltay, T. (2016). Data governance, data literacy and the management of data quality. *IFLA Journal*, *42*(4), 303-312. https://doi.org/10.1177/0340035216672238

Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, *49*(1), 3-14. https://doi.org/10.1177/0961000615616450

Korolov, M. (2020). Why more employees need data literacy skills | TechTarget. *Business Analytics*. Repéré à https://www.techtarget.com/searchbusinessanalytics/feature/Why-more-employees-need-data-literacy-skills

Kõuts-Klemm, R. (2019). Data literacy among journalists: A skills-assessment based approach. *Central European Journal of Communication*, *12*(24), 299-315. Repéré à https://www.ceeol.com/search/article-detail?id=856243

Kremser, W., & Brunauer, R. (2019). Do we have a Data Culture? Dans P. Haber, T. Lampoltshammer, & M. Mayr (Éds), *Data Science – Analytics and Applications* (pp. 83-87). Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-27495-5_11

Kress, G. (2003). *Literacy in the New Media Age*. London: Routledge. https://doi.org/10.4324/9780203299234

Krim, M. (2022). Le data lettrisme devrait être la compétence la plus demandée d'ici 2030. Repéré à https://itsocial.fr/partenaires/mega-international-partenaire/articles-mega/le-data-lettrisme-devrait-etre-la-competence-la-plus-demandee-dici-2030/

Kukutai, T., & Taylor, J. (Éds). (2016). *Indigenous Data Sovereignty: Toward an agenda*. (S.l.): ANU Press. https://doi.org/10.22459/CAEPR38.11.2016

Kuonen, D. (2021). *Advancing data literacy in the post-pandemic world : a primer to catalyse policy dialogue and action*. Paris 21. Repéré à https://paris21.org/sites/default/files/inline-files/Data Literacy_Primer_0.pdf

Kuonen, D. (2022). Littéracie des données: ne laisser personne pour compte! *Unlocking the Power of Data Initiative*. Repéré à https://www.upd-initiative.ch/fr/data-literacy-leave-no-one-behind/

Kuroda, S.-Y. (2012). 5 - La théorie des actes de discours reformulée. Pour une théorie de l'usage du langage. Dans *Pour une théorie poétique de la narration* (pp. 157-174). Paris: Armand Colin. Repéré à https://www.cairn.info/pour-une-theorie-poetique-de-la-narration--9782200272791-p-157.htm

Labelle, S. (2007). *La ville inscrite dans la « société de l'information »: formes d'investissement d'un objet symbolique.* phdthesis. Université Paris IV, Paris. Repéré à https://shs.hal.science/tel-03611133

Labelle, S. (2019). Algorithme. Dans *Comprendre la culture numérique* (pp. 24-30). Paris: Dunod. Repéré à https://www.cairn.info/comprendre-la-culture-numerique--9782100795840-p-24.htm

Labelle, S., Gilliard, A., Népote, C., & Chignard, S. (2017). *Cahier #1 : culture des données. Vers une organisation data-driven : mettre la culture des données au cœur de l'organisation*. Paris: FING. Repéré à https://infolabs.io/sites/default/files/livret1culturedesdonneesvf_weba4.pdf

Labrecque, M.-F. (2016). Féminicide. *Anthropen*. https://doi.org/10.17184/eac.anthropen.011

Ladley, J. (2019). *Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program*. (S.l.): Academic Press.

Ladley, J., & Silverston, L. (2020). The Dangers of Data Illiteracy: Coronavirus and Data Distortions. *DATAVERSITY*. Repéré à https://www.dataversity.net/the-dangers-of-data-illiteracy-coronavirus-and-data-distortions/

Laker, D. R., & Powell, J. L. (2011). The differences between hard and soft skills and their relative impact on training transfer. *Human Resource Development Quarterly*, *22*(1), 111-122. https://doi.org/10.1002/hrdq.20063

Lamizet, B. (1992). *Les lieux de la communication*. (S.l.): Editions Mardaga.

Lanham, R. A. (1995). Digital literacy. *Scientific American*, *3*(273), 198-200. Repéré à https://www.jstor.org/stable/24981795

Lanigan, R. L. (1994). Capta versus Data: Method and Evidence in Communicology. *Human Studies*, *17*(1), 109-130. Repéré à https://www.jstor.org/stable/20011031

Lankshear, C., & Knobel, M. (2006). *New Literacies: Everyday Practices and Classroom Learning* ($ {number}nd édition). Maidenhead; New York: Open University Press.

Lankshear, C., Knobel, M., & Curran, C. (2012). Conceptualizing and Researching « New Literacies ». Dans. https://doi.org/10.1002/9781405198431.wbeal0182

Laprise, J. P. (2006). A State of Constant War: Policy Implications of Data Literacy. Etats-Unis. Repéré à https://www.academia.edu/188757/A_State_of_Constant_War_Policy_Implications_of_Data_Literacy

Latour, B. (2007a). Paris, ville invisible : le plasma. Dans (p. 260). (S.l.): Editions de la Bibliothèque publique d'information. Repéré à https://sciencespo.hal.science/hal-03569660

Latour, B. (2007b). Pensée retenue, pensée distribuée. Dans *Lieux de savoir 1. Espaces et communautés* (Vol. 1, pp. 605-615). Paris: Albin Michel.

Le Deuff, O. (2008). La culture de l'information: Quelles " littératies ' ' pour quelles conceptions de l'information? Dans (pp. 97-116). Communication présentée au VI.ème Colloque international du chapitre français de l'ISKO, IUT de l'Université Paul Sabatier, Toulouse.

Le Deuff, O. (2009). *La culture de l'information en reformation*. These de doctorat. Rennes 2. Repéré à https://theses.fr/2009REN20017

Le Deuff, O. (2012). Littératies informationnelles, médiatiques et numériques : de la concurrence à la convergence ?

Le Deuff, O., & Roumanos, R. (2022). Enjeux définitionnels et scientifiques de la littératie algorithmique : entre mécanologie et rétro-ingénierie documentaire. *tic&société*, (Vol. 15, N° 2-3 | 2ème semestre 2021 - 1er semestre 2022), 325-360. https://doi.org/10.4000/ticetsociete.7105

Le Donut. (2018). DATA Mythes / Atelier de médiation aux données, en partenariat avec la Région Provence-Alpes-Côte d'Azur. *DATA Mythes / Atelier de médiation aux données, en partenariat avec la Région Provence-Alpes-Côte d'Azur*. Repéré à https://www.ledonut-marseille.com/blog/data-mythes-atelier-de-mediation-aux-donnees-en-partenariat-avec-la-region

Le Petitcorps, C., & Desille, A. (2020). La colonialité du pouvoir aujourd'hui : approches par l'étude des migrations. *Migrations Société*, *182*(4), 17-28. https://doi.org/10.3917/migra.182.0017

Leahy, D., & Dolan, D. (2010). History of the European Computer Driving Licence. Dans A. Tatnall (Éd.), *History of Computing. Learning from the Past* (pp. 134-145). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-15199-6_14

Lefebvre, L., Tonnelat, É., Mauss, M., Niceforo, A., & Weber, L. (1930). *Civilisation - le mot et l'idée* (Vol. deuxième fascicule). Paris: La renaissance du livre.

Lehmans, A. (2017). Données ouvertes et redéfinition de la culture de l'information dans les organisations. *Communication et organisation. Revue scientifique francophone en Communication organisationnelle*, (51), 15-26. https://doi.org/10.4000/communicationorganisation.5495

Lehmans, A. (2018). *Pragmatiques et politiques des cultures de l'information* (Rapport No. Volume 2). Université de Rouen. Repéré à https://hal.science/tel-01956567v1/document

Lehmans, A., & Liquète, V. (2021). Les scénarisations pédagogiques comme démarche de médiation entre usagers et dispositifs : l'exemple des données ouvertes en éducation. *Sciences de la société*, (107). https://doi.org/10.4000/sds.13093

Leonelli, S. (2019). Data Governance is Key to Interpretation: Reconceptualizing Data in Data Science. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.17405bb6

Les bonnes pratiques pour développer une culture data dans votre entreprise. (2024, 5 janvier). Repéré à https://www.youtube.com/watch?v=gPVxC7aEwAo

Les échos executive. (2020, 28 janvier). Data : en France, l'absence de compétences fait perdre 9,8 milliards aux entreprises. *Les Echos*. Repéré à https://www.lesechos.fr/idees-debats/leadership-management/data-en-france-labsence-de-competences-fait-perdre-98-milliards-aux-entreprises-1244644

Levien, R., & Maron, M. (1967). A computer system for inference execution and data retrieval. *Communications of the ACM*, *10*, 715-721. https://doi.org/10.1145/363790.363817

Lévi-Strauss, C. (1955). *Tristes tropiques*. Paris: Plon.

Levitin, A. V., & Redman, T. C. (1998). Data as a Resource: Properties, Implications, and Prescriptions. *MIT Sloan Management Review*. Repéré à https://sloanreview.mit.edu/article/data-as-a-resource-properties-implications-and-prescriptions/

Lewandowski, J.-C. (2017). La fièvre des diplômes de big data. *Le Monde.fr*. Repéré à https://www.lemonde.fr/campus/article/2017/02/08/la-fievre-des-diplomes-de-big-data_5076480_4401467.html

Lewison, M., Flint, A. S., & Sluys, K. (2002). Taking on critical literacy: The journey of newcomers and novices. *Language Arts*, *79*, 382-392.

Libre à lire. (2024, 30 avril). Fada de Data avec Joël Gombin. *Libre à lire*. Repéré à https://www.librealire.org/fada-de-data-avec-joel-gombin

Liquete, V. (2014). *Les Cultures de l'information*. Paris: CNRS EDITIONS.

Livingstone, S. (2019). Audiences in an Age of Datafication: Critical Questions for Media Research. *Television & New Media*, *20*(2), 170-183. https://doi.org/10.1177/1527476418811118

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. Dans *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-16). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727

Love, N. (2004). Taking Data to New Depths: There's a Ton of Data Being Collected. The Trick Is to Know How to Use It Effectively. *Journal of Staff Development*, *25*(4), 22-26.

Love, N. (2012). Data Literacy for Teachers, 1.

Lucchesi, L. (2021). Politique de la donnée, des algorithmes et des codes sources : 15 stratégies ministérielles et 500 actions pour accélérer. *Etalab*. Repéré à https://www.etalab.gouv.fr/politique-de-la-donnee-des-algorithmes-et-des-codes-sources-15-strategies-ministerielles-et-500-actions-pour-accelerer

Lupton, D. (2017). Feeling your data: touch and making sense of personal digital data. *New Media & Society*, *19*(10), 1599-1614. https://doi.org/10.1177/1461444817717515

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, *19*(5), 780-794. https://doi.org/10.1177/1461444816686328

Lyu, W., & Liu, J. (2021). Soft skills, hard skills: What matters most? Evidence from job postings. *Applied Energy*, *300*, 117307. https://doi.org/10.1016/j.apenergy.2021.117307

MacFeely, S. (2020). In search of the data revolution: Has the official statistics paradigm shifted? *Statistical Journal of the IAOS*, *36*, 1075-1094. https://doi.org/10.3233/SJI-200662

Mackey, T. P., & Jacobson, T. E. (2017). Reframing Information Literacy as a Metaliteracy | Mackey | College & Research Libraries. https://doi.org/https://doi.org/10.5860/crl-76r1

Maingueneau, D. (2021). Chapitre 1. Quelques éléments d'histoire. Dans *Discours et analyse du discours* (Vol. 2e éd., pp. 3-10). Paris: Armand Colin. Repéré à https://www.cairn.info/discours-et-analyse-du-discours--9782200631710-p-3.htm

Mandinach, E. (2013). Data Literacy vs. Assessment Literacy. *Michael & Susan Dell Foundation*. Repéré à https://www.dell.org/insight/ellen-mandinach-data-literacy-vs-assessment-literacy/

Mandinach, E. B., & Gummer, E. S. (2012). *Navigating the Landscape of Data Literacy: It IS Complex*. (S.l.): WestEd. Repéré à https://eric.ed.gov/?id=ED582807

Mandinach, E. B., & Gummer, E. S. (2016a). *Data Literacy for Educators: Making It Count in Teacher Preparation and Practice*. (S.l.): Teachers College Press.

Mandinach, E. B., & Gummer, E. S. (2016b). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366-376. https://doi.org/10.1016/j.tate.2016.07.011

Mandinach, Ellen., & Gummer, Edith. (2013). A Systemic View of Implementing Data Literacy in Educator Preparation. *Educational Researcher*, *42*(1), 30-37. https://doi.org/10.3102/0013189X12459803

Mann, S. (1998). « Reflectionism » and « Diffusionism »: New Tactics for Deconstructing the Video Surveillance Superhighway. *Leonardo*, *31*(2), 93-102. https://doi.org/10.2307/1576511

Mann, S., Nolan, J., & Wellman, B. (2003). Sousveillance: Inventing and Using Wearable Computing Devices for Data Collection in Surveillance Environments. *Surveillance & Society*, *1*(3), 331-355. https://doi.org/10.24908/ss.v1i3.3344

Mann, T., Lilla, M., & Morris, W. D. (2021). *Reflections of a Nonpolitical Man*. New York City: NYRB Classics.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big Data: The Next Frontier for Innovation, Comptetition, and Productivity.

Markoff, J. (2012). Troves of Personal Data, Forbidden to Researchers. *The New York Times*. Repéré à https://www.nytimes.com/2012/05/22/science/big-data-troves-stay-forbidden-to-social-scientists.html

Marr, B. (2022). The Importance Of Data Literacy And Data Storytelling. *Forbes*. Repéré à https://www.forbes.com/sites/bernardmarr/2022/09/28/the-importance-of-data-literacy-and-data-storytelling/

Martin, A. (2008). Digital literacy and the « digital society ». Dans *Digital Literacies: Concepts, Policies, and Practices* (pp. 151-176). New York: Peter Lang.

Martin, E. W., & Hall, D. J. (1960). Chapter VIII: Data Processing: Automation in Calculation. *Review of Educational Research*, *30*(5), 522-535. https://doi.org/10.3102/00346543030005522

Martin, H. (2007). Mathematical Literacy. *Principal Leadership*, *7*(5), 28-31.

Martín-González, Y., & Iglesias-Rodríguez, A. (2022). Alfabetización en Datos en las bibliotecas-CRAI españolas: Análisis descriptivo y propositivo. *Revista Española de Documentación Científica*, *45*(2), e322-e322. https://doi.org/10.3989/redc.2022.2.1857

Marwick, A. (2012). The Public Domain: Surveillance in Everyday Life. *Surveillance & Society*, *9*. https://doi.org/10.24908/ss.v9i4.4342

Mathematical Association of America. (1994). *Quantitative Reasoning for College Graduates: A Complement to the Standards*. Mathematical Association of America. Repéré à https://maa.org/programs/faculty-and-departments/curriculum-department-guidelines-recommendations/quantitative-literacy/quantitative-reasoning-college-graduates

Mattelart, A. (2018). *Histoire de la société de l'information - Armand Mattelart - Éditions La Découverte* (5e éd.). paris: Repères. Repéré à https://www.editionsladecouverte.fr/histoire_de_la_societe_de_l_information-9782348036170

Matthews, P. (2016). Data literacy conceptions, community capabilities. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3277

Mattmann, C. A. (2013). A vision for data science. *Nature*, *493*(7433), 473-475. https://doi.org/10.1038/493473a

Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1-4. https://doi.org/10.1002/pra2.2015.145052010010 8

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Eamon Dolan / Houghton Mifflin Harcourt.

Mazars. (2023). La Mazars Data School, une école créée par nos équipes pour promouvoir et diffuser la culture Data - Mazars - France. Repéré à https://www.mazars.fr/services/conseil/technology-and-digital-consulting/conseil-en-analyse-de-donnees/la-mazars-data-school

McAuley, D., Rahemtulla, H., Goulding, J., & Souch, C. (2011). How Open Data, data literacy and Linked Data will revolutionise higher education. *Pearson Centre for Policy and Learning*, 88-93. https://doi.org/978-0-997-86378-9

McBride, T. (2010). Organisational theory perspective on process capability measurement scales. *Journal of Software Maintenance and Evolution: Research and Practice*, *22*(4), 243-254. https://doi.org/10.1002/spip.440

McCarthy, J. (2007). From here to human-level AI. *Artificial Intelligence*, *171*(18), 1174-1182. https://doi.org/10.1016/j.artint.2007.10.009

McCosker, A. (2017). Data literacies for the postdemographic social media self. *First Monday*. https://doi.org/10.5210/fm.v22i10.7307

McGlotten, S. (2016). Black Data. https://doi.org/10.1215/9780822373711-014

McGowan, B. S. (2020). OpenStreetMap mapathons support critical data and visual literacy instruction. *Journal of the Medical Library Association*, *108*(4), 649-650. https://doi.org/10.5195/jmla.2020.1070

Mead, M. (1940). Social Change and Cultural Surrogates. *The Journal of Educational Sociology*, *14*(2), 92-109. https://doi.org/10.2307/2262403

Ménissier, T. (2023). Les quatre éthiques de l'intelligence artificielle. *Revue d'anthropologie des connaissances*, *17*(2). https://doi.org/10.4000/rac.29961

Meyran, R. (2009). Genèse de la notion de culture : une perspective globale. *Journal des anthropologues. Association française des anthropologues*, (118-119), 193-214. https://doi.org/10.4000/jda.4188

Meyriat, J. (1985). Information vs communication. Dans *L'Espace social de la communication (concepts et théories)* (pp. 63-89). Paris: Retz & CNRS.

Michelot, F. (2019). Buts et objectifs d'apprentissage de la métalittératie : proposition d'un traduction en français. *Florent Michelot*. Repéré à https://www.michelot.info/education/2019/03/02/buts-et-objectifs-dapprentissage-de-la-metalitteratie-proposition-dun-traduction-en-francais/

Mignolo, W. (2011). Epistemic Disobedience and the Decolonial Option: A Manifesto. *TRANSMODERNITY: Journal of Peripheral Cultural Production of the Luso-Hispanic World*, *1*(2). https://doi.org/10.5070/T4120011807

Mignolo, W. (2013). Géopolitique de la sensibilité et du savoir. (Dé)colonialité, pensée frontalière et désobéissance épistémologique. *Mouvements*, *73*(1), 181-190. https://doi.org/10.3917/mouv.073.0181

Milan, S., & Velden, L. van der. (2016). The Alternative Epistemologies of Data Activism. *Digital Culture & Society*, *2*(2), 57-74. https://doi.org/10.14361/dcs-2016-0205

Miller, S. (2014). Collaborative Approaches Needed to Close the Big Data Skills Gap. [SSRN Scholarly Paper], Rochester, NY. Repéré à https://papers.ssrn.com/abstract=2458930

Miller, sj. (2016). Queer Literacy Framework. Dans N. M. Rodriguez, W. J. Martino, J. C. Ingrey, & E. Brockenbrough (Éds), *Critical Concepts in Queer Studies and Education: An International Guide for the Twenty-First Century* (pp. 259-272). New York: Palgrave Macmillan US. https://doi.org/10.1057/978-1-137-55425-3_26

Miloradov, M., Rakic, S., & Marjanovic, U. (2022). Towards digital economy through data literate workforce. Dans *Companion Proceedings of the Web Conference 2022* (pp. 417-422). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3487553.3524696

Misra, A., Jutting, J., & Kuonen, D. (2021). *Advancing data literacy in the post-pandemic world : A primer to catalyse policy dialogue and action*. The Partnership in Statistics for Development in the 21st Century (PARIS21). Repéré à https://paris21.org/sites/default/files/inline-files/Data Literacy_Primer_0.pdf

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 2053951716679679. https://doi.org/10.1177/2053951716679679

Mokros, J., & Wright, T. (2009). Zoos, Aquariums, and Expanding Students' Data Literacy. *Teaching Children Mathematics*, *15*(9), 524-530. https://doi.org/10.5951/TCM.15.9.0524

Montes, M., & Slater, D. (2019). Issues in Open Data: Data Literacy. Dans *The state of open data : histories and horizons* (African Minds, pp. 274-286). (S.l.): Zenodo. https://doi.org/10.5281/zenodo.2677809

Moreau, A. (2022). Art et technique des données sans fins. Dans *Humains et données. Création, médiation, décision, narration.* (pp. 151-161). Repéré à https://hal.archives-ouvertes.fr/hal-03537951

Morin, O. (2011). Chapitre 1. Transmission et diffusion des traditions. Dans *Comment les traditions naissent et meurent* (pp. 19-63). Paris: Odile Jacob. Repéré à https://www.cairn.info/comment-les-traditions-naissent-et-meurent--9782738127044-p-19.htm

Morning Data Chat. (2023). Adresser les défis organisationnels d'un département data. *Ausha*. Repéré à https://podcast.ausha.co/morning-data-chat/adresser-les-defis-organisationnels-d-un-departement-data

Morrow, J. (2021). *Data Literacy: The Data Literacy Skills Everyone Needs to Succeed*. London ; New York, NY: Kogan Page Ltd.

Morrow, J. (2022). How COVID-19 Exposed Data Literacy Readiness. *The New Stack*. Repéré à https://thenewstack.io/how-covid-19-exposed-data-literacy-readiness/

Mosco, V. (2014). *To the Cloud: Big Data in a Turbulent World*. Boulder: Routledge.

Mosco, V. (2016). Marx in the Cloud. Dans *Marx in the Age of Digital Capitalism* (pp. 516-535). (S.l.): Brill. https://doi.org/10.1163/9789004291393_017

Mourey, D. (2021). Chapitre 6. L'apport de Ian Hacking pour penser les « effets de boucle » des outils de gestion. Dans *Philosophie et outils de gestion* (pp. 142-166). Caen: EMS Editions. https://doi.org/10.3917/ems.gilbe.2021.02.0142

Mumford, D. (2022). Data colonialism: compelling and useful, but whither epistemes? *Information, Communication & Society*, *25*(10), 1511-1516. https://doi.org/10.1080/1369118X.2021.1986103

Nash, C. J. (2016). *Queer Methods and Methodologies: Intersecting Queer Theories and Social Science Research*. London: Routledge. https://doi.org/10.4324/9781315603223

National Academies of Sciences, Engineering, and Medicine. (2018). *Data Science for Undergraduates: Opportunities and Options.* (The National Academies Press.). Washington, USA: (s.n.). https://doi.org/10.17226/25104

National Research Council. (1989). *Everybody Counts: A Report to the Nation on the Future of Mathematics Education*. Washington, D.C.: National Academies Press. https://doi.org/10.17226/1199

Naur, P. (1954). *A view of computers of 1952*. Datalogisk Institut, Københavns Universitet.

Naur, P. (1966). The science of datalogy. *Communications of the ACM*, *9*(7), 485. https://doi.org/10.1145/365719.366510

Naur, P. (1968). « Datalogy », the science of data and data processes. Dans A. J. H. Morrel (Éd.), *Information Processing, Proceedings of IFIP Congress 1968, Edinburgh, UK, 5-10 August 1968, Volume 2 - Hardware, Applications* (pp. 1383-1387).

Naur, P. (1970). Project activity in computer science education. *Calcolo*, *7*(1-2), 1-13. https://doi.org/10.1007/BF02575555

Naur, P. (1974). *Concise Survey of Computer Methods*. (S.l.): Petrocelli Books.

Naur, P. (1992). *Computing, a human activity*. (S.l.): New York : ACM Press ; Reading, Mass. : Addison-Wesley Pub. Co. Repéré à http://archive.org/details/computinghumanacoooo naur

Naur, P. (1996). PARA'96 conference dinner talk 1996 August 20. Dans J. Waśniewski, J. Dongarra, K. Madsen, & D. Olesen (Éds), *Applied Parallel Computing Industrial Computation and Optimization* (Vol. 1184, pp. 529-533). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-62095-8_57

Népote, C. (2016). Organisations : Pourquoi vous devez vous préoccuper de la culture « data » de vos collaborateurs. *L'Usine Digitale*. Repéré à https://www.usine-digitale.fr/article/organisations-pourquoi-vous-devez-vous-preoccuper-de-la-culture-data-de-vos-collaborateurs.N435337

Newman, D., & Logan, D. (2006). Governance Is an Essential Building Block for Enterprise Information Management.

Newton, J. J., & Wahl, D. C. (1993). *Manual for data administration* (Rapport No. NIST SP 500-208). Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.500-208

Ng, T. K., Chu, S., Shen, M., & Leung, J. (2021). AI Literacy: Definition, Teaching, Evaluation and Ethical Issues. Dans.

Nguyen, D. (2021). Mediatisation and datafication in the global COVID-19 pandemic: on the urgency of data literacy. *Media International Australia*, *178*(1), 210-214. https://doi.org/10.1177/1329878X20947563

Niedergang, P. (2023). *Vers la normativité queer*. Toulouse: Blast.

Nieuwbourg, P. (2020). Note de recherche : Glossaire des 70 métiers de la donnée. *Decideo - Actualités sur le Big Data, Business Intelligence, Data Science*. Repéré à https://www.decideo.fr/Note-de-recherche-Glossaire-des-70-metiers-de-la-donnee_a11879.html

Nilsson, N. J. (2009). *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511819346

Nolan, R. L. (1974). *Managing the Data Resource Function* (Subsequent edition). St. Paul: West Group.

Nonnon, É. (2007). Goody Jack. Pouvoirs et savoirs de l'écrit. *Revue française de pédagogie. Recherches en éducation*, (161), 123-126. Repéré à https://journals.openedition.org/rfp/768

Nora, P. (1986). *Les Lieux de mémoire : II - La Nation : 2- Le territoire - L'Etat - Le patrimoine*. Paris: Gallimard.

Nossik, S. (2011). Les récits de vie comme corpus sociolinguistique : une approche discursive et interactionnelle. *Corpus*, (10), 119-135. https://doi.org/10.4000/corpus.2045

Noyes, K. (2024). What's Next for the CDAO Role in an AI Future? *WSJ*. Repéré à https://deloitte.wsj.com/cio/whats-next-for-the-cdao-role-in-an-ai-future-f8fff17c

NSF. (2002). NSF Award Search: Award # 0221415 - Using Data Project: Building Mathematics and Science Education Reform Leaders' Data Literacy. Repéré à https://www.nsf.gov/awardsearch/showAward?AWD_ID=0221415

O'Brien, A. (2023, 10 juillet). Presentation of Conference Paper Submission for CERME-13 TWG-12: The Conditions for the Formation of Numeracy and Related Terms. https://doi.org/10.13140/RG.2.2.21256.80642

O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Reprint édition). New York: Crown.

Obendiek, A. S. (2022). What Are We Actually Talking About? Conceptualizing Data as a Governable Object in Overlapping Jurisdictions. *International Studies Quarterly*, *66*(1), sqab080. https://doi.org/10.1093/isq/sqab080

Oceans of data. (2016). Building Global Interest in Data Literacy: A Dialogue-Workshop Report. Repéré à http://oceansofdata.org/our-work/building-global-interest-data-literacy-dialogue-workshop-report

Okamoto, K. (2017). Introducing Open Government Data. *The Reference Librarian*, *58*(2), 111-123. https://doi.org/10.1080/02763877.2016.1199005

Oliver, G., Cranefield, J., Lilley, S., & Lewellen, M. (2023). Data Cultures: a scoping literature review. *Information Research an international electronic journal*, *28*(1), 3-29. Repéré à https://informationr.net/infres/article/view/324

Ollion, É. (2018). 6. Les sciences sociales, contre la data science ? *Regards croises sur l'economie*, *n° 23*(2), 77-86. Repéré à https://www-cairn-info.proxybib-pp.cnam.fr/revue-regards-croises-sur-l-economie-2018-2-page-77.htm

Olson, D. R. (1998). *L'univers de l'écrit ; comment la culture écrite donne forme à la pensée*. Paris: Editions Retz. Repéré à https://www.boitealivres.com/livre/9782725629926-l-univers-de-l-ecrit-comment-la-culture-ecrite-donne-forme-a-la-pensee-david-r-olson/

Olson, D. R. (2001). Prologue. Dans P. Tynjälä, L. Mason, & K. Lonka (Éds), *Writing as a Learning Tool: Integrating Theory and Practice* (pp. 1-5). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-0740-5_1

Olson, D. R. (2011). Why Literacy Matters, Then and Now. Dans W. A. Johnson & H. N. Parker (Éds), *Ancient Literacies: The Culture of Reading in Greece and Rome* (p. 0). (S.l.): Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199793983.003.0015

Olson, D. R. (2017). Goody et ses critiques. Dans É. Guichard (Éd.), *Écritures : Sur les traces de Jack Goody* (pp. 30-38). Villeurbanne: Presses de l'enssib. Repéré à http://books.openedition.org/pressesenssib/1950

Olson, D. R., & Lejosne, J.-C. (2006). Littératie, scolarisation et cognition. Quelques implications de l'anthropologie de Jack Goody. *Pratiques*, *131*(1), 83-94. https://doi.org/10.3406/prati.2006.2119

Olson, D. R., & Torrance, N. (2009). *The Cambridge Handbook of Literacy* (1er édition). Cambridge, N.Y: Cambridge University Press.

Ouyang, W. (2020). Research on the Role of Algorithm Transparency in Algorithm Accountability. Dans (pp. 234-237). Communication présentée au 2019 3rd International Conference on Education, Economics and Management Research (ICEEMR 2019), Atlantis Press. https://doi.org/10.2991/assehr.k.191221.055

Owens, T. (2011). » Defining Data for Humanists: Text, Artifact, Information or Evidence? *Journal of Digital Humanities*, *1*(1). Repéré à http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/

Padieu, R. (1974). Qu'est-ce que la représentativité ? *Economie et Statistique*, *56*(1), 65-66. https://doi.org/10.3406/estat.1974.1634

Paillé, P., & Mucchielli, A. (2012). Chapitre 12 - L'analyse à l'aide des catégories conceptualisantes. Dans *L'analyse qualitative en sciences humaines et sociales* (pp. 315-374). Paris: Armand Colin. https://doi.org/10.3917/arco.paill.2012.01.0315

Paillé, P., & Mucchielli, A. (2021a). Chapitre 1. Choisir une approche d'analyse qualitative. Dans *L'analyse qualitative en sciences humaines et sociales* (Vol. 5e éd., pp. 13-36). Paris: Armand Colin. Repéré à https://www.cairn.info/l-analyse-qualitative-en-sciences-humaines--9782200624019-p-13.htm

Paillé, P., & Mucchielli, A. (2021b). Chapitre 13. L'analyse à l'aide des catégories conceptualisantes. Dans *L'analyse qualitative en sciences humaines et sociales* (Vol. 5e éd., pp. 359-420). Paris: Armand Colin. Repéré à https://www.cairn.info/l-analyse-qualitative-en-sciences-humaines--9782200624019-p-359.htm

Paillé, P., & Mucchielli, A. (2021c). Chapitre 8. L'examen et l'analyse phénoménologiques des données d'entretien. Dans *L'analyse qualitative en sciences humaines et sociales* (Vol. 5e éd., pp. 183-199). Paris: Armand Colin. Repéré à https://www.cairn.info/l-analyse-qualitative-en-sciences-humaines--9782200624019-p-183.htm

Pangrazio, L., & Selwyn, N. (2019). « Personal data literacies »: A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, *21*(2), 419-437. https://doi.org/10.1177/1461444818799523

Papacharissi, Z. (2015). The unbearable lightness of information and the impossible gravitas of knowledge: Big Data and the makings of a digital orality. *Media, Culture & Society*, *37*(7), 1095-1100. https://doi.org/10.1177/0163443715594103

Papamitsiou, Z., Filippakis, M. E., Poulou, M., Sampson, D., Ifenthaler, D., & Giannakos, M. (2021). Towards an educational data literacy framework: enhancing the profiles of instructional designers and e-tutors of online and blended courses with new competences. *Smart Learning Environments*, *8*(1), 18. https://doi.org/10.1186/s40561-021-00163-w

Papert, S. (1994). *The Children's Machine* (Reprint édition). New York: Basic Books.

Pari, C., & Shor, I. (Éds). (1999). *Critical Literacy in Action: Writing Words, Changing Worlds/A Tribute to the Teachings of Paulo Freire*. Portsmouth, NH: Heinemann.

Parlement Européen. (2006). *Recommandation du Parlement européen et du Conseil du 18 décembre 2006 sur les compétences clés pour l'éducation et la formation tout au long de la vie*. Repéré à http://data.europa.eu/eli/reco/2006/962/oj/fra

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information* (1st edition). Cambridge: Harvard University Press.

Pastoral Times. (1883). The illiteracy of the United States. *Pastoral Times*, p. 3. Repéré à http://nla.gov.au/nla.news-article268024594

Paul, P., & Wang, Y. (2011). *Literate Thought*. (S.l.): Jones & Bartlett Publishers.

Paulos, J. A. (1988). *Innumeracy: Mathematical Illiteracy and Its Consequences* (First Edition). New York: Hill and Wang.

Pearce, R., Erikainen, S., & Vincent, B. (2020). TERF wars: An introduction. *The Sociological Review*, *68*(4), 677-698. https://doi.org/10.1177/0038026120934713

Pedersen, A. Y., & Caviglia, F. (2019). Data Literacy as a Compound Competence. Dans T. Antipova & A. Rocha (Éds), *Digital Science* (pp. 166-173). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-02351-5_21

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming Racially Literate About Data and Data-Literate About Race: Data Visualizations in the Classroom as a Site of Racial-Ideological Micro-Contestations. *Cognition and Instruction*, *34*(4), 361-388. https://doi.org/10.1080/07370008.2016.1210418

Phillips, C. (2021). Teaching Data Journalism. Dans L. Bounegru & J. Gray (Éds), *The Data Journalism Handbook* (pp. 338-343). (S.l.): Amsterdam University Press. https://doi.org/10.2307/j.ctv1qr6smr.49

Piccoli, G., & Pigni, F. (2018). *Information Systems for Managers: Without Cases, Edition 4. 0*. (S.l.): Prospect Press.

Pierre, R. (2003). Entre alphabétisation et littératie : les enjeux didactiques. *Revue française de linguistique appliquée*, *VIII*(1), 121-137. https://doi.org/10.3917/rfla.081.0121

Piggott, S. E. A. (1993). The Virtual Library: Almost There... *Special libraries*, *84*(4), 206-212. Repéré à https://scholarworks.sjsu.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1003&context=sla_sl_1993#page=43

Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, *30*(3), 381-402. https://doi.org/10.1177/0265532213480337

Pin, C. (2023). L'entretien semi-directif. *LIEPP Methods Brief / Fiches méthodologiques du LIEPP*, 6. Repéré à https://sciencespo.hal.science/hal-04087897

Pinède, N. (2020). Les enjeux de la médiation des données a l'heure du numérique L'exemple de la licence professionnelle mind. *Les Cahiers de la SFSIC*. Repéré à http://cahiers.sfsic.org/sfsic/index.php?id=152

Plotkin, D. (2013). *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance* (1st edition). Amsterdam ; Boston: Morgan Kaufmann.

Poirier, L., & Costelloe-Kuehn, B. (2019). Data Sharing at Scale: A Heuristic for Affirming Data Cultures. *Data Science Journal*, *18*, 48. https://doi.org/10.5334/dsj-2019-048

Polizzi, G. (2020a). Digital literacy and the national curriculum for England: Learning from how the experts engage with and evaluate online content. *Computers & Education*, *152*, 103859. https://doi.org/10.1016/j.compedu.2020.103859

Polizzi, G. (2020b). *Digital literacy in theory and practice: Learning from how experts and advocates engage in civic life*. Londres, Royaume-Uni. Repéré à http://etheses.lse.ac.uk/4198/1/Polizzi__Digital-literacy-theory.pdf

Polizzi, G. (2021). Digital and data literacy: comparing children's understanding of data and online privacy with experts' and advocates' data literacy practices [Online resource]. Repéré à https://blogs.lse.ac.uk/parenting4digitalfuture/

Popham, W. J. (2011). Assessment Literacy Overlooked: A Teacher Educator's Confession. *The Teacher Educator*, *46*(4), 265-273. https://doi.org/10.1080/08878730.2011.605048

Pothier, W. G., & Condon, P. B. (2020). Towards data literacy competencies: Business students, workforce needs, and the role of the librarian. *Journal of Business & Finance Librarianship*, *25*(3-4), 123-146. https://doi.org/10.1080/08963568.2019.1680189

Pour une culture critique des données. (2015). Repéré à https://www.youtube.com/watch?v=ek_MjT6J9Y

Precisely Editor, P. (2022). Data Literacy Skills: What Is It and Why It Matters. *Precisely*. Repéré à https://www.precisely.com/blog/data-integrity/data-literacy-what-it-is-and-why-it-matters

Privat, J.-M. (2010). Un bain de littératie. A l'école de la piscine. *Hyper Article en Ligne - Sciences de l'Homme et de la Société*.

Prodromou, T. (2017). Statistical literacy in data revolution era : building blocks and instructional dilemmas. *STATISTICS EDUCATION RESEARCH JOURNAL*, *16*(1), 38-43. https://doi.org/10.52041/serj.v16i1.212

Puschmann, C., & Burgess, J. (2013). The Politics of Twitter Data. [SSRN Scholarly Paper], Rochester, NY. https://doi.org/10.2139/ssrn.2206225

Qin, J., & D'Ignazio, J. (2010). Lessons learned from a two-year experience in science data literacy education. *International Association of Scientific and Technological University Libraries, 31st Annual Conference*. Repéré à https://docs.lib.purdue.edu/iatul2010/conf/day2/5

Qlik. (2018). *Lead with Data : How to Drive Data Literacy in the Enterprise*. Qlik. Repéré à https://www.qlik.com/us/bi/-/media/08F37D711A58406E83BA8418EB1D58C9.ashx?ga-link=datlitreport_resource-library

Qlik. (2022a). *Data Literacy: The Upskilling Evolution Build a workforce ready for the data-driven enterprise*. Qlik. Repéré à https://www.qlik.com/us/-/media/files/resource-library/global-us/direct/datasheets/ds-data-literacy-the-upskilling-evolution-en.pdf

Qlik, C. de. (2022b). La Data Literacy sera la compétence la plus demandée d'ici 2030, en raison de la transformation des environnements de travail par l'IA. *Decideo - Actualités sur le Big Data, Business Intelligence, Data Science*. Repéré à https://www.decideo.fr/La-Data-Literacy-sera-la-competence-la-plus-demandee-d-ici-2030-en-raison-de-la-transformation-des-environnements-de_a12589.html

Quijano, A. (2007). « Race » et colonialité du pouvoir. *Mouvements*, *51*(3), 111-118. https://doi.org/10.3917/mouv.051.0111

Radermacher, W. (2017). The Future Role of Official Statistics. Dans. https://doi.org/10.2785/721672

Rafal, O., & Girard, D. (2023). Maîtriser la #data : un enjeu majeur de 2023. *WENVISION*. Repéré à https://www.wenvision.com/maitriser-la-data-un-enjeu-majeur-de-2023/

Raffaghelli, J. E. (2017). Alfabetizzare ai dati nella società dei big e open data: una sfida formativa. *Formazione & insegnamento*, *15*(3), 299-324. Repéré à https://ojs.pensamultimedia.it/index.php/siref/article/view/2646

Raffaghelli, J. E. (2019). Developing a framework for educator's data literacy in the european context: proposal, implications and debate. Dans (pp. 10520-10530). Communication présentée au 11th International Conference on Education and New Learning Technologies, Palma, Spain. https://doi.org/10.21125/edulearn.2019.2655

Raffaghelli, J. E., Manca, S., Stewart, B., Prinsloo, P., & Sangrà, A. (2020). Supporting the development of critical data literacies in higher education: building blocks for fair data cultures in society. *International Journal of Educational Technology in Higher Education*, *17*(1), 58. https://doi.org/10.1186/s41239-020-00235-w

Raffaghelli, J. E., & Sangrà, A. (Éds). (2023). *Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead* (Vol. 59). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24193-2

Raley, R. (2013). Dataveillance and Countervailance. Repéré à https://escholarship.org/uc/item/2b12683k

Rebouillat, V. (2021). Le partage des données vu par les chercheurs : une approche par la valeur. *Les Enjeux de l'information et de la communication*, *22/1*(1), 35-53. https://doi.org/10.3917/enic.030.0035

Recherche Data Gouv. (2024). Ateliers de la donnée : des services généralistes sur tout le territoire. Repéré à https://recherche.data.gouv.fr/fr/page/ateliers-de-la-donnee-des-services-generalistes-sur-tout-le-territoire

Redfield, R., Linton, R., & Herskovits, M. J. (1936). Memorandum for the Study of Acculturation. *American Anthropologist*, *38*(1), 149-152. https://doi.org/10.1525/aa.1936.38.1.02a00330

Redman, T. C. (1995). Improve Data Quality for Competitive Advantage. *MIT Sloan Management Review*. Repéré à https://sloanreview.mit.edu/article/improve-data-quality-for-competitive-advantage/

Resnick, L. B. (1990). Literacy in School and Out. *Daedalus*, *119*(2), 169-185. Repéré à https://www.jstor.org/stable/20025305

Résovilles. (2022, 8 juin). [Webinaire] Retour sur 3 ans d'expérimentation du programme Data et quartiers | 2019-2022. *Résovilles*. Repéré à https://www.resovilles.com/le-programme-data-et-quartiers-2-et-15-avril-2021/

Résovilles, & Civitéo. (2022). *Data & Quartiers : le livre blanc. Retour sur trois ans d'expérimentation*. ANCT. Repéré à https://www.resovilles.com/data-quartiers-le-livre-blanc/

Rey, O. (2016). *Quand le monde s'est fait nombre*. Paris: Stock.

Ribeiro, J. (2021). A gentle Introduction to Data Literacy. *Medium*. Repéré à https://towardsdatascience.com/a-gentle-introduction-to-data-literacy-98b950ac68fa

Ricaurte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. *Television & New Media*, *20*(4), 350-365. https://doi.org/10.1177/1527476419831640

Richterich, A. (2018). *The Big Data Agenda : Data Ethics and Critical Data Studies*. (S.l.): University of Westminster Press. https://doi.org/10.16997/book14

Rickert, H., & Orth, E. W. (1997). *Science de la culture et science de la nature / Théorie de la définition*. (S.l.): Gallimard.

Ricoeur, P. (2001). *Le Juste 2*. Paris: Points Essais. Repéré à https://www.editionspoints.com/ouvrage/le-juste-2-paul-ric-ur/9782757895702

Ricœur, P. (1989). L'éthique, la morale et la règle. *Autres Temps*, *24*(1), 52-59. https://doi.org/10.3406/chris.1989.1347

Ricœur, P. (2006). Mémoire, Histoire, Oubli. *Esprit*, *Mars/avril*(3-4), 20-29. https://doi.org/10.3917/espri.0603.0020

Ridley, M., & Pawlick-Potts, D. (2021). Algorithmic Literacy and the Role for Libraries. *Information Technology and Libraries*, *40*(2). https://doi.org/10.6017/ital.v40i2.12963

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., ... Ali-Hassan, H. (2015). *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*. https://doi.org/10.13140/RG.2.1.1922.5044

Ríos, M. L. Y. D. L. (2010). Preface: Feminist Keys for Understanding Feminicide: Theoretical, Political, and Legal Construction. Dans *Preface: Feminist Keys for Understanding Feminicide: Theoretical, Political, and Legal Construction* (pp. xi-xxvi). (S.l.): Duke University Press. https://doi.org/10.1515/9780822392644-001

Roberge, J.-F. (2019). *Cadre de référence de la compétence numérique*. Gouvernement du Québec. Repéré à https://www.education.gouv.qc.ca/fileadmin/site_web/documents/ministere/Cadre-reference-competence-num.pdf

Robert, P. (2020). *L'impensé numérique. Tome 2, Interprétations critiques et logiques pragmatiques de l'impensé*. Paris: Archives contemporaines. Repéré à https://www.decitre.fr/livre-pod/l-impense-numerique-9782813003577.html

Rocher, G. (1970). *Introduction à la sociologie générale. Tome 1 : l'action sociale*. Paris: Points/Livre De Poche.

Rocher, G. (2012). *Introduction à la sociologie générale : L'action sociale*. Ville LaSalle: Bibliotheque Québecoise.

Rosenberg, D. (2013). Data before the Fact. Dans L. Gitelman (Éd.), *« Raw Data » Is an Oxymoron* (pp. 15-40). (S.l.): The MIT Press. https://doi.org/10.7551/mitpress/9302.003.0003

Rouvroy, A., & Berns, T. (2013). Gouvernementalité algorithmique et perspectives d'émancipation. *Reseaux*, *177*(1), 163-196. Repéré à https://www.cairn.info/revue-reseaux-2013-1-page-163.htm

Rouvroy, A., & Stiegler, B. (2015). Le régime de vérité numérique. *Socio. La nouvelle revue des sciences sociales*, (4), 113-140. https://doi.org/10.4000/socio.1251

Ruberg, B., & Ruelos, S. (2020). Data for queer lives: How LGBTQ gender and sexuality identities challenge norms of demographics. *Big Data & Society*, *7*(1), 2053951720933286. https://doi.org/10.1177/2053951720933286

Sabar, R. (2021). How Data Literate Is Your Company? *Harvard Business Review*. Repéré à https://hbr.org/2021/08/how-data-literate-is-your-company

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, *6*(1), 2053951718820549. https://doi.org/10.1177/2053951718820549

Saltz, J. S., & Stanton, J. M. (2017). *An Introduction to Data Science*. Los Angeles: SAGE Publications, Inc.

Sampson, D., Papamitsiou, Z., Ifenthaler, D., Giannakos, M., Mougiakou, S., & Vinatsella, D. (2022). *Educational Data Literacy*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-11705-3

Sander, I. (2020). Critical big data literacy tools—Engaging citizens and promoting empowered internet usage. *Data & Policy*, *2*. https://doi.org/10.1017/dap.2020.5

Sander, I. (2023). Critical datafication literacy – a framework for educating about datafication. *Information and Learning Sciences*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/ILS-06-2023-0064

Sander, W. (2019). The Concept of Education (« Bildung ») as a Cultural Heritage: Transcultural Traditions and Perspectives. *Global Education Review*, *6*(4), 19-30. Repéré à https://eric.ed.gov/?id=EJ1242295

Sauvayre, R. (2013). Chapitre 1. La préparation à l'entretien. Dans *Les méthodes de l'entretien en sciences sociales* (pp. 1-47). Paris: Dunod. https://doi.org/10.3917/dunod.sauva.2013.01.0003

Scandurra, C., Mezza, F., Bochicchio, V., Valerio, P., & Vitelli, R. (2019). Health of Non-binary and Genderqueer People: A Systematic Review. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01453

Scates, D. E. (1943). Statistics—The Mathematics for Social Problems. *The Mathematics Teacher*, *36*(2), 68-78. Repéré à https://www.jstor.org/stable/27952725

Schäfer, D. M. T., & Es, D. K. van (Éds). (2017). *The Datafied Society: Studying Culture through Data* (1st edition). Amsterdam: Amsterdam University Press.

Schank, R. C. (1987). What Is AI, Anyway? *AI Magazine*, *8*(4), 59-65. https://doi.org/10.1609/aimag.v8i4.623

Scheets, F. K. (1995). Improving data-literacy among diocesan administrators. *Improving data-literacy among diocesan administrators*, *55*(1), 369-380.

Schembera, B., & Durán, J. M. (2020). Dark Data as the New Challenge for Big Data Science and the Introduction of the Scientific Data Officer. *Philosophy & Technology*, *33*(1), 93-115. https://doi.org/10.1007/s13347-019-00346-x

Schield, M. (1999). Statistical literacy: Thinking critically about statistics. *Of Significance (APDU)*, *1*, 15-20.

Schield, M. (2002). Three Kinds of Statistical Literacy: What Should We Teach? *International Conference on Teaching Statistics (ICOTS-6)*.

Schield, M. (2004a). Information Literacy, Statistical Literacy and Data Literacy. *IASSIST quarterly / International Association for Social Science Information Service and Technology*, *28*, 7-14. https://doi.org/10.29173/iq790

Schield, M. (2004b). Statistical Literacy Curriculum Design, 21.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482-496. https://doi.org/10.1016/j.tate.2009.06.007

Schneider, R. (2013). Research Data Literacy. Dans S. Kurbanoğlu, E. Grassian, D. Mizrachi, R. Catts, & S. Špiranec (Éds), *Worldwide Commonalities and Challenges in Information Literacy Research and Practice* (Vol. 397, pp. 134-140). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-03919-0_16

Schöpfel, J. (2018). *Vers une culture de la donnée en SHS : une étude à l'université de Lille*. Lille: Université de Lille. Repéré à https://hal.science/hal-01846849v1/document

Schüller, K. (2021). *Datenkompetenz als zentraler Baustein einer Datenstrategie.* München, Allemagne: Carl Hanser Verlag GmbH & Co. KG.

Schüller, K., Rampelt, F., Koch, H., & Schleiss, J. (2023). Better ready than just aware: Data and AI Literacy as an enabler for informed decision making in the data age. Dans.

Seaver, N. (2021). Everything lies in a space: cultural data and spatial reality. *Journal of the Royal Anthropological Institute*, *27*(S1), 43-61. https://doi.org/10.1111/1467-9655.13479

Sedgwick, E. K. (1994). *Tendencies*. Etats-Unis: Duke University Press.

Seignour, A. (2011). Méthode d'analyse des discours. L'exemple de l'allocution d'un dirigeant d'entreprise publique. *Revue française de gestion*, *211*(2), 29-45. https://doi.org/10.3166/rfg.211.29-45

Selwyn, N., & Pangrazio, L. (2018). Doing data differently? Developing personal data tactics and strategies amongst young mobile media users. *Big Data & Society*, *5*(1), 2053951718765021. https://doi.org/10.1177/2053951718765021

Sfez, L. (2002). *Technique et idéologie : Un enjeu de pouvoir*. Paris: Seuil.

Sherzer, J. (2012). Langage et culture : une approche centrée sur le discours. *Langage et societe*, *n° 139*(1), 21-45. Repéré à https://www.cairn.info/revue-langage-et-societe-2012-1-page-21.htm

Shin, D., Rasul, A., & Fotiadis, A. (2021). Why am I seeing this? Deconstructing algorithm literacy through the lens of users. *Internet Research*, *32*(4), 1214-1234. https://doi.org/10.1108/INTR-02-2021-0087

Siefkes, D. (1974). *Lecture notes in computer science*. (S.l.): Springer-Verlag.

Silver, N. (2013). *The Signal and the Noise: The Art and Science of Prediction* (1er édition). London: Penguin.

Simondon, G. (2012). *Du mode d'existence des objets techniques*. Paris: AUBIER.

Simonnot, B. (2018). Conduire des recherches en régime numérique : vers un cadre conceptuel de réflexion éthique. Dans *L'éthique en contexte info-communicationnel numérique* (pp. 11-21). Louvain-la-Neuve: De Boeck Supérieur. https://doi.org/10.3917/dbu.balic.2018.01.0011

Skovsmose, O. (1994). Towards a Critical Mathematics Education. *Educational Studies in Mathematics*, *27*(1), 35-57. Repéré à https://www.jstor.org/stable/3482665

Smolnikova, M. (2022). Building a Data Literate Business Workforce: Dans *Proceedings of the 14th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management* (pp. 221-230). Valletta, Malta: SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0011587600003335

Snipp, C. M. (2016). What does data sovereignty imply: what does it look like? Dans T. Kukutai & J. Taylor (Éds), *Indigenous Data Sovereignty* (Vol. 38, pp. 39-56). (S.l.): ANU Press. Repéré à https://www.jstor.org/stable/j.ctt1q1crgf.10

Souchier, E. (2012). La « lettrure » à l'écran. Lire & écrire au regard des médias informatisés. *Communication & langages*, *174*(4), 85-108. https://doi.org/10.4074/S0336150012014068

Soulages, J.-C. (2015). *L'analyse de discours. Sa place dans les sciences du langage et de la communication* (PU Rennes). (S.l.): Rivages linguistiques. Repéré à https://www.decitre.fr/livres/l-analyse-de-discours-9782753542815.html

Souter, D. (2021). Inside the Digital Society: Data, data everywhere.... *Association for Progressive Communications*. Repéré à https://www.apc.org/en/blog/inside-digital-society-data-data-everywhere

Spinoza. (1677). *Ethique*. Paris: Eclat.

Špiranec, S., Kos, D., & George, M. (2019). Searching for critical dimensions in data literacy. [text], University of Borås. Repéré à http://informationr.net/ir/24-4/colis/colis1922.html

Splunk. (2019). What Is Dark Data? An Introduction to Dark Data. *Splunk*. Repéré à https://www.splunk.com/en_us/data-insider/what-is-dark-data.html

Stair, R., & Reynolds, G. (2020). *Principles of Information Systems* (14th ed). Mason, OH: Cengage. Repéré à https://public.ebookcentral.proquest.com/choice/PublicFullRecord.aspx?p=6914086

Star, S. L. (2010). Ceci n'est pas un objet-frontière ! *Revue d'anthropologie des connaissances*, *4*(1). Traduction par M. B. Bah. https://doi.org/10.3917/rac.009.0018

Steele, B., & Kiliç-Bahi, S. (2008). Quantitative Literacy Across the Curriculum: A Case Study. *Numeracy*, *1*(2). https://doi.org/http://dx.doi.org/10.5038/1936-4660.1.2.3

Stephenson, E., & Schifter Caravello, P. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review*, *35*(4), 525-540. https://doi.org/10.1108/00907320710838354

Sternkopf, H., & Mueller, R. M. (2018). Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations. Dans. https://doi.org/10.24251/HICSS.2018.630

Stiegler, B. (1998). Leroi-Gourhan : l'inorganique organisé. *Les cahiers de mediologie*, *N° 6*(2), 187-194. Repéré à https://www.cairn.info/revue-les-cahiers-de-mediologie-1998-2-page-187.htm

Stiegler, B. (2008). Psychopouvoir et guerre métapsychologique : la question du pharmakon. Dans *Enfants turbulents : l'enfer est-il pavé de bonnes préventions ?* (pp. 119-130). Toulouse: Érès. https://doi.org/10.3917/eres.colle.2008.01.0119

Stiggins, R. (1995). Assessment Literacy for the 21st Century. *Phi Delta Kappan*. Repéré à https://www.semanticscholar.org/paper/Assessment-Literacy-for-the-21st-Century.-Stiggins/4c3cfee71bbe35a7f911a46c668251cca47d1f70

Street, B. (2006). Autonomous and Ideological Models of Literacy : approaches from New Literacy Studies. Repéré à https://www.semanticscholar.org/paper/Autonomous-and-Ideological-Models-of-Literacy-%3A-New-Street/1957884a4cad853a1c6eff5bf148671e45f6af4f

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, *62*(4), 271-286. https://doi.org/10.1037/0003-066X.62.4.271

Suen, L. W., Lunn, M. R., Katuzny, K., Finn, S., Duncan, L., Sevelius, J., … Obedin-Maliver, J. (2020). What Sexual and Gender Minority People Want Researchers to Know About Sexual Orientation and Gender Identity Questions: A Qualitative Study. *Archives of Sexual Behavior*, *49*(7), 2301-2318. https://doi.org/10.1007/s10508-020-01810-y

Sveinsdottir, E., & Frøkjær, E. (1998). Datalogy - The Copenhagen Tradition of Computer Science. *BIT*, *28*, 450-472. https://doi.org/10.1007/BF01941128

Swan, K., Patton, C., & Yarnall, L. (2006). Mathematizing middle school: Results from a cross-disciplinary study of data literacy. *American Educators Research Association Annual Conference 2006*. Repéré à https://www.academia.edu/20169905/Mathematizing_middle_school_Results_from_a_cross_disciplinary_study_of_data_literacy

Swan, K., Vahey, P. J., Kratcoski, A., van, M. A., Rafanan, K., & Stanford, T. (2009). Challenges to Cross-Disciplinary Curricula: Data Literacy and Divergent Disciplinary Perspectives. *Educational Studies in Mathematics*, 14. https://doi.org/10.1007/s10649-012-9392-z

Talend. (2021). Tout savoir sur la culture des données Comment réussir un changement de culture ? *Talend - A Leader in Data Integration & Data Integrity*. Repéré à https://www.talend.com/fr/resources/what-is-data-culture/

Tarrant, D., Maddison, J., & Thereaux, O. (2021). Data literacy: what is it and how do we address it at the ODI? Repéré à https://theodi.org/article/data-literacy-what-is-it-and-how-do-we-address-it-at-odi/

Taylor, L. (2024). Black data: higher education, datafication, and the Black student body. *Higher Education*. https://doi.org/10.1007/s10734-023-01141-6

Tedesco, M. (2002). Using the 2001 Census Products in Data Literacy Programs, 32.

The Data Literacy Revolution - Dr. Eugene Dubossarsky. (2019). Repéré à https://www.youtube.com/watch?v=RnrQ1nTQMok

The Economist. (2010). The data deluge. Repéré à https://www.economist.com/leaders/2010/02/25/the-data-deluge

The Economist. (2020). Are data more like oil or sunlight? *The Economist*. Repéré à https://www.economist.com/special-report/2020/02/20/are-data-more-like-oil-or-sunlight

Thessen, A., & Patterson, D. (2011). Data issues in the life sciences. *ZooKeys*, *150*, 15-51. https://doi.org/10.3897/zookeys.150.1766

Theviot, A. (2023). Saisir les données en contexte. Dans *Gouverner par les données ? Pour une sociologie politique du numérique* (pp. 13-32). Lyon: ENS Éditions. Repéré à http://books.openedition.org/enseditions/44838

Thomas, G. (2006). *Alpha Males and Data Disasters: The Case for Data Governance*. (S.l.): Brass Cannon Press.

Thompson, C. (2016). 8 Most In-Demand Jobs Every Company Will Be Hiring For in 2021. *World Economic Forum*. Repéré à https://www.weforum.org/agenda/2016/01/most-in-demand-jobs-every-company-will-be-hiring-for/

Tonnelat, É. (1930). Kultur : histoire du mot, évolution du sens. Dans *Civilisation - le mot et l'idée* (pp. 64-75). Paris: La renaissance du livre.

Traub, J. F. (1966). American standard and IFIP / ICC vocabularies compared. *Communications of the ACM*, *8*(6), 361-362. https://doi.org/10.1145/364955.364961

Trompette, P., & Vinck, D. (2009). Retour sur la notion d'objet-frontière. *Revue d'anthropologie des connaissances*, *3, 1*(1), 5-27. https://doi.org/10.3917/rac.006.0005

Trudel, D., Le Deuff, O., & Averbeck-Lietz, S. (2023). L'histoire digitale et la recherche en communication. *Communication. Information médias théories pratiques*, (Vol. 40/1). https://doi.org/10.4000/communication.17010

Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2022). The ethics of algorithms: key problems and solutions. *AI & SOCIETY*, *37*(1), 215-230. https://doi.org/10.1007/s00146-021-01154-8

Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, *33*(1), 1-67. Repéré à https://www.jstor.org/stable/2237638

Tukey, J. W. (1977). *Exploratory Data Analysis* (1er édition). Reading, Mass: Pearson.

Tuomi, I. (1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems*, *16*, 103-117. https://doi.org/10.1080/07421222.1999.11518258

Turing, A. (1948). Intelligent Machinery (1948). Dans B. J. Copeland (Éd.), *The Essential Turing* (p. 0). (S.l.): Oxford University Press. https://doi.org/10.1093/oso/9780198250791.003.0016

Twidale, M., Blake, C., & Gant, J. (2013). Towards a Data Literate Citizenry. Dans. https://doi.org/10.9776/13189

Tygel, A. F., & Kirsch, R. (2016). Contributions of Paulo Freire for a Critical Data Literacy: a Popular Education Approach. *The Journal of Community Informatics*, *12*(3). https://doi.org/10.15353/joci.v12i3.3279

Tylor, E. B. (1871). *Primitive Culture: Researches Into the Development of Mythology, Philosophy, Religion, Art, and Custom* (Vol. 1). (S.l.): J. Murray.

Umbach, G. (2022). Statistical and data literacy in policy-making. *Statistical Journal of the IAOS*, *38*(2), 445-452. https://doi.org/10.3233/SJI-220962

UNESCO. (1968). *Alphabétisation, 1965-1967*. Paris: UNESCO. Repéré à https://unesdoc.unesco.org/ark:/48223/pf0000137447

UNESCO. (1997). Education des adultes, la déclaration de Hambourg, l'agenda pour l'avenir. Dans (p. 36). Communication présentée au 5e conférence internationale sur l'éducation des adultes. Repéré à https://unesdoc.unesco.org/ark:/48223/pf0000116114_fre

UNESCO. (2000). *Cadre d'action de Dakar: l'Education pour tous: tenir nos engagements collectifs (avec les six cadres d'action régionaux)*. Dakar, Sénégal: UNESCO. Repéré à https://unesdoc.unesco.org/ark:/48223/pf0000121147_fre

UNESCO. (2004). Déclaration universelle de l'unesco sur la diversité culturelle. *Diogene*, *n° 205*(1), 166-169. Repéré à https://www.cairn.info/revue-diogene-2004-1-page-166.htm

UNESCO. (2011). *Déclaration de Jomtien*. Jomtien, Thaïlande. Repéré à https://unesdoc.unesco.org/ark:/48223/pf0000191931_fre

UNESCO, UNICEF, Banque Mondiale, FNUAP, PNUD, ONU-Femmes, & HCR. (2017). Déclaration d'Incheon et cadre d'action pour la mise en oeuvre de l'objectif de développement durable 4. Assurer à tous une éducation équitable, inclusive et de qualité et des possibilités d'apprentissage tout au long de la vie. Dans *Construire l'avenir* (pp. 24-28). (S.l.): UN. https://doi.org/10.18356/5b6f4c9e-fr

Upward Data. (n.d.). Comment créer un département data ? *Upward Data*. Repéré à https://www.upwarddata.fr/blog/creer-un-departement-data/

US Census Webinar: Economic Programs Webinar Series: Data Literacy — Exploring Economic Data. (2018). US Census Studio.

US Chamber of commerce fundation. (2014). *The future of data-driven innovation*. Washington, USA: US Chamber of commerce fundation. Repéré à https://www.uschamberfoundation.org/sites/default/files/The%20Future%20of%20Data-Driven%20Innovation.pdf

Usine Digitale. (2016). Organisations : Pourquoi vous devez vous préoccuper de la culture « data » de vos collaborateurs. Repéré à https://www.usine-digitale.fr/article/organisations-pourquoi-vous-devez-vous-preoccuper-de-la-culture-data-de-vos-collaborateurs.N435337

Valenduc, G., & Vendramin, P. (2006). Fractures numériques, inégalités sociales et processus d'appropriation des innovations. *Terminal*, (95-96), 137-154.

Van Meter, H. J. (2020). Revising the DIKW pyramid and the real relationship between data, information, knowledge, and wisdom. *Law, Technology and Humans*, *2*(2), 69-80. https://doi.org/10.3316/agispt.20210112042035

Vandevelde-Rougale, A., & Fugier, P. (2016). Acculturation. Dans *Dictionnaire de la fatigue* (pp. 20-27). Genève: Librairie Droz. https://doi.org/10.3917/droz.zawie.2016.01.0020

Verbiest, E., Pol, M., Vanlommel, K., Mahieu, P., Lazarová, B., Malmberg, K., … Hortlund, T. (2014). Becoming a data-wise school leader: Developing leadership capacity for data-informed school improvement. *Sodobna pedagogika*, (4). Repéré à https://is.muni.cz/publication/1216772/en/Becoming-a-data-wise-school-leader-Developing-leadership-capacity-for-data-informed-school-improvement/Verbiest-Pol-Vanlommel-Mahieu

Verdi, U. (2023a). L'éthique des données dans les chartes éthiques des collectivités territoriales. *Communication et organisation. Revue scientifique francophone en Communication organisationnelle*, (64), 51-66. https://doi.org/10.4000/communicationorganisation.12626

Verdi, U. (2023b). Quelle(s) réponse(s) à l'enjeu d'acculturation aux données ? Un état de l'art des caractéristiques de la data literacy. *Revue française des sciences de l'information et de la communication*, (26). https://doi.org/10.4000/rfsic.14589

Verdi, U., & Desfriches Doria, O. (2019). Information Literacy : une exploration hypertextuelle. Dans I. Editions (Éd.), *H2PTM'19* (pp. 22-46). Montbéliard (Doubs), France. Repéré à https://hal.science/hal-04300485

Verdi, U., & Le Deuff, O. (2020). La data literacy distribuée. Périmètres définitionnels, origines documentaires, perspectives réticulaires. *Les Cahiers du numérique*, *16*(2-3-4), 137-173. https://doi.org/10.3166/LCN.2020.006

Verdier, H. (2015a). *Administrateur général des données - Rapport au Premier ministre sur la gouvernance de la donnée 2015 : les données au service de la transformation publique*. Paris: Secrétariat général pour la modernisation de l'action publique. Repéré à https://www.gouvernement.fr/sites/default/files/document/document/2016/01/rapport_au_premier_ministre_sur_la_gouvernance_de_la_donnee_2015.pdf

Verdier, H. (2015b). *Rapport au Premier Ministre sur la gouvernance de la donnée 2015 : les données au service de la transformation publique*. Paris: Secrétariat général pour la modernisation de l'action publique. Repéré à https://www.gouvernement.fr/sites/default/files/document/document/2016/01/rapport_au_premier_ministre_sur_la_gouvernance_de_la_donnee_2015.pdf

Verdier, H. (2018). *La donnée comme infrastructure essentielle : rapport au premier ministre sur la donnée dans les administrations 2016 - 2017*. Paris: Direction interministérielle du numérique et du système d'information et de communication de l'Etat. Repéré à https://www.etalab.gouv.fr/wp-content/uploads/2018/04/RapportAGD_2016-2017_web.pdf

Vesnic-Alujevic, L., Stoermer, E., Rudkin, J., Scapolo, F., & Kimbell, L. (2019). *The Future of Government 2030+ : A Citizen Centric Perspective on New Government Models*. Luxembourg: Joint Research Center, Commission Européenne. Repéré à https://digital-strategy.ec.europa.eu/en/library/future-government-2030-policy-implications-and-recommendations

Vigne, A. (2023). Un guichet unique pour la gestion de vos données de recherche. *Bibliothèque universitaire Nantes Université*. Repéré à https://bu.univ-nantes.fr/science-ouverte/un-guichet-unique-pour-la-gestion-de-vos-donnees-de-recherche

Vitali-Rosati, M. (2012). Une éthique appliquée ? *Éthique publique. Revue internationale d'éthique sociétale et gouvernementale*, (vol. 14, n° 2). https://doi.org/10.4000/ethiquepublique.995

Wagner, D. A. (1995). Literacy and development: Rationales, myths, innovations, and future directions. *International Journal of Educational Development*, *15*(4), 341-362. https://doi.org/10.1016/0738-0593(95)00019-y

Walker, H. M. (1951). Statistical Literacy in the Social Sciences. *The American Statistician*, *5*(1), 6-12. https://doi.org/10.2307/2685917

Wallman, K. K. (1993). Enhancing Statistical Literacy: Enriching Our Society. *Journal of the American Statistical Association*, *88*(421), 1-8. https://doi.org/10.2307/2290686

Walter, M. (2016). Data politics and Indigenous representation in Australian statistics. Dans. https://doi.org/10.22459/CAEPR38.11.2016.05

Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, *22*(3), 233-243. https://doi.org/10.1080/13645579.2018.1531228

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, *10*(2), 1-37. Repéré à https://sciendo.com/fr/article/10.2478/jagi-2019-0002

Warschauer, M. (2012). A Literacy Approach to the Digital Divide. Repéré à https://www.semanticscholar.org/paper/A-Literacy-Approach-to-the-Digital-Divide-Warschauer/7d7f716237l2ee80254454a2bd1dad0dfd255c28

Webb, N. (2002). Assessment Literacy in a Standards-Based Education Setting.

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All---A Contingency Approach to Data Governance. *Journal of Data and Information Quality*, *1*(1), 4:1-4:27. https://doi.org/10.1145/1515693.1515696

Weber, N. (2017). Introducing the « Open Data Literacy » project at the UW iSchool. *Open Data Literacy*. Repéré à https://medium.com/open-data-literacy/introducing-the-open-data-literacy-project-at-the-uw-ischool-7a2555b3cf94

Wende, K., & Otto, B. (2007). A contingency approach to data governance. Dans. Communication présentée au 12th International Conference on Information Quality (ICIQ-07), Cambridge, MA.

Wilkinson, N. (2014). *A framework for organisational governance maturity : an internal audit perspective*. Manuscrit soumis pour publication. Repéré à https://repository.up.ac.za/bitstream/handle/2263/43563/Wilkinson_Framework_2014.pdf;sequence=4

Williams, E. (2004). Literacy Studies. Dans *The Handbook of Applied Linguistics* (pp. 576-603). (S.l.): John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470757000.ch23

Williamson, B., & Hague, C. (2009). Digital participation, digital literacy, and school subjects: A review of the policies, literature and evidence. Dans. Repéré à https://www.semanticscholar.org/paper/Digital-participation%2C-digital-literacy%2C-and-school-Williamson-Hague/d0227e50bd643ee61d7c4459023422l88c511f519

Wise, A. F. (2020). Educating Data Scientists and Data Literate Citizens for a New Generation of Data. *Journal of the Learning Sciences*, *29*(1), 165-181. https://doi.org/10.1080/10508406.2019.1705678

Wolff, A., Gooch, D., & Cavero Montaner, J. J. (2016). Creating an Understanding of Data Literacy for a Data-driven Society. *The Journal of Community Informatics*, (810), 205-214. https://doi.org/https://doi.org/10.1007/978-3-319-74334-9_22

Wolff, A., Hudson, L., & Kortuem, G. (2016). The Role of Data Literacy within a MOOC Analysis. Dans. Communication présentée au LAK'16: The 6th International Learning Analytics & Knowledge Conference, University of Edinburgh, Edinburgh. Repéré à http://events.kmi.open.ac.uk/dataliteracy4la/papers/DLitLA-16_paper_1.pdf

Wolff, A., Knutas, A., Palacin, V., Seidelin, C., Kun, P., & Mulder, I. (2019). Civic Data Literacies for Bottom-up Data Innovation. Dans *Proceedings of the 9th International Conference on Com-*

munities & Technologies - Transforming Communities (pp. 306-309). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3328320.3328381

Wolff, A., Moore, J., Zdrahal, Z., Hlosta, M., & Kuzilek, J. (2016). Data literacy for learning analytics. Dans *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 500-501). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2883851.2883864

Wong, C. (2009). *Kid's Survey Network : teaching data literacy with multiplayer online games*. Thesis. Massachusetts Institute of Technology. Repéré à https://dspace.mit.edu/handle/1721.1/53130

Wu, J. (1998). Statistics = Data science ? [Leçon inaugurale]. University of Michigan.

Wynter, S. (1989). Beyond the Word of Man: Glissant and the New Discourse of the Antilles. *World Literature Today*, *63*(4), 637-648. https://doi.org/10.2307/40145557

Wynter, S. (2003). Unsettling the Coloniality of Being/Power/Truth/Freedom: Towards the Human, After Man, Its Overrepresentation—An Argument. *CR: The New Centennial Review*, *3*(3), 257-337. Repéré à https://www.jstor.org/stable/41949874

Wynter, S., & McKittrick, K. (2015). Unparalleled Catastrophe for Our Species?: Or, to Give Humanness a Different Future: Conversations. https://doi.org/10.1215/9780822375852-002

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162. https://doi.org/10.1016/j.tate.2016.05.010

Yang, N., & Li, T. (2023). How Stakeholders' Data Literacy Contributes to Quality in Higher Education: A Goal-Oriented Analysis. Dans J. E. Raffaghelli & A. Sangrà (Éds), *Data Cultures in Higher Education : Emergent Practices and the Challenge Ahead* (pp. 313-327). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24193-2_13

Zacklad, M., & Rouvroy, A. (2022). L'éthique située de l'IA et ses controverses. *Revue française des sciences de l'information et de la communication*, (25). Repéré à https://journals.openedition.org/rfsic/13204

Zielinska, A. C. (2022). « Nous avons besoin d'une conversation sur la sexualité ». Politiques de la sexualité : autour de la notion de queer. *Revue française d'éthique appliquée*, *13*(2), 17-20. https://doi.org/10.3917/rfeap.013.0017

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (Main édition). London: Profile Books Ltd.

Zumthor, P. (1985). Litteratus/Illitteratus: Remarques Sur Le Contexte Vocal De L'écriture Médiévale. *Romania*, *106*(421 (1)), 1-18. Repéré à https://www.jstor.org/stable/45040841

Zurkowsky, P. G. (1974). *The information service environment relationships and priorities. Related paper No.5.* Washington, USA: National Commission on librairies and information science. Repéré à https://files.eric.ed.gov/fulltext/ED100391.pdf

Index

A

acculturation, 3, 233, 289
    aux données, 231
algorithm(ic) literacy, 140, 141
algorithme, 141
apophénie, 132
artificial intelligence literacy, 143
assessment literacy, 155

B

Barthes, Roland, 301
Bertin, Jacques, 90, 91
big data, 97, 119, 168
    caractéristiques du, 174
    critiques du, 174, 175, 179
    fondamentalisme du, 98, 168
    literacy, 165
big data fundamentalism, *voir*
        fondamentalisme du big data
bildung, 236, 239
black data, 185
    définition de la, 185
blackness, 186
Borne, Kirk D., 55, 60
Bothorel, Eric, 200
Brugière, Amandine, 211
Bruyas, Dorie, 218

C

capta, 83
catégorisation, 252
champ sémantique, 10
Chignard, Simon, 200, 208, 211
civic data literacy, 167
civilisation, 237
Cleveland, William S., 118
colonialité, 184
concept
    composantes du, 116
    définition du, 116
convergence, 47, 132, 133, 135, 136, 145, 206,
        302
covid-19, 18, 19, 199
creative data literacy, 164
critical
    big data literacy, 165, 178
    data literacies, 177
    data studies, 174
    datatafication literacy, 178
    literacy, 176, 177
culture
    civilisation et, 237
    data, 202, 204, 205, 299
    data-driven, 202
    de l'écrit, 27, 230
    définition de la, 235
    institutionnelle, 9
    literacy et, 240, 260
    nature et, 236
    technique et, 134, 240
    transmission et, 239
    valeur et, 238
    weltanschauung et, 238

D

dark data, 185
data, 82
data acumen, 15, 300
data administration, 95, 125
data administrator, 126
data and IA literacy, 144, 145
data antennae, 30
data assemblage, 173
data biographies, 164
data citizenship, 167
data culture, 241, 261
    caractéristiques de la, 242, 243
    data literacy et, 242, 244
    fair, 245
data friction, 230
data governance, *voir* gouvernance des données
data information literacy, 162
data infrastructure literacy, 178
data journalisme, 28, 156, 165
data justice, *voir* justice des données
data lettrisme, 202
data literacy
    assessment literacy et, 156
    cadre de compétences en, 35, 37
    caractéristiques de la, 120, 143, 298, 301, 302
    définitions de la, 23
    digital, 44
    digital literacy et, 148, 152, 153
    formation en, 34, 35
    gap, 17
    information literacy et, 150, 151
    niveaux de formation en, 34
    numeracy et, 156
    référentiel de compétences en, 38
    statistical literacy et, 152, 156
data literate, 27–29
    educator, 31, 33
data littératie, 224, 226
data management, 123
data practices, 243
data privacy literacy, 180
data quality, 123
data science
    caractéristiques de la, 117
    data literacy et, 120, 268
    et autres disciplines, 118
    histoire de la, 116
    objectifs de la, 118
    statistiques et, 120
data scientist, 17, 18, 119
data sense, 31, 300
data state of mind, 30
data stewardship, 123
data structure, 94, 134
data vizualisation literacy, 165
data wise, 15, 300
datacy, 15
datafication, 2, 98
    caractéristiques de la, 169
    définition de la, 169
datalogie
    concepts de la, 102
    datamatique, 104
    définition de la, 100
    informatique et, 101, 106
    la datamatique, 104
    langage et, 109
    mathématiques et, 108, 111
    objectifs de la, 104, 107
    ordinateur et, 106
    pédagogie et, 110
    traductions de la, 102
datalphabétisation, 202
data-philosophie, 173
dataveillance, 2, 122
    aspects de la, 171
    dangers de la, 172

définition de la, 171
datavizualisation, 189
literacy, 165
dedomena, 82
démasquer l'universalisme, 183
désobéissance épistémique, 184
diaphora, 85
digital data literacy, 44, 298
digital literacy, 147
discours
définition du, 8
d'instruction, 26
discrimination, 181
donnée
brute, 83, 265
caractéristiques de la, 86
culture de la, 224, 231
définition de la, 85
définitions de la, 82
essentialisme de la, 261
et trace, 86
éthique des, 268
information et, 84, 95, 103, 265
numérique, 103
pouvoir et, 174
donnée ouverte liée, *voir* linked open data
donnée sombre, *voir* dark data
données
activisme des, 179
algorithmes et, 94, 140
centralisation des, 95
certification en, 44
colonialisme des, 141, 171, 184
comme atout, 95, 96
corps et, 171
cultivation des, 230
culture des, 224, 226, 229, 230, 246
enjeux des, 274
esprit critique des, 121
état d'esprit des, 30, 212
littératie des, 224, 246, 301, 302
manquantes, 181
médiation aux, 224
monde de, 3
numérique et, 44, 49
personnelles, 179, 180
référentiels des, 43
révolution des, 2, 5, 98, 168
souveraineté des, 188
types de, 276
vision numérique des, 44
dătŭs, 82
Drucker, Peter Ferdinand, 27
dysselection, 186

E

écriture
fonctions de l', 69–71
langage et, 69
oralité et, 69, 70
educational data literacy, 163
eigencapital, 170
enculturation, 127
enquête, 32
qualitative, 249
enseignant formateur, 31–33
cadre de formation de l', 33
entretien semi-directif, 250
erziehung, 239
e-science, 166
espace conceptuel, 132
esprit lettré, 25
éthique
de l'intelligence artificielle, 130
définition, 128
des algorithmes, 131
des données, 33, 127, 129
du numérique, 129, 130
située de l'intelligence artificielle, 130
évitement quantitatif, 187

F
féminicide, 184
féminisme
    caractéristiques du, 182
    des données, 182, 183
    intersectionnel, 183
    noir, 181
Freire, Paulo, 121, 301

G
Gärdenfors, Peter, 132
Gilliard, Armelle, 212
Goëta, Samuel, 4, 98, 200, 208, 215
Gombin, Joël, 215
Goody, Jack, 68, 70
gouvernance des données, 200, 212
    caractéristiques de la, 123, 267
    data literacy et, 126, 127
    définition de la, 123
    objectifs de la, 124
    visions de la, 124
gouvernementalité algorithmique, 169
grand partage, 68

H
Hayashi, Chikio, 116
healthcare data literacy, 166
hétéroglossie, 177

I
illettré, *voir* illiterate
illettrisme, *voir* illiteracy
illiteracy, 72
illiterate, 77, 78
illiteratus, 77
imaginaire, 9, 238, 250
impensé technique, 2
impératif des données, 169, 170
indigenous data, 187, 188
information
    définition de l', 150
    literacy, 149, 150
    literate, 150
    patterns d', 85
informatique, 92
    comme science de l'information, 93
    et mathématiques, 93
    programmes en, 92
innumeracy, 153
intelligence artificielle, 142
    mythe de l', 142, 143
intersectionnalité, 181

J
justice des données, 182

K
Kaplan, Daniel, 211
kultur, 236, 237

L
langage lettré, 26
letteracy, 79
lettré, *voir* literate
lettré des données, *voir* data literate
Lévi-Strauss, Claude, 71
Levy-Bruhl, Lucien, 68
illettré, *voir* illiterate
linked open data
    définition, 166
    literacy, 166
literacy
    crise de la, 297
    culture et, 234
    in the age of data, 192
    thesis, 238
    traduction de la, 233, 234
literate, 25, 26, 77, 78
littératie
    besoin de, 74
    campagne de, 73, 75, 76
    courant autonome de la, 80

courant idéologique de la, 80
crise de la, 80, 81
définition de la, 68
définitions de la, 25
modèle idéologique de la, 81
mythes de la, 75, 76
nouvelle, 81, 159
spécialisée, 159–161
vision maximale de la, 68
vision minimaliste de la, 68
littératie de l'information, *voir* information literacy
littératie de l'intelligence artificielle, *voir* artificial intelligence literacy
littératie des algorithmes, *voir* algorithm(ic) literacy
littératie des données, *voir* data literacy
littératie des mathématiques, *voir* mathematical literacy
littératie des statistiques, *voir* statistical literacy
littératie spécialisée, 301
litteratus, 77

M

Marchandise, Jacques-François, 200, 208, 211
mathemacy, 154
mathematical literacy, 153
matheracy, 154
McCarthy, John, 142
Medjek, Sarah, 214
mégadonnée, *voir* big data
métallittératie, 137
métriques, 152
micro-constestation, 188
monde, 3
Morrow, Jordan, 46, 126, 143
mündigkeit, 179
mythe, 2

N

Naur, Peter, 100
Népote, Charles, 211, 212, 217
numeracy, 153
numératie, *voir* numeracy
numérique
compétence, 38, 40–42, 148
fracture, 40
positivisme, 174

O

objet frontière, 6, 298
oligoptique, 171
Olson, David, 69
open data, 99
literacy, 165
ordinateur, 91

P

Papert, Seymour, 79
pedagogical data literacy, 163
pensée
critique, 122
pensée lettrée, 25
personal data literacies, 180
pharmakon, 134
pouvoir
colonialité du, 184
Priol, Jacques, 208
pyramide d'Ackoff, 84, 85

Q

quaesita, 82
qualité des données, *voir* data quality
quantitative literacy, 154
queer
caractéristiques du, 191
définition de, 190
données, 190
queerisation des données, 191

R
racial data literacy, 188
raw data, *voir* donnée brute
réflexionisme, 179
représentation, 8
représentativité, 251
research data literacy, 166

S
savoir
    assujetti, 99
    situé, 183
science des données, *voir* data science
science/scientific data literacy, 166
sémiologie graphique, 90, 91
social data literacy, 167
sousveillance, 179
    digital, 180
statistical
    literacy, 145, 146
    literate, 146
statistiques
    arithmétie anglaise, 89
    définition des, 88
    doubles statistiques, 169
    statistique allemande, 88
    statistique américaine, 90
    statistique française, 89
Street, Brian, 80
sublata, 83
substance sémantique, 9
surveillance
    dark, 180
    oligoptique, 171
    participante, 171
    sociale, 171
système, 133, 302
    de données, 133, 134
    définition du, 133
    d'information, 135

T
techno-discours, 2, 99
théorie critique de la race, 189
Thomas, Gwen, 123
tradition lettrée, 27
Tukey, John, 90
Turing, Alan, 141

U
übercapital, 170
utopie, 2

V
Verdier, Henri, 200

W
weltanschauung, 71
Wu, Jeff, 117

Y
youth data literacy, 164

Z
zivilisation, 237

Annexes

Annexe 1 : les définitions de la *data literacy*

Tableau des définitions de la *data literacy*

Auteur	Définition
(Schield, 2004a)	And students must be data literate: they must be able to access, assess, manipulate, summarize, and present data.
(Love, 2004)	The ability to examine multiple measures and multiple levels of data, to consider the research, and to draw sound inferences.
(Swan, Patton, & Yarnall, 2006)	Data literacy includes the ability to: formulate and answer questions using data as part of evidence-based thinking; use appropriate data, tools, and representations to support this thinking; interpret information from data; develop and evaluate data-based inferences and explanations; and use data to solve real problems and communicate their solutions.
(Wong, 2009)	Data literacy is the ability to comprehend and utilize data to correctly and advantageously draw inferences in solving a problem or forming an opinion.
(Swan et al., 2009)	The ability to ask and answer meaningful questions by collecting, analyzing and making sense of data encountered in our everyday lives.
(McAuley, Rahemtulla, Goulding, & Souch, 2011)	The ability to identify, retrieve, evaluate and use information to both ask and answer meaningful questions.

Auteur	**Définition**
(Carlson et al., 2011)	Understanding what data mean, including how to read graphs and charts appropriately, draw correct conclusions from data, and recognize when data are being used in misleading or inappropriate ways.
(Bounegru, Chambers, & Gray, 2012)	Data-literacy is the ability to consume for knowledge, produce coherently and think critically about data.
(Harris, 2012)	You might call this "data literacy": competence in finding, manipulating, managing, and interpreting data, including not just numbers, but also text and images. Data literacy skills must spread far beyond their usual home, the IT function, and become an integral aspect of every business function and activity.
(Johnson, 2012)	The ability to process, sort, and filter vast quantities of information, or data literacy.
(Love, 2012)	Data literacy is the ability to accurately observe, analyse and respond to a variety of different kinds of data for the purpose of continuously improving teaching and learning in the classroom and school.
(Harris, 2012)	Competence in finding, manipulating, managing, and interpreting data, including not just numbers but also text and images.
(Calzada-Prado & Marzal, 2013)	Data literacy can be defined, then, as the component of information literacy that enables individuals to access, interpret, critically assess, manage, handle and ethically use data ;

Auteur	**Définition**
(Mandinach & Gummer, 2013)	The ability to understand and use data effectively to inform decisions. It is composed of a specific skill set and knowledge base that enables educators to transform data into information and ultimately into actionable knowledge. These skills include knowing how to identify, collect, organise, analyse, summarise and prioritise data. They also include how to develop hypotheses, identify problems, interpret the data, and determine, plan, implement, and monitor courses of action.
(Twidale et al., 2013)	The ability to discover, analyze, think critically about data encountered in daily life and become comfortable with questioning sources and accuracy rather than feeling intimidated.
(Athanases et al., 2013)	The capacity to manage, understand, and critique proliferation of data in a high-tech, networked information age [...] data literacy also includes capacity to conduct inquiry: formulating questions, collecting and organizing data systematically, using appropriate tools to analyze and display findings, and drawing conclusions.
(Gemignani, Gemignani, Galentino, & Schuermann, 2014)	The ability of individuals to understand and draw meaning from data...the abilities necessary to thoughtfully consume data.
(Bhargava et al., 2015)	The desire and ability to engage constructively in society through and with data.
(Koltay, 2015)	Access, interpret, critically assess, manage, handle and ethically use data.
(Ridsdale et al., 2015)	Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner.
(Frank et al., 2016)	The ability to understand and use data, particularly in the context of the Internet

Auteur	**Définition**
(D'Ignazio & Bhargava, 2016)	Data literacy includes the ability to read, work with, analyze and argue with data as part of a larger inquiry process.
(Crusoe, 2016)	Data literacy is the knowledge of what data are, how they are collected, analyzed, visualized and shared, and is the understanding of how data are applied for benefit or detriment, within the cultural context of security and privacy.
(Wolff, Gooch, et al., 2016)	Data literacy is the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyse, visualise, critique and interpret data, as well as to communicate stories from data and to use data as part of a design process.
(Okamoto, 2017)	Data literacy is the ability to access, critically assess, interpret, manipulate, manage, summarize, handle, present, and ethically use data.
(*US Census Webinar: Economic Programs Webinar Series: Data Literacy — Exploring Economic Data*, 2018)	The ability to derive meaningful information from data.
(Pangrazio & Selwyn, 2019)	Individuals might better engage with and make use of the 'personal data' generated by their own digital practices.
(Guler, 2019)	Data literacy is the ability to collect, transform, evaluate/analyze, and communicate the results to others.
(Keller et al., 2020)	Data acumen: the ability to make good judgements about the use of data to support problem solutions.
(Morrow, 2021)	Data literacy is the ability to read, work with, analyze, and communicate with data.

Auteur	Définition
(Schüller, 2021)	To carry out the process of transforming data into information, knowledge, and wisdom/"power", i.e. action knowledge and capacity to act in a controlled and responsible manner, data literacy is required. This competence enables people to address real problems through the use, analysis and interpretation of data that measuresunderlying phenomena. It is essential for understanding complex, socially relevant phenomena such as global economic and financial interdependence, migration, or climate change.
(Tarrant, Maddison, & Thereaux, 2021)	The ability to think critically about data in different contexts and examine the impact of different approaches when collecting, using and sharing data and information.

Bibliographie de l'annexe 1

Athanases, S., Bennett, L., & Michelsen Wahleithner, J. 2013. "Fostering Data Literacy Through Preservice Teacher Inquiry in English Language Arts." *The Teacher Educator* 48 (1): 8–28. https://doi.org/10.1080/08878730.2012.740151.

Bhargava, R., Deahl, E., Letouzé, E., Noonan, A., Sangokoya, D. & Shoup, N. 2015. "Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data." https://dspace.mit.edu/handle/1721.1/123471.

Bounegru, L., Chambers, L. & Gray, J. 2012. *The Data Journalism Handbook 1*. European Journalism Centre. https://s3.eu-central-1.amazonaws.com/datajournalismcom/handbooks/The-Data-Journalism-Handbook-1.pdf.

Calzada-Prado, F-J., & Marzal, M. 2013. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents." *Libri* 63. https://doi.org/10.1515/libri-2013-0010.

Carlson, J., Fosmire, M., Miller, C., & Sapp Nelson, M. 2011. "Determining Data Information Literacy Needs: A Study of Students and Research Faculty." *Libraries Faculty and Staff Scholarship and Research*. https://docs.lib.purdue.edu/lib_fsdocs/23.

Crusoe, D. 2016. "Data Literacy Defined Pro Populo: To Read This Article, Please Provide a Little Information." *The Journal of Community Informatics* 12 (3). https://doi.org/10.15353/joci.v12i3.3276.

D'Ignazio, C., & Bhargava, R. 2016. "DataBasic: Design Principles, Tools and Activities for Data Literacy Learners." *The Journal of Community Informatics* 12 (3). https://doi.org/10.15353/joci.v12i3.3280.

Frank, M., Walker, J., Attard, J., & Tygel, A. 2016. "Data Literacy - What Is It and How Can We Make It Happen?" *The Journal of Community Informatics* 12 (3). https://doi.org/10.15353/joci.v12i3.3274.

Gemignani, Z., Gemignani, C., Galentino, R., & Schuermann, P. 2014. *Data Fluency: Empowering Your Organization with Effective Data Communication*. John Wiley & Sons.

Guler, G. 2019. "Data Literacy from Theory to Reality: How Does It Look?"

Harris, J. 2012. "Data Is Useless Without the Skills to Analyze It." *Harvard Business Review*, 2012. https://hbr.org/2012/09/data-is-useless-without-the-skills.

Johnson, C. 2012. *The Information Diet: A Case for Conscious Comsumption*. 1st edition. Sebastopol, CA: O'Reilly Media.

Keller, S-A., Shipp,S., Schroeder, A., & Korkmaz, G. 2020. "Doing Data Science: A Framework and Case Study." *Harvard Data Science Review* 2 (1). https://doi.org/10.1162/99608f92.2d83f7f5.

Koltay, T. 2015. "Data Literacy: In Search of a Name and Identity." *Journal of Documentation* 71 (2): 401–15. https://doi.org/10.1108/JD-02-2014-0026.

Love, N. 2004. "Taking Data to New Depths: There's a Ton of Data Being Collected. The Trick Is to Know How to Use It Effectively." *Journal of Staff Development* 25 (4): 22–26.

Love, N. 2012. "Data Literacy for Teachers," 1.

Mandinach, E-B., & Gummer, E-S. 2013. "A Systemic View of Implementing Data Literacy in Educator Preparation." *Educational Researcher* 42 (1): 30–37. https://doi.org/10.3102/0013189X12459803.

McAuley, D., Rahemtulla, H., Goulding, J., & Souch, S. 2011. "How Open Data, Data Literacy and Linked Data Will Revolutionise Higher Education." *Pearson Centre for Policy and Learning*, 88–93. https://doi.org/978-0-997-86378-9.

Morrow, J. 2021. *Data Literacy: The Data Literacy Skills Everyone Needs to Succeed*. London ; New York, NY: Kogan Page Ltd.

Okamoto, K. 2017. "Introducing Open Government Data." *The Reference Librarian* 58 (2): 111–23. https://doi.org/10.1080/02763877.2016.1199005.

Pangrazio, L., & Selwyn, N. 2019. " 'Personal Data Literacies': A Critical Literacies Approach to Enhancing Understandings of Personal Digital Data." *New Media & Society* 21 (2): 419–37. https://doi.org/10.1177/1461444818799523.

Ridsdale, C., Rothwell, J., Smith, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B., & Ali-Hassan, H. 2015. *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*. https://doi.org/10.13140/RG.2.1.1922.5044.

Schield, M. 2004. "Information Literacy, Statistical Literacy and Data Literacy." *IASSIST Quarterly / International Association for Social Science Information Service and Technology* 28: 7–14. https://doi.org/10.29173/iq790.

Schüller, K. 2021. *Datenkompetenz als zentraler Baustein einer Datenstrategie.* Edited by Philipp Ramin. München, Allemagne: Carl Hanser Verlag GmbH & Co. KG.

Sperry, J. 2018. *US Census Webinar: Economic Programs Webinar Series: Data Literacy — Exploring Economic Data*. US Census Studio.

Swan, K., Patton, C., & Yarnall, L. 2006. "Mathematizing Middle School: Results from a Cross-Disciplinary Study of Data Literacy." *American Educators Research Association Annual Conference 2006*. https://www.academia.edu/20169905/Mathematizing_middle_school_Results_from_a_cross_disciplinary_study_of_data_literacy.

Swan, K., Vahey, P., Kratcoski, A., Van, M., Rafanan, K., & Stanford, T. 2009. "Challenges to Cross-Disciplinary Curricula: Data Literacy and Divergent Disciplinary Perspectives." *Educational Studies in Mathematics*, 14. https://doi.org/10.1007/s10649-012-9392-z.

Tarrant, D., Maddison, J., & Thereaux, O. 2021. "Data Literacy: What Is It and How Do We Address It at the ODI?" 2021. https://theodi.org/article/data-literacy-what-is-it-and-how-do-we-address-it-at-odi/.

Twidale, M., Blake, C., & Gant, J. 2013. "Towards a Data Literate Citizenry.". https://doi.org/10.9776/13189.

Wolff, A., Gooch, D. & Cavero Montaner, J. 2016. "Creating an Understanding of Data Literacy for a Data-Driven Society." *The Journal of Community Informatics*, Springer, no. 810: 205–14. https://doi.org/https://doi.org/10.1007/978-3-319-74334-9_22.

Wong, C. 2009. "Kid's Survey Network : teaching data literacy with multiplayer online games." Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/53130.

Annexe 2 : les formules R

Importer un fichier CSV

```
library(readr)
Data_literacy_complet <- read_delim("Data_literacy_complet.csv",
    delim = ";", escape_double = FALSE, trim_ws = TRUE,
    locale=locale(encoding="UTF-8"))
View(Data_literacy_complet)
```

Ouvrir les libraries

```
library(dplyr)
library(readr)
library(here)
library(rtweet)
library(ggplot2)
library(tidytext)
library(tidyr)
library(magrittr)
library(stringr)
library(meta)
library(ngram)
```

```
library(quanteda)
library(quanteda.textplots)
library(quanteda.textstats)
library(igraph)
library(ggraph)
library(stopwords)
library(ngram)
library(tm)
```

Récupérer les termes

```
termes_pat <- "[a-zA-Z0-9_ -]+"
termes <- str_extract_all(Data_literacy_complet$description, termes_pat)
termes_complet <- unlist(termes)
termes_complet <- tolower(termes_complet)
termes_complet <- gsub("[[:punct:] ]", "", termes_complet)
termes_count <- table(termes_complet)
top_20_freqs_termes <- sort(termes_count, decreasing = TRUE)
View(top_20_freqs_termes)
```

Description

```
Datalitdes <- Data_literacy_complet$description
```

```
Corpus(VectorSource(Datalitdes))
```

```
    tm_map(Corpus, removeWords, stop_words("english"))
```

Récupérer les hashtags

```
hashtag_pat <- "#[a-zA-Z0-9_ -]+ "
hashtag <- str_extract_all(Data_literacy_complet$description, hashtag_pat)
hashtag_word2 <- unlist(hashtag)
hashtag_word3 <- tolower(hashtag_word2)
hashtag_word4 <- gsub("[[:punct:] ]", "", hashtag_word3)
View(hashtag_word4)
```

```
hashtag_count <- table(hashtag_word4)
top_20_freqs <- sort(hashtag_count, decreasing = TRUE)
View(top_20_freqs)
```

Grams

```
Hashtag_word5 <- Corpus(VectorSource(Datalitdes))
PDFDATA <- data.frame(text = sapply(Hashtag_word5, as.character),
stringsAsFactors = FALSE)
```

```
New_bigram <- PDFDATA %>%
unnest_tokens(bigram, text, token = "ngrams", n = 2)
```

```
New_bigram %>%
  count(bigram, sort = TRUE)
```

```
hashtag_datalit <- table(New_bigram)
hashtag_datalit1 <- sort(hashtag_datalit, decreasing = TRUE)
View(hashtag_datalit1)
```

Bigrammes séparés

```
word1 <- "#dataliteracy"
  word2 <- "#data"
bigram_separ <- New_bigram %>%
  separate(bigram, c("word1", "word2"), sep = " ")
```

```
bigram_separ %>%
filter(word1 == "dataliteracy") %>%
count(word2 == "data", sort = TRUE)
```

```
bigram_count <- bigram_separ %>%
count(word1, word2, sort = TRUE)
```

```
bigram_graph <- bigram_count %>%
filter(n > 1000) %>%
graph_from_data_frame()
```

```
set.seed(2017)
ggraph(bigram_graph, layout = "fr") +
  geom_edge_link() +
  geom_node_point() +
  geom_node_text(aes(label=name), vjust= 2, hjust = 2)
```

Nombre de retweets

```
test <- Data_literacy_complet %>%
  arrange(desc(retweet_count))
  test %>%
  select(author, retweet_count)%>%
  head(100)
```

Termes employés par KirKDBorne

```
KirkDBorne <- read_delim("KirkDBorne.csv",
    delim = ",", escape_double = FALSE, trim_ws = TRUE)
View(KirkDBorne)
```

```
termes_pat <- "[a-zA-Z0-9_ -]+"
termes <- str_extract_all(KirkDBorne$description, termes_pat)
termes_complet <- unlist(termes)
termes_complet <- tolower(termes_complet)
termes_complet <- gsub("[[:punct:] ]", "", termes_complet)
termes_count <- table(termes_complet)
Kirkborne_termes <- sort(termes_count, decreasing = TRUE)[1:100]
write.csv2(Kirkborne_termes)
```

hashtags employés par KirKDborne

```
KirkDBorne <- Data_literacy_complet %>%
filter(author == 'KirkDBorne')
View(KirkDBorne)
```

```
termes_pat <- "#[a-zA-Z0-9_ -]+"
termes <- str_extract_all(KirkDBorne$description, termes_pat)
termes_complet <- unlist(termes)
termes_complet <- tolower(termes_complet)
termes_complet <- gsub("[[:punct:] ]", "", termes_complet)
termes_count <- table(termes_complet)
top_20_freqs_hashtags <- sort(termes_count, decreasing = TRUE)[1:100]
write.csv2(top_20_freqs_hashtags)
```

Termes employés par Williamstyles23

```
williamstyles23 <- Data_literacy_complet %>%
filter(author == 'williamstyles23')
View(williamstyles23)
```

```
termes_pat <- "[a-zA-Z0-9_ -]+"
termes <- str_extract_all(williamstyles23$description, termes_pat)
termes_complet <- unlist(termes)
termes_complet <- tolower(termes_complet)
termes_complet <- gsub("[[:punct:] ]", "", termes_complet)
termes_count <- table(termes_complet)
williamstyles_termes <- sort(termes_count, decreasing = TRUE)[1:100]
print(williamstyles_termes)
```

Hashtags employés par Williamstyles23

```
williamstyles23 <- Data_literacy_complet %>%
filter(author == 'williamstyles23')
View(williamstyles23)
```

```
termes_pat <- "#[a-zA-Z0-9_ -]+"
termes <- str_extract_all(williamstyles23$description, termes_pat)
termes_complet <- unlist(termes)
termes_complet <- tolower(termes_complet)
termes_complet <- gsub("[[:punct:] ]", "", termes_complet)
termes_count <- table(termes_complet)
top_20_freqs_termes <- sort(termes_count, decreasing = TRUE)[1:100]
print(top_20_freqs_termes)
```

Lister les bots

```
df_filtre <- filter(Data_literacy_complet, grepl("\\b(bot|bots)\\b",
Data_literacy_complet$user_bio))
liste_bots <- select(df_filtre, author, user_bio, tweet_type, retweet_count)
View(liste_bots)
```

Annexe 3 : les tweets de William D Styles

Nous avons renouvelé l'expérience avec le second compte le plus retweeté, à savoir @Williamstyles23. Nous débutons par les termes les plus employés (>100 occurences) tout en supprimant les mots vides :

Tableau 6.9.: Termes les plus employés par @Williamstyles23

Termes	Occurrences
dataliteracy	165
ai	163
100daysofcode	162
bigdata	159
dvcommunity	157
gamedev	152
indiedev	152
codenewbie	149
gametheory	149
cube	147
indiegamedev	147
newzealand	143

Ici, si les premiers termes renvoient aux usages de la *data science* et à l'usage du code, les termes intégrant le développement de jeux (*gamedev*) sont dissociés de la profession de l'auteur (*graphic & web designer*).

Tableau 6.10.: Les hashtags les plus employés par @Williamstyles23

Hashtag	Occurrences
dataliteracy	165
ai	163
100daysofcode	162
bigdata	159
dvcommunity	157
gamedev	152
indiedev	152
codenewbie	149
gametheory	149
cube	147
indiegamedev	147
newzealand	143

De manière surprenante, les termes repertoriés précédemment s'avèrent être tous des *hashtags*, laissant supposer la présence des contenus peu développés.

En analysant les tweets les plus retweetés, nous constatons que la totalité renvoient soit à des citations d'auteurs divers comme l'auteur de science-fiction Robert Heinlein et les hommes d'affaires Stephen Covey et Ray Croc. La *data literacy* n'est ici aucunement développée : elle s'avère même non pertinente vis-à-vis du contenu véhiculé.

Tableau 6.11.: Les tweets les plus retweetés de @Williamstyles23

Description	Date de publication	Nombre de retweets
There is a big difference between making a simple product and making a product simple. - Des Traynor #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #design #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev https://t.co/5CpWC3SObd	2022-02-27 02:17:44	93

Description	Date de publication	Nombre de retweets
Luck is the result of careful preparation, failure is the result of negligence. - Robert Heinlein #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev #CodeNewbie #NewZealand https://t.co/AwrQCNnkEg	2022-02-10 02:06:40	69
There is a big difference between making a simple product and making a product simple. - Des Traynor #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #design #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev https://t.co/5CpWC3SObd	2022-02-04 03:12:16	62
"Adversity can strengthen you if you have the will to grind it out." – Ray Kroc #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev #CodeNewbie #NewZealand https://t.co/LoXz3EdZQb	2022-02-07 06:00:21	54
Moral superiority is the belief or attitude that one's position & actions are justified by having higher moral values than others. #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev #CodeNewbie #NewZealand	2022-02-12 00:15:53	53

Description	Date de publication	Nombre de retweets
This is the core logic being used by @FavourableGroup in their dynamic bracket system #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev #CodeNewbie #NewZealand https://t.co/noD4vb6VyW https://t.co/2AmcheRdz9	2022-02-28 10:02:25	52
These 81 codes determine round results in our #fundraising template and were found by using a permutation generator. #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev #CodeNewbie #NewZealand https://t.co/15WZ6PKpAI	2022-02-02 17:58:03	51
I channel my anger into my project, the whole point is to create an unbiased impartial transparent unpredictable fundraising tool. #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #InfoSec #Startup #IndieGameDev #CodeNewbie #NewZealand https://t.co/OtEUeXdVud	2022-02-11 23:43:05	50
“There is no such thing as a self made man. You will reach your goals only with the help of others.” https://t.co/8u9N93Jc2h #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #PatreonCreator #IndieGameDev #CodeNewbie #NewZealand https://t.co/GZtS9qkzk9	2021-11-04 09:21:49	49

Description	Date de publication	Nombre de retweets
The key to life is not accumulation. It's contribution. – Stephen Covey #GameTheory #GameDev #IndieDev #BigData #AI #DataLiteracy #Cube #DEVCommunity #100DaysOfCode #PatreonCreator #IndieGameDev #CodeNewbie #NewZealand https://t.co/Crhl7q4PBu	2021-09-09 06:30:40	48

Concernant les tweets des derniers auteurs les plus retweetés, nous ne présenterons pas ici les résultats obtenus du fait qu'à nouveau la *data literacy* ne tient aucune place dans les contenus des tweets, ces derniers servant simplement de promotion aux activités des auteurs.

Annexe 4 : l'agencement des données

Ce concept a été défini par Kitchin & Lauriault (2014) comme « un système socio-technique complexe composé de nombreux dispositifs et éléments étroitement liés dont la préoccupation centrale est la production d'une donnée »[10]. Ces dispositifs (*apparatuses*), à la fois technologiques, politiques, sociaux et économiques, interagissent entre eux et encadrent la production, la circulation et le déploiement des données.

Tableau 6.12.: Les dispositifs d'un agencement de données - extraits et traduits de Kitchin & Lauriault (2014)

Dispositif	Eléments
Modes de pensée	Modes de penser, philosophies, théories, modèles, idéologies, rationalités, etc.
Formes de savoir	Recherche textuelle, manuels, magazines, sites web, expérience, bouche à oreille, chats et forums, etc.
Finance	Modèles commerciaux, investissements, capital de risque, suventions, philanthropie, profit, etc.
Economie politique	Stratégie, régime fiscal, instrument incitatif, opinions politiques et publiques, etc.

[10] « *a complex socio-technical system, composed of many apparatuses and elements that are thoroughly entwined, whose central concern is the production of a data* ».

Dispositif	**Eléments**
Gouvernementalités et réglementations	Standards de données, formats de fichiers, configuration système, protocoles, régulations, lois, licences, régimes de propriété intellectuelle, considérations éthiques, etc.
Matérialités et infrastructures	Papier/stylos, ordinateurs, appareils numériques, capteurs, scanners, bases de données, réseaux, serveurs, bâtiments, etc.
Méthodes	Techniques, façons de faire, comportement appris, conventions scientifiques, etc.
Organisations et institutions	Archives, corporations, consultants, fabriquants, détaillants, agences gouvernementales, universités, conférences, clubs et sociétés, comités and conseils, communautés de pratiques, etc.
Subjectivités et communautés	de producteurs, experts, conservateurs, managers, analystes, scientifiques, politiques, utilisateurs, citoyens de données, etc.
Lieux	Laboratoires, bureaux, terrains, centres de données, grappes de serveurs, parcs commerciaux, etc., et leur agglomération.
Marché	pour les données, ses dérivés (ex : les textes, les tableaux, les graphiques, les cartes), les analystes, les logiciels analytiques, les interprétations, etc.

Les auteurs relient la notion d'agencement au dispositif de Foucault basé sur la définition suivante « un ensemble résolument hétérogène comportant des discours, des institutions, des aménagements architecturaux, des décisions réglementaires, des lois, des mesures administratives, des énoncés scientifiques, des propositions philosophiques, morales, philanthropiques » [11]. Du fait de l'inscription du dispositif dans des logiques de pouvoir, les données, dans le cadre de l'agencement de données, sont elles-mêmes non-neutres et non-objectives, intégrées dans des jeux de pouvoir et héritières des logiques des organisations qui les produisent. Reprenant les travaux du philosophe des sciences Ian Hacking (1936 - 2023), ils précisent que deux processus interdépendants sont à l'oeuvre dans un agencement qui façonne la manière dont les données interagissent avec le monde qui, en retour, influence les futures versions des données et la constitution de l'agencement lui-même.

Le premier processus est « l'effet de boucle [12] » (*looping effect*) d'Hacking qui souligne la manière dont les données sont classées et organisées, comment une ontologie de données émerge et comment peut réordonné ce qui a déjà été classé. Cinq étapes la composent : (1) la classification (où les objets sont considérés comme porteurs de caractéristiques et rassemblés en fonction de ces dernières), (2) les objets de recherche (dans le cas des êtres humains, ils s'identifient à la classe qui leur a été assi-

[11] L'article mentionne seulement une traduction, nous avons repris le texte original en français retrouvé dans « le jeu de Michel Foucault » du tome III des dits et écrits (Foucault, 1994, p. 299).

[12] Nous reprenons ici la traduction de Mourey (2021)

gnée ; dans le cas des objets non-humains, les individus comprennent et interagissent avec ceux-ci en accord avec leur classification), (3) les institutions (institutionnalisent les classifications et gèrent les infrastructures de données), (4) le savoir (utilisé pour formuler, reproduire et améliorer les classifications) et (5) les experts (les personnes qui dans les institutions produisent, testent le savoir et mettent en place les classifications).

Le second processus, « les machines de découverte » (*engines of discoverability*), réalise le travail de l'agencement et contribue à sa légitimé et sa reproduction. Il est composé de six machines principales et de trois machines dérivées. Les machines principales sont (a) le comptage (compter les volumes d'un phénomène), (b) la quantification (transformer les comptages en mesures, ratios, classifications), (c) la création de normes (établir ce qui pourrait ou devrait être attendu), (d) la corrélation (déterminer des relations entre des mesures), (e) la prise de mesure (utiliser le savoir pour traiter des problèmes), (f) la scientifisation (définir et adopter les connaissances scientifiques). Les machines dérivées sont (g) la normalisation (façonner le monde pour qu'il soit conforme aux normes), (h) la bureaucratisation (mettre en place des institutions et des procédures pour gérer la production des attentes et entreprendre des actions) et (i) la résistance aux formes de savoir, aux normes, à la bureaucratie de la part de ceux qui sont concernés de manières négatives. Dans les cas précités, un « nomalisme dynamique » (*dynamic nomalism*) [13] est à l'oeuvre : les données et ce qu'elles représentent interagissent, conduisant à des changements mutuels.

Selon Kitchin (2014b), sur la base de ces éléments, il est nécessaire d'employer un panel varié de techniques telles que « les ethnographies, les entretiens, les groupes de discussion et l'observation participante pour approfondir le fonctionnement des agencements, retracer la manière dont le paysage des données a changé au fil du temps et de l'espace, pour cartographier les matérialités et les infrastructures qui constituent les infrastructures de données, et à déconstruire le régime discursif qui accompagne les initiatives basées sur les données [14] ».

Bibliographie de l'annexe 4

Foucault, M. 1994. *Dits et Ecrits, 1954-1988. Tome III. 1976-1979*. Gallimard. https://www.amazon.fr/Dits-Ecrits-1954-1988-III-1976-1979/dp/2070739880.

[13] « Je me considère comme un "nominaliste dynamique" intéressé par la manière dont nos pratiques de dénomination interagissent avec les choses que nous nommons – mais je pourrais tout aussi bien être qualifié de réaliste dialectique, préoccupé par les interactions entre ce qui est (et ce qui vient au monde) et les conceptions que nous en avons. » traduction de Mourey (2021) réalisée à partir de Hacking (2002).

[14] « *... such as ethnographies, interviews, focus groups, and participant observation to delve into the workings of assemblages, to trace out genealogies of how the data landscape has changed over time and space, to map the materialities and infrastructures that constitute data infrastructures, and to deconstruct the discursive regime accompanying data-driven initiatives* ».

Hacking, I. 2002. *Historical Ontology*. Revised edition. Cambridge, Mass. Harvard University Press.

Kitchin, R. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. 1st edition. Los Angeles, California: SAGE Publications Ltd.

Kitchin, R., & Lauriault, T. 2014. "Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work." SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=2474112.

Mourey, D. 2021. "Chapitre 6. L'apport de Ian Hacking pour penser les « effets de boucle » des outils de gestion." In *Philosophie et outils de gestion*, 142–66. Lectures / relectures. Caen: EMS Editions. https://doi.org/10.3917/ems.gilbe.2021.02.0142.

Annexe 5 : La data-philosophie

Développée assez récemment par Sonia Bressler dans (Bressler, 2020) et (Bressler, 2023), la data-philosophie se présente comme « un domaine de recherche interdisciplinaire qui explore les implications philosophiques, éthiques, épistémologiques et sociopolitiques de l'omniprésence des données numériques dans notre société. Nous devons la considérer comme une discipline à part entière en raison de la rapide évolution des technologies de l'information et de la communication, ainsi que de la prolifération des données massives et de l'intelligence artificielle (mais également pour faire face à l'émergente de l'informatique quantique) » (Bressler, 2023, introduction). La data-philosophie doit en ce sens explorer « les questions éthiques, épistémologiques et ontologiques soulevées par la création, la gestion, l'analyse et l'utilisation des données à grande échelle ».

Elle dissocie cette philosophie de celle de l'information développée par Floridi : « Ici, nous dirons juste que la philosophie de l'information (selon Floridi) a pour objectif d'étudier les questions conceptuelles qui se posent à l'intersection de la science du calcul (base de la technologie de l'information) et de la philosophie. Tandis que la "data-philosophie" opère sur la question du "réel" présupposé par la philosophie de l'information » (Bressler, 2020). Elle a pour concept fondamental la relation entre les données, la connaissance et la réalité. En effet, pour cette auteur, la donnée est un « critère du réel »

Basé sur cette vision des données comme critère du réel, quatre champs globaux intègre la data-philosophie : l'ontologie, l'épistémologie, la praxis et l'éthique de la data. Ceux-ci doivent répondre

à quatre questions centrales : que sont précisément les données ? Les datas ont-elle un genre ? Dans quelles conditions ?

Tableau 6.13.: Les champs d'investigation de la data-philosophie - compilés à partir de Bressler (2020) et Bressler (2023)

Champ	**Questionnements**	**Terrains d'investigation**
Ontologie	que sont précisément les données ? ; les datas ont-elles un genre ? ; dans quelles conditions une chose peut-elle être considérée comme une donnée ? ; quelles propriétés doit avoir une chose pour être ou pour jouer le rôle d'une donnée ?	Connaissance et vérité (comment les données massives peuvent être utilisées pour découvrir des connaissances significatives et comment distinguer les vérités solides des conclusions erronées ou trompeuses)
Epistémologie	quels sont les objets et les méthodes des data-sciences ? ; quels types de connaissance pouvons-nous avoir sur les données, de quelle façon pouvons-nous acquérir ou produire ces connaissances ? ; les sciences des données sont-elles un mouvement graduel ou discontinu ? ; les sciences des données obéissent-elles à leur propre logique ou bien si elles subissent l'influence de ce qui se déroule en dehors d'elles, et si les datas sciences répondent à une demande sociale ? ;	Connaisance et vérité (relever le défi de la fiabilité et de la validité des informations en contexte de données massives), limites des méthodes d'analyse (comprendre les limites et biais des méthodes d'analyse de données pour améliorer la fiabilité et la validité des connaissances tirées des données massives), biais inhérents (déceler et corriger les biais pour garantir une représentation plus juste et équilibrée de la réalité), qualité, validité et représentation des données (développer des méthodes et des critères pour évaluer la qualité des données, et promouvoir la transparence et l'ouverture dans le partage des données pour assurer la vérification et la réplication des résultats).

Champ	Questionnements	Terrains d'investigation
Praxis	quand une donnée devient une information et quand la data devient une information exploitable ?	l'analyse descriptive et exploratoire des données, l'analyse des données inférentielles, l'analyse prédictive des données, et l'analyse des données causales
Ethique	à quel type de valeur pouvons-nous ou devons-nous nous soumettre ? ; dans quelle pratique utilisons-nous les données ? ; quelles sont l'ensemble des possibilités de ces pratiques ?	respect de la vie privée (comment protéger les individus tout en permettant l'exploitation des données), la sécurité des données (analyse des normes éthiques et des meilleures pratiques pour garantir la sécurité et l'intégrité des données), la justice algorithmique (garantir l'équité et la justice dans le développement et l'utilisation des algorithmes), la transparence (la promouvoir pour renforcer la confiance et la responsabilité) et la responsabilité des entreprises et des individus impliqués dans la manipulation des données (comment les normes éthiques et les réglementations peuvent guider les pratiques en matière de données, ainsi que comment responsabiliser les acteurs pour leurs actions liées aux données).

Bibliographie de l'annexe 5

Bressler, S. 2020. "Fondements et enjeux de la data-philosophie." *Management & Datascience* 4 (4). https://doi.org/10.36863/mds.a.13148.

Bressler, S. 2023. *Data-Philosophie*. Paris: LA ROUTE.

Annexe 6 : les *critical data literacies*

Tableau des *critical data literaries*

Auteur	Conception de la littératie critique des données
(Tygel & Kirsch, 2016)	Elle est une littératie sur les données dont l'émancipation est une visée première. Une émancipation qui passe par une compréhension du monde, la remise en question des concepts hégémoniques (comme le PIB) et la nécessité de développement de nouveaux concepts (comme l'économie solidaire). La prise en compte des éléments de la vie quotidienne des personnes est vitale, notamment pour soulever les limites vécues (souvent perçues par Freire, selon ces auteurs, comme des situations d'oppression). Quatre compétences globales composent cette littératie : la lecture (savoir lire les données et comprendre la manière dont elles ont été conçues), la manipulation (savoir traiter techniquement les données), la communication (déterminer la bonne représentation et communication des données) et la production (la capacité de produire des données grâce à une bonne connaissance des formats et des outils de publication).
(Hautea, Dasgupta, & Hill, 2017)	Cette littératie est centrée sur les jeunes publics et la nécessité qu'ils prennent part à la conception d'artefacts de données en contexte social, de comprendre et questionner les effets des données numériques dans leur vie. Cinq nécessités sont intégrées dans cette littératie : le fait que la collection et la conservation des données ont un impact sur la vie privée, que l'analyse des données nécessite des capacités d'interprétation et un recours au scepticisme, que les données sont biaisées et porteuses de décisions cachées (hidden decisions), que les algorithmes sont exclusifs, et que la mesure et la communication des données peut nuire au système qui a conçu les données.

Auteur	**Conception de la littératie critique des données**
(Špiranec, Kos, & George, 2019)	La data literacy, en tant que pratique pédagogique située (situated educationnal praxis) doit promouvoir la justice sociale et bien commun, comprendre les relations et les inégalités de pouvoir, et œuvrer à réduire les inégalités (sociales, économiques, politiques, etc). Cinq thématiques la composent. D'abord, les traitements ontologiques des données et la critique du statut épistémologique des données où les données sont définies comme contextuelles et lieu d'interprétations. Puis la justification d'une littératie située dans les accentuations de problèmes clés, au sens d'une aide à la compréhension de certains aspects des sociétés contemporaines (ex : les algorithmes, le big data, la science des données, etc.). Ensuite, l'articulation pédagogique qui doit se faire à partir de véritables expériences et au sein de la réalité des apprenants. Et enfin, l'éthique au sens du respect de pratiques éthiques vis-à-vis des données comme le respect de la vie privée.
(Irgens et al., 2020)	Centrée sur une approche citoyenne liée à la participation, cette littératie met au cœur de ses analyses la reconnaissance et le questionnement des inégalités à l'œuvre dans l'utilisation des jeux de données, et la nécessité de prendre en compte les dimensions idéologiques, historiques et politiques embarquées dans la production et l'usage des données. Elle est l'affaire de compréhension (comprendre pourquoi la production des données n'est pas neutre mais politiquement, historiquement et matériellement située), de critique (questionner les usages et les biais) et de participation (employer les données pour conduire les changements nécessaires afin d'œuvre faire une société plus juste et durable).
(Claes & Philippette, 2020)	Cette littératie s'inscrit dans les systèmes de recommandation: elle a pour but la mise en place d'une pédagogie offrant la possibilité de remettre en cause et d'influencer les manipulations des données par les plateformes technologiques.

Auteur	**Conception de la littératie critique des données**
(McGowan, 2020)	Cette littératie, nommée critical sociotechnical literacy, se situe à l'intersection de la théorie critique du design qui interroge les conceptions ingénieriques et des pédagogies critiques du lieu (critical pedagogy of place) qui intègre un processus politique scindé entre une action de décolonisation (déconstruire et resituer le monde dans son contexte socio-historique) et une action de « réhabitation » (reconstruire l'histoire et les connaissances d'un lieu qui a souffert d'inégalités sociales et technologiques).
(Raffaghelli, Manca, Stewart, Prinsloo, & Sangrà, 2020)	Cette littératie critique, intégrée dans la pédagogie de l'enseignement supérieur, prend pour parti de réinterroger les pratiques des données et la techno-structure, de rendre visible grâce à la data literacy les conséquences politiques et éthiques de la datafication à travers les plateformes, de savoir transformer les données en information par l'intermédiaire de certaines techniques (comme la visualisation de données) et d'être capable de mesurer les effets des pratiques pédagogiques.
(Fotopoulou, 2020)	Cette littératie est centrée sur une approche devant permettre aux citoyens de pouvoir prendre part aux débats sur les aspects éthiques, idéologiques et de pouvoir des données, en particulier dans un contexte de datafication. Pour cet auteur, les data literacies sont nécessairement des littératies sociales (social literaties), contextuelles et interactionnistes du fait que c'est au sein des conditions matérielles d'existence que l'apprentissage a lieu et que les pratiques de données prennent leur sens. La prise en compte du contexte permet de comprendre les interactions entre acteurs ainsi que la manière dont les données ont été produites. Elles sont aussi « agentiques » (agentics) au sens que les données n'existent et n'ont d'influence que du fait de leur manipulation par les acteurs.

Auteur	Conception de la littératie critique des données
(Dangol & Dasgupta, 2023)	Dans cette publication, l'étude de Dangol et Dasgupta se centre sur une approche constructionniste appliquée à la littératie critique des données pour permettre à un jeune public de comprendre les données et les enjeux associés. Elle intègre plusieurs principes : le principe de continuité où les participants sont libres du mode de questionnement à appliquer aux données, le principe de pouvoir où les participants s'engagent dans des projets ayant de l'importance pour eux leur permettant de dépasser ainsi leurs craintes initiales (notamment l'appréhension technique), et le principe de la résonance culturelle où les pratiques des données se font dans des contextes socio-culturels élargis, ce qui permet l'émergence d'un processus d'apprentissage itératif et cumulatif.
(Atenas et al., 2023)	Cette littératie a pour but d'améliorer la pensée critique à travers la citoyenneté ainsi que des compétences de recherche. Elle promeut les principes en science et éducation ouverte, ce qui comprend l'utilisation de l'open data comme ressource en éducation ouverte.

Bibliographie de l'annexe 6

Atenas, J., Havemann, L., Rodés, V., & Podetti, M. 2023. "Critical Data Literacy in Praxis: An Open Education Approach for Academic Development." *Edutec. Revista Electrónica de Tecnología Educativa*, no. 85: 49–67. https://doi.org/10.21556/edutec.2023.85.2851.

Claes, A., & Philippette, T. 2020. "Defining a Critical Data Literacy for Recommender Systems: A Media-Grounded Approach." *Journal of Media Literacy Education* 12 (3): 17–29. https://doi.org/https://doi.org/10.23860/JMLE-2020-12-3-3.

Dangol, A., & Dasgupta, S. 2023. "Constructionist Approaches to Critical Data Literacy: A Review." In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 112–23. IDC '23. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3585088.3589367.

Fotopoulou, A. 2020. "Conceptualising Critical Data Literacies for Civil Society Organisations: Agency, Care, and Social Responsibility." *Information, Communication & Society* 0 (0): 1–18. https://doi.org/10.1080/1369118X.2020.1716041.

Hautea, S., Dasgupta., & Mako Hill, B. 2017. "Youth Perspectives on Critical Data Literacies." In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 919–30. CHI '17. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3025453.3025823.

Irgens, G-A., Knight, S., Wise, A,. Philip, T., Olivares, M-C., Van Wart, S., Vakil, S., & al. 2020. "Data Literacies and Social Justice: Exploring Critical Data Literacies Through Sociocultural Perspectives." https://repository.isls.org//handle/1/6665.

McGowan, B. 2020. "OpenStreetMap Mapathons Support Critical Data and Visual Literacy Instruction." *Journal of the Medical Library Association* 108 (4): 649–50. https://doi.org/10.5195/jmla.2020.1070.

Raffaghelli, J-E, Manca, S., Stewart, B., Prinsloo, P., & Sangrà, A. 2020. "Supporting the Development of Critical Data Literacies in Higher Education: Building Blocks for Fair Data Cultures in Society." *International Journal of Educational Technology in Higher Education* 17 (1): 58. https://doi.org/10.1186/s41239-020-00235-w.

Špiranec, S., Kos, D., & George, M. 2019. "Searching for Critical Dimensions in Data Literacy." Text. University of Borås. 2019. http://informationr.net/ir/24-4/colis/colis1922.html.

Tygel, A-F., & Kirsch, R. 2016. "Contributions of Paulo Freire for a Critical Data Literacy: A Popular Education Approach." *The Journal of Community Informatics* 12 (3). https://doi.org/10.15353/joci.v12i3.3279.

Annexe 7 : Les réglementations et organismes sur les données

- **1978** : la Loi Informatique et Liberté a amené à la création de la *CADA* et de la *CNIL*. La Commission d'accès aux documents administratifs (*CADA*) est chargée de veiller à la liberté d'accès aux documents administratifs et aux archives publiques ainsi qu'à la réutilisation des informations publiques[15], tandis que la Commission nationale de l'informatique et des libertés (*CNIL*) est chargée de veiller à ce que l'informatique soit au service du citoyen et qu'elle

[15] https://www.cada.fr/

ne porte atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques[16].

- **2007** : la directive *INSPIRE* de la Direction générale de l'environnement de la Commission européenne, « vise à établir en Europe une infrastructure de données géographiques pour assurer l'interopérabilité entre bases de données et faciliter la diffusion, la disponibilité, l'utilisation et la réutilisation de l'information géographique en Europe »[17].

- **2011** : création de la Direction interministérielle du numérique et de communication de l'État (*DISIC*)[18] ainsi que de la mission « Etalab » sous l'autorité du Premier ministre. La *DISIC*, sous l'autorité du Premier Ministre, « oriente, anime et coordonne les actions des administrations de l'État visant à améliorer la qualité, l'efficacité, l'efficience et la fiabilité du service rendu par les systèmes d'information et de communication ». *Etalab* coordonne la conception et la mise en oeuvre de la stratégie de l'État dans le domaine de la donnée, dont ses composantes juridique et sociétale, et avait pour mission de créer le portail data.gouv.fr [19], le portail unique interministériel des données publiques.

- **2013** : adoption de la charte sur « l'ouverture des données publiques » qui « marque l'ambition collective des États Membres de promouvoir une gouvernance plus ouverte et plus transparente : en établissant un principe d'ouverture par défaut des données publiques, en affirmant le principe de gratuité de leur réutilisation et en privilégiant les formats ouverts et non-propriétaires, elle encourage l'accès de tous à l'information et promeut l'innovation entrepreneuriale, citoyenne et sociale »[20]. La même année, une feuille de route souligne la nécessité d'une ouverture des données « stratégiques et de qualité », d'accompagner la réutilisation des données publiques ainsi que de la refonte du portail data.gouv.fr [21].

- **2014** : le décret n°2014-1050 entérine la création de la fonction d'Administrateur général des données (*chief data officer*) qui « coordonne l'action des administrations en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données par les administrations. Il organise, dans le respect de la protection des données personnelles et des secrets protégés par la loi, la meilleure exploitation de ces données et leur plus large circulation, notamment aux fins d'évaluation des politiques publiques, d'amélioration et de transparence de l'action publique et de stimulation de la recherche et de l'innovation »[22].

[16]https://www.cnil.fr/professionnel
[17]https://cnig.gouv.fr/presentation-de-la-directive-a8991.html
[18]https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000023619023
[19]https://www.etalab.gouv.fr/
[20]https://www.etalab.gouv.fr/leschefsdetatdug8signentunechartepourlouverturedesdonneespubliques/
[21]https://www.etalab.gouv.fr/lafeuillederoutedugouvernementenmatieredouvertureetdepartagedesdonneespubliques/
[22]https://www.legifrance.gouv.fr/loda/id/JORFTEXT000029463482

Henri Verdier sera nommé à cette fonction la même année[23] et sera suivi par Nadi Bou Hanna en 2018[24] puis par Isabelle Blanc en 2021. La fonction englobe aujourd'hui les données, les algorithmes ainsi que les codes sources.

- **2015** : intégration de la *DISIC* dans une Direction interministérielle du numérique et du système d'information et de communication de l'État (*DINSIC*), qui inclut aussi l'Administrateur général des données[25]. Par ailleurs est publiée la loi 2015-1779, dite loi Valter, relative à la gratuité et aux modalités de la réutilisation des informations du secteur public[26].

- **2016** : la loi pour une République numérique, dite « loi Lemaire », prévoit une ouverture des données publiques « par défaut » et une diffusion en ligne des bases de données et les données qui présentent un intérêt économique, social, sanitaire ou environnemental, mises à jour et « dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé » [27]. Par l'intermédiaire de son article 14, elle instaure le « service public de la donnée » qui « vise à mettre à disposition, en vue de faciliter leur réutilisation, les jeux de données de référence qui présentent le plus fort impact économique et social »[28]. Ce service public de la donnée sera lancé en 2017.

- **2018** : le Règlement Général Pour la Protection des Données (*RGPD*) est institué en France « établit des règles relatives à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et des règles relatives à la libre circulation de ces données ; protège les libertés et droits fondamentaux des personnes physiques, et en particulier leur droit à la protection des données à caractère personne »[29]. Dans ce cadre, les données personnelles sont définies comme « toute information se rapportant à une personne physique identifiée ou identifiable ». Une obligation pour les organismes manipulant les données personnelles de rendre compte des mesures prises est instaurée : elle comprend la minimisation des données recueillies, la sécurisation des données, de tenir un registre à jour des fichiers, d'informer les personnes du recueil de leurs données, et de signaler tout manquement à la *CNIL*[30].

- **2019** : création de la Direction Interministérielle du Numérique (*DINUM*) qui incorpore Etalab comme département et rattache l'administrateur des données à son organisation [31]. Son rôle est d'élaborer la stratégie numérique de l'État et de piloter sa mise en œuvre.

[23]https://www.silicon.fr/henri-verdier-chief-data-officer-etat-france-96874.html
[24]https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037524143
[25]https://www.legifrance.gouv.fr/loda/id/JORFTEXT000031194412/
[26]https://www.legifrance.gouv.fr/loda/id/JORFTEXT000031701525
[27]https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746
[28]https://www.data.gouv.fr/fr/pages/spd/reference/
[29]https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32016R0679
[30]https://www.cnil.fr/fr/reglement-europeen-protection-donnees
[31] https://www.numerique.gouv.fr/dinum/

- **2022 - 2024** : s'inscrivant dans la stratégie européenne pour la donnée, ont été proposés le *data governance act* ainsi que le *data act* au cours de l'année 2022 avant leur adoption l'année suivante.

  – Le *Data governance act*, adopté en 2022 pour une application en septembre 2023 vise à favoriser le partage des données (personnelles et non personnelles) par la mise en place de structures d'intermédiation[32]. Il « fournit un cadre pour renforcer la confiance dans le partage volontaire de données dans l'intérêt des entreprises et des citoyens. Il est un « instrument intersectoriel qui vise à réglementer la réutilisation des données publiques/détenues et protégées, en stimulant le partage de données grâce à la réglementation des intermédiaires de données nouveaux et en encourageant le partage de données à des fins altruiste »[33].

  – Le *Data act*, adopté en novembre 2023[34] pour une application prévue en septembre 2025[35], complète le *data governance act* et vise à une meilleure répartition de la valeur issue de l'utilisation des données. Il a pour but de (1) donner aux utilisateurs de produits connectés l'accès aux données qu'ils génèrent, (2) de permettre au secteur public de disposer des données détenues par le secteur privé, (3) de protéger les entreprises contre les clauses contractuelles abusives, (4) de permettre aux clients de changer de fournisseur de service de *cloud*, (5) d'inscrire plus durablement l'interopérabilité des données[36].

  – Dernièrement l'*IA Act*[37], supervisé par l'Office européen de l'IA créé en février 2024[38], devra classifier l'intelligence artificelle selon les risques qu'elle représente, faire retomber les obligations sur les fournisseurs et développeurs d'IA à haut risque, interdire certains types d'IA jugées trop dangereuses vis-à-vis de la sécurité des personnes, et de manière plus globale instaurer des règles de gouvernance.

Annexe 8 : les feuilles de route des ministères français

[32]https://www.cnil.fr/fr/strategie-europeenne-pour-la-donnee-la-cnil-et-ses-homologues-se-prononcent-sur-le-data-governance
[33]https://digital-strategy.ec.europa.eu/fr/policies/data-governance-act-explained
[34]https://www.dalloz-actualite.fr/flash/strategie-europeenne-pour-donnees-adoption-du-data-act-par-conseil-de-l-union-europeenne
[35]https://digital-strategy.ec.europa.eu/en/policies/data-act
[36]https://digital-strategy.ec.europa.eu/fr/news/european-data-act-enters-force-putting-place-new-rules-fair-and-innovative-data-economy
[37]https://digital-strategy.ec.europa.eu/fr/policies/european-approach-artificial-intelligence
[38]https://digital-strategy.ec.europa.eu/fr/policies/regulatory-framework-ai

Tableau 6.15.: Compilation des feuilles de route des ministères français

Ministère	**Objectifs visés**
Ministère du Travail, de l'Emploi et de l'Insertion	Le développement d'une culture commune de la donnée, à toutes les échelles de la décision et de l'action publique, participe de la dynamique de transversalité sous-jacente à la stratégie data du ministère et de sa volonté de monter en compétence et en qualité sur les services publics enrichis par la donnée. (p. 13)
Ministère de la Transition écologique, de la Cohésion des territoires et de la Mer	Diffuser la culture et développer les compétences de la donnée (p. 19)
Ministère des Solidarités et de la Santé	Le développement d'une « culture de la donnée » suffisante pour permettre les interactions au sein de la sphère Solidarités-Santé et avec les partenaires sur les sujets data est une priorité. Il s'agit à la fois de démystifier les sujets (« Qu'est-ce que l'IA ? », « Que sont l'open data et l'open source ? », « Qu'est-ce qu'une Dataviz ? », etc.) et de faire en sorte que chacun puisse mieux réaliser ce que l'on peut obtenir ou non des données dont on dispose (p. 29)
Ministère de la Transformation et de la Fonction Publiques	Le développement d'une culture de la donnée dans l'administration doit permettre une prise de conscience des agents publics (à commencer par les cadres dirigeants) du potentiel des données dans le cadre de leurs missions de service public (p. 36).
Ministère de l'Économie, des Finances et de la Relance	Développer « une culture commune de la donnée au sein de Bercy », à savoir des « des moments d'acculturation, des ateliers techniques, des nouvelles formations dédiées aux traitements des données ou des séquences de mentoring adaptées aux managers et l'encadrement supérieur » (p. 5)

Ministère	**Objectifs visés**
Ministère de la Justice	Deux facteurs-clés du succès des initiatives de valorisation de la donnée sont la connaissance par tous de ces initiatives et la capacité de s'approprier les données et les outils. Cette appropriation de ce qu'est la data, quelles en sont les possibilités et quels sont les moyens offerts par le ministère est indispensable tant au niveau des cadres dirigeants que de l'ensemble des agents (p. 11).
Ministère de l'Enseignement Supérieur et de la Recherche	« Le développement d'une « culture de la donnée » suffisante est une priorité pour favoriser l'émergence d'une dynamique de maîtrise et d'utilisation des données au sein du ministère. Cette acculturation à la gestion maîtrisée des données est aussi une priorité pour les établissements d'enseignement supérieur et de recherche dans lesquels le soin à apporter aux données à chaque étape du cycle de la recherche doit devenir la pratique. Les dispositifs d'acculturation et de formation doivent être élaborés et se décliner au plus près des enjeux spécifiques de chacun des domaines de données : administration, enseignement supérieur et recherche. » (p. 25)
Ministère des Armées	La compréhension par l'ensemble des agents des possibilités offertes par les données et par les technologies comme l'intelligence artificielle est une priorité. Promouvoir et structurer progressivement une filière « données » et la rendre attractive constitue un objectif essentiel pour garantir la maîtrise des technologies. Ces deux orientations concourent à la réalisation de l'axe de renforcement de la culture de la donnée au sein du ministère (p. 8).
Ministère de l'Intérieur	Cette acculturation est nécessaire car l'encadrement doit être en mesure de s'adapter aux évolutions de l'environnement dans lequel ils évoluent, et de tirer pleinement parti de toutes les innovations rendues possibles par une meilleure exploitation de la donnée. (p. 32)

Ministère	**Objectifs visés**
Ministère de l'Europe et des Affaires étrangères	L'acculturation des agents est l'une des priorités de la feuille de route. Il est en effet indispensable face aux grands changements occasionnés par la « révolution des données » de mettre les agents au cœur du dispositif [...] transmettre des éléments de contexte et de communication aux directeurs d'administration afin de faciliter leur rôle de sponsor, inspirer les agents et faciliter l'émergence de nouveaux cas d'usage au sein des services (p.16)
Ministère de la Cohésion des territoires et des Relations avec les collectivités territoriales	Objectif de l'acculturation : inclure le numérique dans le socle de compétences de tous les agents publics, anticiper les évolutions des métiers, impliquer les élus, cadres dirigeants et les managers, former en favorisant l'expérimentation (p. 23).
Ministère de la Culture	« La connaissance et les compétences techniques, méthodologiques ou juridiques sur les données sont centrales pour comprendre les évolutions des métiers impactés par la donnée et les innovations numériques, pour identifier et développer des usages innovants ou une amélioration des politiques culturelles et des services aux publics.Le ministère souhaite développer une culture commune de la donnée pour rendre familier aux agents les nouvelles possibilités permises par la technologie et l'analyse des données. À différents niveaux, de la conception des politiques publiques à l'encadrement des services comme pour la réalisation opérationnelle des activités, une maîtrise pertinente des enjeux et des possibilités est une priorité pour permettre aux agents de se projeter dans l'évolution des métiers » (p. 24)
Ministère de l'Agriculture et de l'Alimentation	Non définie

Ministère	Objectifs visés
Ministère de la Jeunesse et des Sports	Le rôle des données dépasse le cadre des métiers et des compétences professionnelles individuelles. Tous concernés par les données, que ce soit en tant que propriétaire ou responsable des données, propriétaire d'un processus de traitement des données, prescripteur d'applications embarquant des données ou en tant que simples utilisateurs, les agents ont besoin d'une culture commune. En particulier du fait de l'ouverture des codes et des données, l'administration doit présenter un visage unique et doit organiser une collaboration entre ses services pour « dé-siloter » les données et assurer ainsi leur cohérence voire leur unicité. Cette culture commune doit donc reposer sur les notions de collaboration et de transversalité. Développer un esprit « collaboratif » et reconnaître la dépendance vis-à-vis d'autres services, qu'ils soient en charge de la collecte, du traitement, de l'exploitation ou de l'utilisation, ou anticiper les coopérations interservices sont les objectifs à poursuivre (p. 20)
Services du Premier Ministre	« Placer l'ouverture des données, des algorithmes et des codes sources des SPM au cœur de la stratégie d'amélioration du service public et du débat public implique de développer, à tous les niveaux, une culture commune de la donnée et des enjeux associés en termes d'efficacité et de transparente (sic) de l'action publique. Pour y parvenir, la sensibilisation et la formation des agents constituent les deux principaux chantiers à lancer au sein des SPM » (p. 9)

Annexe 9 : les définitions de la *data culture*

Tableau des définitions de la *data culture*

Auteurs	**Définition (traduction française)**	**Définition (langue originelle)**
Thessen & Patterson (2011)	La phrase « data culture » renvoie aux pratiques de données explicites et implicites et aux attentes qui détermine la destiné des données. Elle est liée aux conventions sociales de l'acquisition, la conservation, la préservation, le partage et la réutilisation des données. Si l'objectif visé est de créer des données numériques, standardisées et en accès ouvert sous format réutilisable, alors les data cultures sont des points de départ pour déterminer les changements qui seront nécessaires avant la possibilité de réalisation de cette vision.	The phrase "data culture" refers to the explicit and implicit data practices and expectations that determine the destiny of data. It relates to the social conventions of acquisition, curation, preservation, sharing, and reuse of data. If the goal is to make data digital, standardized and openly accessible in a reusable format, then current data cultures provide starting points to determine the changes that will be needed before that vision can be realized.
Aragona & Zindato (2016)	La data culture renvoie à la connexion entre le moment où les données sont construites et le moment où elles sont utilisées pour produire du savoir dans un domaine spécifique.	Data culture refers to the connection between the moment when data are constructed and the moment when they are used to produce knowledge in a specific domain.
Albury, Burgess, Light, Race, & Wilken (2017)	Le terme « data cultures » est conçu pour être productif et dynamique… pour démêler la complexité des donnée dans les cultures numériques de rencontres et de drague , et de dépasser la vision simpliste « top-down, bottom-up » du pouvoir des données.	The term "data cultures" is intended to be generative and dynamic… to tease out the complexity of data within digitally mediated dating and hookup cultures, and to move beyond simplistic "top-down, bottom-up" understandings of data power.
Hora, Bouwma-Gearhart, & Park (2017)	… une « data culture » est développée dans laquelle les usages de données sont intégrés dans les normes et routines organisationnelles.	…a "dataculture" is developed in which data use is embedded in organizational norms and routines.

Auteurs	**Définition (traduction française)**	**Définition (langue originelle)**
Radermacher (2017)	Il existe une dimension sociale et politique dans la littératie contenue dans les données et les statistiques. Appelons-la data culture.	There is a social and political dimension to literacy in data and statistics. Let's call it data culture.
Kremser & Brunauer (2019)	Les valeurs, comportements et normes partagés par la plupart des individus d'une organisation en lien avec des problématiques liées aux données.	The values, behaviors and norms shared by most individuals within an organization with regards to data-related issues.
Elisa Raffaghelli (2020)	Une data culture est perçue comme une expression collective située qui recouvre des identités professionnelles, des règles et des pratiques spécifiques liées aux données, intégrée dans une culture institutionnelle.	A data culture is seen as a situated, collective expression which encompasses professional identities, policies and specific practices relating to data, as part of an institutional culture.

Auteurs	**Définition (traduction française)**	**Définition (langue originelle)**
Raffaghelli & Sangrà (2023)	La *data culture* décrit une institution et ses traditions, ses publics, ses histoires et ses symboles autour des données et de la datafication (p. 1). Nous proposons l'idée d'une *data culture* comme un dispositif conceptuel nous permettant de repenser les pratiques de données comme un phénomène relationnel, interdisciplinaire et multi-voix. Et ainsi de promouvoir une compréhension et une réflexion sur la datafication comme processus situé nécessitant de nouvelles littératies (critiques) au sein et à partir de l'enseignement supérieur pour questionner, résister ou transformer les pratiques et histoires des données (p. 1). Nous introduisons le concept de data culture comme un dispositif potentiel permettant d'explorer et de comprendre la complexité. Une data culture est située et requiert des acteurs une connaissance des contextes socio-historiques des institutions et des communautés. De plus, les acteurs peuvent transformer la *data culture*, créer des espaces et des pratiques plus justes (p. 3).	Data cultures characterise an institution and its tradition, people, narratives, and symbols around data and datafication (p.1). We offer the idea of data cultures as a conceptual apparatus, enabling us to rethink data practices as a relational, interdisciplinary and multivoiced phenomenon. And hence, promoting understanding and reflection around datafication are situated processes that require new (critical) literacies within and from higher education, to question, resist or transform data practices and narratives (p. 1). We introduce the concept of data culture as a potential apparatus to explore and understand complexity. A data culture is situated and requires actors' acknowledgement of the institutions and communities' socio-historical context. Moreover, the actors can transform the data culture, building fairer spaces and practices (p. 3).
Correll & Garrison (2024)	Les manières dont les données, le savoir et l'information sont conceptualisés, collectés, structurés et partagés.	ways that data, knowledge, and information are conceptualized and collected, structured and shared.

Bibliographie de l'annexe 9

Albury, K., Burgess, J., Light, B., Race, K., & Wilken, R. 2017. "Data Cultures of Mobile Dating and Hook-up Apps: Emerging Issues for Critical Social Science Research." *Big Data & Society* 4 (2): 2053951717720950. https://doi.org/10.1177/2053951717720950.

Aragona, B., & Zindato, D. 2016. "Counting People in the Data Revolution Era: Challenges and Opportunities for Population Censuses*." *International Review of Sociology* 26 (3): 367–85. https://doi.org/10.1080/03906701.2016.1244926.

Correll, M., & Garrison, L-A. 2024. "When the Body Became Data: Historical Data Cultures and Anatomical Illustration." arXiv. https://doi.org/10.48550/arXiv.2402.05014.

Raffaghelli, E-J. 2020. "Is Data Literacy a Catalyst of Social Justice? A Response from Nine Data Literacy Initiatives in Higher Education." *Education Sciences* 10 (9): 233. https://doi.org/10.3390/educsci10090233.

Hora, M., Bouwma-Gearhart, J., & Park, H. 2017. "Data Driven Decision-Making in the Era of Accountability: Fostering Faculty Data Cultures for Learning." *The Review of Higher Education* 40 (January): 391–426. https://doi.org/10.1353/rhe.2017.0013.

Kremser, W., & Brunauer, R. 2019. "Do we have a Data Culture?" In *Data Science – Analytics and Applications*, edited by Peter Haber, Thomas Lampoltshammer, and Manfred Mayr, 83–87. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-27495-5_11.

Radermacher, W. 2017. "The Future Role of Official Statistics." In. https://doi.org/10.2785/721672.

Raffaghelli, J-E., & Sangrà, A. 2023. *Data Cultures in Higher Education: Emergent Practices and the Challenge Ahead*. Vol. 59. Higher Education Dynamics. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-24193-2.

Thessen, A., & Patterson, D. 2011. "Data Issues in the Life Sciences." *ZooKeys* 150 (November): 15–51. https://doi.org/10.3897/zookeys.150.1766.

Annexe 10 : le guide d'entretien

1. Le cadre contextuel

Définir la data literacy

1.1 Quelle définition donnez-vous de la *data literacy*, employez-vous ce terme et pourquoi ?

1.2 Où situez-vous la *data literacy* vis-à-vis d'autres concepts liés aux données (donner un exemple comme la *data science*)?

1.3 Quelle « nature » de la *data literacy* pourriez-vous donner ?

Les enjeux de la data literacy

1.4 Quelle est l'importance de la *data literacy* pour une organisation comparée à d'autres nécessités (ex : *information literacy*, technicité informatique...) ?

1.5 Depuis quand est-ce devenu une nécessité (est-ce que vous pourriez le dater) et pour quelles raisons ?

2. Les enjeux des données

2.1 Qu'est-ce qu'une donnée pour vous ?

2.2 Quels en sont les enjeux actuels et est-ce que la *data literacy* permet d'y répondre ?

2.3 Y a-t-il des types de données en particulier qui sont au coeur des enjeux précités ?

2.4 Quelle est l'influence du cadre législatif (ex : *RGPD*) sur vos pratiques? Constatez-vous une différence ?

3. Application de la *data literacy*

La mise en application

3.1 Comment met-on en place la *data literacy* dans une organisation ?

3.2 Doit-elle concerner tout le monde ou seulement des acteurs spécifiques ?

Les compétences

3.3 Que doivent développer les personnes à l'issue de cette apprentissage de la *data literacy* ?

3.4 Est-ce que le terme d'esprit critique vis-à-vis des données vous parle et si oui que recouvre t-il pour vous ? Même question pour l'éthique des données.

Les formations

3.5 Quel est le niveau général en culture des données des personnes (collègues, étudiants, salariés, etc.) que vous fréquentez ?

3.6 Quelles sont les difficultés actuelles de cette mise en place ?

Les projets

3.7 Avez-vous en tête un projet *data literacy* en France qui a bien fonctionné et pourquoi ?

Conclusion

Quelles seraient vos recommandations pour une meilleure appropriation de la *data literacy* en France ?

Annexe 11 : les entretiens avec le public « associatif »

Associatif 1 (04/03/22)

Qu'est-ce que pour vous la *data literacy* et employez-vous ce terme ?

Cela dépend de mon interlocuteur mais non la plupart du temps j'utilise le terme de « culture des données ». Après comme je vous l'ai dit j'ai changé de profession donc il y a des choses que j'ai un peu oubliées, mis de côté. J'avais bien des définitions sur tout ce qui est culture des données qu'on avait mises en place dans nos différents travaux, mais je ne suis plus capable de vous la ressortir. Grossièrement, c'est pouvoir être à l'aise avec les données. On avait fait plusieurs niveaux dans le travail qu'on avait fait avec [nom de personne], mais globalement c'est pouvoir être à l'aise avec les données qui nous concernent et qu'on manipule.

Il y a tout un ensemble de concepts autour des données : on parle de gouvernance de données, de *data science*... où peut-on situer la *data literacy* ? Est-ce la même chose ? Est-elle séparée ?

Dans la transformation numérique territoriale on parle aussi pas mal de culture des données. Donc en fait pour répondre à votre question : j'avais décidé qu'il y avait sept pôles dans le travail qu'on avait à faire avec les collectivités territoriales sur les données. Il y avait le pôle « gouvernance », il y avait la partie technique, tout ce qui était législation, etc. Enfin, techno-législatif. Ensuite, il y avait les données elles-mêmes. Il y avait la partie « culture des données ». Et puis après il y avait la partie « expérimentation ». Et ensuite, il y avait la partie « communication » : storytelling plus exactement, c'est comme ça en général qu'on le racontait. Donc la partie *data science* c'est plutôt dans la sphère technico-donnée. La partie « qu'est-ce qu'on fait ? Comment on le fait ? », c'est la partie gouvernance. « Expérimentation », ce sont les prises de risques, les tentatives qu'on peut faire par rapport à ça. Et la partie *storytelling* qui est souvent très oubliée de nos amis les « data litérateurs », bien que quand ils sont journalistes, ils arrivent à bien le faire. Selon les cas, on s'appuie sur un élément ou un autre.

Dans mon travail, quand je rencontrais un nouveau client, j'essayais déjà de voir quels étaient ses points forts, ses points faibles dans ces différents éléments. Et puis, j'essayais de remettre en équilibre : parce qu'on peut être hyper puissant dans la partie « données » par exemple ou hyper puissant dans la partie « expérimentation » mais avoir laissé un peu en souffrance les autres niveaux. Et tant qu'on n'a pas mis tout ça à peu près sur le même plan, eh bien on n'arrive pas à faire l'effet de transformation qui permet d'aller au niveau supérieur. Donc quand on menait un cours spécifique, par exemple du grand public ou pour mener des cours auprès des bibliothécaires, on faisait attention à ce type d'équilibre pour leur donner un niveau d'information équilibré qui leur permette d'être en capacité d'agir.

Pour parler des caractéristiques de la *data literacy*, est-ce que l'on va parler d'une culture, d'un ensemble de pratiques et de techniques, d'un état d'esprit ou tout ça à la fois ?

Pour nous, du point de vue du pédagogue, c'était une culture et une mise en pratique, forcément. Le plus possible on se disait que tant qu'on parlait des données c'était du brouhaha. C'est le fait de pouvoir donner à voir, que les gens puissent prendre, puissent manipuler, puissent changer. On sortait le plus possible de l'ordinateur pour qu'on puisse avoir d'autres appréhensions, les plus sensorielles possibles, par rapport aux données afin que les gens puissent se les approprier.

Pour rebondir sur ce point, la donnée pose des problèmes d'un point de vue conceptuel. Aussi, je voudrais vous poser la question : qu'est-ce qu'une donnée pour vous ?

Alors on passait pour certains de nos cours quasiment trois quart d'heure à le définir avec nos participants. La technique pédagogique qu'on avait employée, c'était de poser la question que vous venez de poser et on leur disait par exemple : « ce matin vous êtes venus en formation, vous êtes venus en atelier, quelles sont les données dont vous vous êtes servis ? quelles sont les données que vous avez générées ? » ou « quelles sont les données qu'il y a dans cette salle ? Quelles sont les données qu'il y a sur vous ? ». On leur posait ce genre de questions et ils se mettaient à chercher sur les radiateurs, ils se mettaient à chercher dans leurs fringues ou sur les bouteilles

d'eau ou ils nous parlaient de la température, la circulation, et ce genre de choses. En tout cas, avec un groupe de cinq ça marche déjà très bien. Avec un groupe de trente, il n'y a pas de soucis : ils arrivent en moins de dix minutes à nous trouver une cinquantaine de données. Et après on faisait de la déconstruction en disant : « ces données elles sont tiers, propriétaires, comment on agit dessus ? est-ce qu'on peut agir dessus ? comment on s'en sert ? dans quel système d'informations elles sont ? ».

Donc, par exemple, si on me donnait la référence de la table, je disais : « elles sont dans le système d'informations du fabricant, dans le système d'informations du livreur, dans le système d'informations de celui qui nous hébergeait, dans les processus d'inventaire, dans les processus de comptabilité, etc. ». On expliquait que ces données-là étaient ressaisies et recaptées à plusieurs endroits et on essayait de mettre un point fort sur pourquoi on captait ces informations-là et quelles étaient les finalités.

Par exemple, si je demandais quelle était la température de la pièce et que je suis technicien en réglage de températures ou en économie d'énergies ou en confort dans un hôtel de luxe pour mes participants, je ne vais pas capter la température au même endroit et je ne vais pas être vigilant à la même finesse de la température que je captais. En s'appuyant sur cet exemple, on essayait de leur faire toucher pour la première fois du doigt, parce que c'est une notion qui est difficile à comprendre, que la donnée elle est décidée, elle est construite : il y a une intention puissante qui est derrière.

Parfois on voyait des *data scientists* qui prenaient des données de plusieurs endroits différents et qui faisaient une belle salade niçoise avec : mais il fallait bien regarder ce qu'on saisissait comme données pour pouvoir en sortir de l'intelligence et que l'intelligence elle était toujours du côté de l'humain. On insistait énormément là-dessus pour que nos interlocuteurs se sentent moins dépossédés par rapport aux données parce que souvent il y avait un complexe d'infériorité, d'informatique, de mathématiques et j'en passe. Plus la suprématie technocratique usuelle dans les discours.

On essayait de casser un peu les pattes à quelques idées préconçues sur cette question de définition des données. Et du coup, ça nous demandait pas mal de temps et j'avoue qu'on s'est beaucoup battues avec les *data scientists*, même avec nos collègues pédagogues parce que ça ils avaient du mal à entendre qu'il fallait vraiment passer du temps sur cette partie-là. Les *data scientists* nous disaient qu'on étaient peu technicien(ne)s et qu'on n'avait qu'à fermer nos gueules. J'exagère un peu mais à peine. Il y avait une domination technocratique sur certains aspects de notre ignorance et de nos postures et on oubliait un petit peu notre hyper technicité de pédagogue.

Est-ce que cela voulait dire que de leur côté il n'y avait pas de grandes interrogations sur ce que pouvait être la donnée, qu'ils l'utilisaient comme une application pratique sans grand questionnement ?

Il y avait en partie cela et en partie « si tu sais pas coder, ferme ta gueule ». Un peu cette idée-là où comme parfois dans les sphères d'informaticiens il faut un peu avoir son passe-droit d'hyper compétence technique pour pouvoir avoir droit à la parole. Un peu de méritocratie. Il faut gagner des galons pour trouver sa place et que notre parole soit entendue. On peut toujours parler : les tours de table se font mais les inscriptions dans les compte-rendus ou les prises en considération de certaines paroles étaient plus difficiles du fait que

nous n'étions pas codeuses. Alors que nous, nous étions dans l'affirmation qu'on voulait être du côté du grand public et qu'on voulait savoir : on était des curieux/ses et on voulait être des passeurs/seuses.

Sur les enjeux actuels autour des données, lesquels pourrions-nous citer et comment la culture des données peut y répondre ?

C'est un peu autour de remettre l'humain au centre des données et pas les données au centre de la décision humaine. Et de bien revaloriser l'intelligence humaine. Certes, il y a des choses qu'on ne sait pas faire, notre cerveau est limité et c'est sympa de faire appel à un ordinateur. Moi je me sers des ordinateurs tous les jours, je suis bien content(e) de pouvoir utiliser un ordinateur pour vous parler aujourd'hui. Je suis hyper favorable à l'informatique. Mais je pense aussi que l'humain a toute sa place. Donc de ne pas être dans un sentiment et dans une relation de domination par rapport aux données et de se dire « ce n'est pas *Google* qui a la réponse, c'est bien moi qui a la réponse ». Oui, *Google* a plein de réponses mais je dois toujours être lucide quand j'ai affaire à une donnée. De savoir qui la produit, dans quelles intentions, de me poser les bonnes questions pour savoir quelles sont les sources, qu'est-ce qu'on essaye de me faire comprendre, de bien être au cœur de tout cela. Et pour moi la *data literacy* donne des clés, elle donne des outils pour pouvoir être autonome, pour pouvoir le faire. Mais il faut un accompagnement assez structuré pour que nos interlocuteurs puissent être autonomes.

Et cette nécessité de maîtriser les données, en tout cas de la comprendre, d'avoir une certaine culture des données, est-ce qu'on peut la dater ? Et si on peut expliquer le besoin actuel de maîtriser ces données ?

Ce que je vous raconte, la plupart des interlocuteurs ne le comprennent pas. Je me souviens avoir travaillé pour [nom d'organisation] sur justement la question des données sur une problématique concernant la base de données sur l'inclusion numérique. A [nom d'organisation] on pourrait se dire que les gens sont un peu éclairés sur la culture des données, qu'ils savent à peu près ce qu'est une base de données. Mais quand j'ai fait ce travail, ils ne se rendaient pas du tout compte des enjeux qui étaient liés à cette notion de rendre les gens autonomes par rapport à l'informatique et de pouvoir leur donner cette capacité de prendre le pouvoir. Je dévie un peu de votre question mais c'est un point que je voudrais mentionner dans cet entretien : les gens qui sont dans l'inclusion numérique, on pense qu'ils doivent savoir ouvrir un fichier *Excel* ou savoir rédiger une lettre de motivation. Or, ils ont aussi besoin de savoir se défendre quand il y a un commercial qui a un ordinateur qui l'aide à guider l'entretien pour leur vendre un produit, qu'ils puissent savoir de quoi il en retourne et qu'ils soient un peu armé pour pouvoir aller dans le sens où ils veulent aller et ne pas être pilotés par des systèmes de manipulation commerciaux et informatiques. Donc la culture des données, elle doit surtout aider les plus fragiles à pouvoir se défendre sur ces facettes-là.

Est-ce que cette culture des données est devenue plus importante, aussi importante, ou moins importante que lorsqu'on parlait de culture de l'information que l'on entend maintenant surtout du point de vue de la désinformation et du complotisme, qu'on parlait aussi de l'informatique... Est-ce que c'est devenu aussi important que tout ce que j'ai pu vous citer ?

Oui. On voit bien par exemple : « pour ou contre le nucléaire ? ». Si on est bien informés, qu'on sait décrypter les données qu'on nous présente, on va savoir répondre : « moi par rapport à ce que j'ai envie et les différents éléments, les différentes données que l'on me présente, je vois qu'on essaye de m'embrouiller la tête, je vois bien que ces informations elles sont cohérentes avec les informations que j'ai collectées ». Donc oui, comme on est vraiment envahi de données, c'est vraiment important. On est vraiment englué dans notre vie quotidienne dans la capture de nos données personnelles, donc doublement oui. Parce que si on ne veut pas trop subir cet engluement, on doit être des citoyens éclairés, actifs et vivants : il faut qu'on ait beaucoup de lucidité sur ces sujets. C'est déjà difficile pour des personnes qui sont un peu éclairées, qui sont désenglués. Donc plus on va apporter des éléments pour que les personnes soient au fait de ces éléments-là, plus ça sera important.

Cela me fait penser à une chose : est-ce qu'il y a des types de données qui sont au cœur des enjeux actuels ?

Si on se met dans la posture humaniste bienveillant, comme je l'ai laissé transparaître dans mon discours, protéger le plus faible me semble vraiment un élément important donc à ce moment j'irai plutôt sur les données les plus importantes qui sont celles qui concernent toute la sphère identifiée par la *CNIL* autour du *RGPD*. De bien savoir ce que je peux en faire, ce que les autres peuvent en faire et quels sont les impacts. Ça nous donne un peu de liberté économique et cela peut permettre d'avoir plus de liberté vis-à-vis de la manipulation. Il faut bien commencer quelque part donc c'est une base, c'est un socle : je dirais que c'est un peu le socle minimum mais ce n'est pas le socle le plus facile à approcher.

Parce que notre interlocuteur qui va vouloir faire sa montée en compétences dans la culture des données, il va venir avec un bagage, une préoccupation. Il va être soit militant écologiste, soit militant politique, soit avoir envie de baisser sa facture d'eau. Et donc il va plutôt falloir apporter la culture des données sur la problématique du moment, le truc qui le préoccupe, parce qu'on aura un taux de pénétration, un taux de pédagogie, avec des effets de levier beaucoup plus intéressants. Le point d'entrée principal c'est la préoccupation du moment de notre interlocuteur. Que ça soit de la randonnée, des questions alimentaires, des aspects de santé, de choisir la maison de retraite pour son père ou sa mère, cela n'a pas d'importance. Le point d'intérêt qui est déjà éveillé chez l'autre, c'est ce fil-là que l'on va tirer avec lui petit à petit. Et à nous de voir si on l'amène sur des données cartographiques, des données politiques, des données sur la vie privée et ainsi de suite. Le but c'est de répondre aux attendus de notre interlocuteur.

Et est-ce qu'en tant que professionnel(le) vous avez constaté une différence dans vos pratiques du fait de l'arrivée de la donnée ?

Quand on a commencé à beaucoup parler du *RGPD* ?

Même avant en fait.

Est-ce que vous pouvez reformuler la question ?

Est-ce que l'arrivée du numérique, l'usage nouveau des données numériques a changé quelque chose ? Par exemple, on a vu émerger le courant des data journalistes. L'arrivée des données a changé les pratiques et les compétences. Et je voulais savoir si vous aviez eu la même chose ou pas du tout.

J'ai travaillé 30 ans dans l'innovation donc j'ai toujours été sur le fil du rasoir. J'ai toujours transformé toutes mes pratiques. J'ai commencé à donner des cours en informatique sur comment se servir de *Word* pour la chambre de commerce donc j'avais des commerçants, coiffeurs, charcutiers, qui voulaient simplement ouvrir leur courrier et savoir se servir d'internet.

Et puis après, petit à petit, de fil en aiguille, je suis devenu(e) chef(fe) de projet en communication numérique. Je m'étais déjà questionné(e) sur ce type de sujets. Après j'ai glissé dans l'*open data*: j'en ai fait sept ans. Donc mes pratiques ont toujours changé. Il n'y a pas une année où je n'ai pas transformé ma connaissance, ma culture des données. Donc oui, ça ne fait que bouger tout le temps, même encore aujourd'hui. Je vois comment mes amis, mes collègues utilisent des données pour pouvoir faire leur boulot, je suis complètement en décalage par rapport à eux sur mes pratiques. Parce que je vais chercher directement sur les portails *open data* parce que c'est plus pratique pour moi : c'est plus simple, je sais où je vais trouver les infos. Donc même sur des choses qui n'ont plus rien à voir cela influence complètement ma vie quotidienne et ma vie professionnelle. Et même ma vie familiale parce que ce que j'échange avec mes enfants sur ce qu'il se passe sur les réseaux sociaux, c'est pareil : les données reviennent souvent en discussion.

Si on revient sur ce point : la loi Lemaire, le RGPD, on a eu beaucoup de lois en lien avec les données ces dernières années, est-ce que vous ressentez un impact sur les pratiques professionnelles ou pas du tout ?

La loi Lemaire est sortie parce qu'il y avait [nom d'organisation] et on a eu la chance de rencontrer Axelle Lemaire. On lui a dit « l'open data en France, ça ne va pas du tout ». Elle nous a dit « ah bon ? Vous m'intéressez ». On lui a répondu « on aurait beaucoup de choses à vous dire là-dessus ». Donc elle nous dit « écrivez-moi un rapport ». On lui a écrit un rapport de 100 pages. Et elle nous a dit, en réponse à ce rapport « eh bien je vais vous donner [somme d'argent] pour que vous accompagniez les acteurs publics à monter en compétences dans l'ouverture des données ». Donc non seulement la loi est sortie mais elle a aussi financé notre secteur pendant un petit moment puis les fonds sont partis.

En termes de discours, on est dans le petit schéma que je vous ai fait tout à l'heure, c'est-à-dire que le fait qu'il y ait des lois qui soient réelles et sérieuses, ça a participé à équilibrer notre discours technique. Ça donnait des directives sur les données cartographiques dans les collectivités territoriales pour que nos interlocuteurs nous entendent. Ou même pour aider les associations qui étaient militantes sur le développement durable. Il y a beaucoup de lois pour ouvrir les données dans la thématique du développement durable. Et toutes ces données-là qui concernent le développement durable, il y a beaucoup de facilitation qui sont offertes aux militants. Une fois que la loi est en place et nous accompagne, ça donne quand même de la puissance au discours auprès de certains interlocuteurs. Le citoyen en général il s'en moque la plupart du temps. Mais les militants, les associations, les acteurs publics, ça vient compléter notre argumentaire commercial. Il n'y a pas que ça mais ça aide un peu.

Je vais entrer dans la dernière partie de l'entretien. Sur le côté pratique, si on veut vraiment aujourd'hui mettre en place une culture des données dans un organisme, comment est-ce que cela va se définir ? Est-ce que l'on va avoir quelqu'un en charge ? Des moyens particuliers ? Et est-ce que ça doit concerner tout le monde ?

On a un petit schéma où on a un système d'étages qui dit qu'en bas on diffuse : on infuse la culture des données par quelques actions. C'est mieux si une ou deux personnes qui sont missionnées spécifiquement dans l'entreprise pour pouvoir le faire. Et puis après, petit à petit, on reconstruit le schéma dont je vous ai parlé où on commence à avancer, à sensibiliser en interne. L'idéal en entreprise c'est de pouvoir aller jusqu'à l'expérimentation. Une fois qu'on a fait une expérimentation, on fait du *storytelling*. Une fois qu'on a fait du *storytelling* dessus, on commence à embarquer des communautés plus grandes autour de sujets autour de l'ouverture des données.

Au début, cela s'est exactement passé comme ça : on avait trois interlocuteurs de l'informatique qui étaient plutôt en recherche sur la manière de mettre en place une gouvernance des données. Notre porte d'entrée, c'était ça. Et en particulier, ils étaient un peu dans l'embarras parce qu'ils pouvaient savoir que Monsieur Tartempion ou Madame Tartempion les appelaient à un moment donné, mais le commercial n'avait pas l'historique de son interlocuteur qu'il avait au bout du téléphone parce qu'il y avait des données qui concernaient les clients, etc. Ils voulaient améliorer un peu les interfaces par rapport aux données qui étaient produites par les différents services. Ensuite, il y a un des informaticiens qui s'est emparé de cette question, donc on a fait des ateliers au niveau des cadres au moment des rencontres pour les cadres. Et on a fait petit à petit l'accompagnement de projet pour faire émerger des projets. Après ils sont passés en stade d'expérimentation. Et après évidemment ils ont fait de la diffusion de l'information, du *storytelling*, etc. C'est un bon exemple de bonne appropriation de culture des données.

Que doivent développer les personnes qui suivent une formation en culture des données ? Un esprit critique, une certaine culture, des compétences ?

En fait, quand vous me parlez de culture des données, c'est comme avoir la culture de la musique. C'est impossible d'avoir la culture complète des données. On peut parler d'historique, de bases en statistiques, comme on peut parler de trucs hyper pointus avec les *data scientists*. Et tout ce qu'il y a entre. La culture des données, c'est un investissement temps. Mais par contre, il y a quand même un peu de méthodologie qu'on peut avoir pour aider les personnes à structurer leur culture des données et qu'elles sachent qu'elles ont évolué, un peu comme le *B2i* avant. Avec le *B2i* on vous disait « est-ce que tu sais ce que c'est qu'un point xls ? ». Si je sais ce que c'est qu'un point xls, je sais ce que c'est qu'un tableur, que je sais l'enregistrer, que je peux enregistrer sous un autre format de fichier, etc. Donc petit à petit, je sais que j'ai quelques marqueurs qui vont faire que j'ai acquis des compétences dans un domaine. Et ce qui est toujours très agréable et très intéressant, c'est savoir quelles sont tous les possibilités d'acquisition dans ce domaine-là et comment je me situe par rapport à ce niveau d'acquisition. Et donc, selon mon métier, est-ce que j'ai besoin d'être fort en données géographiques, en données *RGPD*, en datavisualisation, savoir comment je me situe par rapport à ces sujets-là. Et me dire « je suis commercial, je fais du marketing et je me rends compte que je suis nul en cartographie des données. Eh bien dans mon métier, peut être que pour explorer, pour pouvoir mieux structurer ma technique commer-

ciale, ça serait vraiment intéressant de savoir représenter mes données sur une cartographie. Et du coup, ça vaut le coup que j'investisse du temps dans cette partie-là de la culture des données parce que je sais que je vais avoir un retour sur investissement intéressant ». C'est vraiment le travail qu'on avait voulu faire avec [nom de personne] sur ces aspects-là.

Est-ce que quelqu'un qui sait à peine écrire une lettre sur son ordinateur n'a pas besoin de bien se repérer dans la sphère informationnelle des collectivités territoriales et savoir qu'à un moment il peut aller chercher des données ouvertes sur une problématique qui l'intéresse ? Je ne crois pas.

Je crois que petit à petit il faut que mon interlocuteur, qui n'est pas à l'aise avec l'informatique, sache se débrouiller pour pouvoir aller sur les sites internet des collectivités territoriales, pour aller chercher des informations qui lui sont utiles, pour pouvoir trouver des financements qui lui sont utiles, pour pouvoir faire les sorties pleine nature dont il a besoin pour qu'il soit en pleine santé et que petit à petit, il puisse s'intéresser à la concertation et que petit à petit il puisse s'intéresser aux données qui lui sont proposées. Pour moi, c'est l'encapacitation.

Après il y a des gens qui vont s'arrêter à un certain stade. Mais il ne faut jamais oublier qu'un citoyen est capable de faire beaucoup d'efforts si le sujet le passionne. Donc à nous de lui proposer l'éventail qui l'intéresse et si le sujet le passionne ça devient un administrateur d'*open street map* parce qu'il a trouvé des copains sympas, parce que ça lui sert pour ses randonnées, parce qu'il est prêt à aller sauver d'autres personnes qui ont vécu un tsunami à Hawaï et j'en passe. L'acteur public doit pouvoir offrir son champs des possibles pour que l'encapacitement se mette en place. Et c'est ça que je regrette beaucoup dans nos politiques : elles ne se rendent pas compte qu'il n'y a pas plus important que de rendre nos citoyens les plus actifs possibles. Voilà où j'en étais arrivé(e) dans mes réflexions sur pourquoi je faisais de la *data literacy*.

Justement par rapport à cela, est-ce que l'on peut estimer la situation au niveau des compétences et des connaissances des personnes, et j'emploie le terme de personne de manière très globale, cela peut être le citoyen lambda, sur la culture des données. Est-ce qu'on est dans une situation catastrophique ou est-ce que ça évolue au fil du temps ?

Tant que le projet ne sera pas porté de façon éclairée, cela ne marchera pas. Il y a eu des tentatives de la part de beta.gouv qui a fait, je sais pas comment ça s'appelle, ça a changé de nom plusieurs fois, une espèce d'autotest, qui remplace le *B2i* pour savoir où est-ce que j'en suis dans ma culture des données.

Le PIX.

PIX ça devrait être d'une banalité des plus banales auprès de tous les enseignants. Au moins ceux qui portent la culture des données, la culture numérique et la culture informatique. On n'en est très loin. On a quand même eu sur le plan pédagogique une porte d'entrée fabuleuse où dans les programmes était indiqué au niveau des lycées une appropriation d'*open street map*. C'était marqué noir sur blanc. *Open street map* était heureux. Mais ce n'était pas accompagné. C'est-à-dire que c'était une ou deux lignes dans un programme. C'est quand même bien précisé que les gens doivent savoir placer un point sur une carte, s'approprier *Open Street Map*, qu'ils

doivent être contributeurs, qu'ils doivent s'en servir dans les cours, etc. Qu'ils doivent comprendre tout ce qui relevait de la cartographie. Sur le plan pédagogique, il y a beaucoup à faire mais qu'est-ce qui est mis en place pour que ça soit suivi des faits ? Quelle formation propose t-on aux profs pour qu'ils puissent vraiment eux-mêmes être *open street map friendly* et être en capacité de le transmettre aux élèves ? Rien. Ce sont des petites pierres mais pour l'instant on a des petites miettes qui sont mises comme ça par petites bouts : on a *PIX*, *open street map*, etc. Mais pour que ça marche, il faut un vrai plan cohérent pour monter en compétence en culture des données. Il faut quelque chose de structuré.

** Les difficultés actuelles viendraient donc d'un manque de cohérence, d'un manque de politiques publiques ?**

C'est un peu tout ça oui. Je reprends mon petit schéma : c'est-à-dire que la gouvernance sur ce sujet-là, elle n'est pas du tout en place. On en est à l'âge de pierre. Sur le plan technique, on va dire sur la pédagogie, il y a quelques acteurs qui savent faire et qui ont fait et qui continuent. Mais ils sont quand même sous-représentés. La partie expérimentation : est-ce qu'il y a un avant un après quand le citoyen a une culture des données un minimum construite ? Pour les élèves mêmes : « est-ce que ça transforme l'élève ? Est-ce qu'il est capable de faire des exposés différemment ? Est-ce qu'il est capable de faire des datavisualisations dans ses exposés ? Est-ce qu'il est capable de se dire "tiens pour cet exposé il faut que j'aille chercher les données chez *Greenpeace.* Est-ce que je peux leur demander en *open data* ?" ». Donc tout ça, vous voyez bien qu'on est à la ramasse.

** Je pensais à des projets en data literacy qui a bien fonctionné. Tu en as cité, et il y a le tien aussi j'imagine ?**

Les travaux dont je suis le plus fier(e) ce sont mes travaux avec [nom de personne]. Et je suis fier(e) d'avoir fait mon jeu sérieux sur le *RGPD*, les données personnelles parce que je l'ai fait tout(e) seul(e). Donc je ne savais pas que j'étais capable de faire un jeu sérieux. Mais tout le travail que j'ai fait avec [nom de personne] c'est vraiment super et ça serait super que tu l'interroges parce que tout le travail que l'on a structuré sur la compétence des données était basé sur un travail qui avait été fait par la [nom d'organisation] qu'on a repris. Et moi j'étais la petite main de [nom de personne] et vraiment c'est du super boulot. Cela structure comment on peut monter de façon graduelle, éclairée sur la culture des données et comment on peut avoir le sentiment de ne pas être submergé par le sujet. Au contraire, d'avoir un peu de prise de pouvoir.

En dernière question, quelles seraient tes recommandations pour une meilleure appropriation de la culture des données en France ?

Un portage politique cohérent. Par exemple à la région, j'avais rencontré l'élu qui est en charge de tout ce qui est numérique et avec [nom de personne], on a déposé plusieurs fois des projets avec [nom de personnes], des projets particulièrement pertinents où on racontait tout ce qu'on voulait faire pour pouvoir sensibiliser des enseignants, les embarquer, transformer les élèves, transformer les citoyens. On avait trouvé des partenariats avec d'autres universitaires sur [nom de ville]. On avait trouvé des partenariats avec des entreprises. On avait un super beau projet qui était bien ficelé, bien pensé, qui était tant efficace que simple. En plus, on était déjà une équipe constituée, on avait déjà travaillé, on était une douzaine d'universitaires. Moi je me mets dedans

mais je n'étais pas du tout universitaire. En tout cas, les universitaires m'avaient accueilli dans leur giron et on avait constitué un travail pour pouvoir faire intelligemment ce dossier. L'élu de la région on lui raconte tout ça, pendant une heure, il nous dit « oui c'est génial, je comprends bien, pour la population c'est important, tout ça je comprends bien, etc ». Et je lui dis « si c'est pas vous qui payez, personne paiera. Alors que c'est nécessaire, c'est utile pour les citoyens. C'est socialement pertinent, c'est économiquement pertinent. Sur la transformation, on va gagner de l'argent. Pour les acteurs publics aussi parce que les petites communes elles sont bien évidemment à la ramasse sur ces sujets-là. Et donner des moyens aux communes, donner des citoyens autour d'elles vivants qui sont en capacité d'agir tant physiquement que numériquement, on peut le lever avec ces sujets-là si on a des communes, si on allume l'étincelle, le feu peut prendre. Donc c'est stupide de ne pas le faire ».

Il avait reconnu que c'était utile mais il nous a jamais sélectionné pour notre dossier de subvention. C'est vrai qu'à un moment il faut un sacré sacerdoce alors qu'ils comprennent nos éléments de discours, ils comprennent nos enjeux, mais ils mettent l'argent ailleurs parce qu'ils ne voient pas comment ça peut leur servir. Et on a fait ça avec l'élu au numérique, et on a fait ça avec l'élu aux transports, parce que dans la thématique du transport sur l'ouverture des données, il y a mille milliards de choses à faire. Et beaucoup aussi en encapacitation du citoyen de pouvoir mieux transformer les données. Il y a un travail sur l'ouverture des données qui a été fait dans [nom de lieu], c'est un super boulot et ils ont fait pas mal de culture des données : ils ont une offre de transport mais tout est en mode puzzle. C'est-à-dire que t'as des acteurs publics, des acteurs privés, des acteurs ultra locaux, des acteurs nationaux, des acteurs régionaux. Tout ça fait du transport. Mais pour aller d'un point A [nom de pays] à un point B [nom de pays] dans [nom de lieu] c'était l'enfer. Tu étais obligé en tant que citoyen d'aller sur une quinzaine de sites pour avoir un bon niveau d'information et pouvoir organiser ton déplacement si tu n'avais pas de bagnole. Donc du coup, la région a dit « on va essayer de mettre tout ça à plat » et c'est [nom de personne], un mec génial, il a décidé de mettre tout le monde d'accord, il a un super pouvoir : je mets les gens qui ne sont pas censés s'entendre ensemble pour qu'ils s'entendent ensemble et mettent leurs données sous la même forme et qu'en plus ils l'ouvrent en *open data*. Je ne sais pas comment il l'a fait. Il est quand même dans [nom de lieu], avec des gens qui ont un tempérament, un égo, des affirmations. En plus il a du mettre ensemble des acteurs privés avec des acteurs publics. Des concurrents privés ensemble. Et il a réussi à faire une super plateforme d'échange de données de transports pour les citoyens qui marche bien et tout est en *open data*. La culture des données ça peut faire ça, donc c'est un super exemple de transformation des acteurs parce que là il a fait de la culture des données à tous les niveaux, ça donne un tout cohérent et c'est plus pratique pour le citoyen, c'est plus efficace, c'est plus cohérent, c'est plus harmonieux. J'ai l'impression de parler d'un truc religieux mais non. C'est la beauté du truc. Le bon storytelling peut amener à ces transformations-là.

On a dit à l'élu de la région qui était sur le transport : « on veut continuer à faire ce type de travail sur la question du vélo, sur la question du train, sur toutes ces questions-là ». Le citoyen aura sa place de discussion si on lui donne les bonnes visualisations des données, si on pouvait lui montrer plus de choses. L'acteur public est là pour nous faciliter la vie : tout ça on pourrait le valoriser avec les données. Même en ayant ce discours pertinent et intelligent sur la culture des données qu'on a pu avoir dans nos dossiers de demandes de subventions, cela n'a pas marché. Qu'est-ce qu'on peut faire ? On peut continuer et au bout d'un moment on va réussir à convaincre. Mais il faut pas mal d'énergie. En tout cas, je suis de la génération qui a essayé et aujourd'hui je passe le relais.

Associatif 2 (21/03/22)

Est-ce que tu emploies le terme de *data literacy* et comment le définis-tu ?

Je ne sais pas trop parce que ce n'est pas vraiment un terme que j'utilise. Je suis plus à l'aise avec les notions de médiation et de médiation aux données. *Data literacy*, je trouve que c'est un peu un terme de novlangue qu'on utilise beaucoup dans les milieux où « ça fait très bien ». Après, j'aime bien aussi le fait qu'il y ait *literacy*, c'est un mélange. « Data » m'ennuie : il y a un néologisme que je n'aime pas beaucoup. Mais après dans l'idée de développer une culture des données, de développer une médiation aux données, de développer un travail autour des enjeux autour des données avec des D majuscules à D et à Données vu la quantité, vu la diversité, vu que cela ne veut rien dire en partie aussi, parce que c'est tellement large. Est-ce que c'est un tableur *Excel* ou est-ce que c'est du *big data* d'acquisition de données sans consentement ? Voilà, il y a tellement du personnel, du pro, du privé et ainsi de suite.

Donc, j'aime bien la question de l'éducation : c'est pour moi un enjeu quel que soit le sujet. Et dans ce monde-là, la question de l'éducation aux données où il y a une vraie difficulté aujourd'hui à appréhender cette question-là de l'éducation aux données, de la médiation aux données. Alors, éducation, médiation, je ne rentre pas dans les détails mais globalement c'est la manière dont on monte tous en compétences, en compréhension, en appréhension des enjeux autour de la question des données.

Il y a un écosystème conceptuel sur les données : on parle de gouvernance de la donnée, de sciences des données… où est-ce que l'on peut situer la *data literacy*? Est-ce que c'est la même chose ? Est-ce que c'est plus situé en surplomb ?

Sur le même plan, je ne sais pas. Après il faudrait voir l'historique du terme mais je pense que les *data sciences* ont leur histoire. Les géosciences, la géomatique par exemple, en termes de gestion de données, c'est un monde en soi avec toute une histoire de la sémiologie des cartes : donc de la représentation des cartes à travers des données géographiques. C'est un monde qui a quelques siècles en plus. Donc on n'est pas sûrs quand on parle de toutes ces questions-là. On parle pas de quelque chose qui a 15 ans où d'un coup on s'est aperçu qu'il y avait de la donnée numérique. Toutes ces questions-là, les travaux sur la cartographie qui ont été fait il y a 200 ans c'est juste hyper intéressant : c'est une forme de *data literacy*. Mais quand on a l'exploration et l'avancée des armées napoléoniennes avec une représentation de la perte, des morts, c'est tout un travail autour de la donnée, de la représentation, de ce que l'on peut en faire et c'est intéressant. C'est pour ça que je fais attention à toutes ces choses qu'on a l'impression de redécouvrir. Alors il y a une forme de redécouverte sans doute, il y a certainement des spécificités, je ne suis peut-être pas suffisamment au fait des spécificités qui sont des fois glissées dans le sens de ces mots mais j'ai quand même des fois l'impression qu'on réinvente aussi des choses en se disant « wow, ce qu'on fait depuis 15 ans c'est tout neuf ». Non ! On utilise d'autres types de données, d'autres outils, d'autres résultats, mais cela existe depuis longtemps dans plein de domaines. Donc après dire, est-ce que c'est gouvernance de données, est-ce que c'est un autre terme ? Quand le mec met des petites croix pour dire les cas de choléra autour des points d'eau [note : la personne fait référence aux travaux de John Snow], je ne les connais pas bien ces exemples, on avait fait des présentations sur ces questions sur justement des questions de

datavisualisation. On remonte à quelques siècles quand même. Donc je ne sais pas si je réponds mais je ne sais pas où me placer parce que des fois j'ai l'impression qu'il y a un peu de « c'est du nouveau, c'est du tout chaud, c'est du tout beau ». Ou pas !

Mais qu'est-ce qu'on veut faire, qu'est-ce qu'on veut montrer, quels sont les enjeux ? Pour qui ? De quoi on parle ? Quel est le public que l'on va viser ? Ce sont des gamins parce qu'on va bosser dans des parcours pédagogiques de compréhension des données ? Est-ce que c'est à destination d'élus qui ne comprennent rien à la question de la donnée ? Mais qui en même temps ont envie d'avoir des synthèses hyper claires de projets et donc d'avoir des représentations très efficaces ? Dans les collectivités, il y a des demandes très fortes d'avoir des tableaux de bord, d'avoir des choses comme ça sur les données, sur les remontées, sur les visualisations. Sauf que si on ne sait pas ce qu'il y a derrière, cela n'a pas de sens. Donc toutes ces histoires-là, elles vont dépendre vraiment de publics cadres. On avait fait des choses dans le mouvement associatif : on avait formé un ensemble de coordo de différentes assos de réseaux nationaux sur la *data literacy* avec justement une réflexion sur « où se positionnent les assos par rapport à cela et qu'est-ce que ça veut dire pour eux ? Comment ils appréhendent la question de la gestion de leurs données ? ». On parlait open data, RGPD, logiciels libres, logiciels propriétaires, réappropriation, enjeux politiques au sens éthique, j'ai envie de dire autour de la donnée, de ce que l'on en fait. Du coup je ne sais pas le définir.

Tu as parlé tout à l'heure de médiation à la donnée. Et moi la question que je me pose, je vais la poser de manière très sommaire : mais si on devait définir la « nature », j'emploie des guillemets en disant ça, de la *data literacy*, est-ce qu'on va parler de culture, d'un ensemble de pratiques, de techniques, d'état d'esprit, tout ça à la fois ? Comme tu le définirais ?

Je pense qu'il y a une question de culture mais aussi d'environnement, au sens que quand aujourd'hui on parle *data literacy*, ce que j'entends ce sont des questions autour du numérique. Déjà, ça pose un cadre très spécifique sur la question, sur ce dont on parle et avec qui. Et après j'y vois effectivement quelque chose autour d'une culture et d'un travail autour de la manière dont on peut utiliser les données pour raconter des histoires, pour faire des représentations, faire des beaux graphiques, ce que veut dire *literacy*. Aussi, comment on peut utiliser ce nouveau, nouveau au sens numérique, ce nouveau support de donnée numérique. Qu'est-ce qu'on peut en faire pour l'utiliser après pour divers usages ? Ça pourrait être des projets journalistiques comme pour un travail de CE2. Je pense qu'il y a une grande partie autour de la manière dont on s'approprie ces données numériques.

Alors après certainement qu'il y a des outils derrière. Comment on utilise tous ces outils numériques en ligne qui ont une capacité de traitement de ces données, de facilitation. On peut faire quelque chose qui a vraiment de la gueule. Mais si on ne réfléchit pas à ce que l'on veut transmettre, à ce que l'on souhaite dire et sur quelles bases de représentation, cela ne ressemble à rien. Donc c'est un peu tout ça pour moi ce qui est englobé sous le terme *data literacy*. Et encore une fois, je ne me sens pas dans le mouvement de la *data literacy*. Je ne m'y sens pas en y étant, dans le sens que je vois bien que certains des projets que l'on peut mener : je fais beaucoup de formation aux données cartographiques, autour de la contribution aux données et ainsi de suite. Donc je vois pas mal de choses autour de ça et j'imagine que ça en fait un peu partie.

Sur la question des représentations des données cartographiques, j'ai plein d'entrées qui sans doute s'en rapprochent beaucoup et en même temps j'ai un peu de mal à me revendiquer de ce champ-là. Quand on avait fait la formation c'était la demande du mouvement associatif de travailler cette question-là, on est intervenus avec une collègue qui était plus à l'aise avec le fait d'affirmer qu'on fait de la *data literacy*. Après moi en pratique j'avais fait du *Datawrapper*, j'avais fait du *Umap*, j'avais fait un travail sur les représentations autour des données : est-ce que c'est de la donnée ? de quoi on parle ? structurer des données, de l'encodage UTF-8 ? plein de trucs comme ça. Mais pour moi, je pense que c'est la première fois où j'imaginais que ça rentrait dans un truc qui s'appelait *data literacy*. Pour moi c'est vraiment la médiation aux données. Après je suis prêt à l'appeler *data literacy*. Pendant 3 ans, 4 ans, j'intervenais sur [nom de formation] et j'encadrais le module « données ». Mais moi je parlais de médiation aux données. Je pense que je n'ai jamais utilisé sur tout le module sur les 3 ans, je pense que j'ai jamais utilisé le terme de *data literacy*. Par contre oui on a touché des outils, on a touché des données, on a touché la donnée perso, pro, et comment se positionner en tant que médiateur : c'était de la médiation aux données. Donc comment on se positionne en tant que médiateur sur la question des données : médiation plutôt numérique ? A quel endroit ? De quoi on parle ? Quels sont les enjeux ?.

Si on devait qualifier l'importance aujourd'hui de la médiation aux données vu que c'est la vision dans laquelle tu t'inscris : on a eu un ensemble de nécessité au fil du temps comme l'*information literacy*, tout ce qui est technicité informatique... la *data literacy* semble être la dernière en date. Et la question que je me pose : quelle est l'importance de cette nécessité de médiation de la donnée comparée aux autres nécessités aujourd'hui ?

Cela dépend de quoi on parle. Si on parle de manger, dormir, boire, on s'en fout. Là je ne sais pas ce que ça veut dire de comparer des enjeux sur ces questions-là. On est en plein dedans mais on est proche du niveau zéro en termes d'enjeux. C'est ma vie de bosser sur ces questions de communs numériques. Quand je dis ça, je ne suis pas du tout en train de dénigrer la *data literacy* mais il faut faire gaffe aussi à cela vis-à-vis des enjeux. Après, si on rentre dans ce monde-là et qu'on rentre dans un monde d'une forme de citoyenneté numérique, une sorte de réflexion sur une identité numérique, sur ces questions-là, je pense qu'aujourd'hui les enjeux autour des données sont immenses parce que si on prend le cas des GAFAM, on a parmi les dix plus grandes entreprises en bourse dont la plus grosse activité tourne autour de la gestion de données : *Google*, *Apple*, *Facebook*, *Amazon* et *Microsoft* sont entrain de monter leur *business* sur l'acquisition de données, l'utilisation de ces données, la construction des données.

Les enjeux aujourd'hui de compréhension, de ce que l'on fait de nos données, de savoir ce que sont les données, à quel moment je produis des données, qu'est-ce qui est fait de ces données : tous ces enjeux-là sont des enjeux qui sont juste immenses au regard de se dire que si tu regardes ton profil, si tu as un compte *Google* et que tu regardes ton profil et que tu vois où les gens sont passés, leurs habitudes de vie et tout ça. 80% des gens ont une adresse *Google* dans nos mondes-là, ce qui est juste une aberration absolue. Les gens ne le savent pas ; ils ont un compte_ Google_ qui a été créé pour eux sur leur téléphone portable pour *Android*. Donc aujourd'hui sur ces enjeux autour des données, je pense que l'enjeu est immense.

Alors est-ce que c'est la *data literacy* ou la médiation aux données, et la prise en jeu de ces questions numériques comme étant des questions centrales de compréhension de ce qui est fait de nos données ? Cela, j'en suis sûr.

Mais cela rejoint pour moi le monde des logiciels libres qui rejoint la question des communs numériques qui rejoint tout ça. A un moment donné, qu'est-ce qu'on a comme modèle, quasi de société, autour des identités numériques ? Et je ne suis pas dans du transhumanisme. Je ne suis pas du tout dans ces endroits-là. Loin de là. Mais voilà, ce sont vraiment ces enjeux-là de données perso, de données en commun, de données privées qui je pense sont des enjeux aujourd'hui, dans la question de ce double numérique qui se construit, qui sont juste immenses.

Je pense que c'est devenu une nécessité toute cette médiation à la donnée : est-ce qu'on peut dater depuis quand c'est vraiment devenu prégnant ?

C'est un peu difficile de dater. Je pense qu'il y a des marqueurs révélateurs selon les milieux. Par exemple, 2018 dans le milieu de la cartographie, de la géomatique : *Google Maps* annonce qu'il ne va plus être gratuit. Tout d'un coup, des dizaines de centaines de milliers de sites se retrouvent sans cartes, sans données, et se rendent compte qu'ils doivent payer s'ils veulent avoir des fonds de cartes *Google*. Ils se rendent compte que c'est aussi plein de données qui ont été appropriées par *Google* à travers leur fonds de cartes. Par exemple, dans le milieu de la géomatique, il y a une date où ça a été révélateur. Pour avoir bossé dans le milieu du tourisme par exemple : tous les offices de tourisme de France se sont posés la question tout d'un coup de leurs rapports à *Google* et de leurs rapports à *Google Maps* et de leur dépendance, et ainsi de suite. Donc, dans ce milieu de la géomatique, de la cartographie par exemple, il y a des dates comme ça.

Après le « dans tes villes » qui disparaît de *Google* on est dans les années 2012 ou 2013 : à un moment donné il y a une bascule de ce qu'est *Google*. Je pense qu'il y a beaucoup de choses autour de grandes enseignes comme ça qui sont marquantes. Les *Cambridge Analytica* de *Facebook*, je pense que ce sont des marqueurs à un moment dans la société de choses où « ils se foutent de nous et ils font ce qu'ils veulent pour se faire de l'argent ». Des marqueurs de choses identifiées, qui ont été publiques. Tu prends les questions de Snowden sur toutes les données qui sont accumulées par exemple. Tout cela, ce sont des marqueurs pour moi à un moment donné de prise de conscience : il se passe des trucs avec nos données sur lesquels on est en train de perdre la main. Assange et toutes les révélations sur toutes les données qui ont été publiées qui révélaient tout le bordel sur la surveillance généralisée du monde. On est sur une prise de conscience de ces enjeux. Alors qu'après on ne comprenne rien, qu'on réfléchisse à de la médiation aux données, de la *data literacy*, tout cela se renvoie un peu.

Du coup, puisqu'on est sur le volet « données » : toi, comment est-ce que tu définis ce qu'est une donnée ?

Eh bien, j'ai toujours beaucoup de mal parce que je pense qu'il y a plein de représentations selon les cadres. L'image que j'utilise le plus souvent quand je donne des cours, c'est celle-ci : « dans un tableur, il y a dans la petite cellule, il faut faire en sorte que cette petite cellule puisse être utilisée de manière unique, automatique et qu'elle n'agrège pas d'autres données ». Que ça soit un côté, un peu unique. Je donne l'exemple de l'adresse : 18 rue Victor Hugo. Là, on a « 18 », on a « rue » et on a « Victor Hugo ». Si on a tout ça dans une cellule, là on a un bug de données parce que ce n'est pas quelque chose que l'on peut utiliser comme ça. Si on a « 18 », si on a « rue » et si on a « Victor Hugo », dans 3 cellules du tableur, là on commence à avoir de la donnée. Parce

que chacune de ces données est unique et on commence effectivement à avoir une donnée. Il y a un caractère un peu unique, un peu singulier. Après si on parle de données personnelles, de données...ça va dépendre de quoi on parle : données volontaires, pas volontaires, sensibles, etc. Il y a plein de choses pour ramener à ça mais on va dire que c'est un ensemble dont on connaît à peu l'origine. Ce n'est pas mal ça aussi. Si je fais « 18 rue Victor Hugo », j'ai une adresse. Pour moi, c'est un peu ça. Mais je trouve que c'est compliqué, si ce n'est en donnant des exemples. Les gens ils arrivent de très loin souvent. Et moi le premier, je suis chimiste au départ. Ce n'est pas mon monde. Et en même temps dans une démarche scientifique, quand tu fais un tableur, quand tu fais des manips, dans ton tableur tu vas mettre ta quantité, ton type de produit : tu fais cela pour avoir des automatismes.

Au niveau des enjeux des données. Tu les as pas mal exprimés tout à l'heure. Comment la médiation aux données permet de répondre aux enjeux actuels ?

Pour moi, la question de la médiation aux données telle que je l'imagine, c'est que les gens soient au fait, soient au courant, dans le sens qu'ils sachent que par exemple quand ils allument leur téléphone portable le matin et qu'ils se connectent sur un site, il y a de la donnée qui est créée. A quoi elle sert, à quoi elle est utilisée. Je n'ai pas envie de rentrer dans une diabolisation de quoi que ce soit mais c'est juste d'avoir des éléments un peu factuels sur ce qu'on est en train de faire.

Par exemple, les cours de médiation aux données avec les étudiants de licence pro : à un moment donné si tu fais de la médiation aux données et que tu demandes aux gens d'utiliser des adresses *Google*, que tu saches qu'en tant que médiateur à l'endroit où tu es en faisant ça, quelle est ta posture à cet endroit-là si tu fais des animations en utilisant *Google Doc* ? Après si tu acceptes d'être à cet endroit-là, que tu le sais pour diverses raisons. Il y a cette idée aussi de conscience, de ce qui est fait aujourd'hui des données et pas d'une espèce de désinvolture : « non on ne peut rien maîtriser ». Bah si ! Déjà il faut savoir de quoi on parle. Il faut savoir quelles sont les possibilités alternatives. C'est aussi reprendre la main sur ces questions-là. Et oui, la médiation aux données, c'est ça aussi. Pour moi c'est de reprendre, surtout pour des médiateurs, c'est le biais que j'ai, de dire « ah mais vous ne pouvez pas ne pas savoir. Vous pouvez décider d'utiliser *Google* mais vous ne pouvez pas ne pas savoir ».

Est-ce qu'il y a des types de données qui sont plus au cœur des enjeux ?

Pour moi il y a un énorme enjeu sur les données persos. Sur la privatisation qui est faite des données personnelles en fait. A travers les données de santé par exemple. A travers la création d'un espace de santé par le gouvernement et ainsi de suite. Il y a des gros enjeux sur les données personnelles de santé, leur appropriation. On sait aujourd'hui qu'il y a des gens qui se penchent, en termes d'assureurs, sur la question de comment réfléchir à « si t'es riche et en bonne santé, tu paieras moins que si t'es pauvre et malade ». Et cela est basé sur la santé. Donc c'est basé sur les données de santé, sur l'acquisition de ces données. Je pense qu'il y a des énormes enjeux effectivement autour des données de santé.

Mais je crois qu'il y a surtout un énorme enjeu de comprendre ce qui est en train de se faire autour de toutes ces données et leur appropriation aujourd'hui par des boites. Que toutes les données soient sur des serveurs

Amazon, que les serveurs des *data hubs* santé, *health data hub* et compagnie, soient hébergés par des serveurs *Microsoft*, ça pose plein de questions sur : à qui sont ces données ? Qui y a accès ? Et je suis pas anti-américain, ce n'est pas de cet ordre-là aussi, je préfère le préciser, ce n'est pas à cet endroit-là mais est-il normal que le *Patriot Act* permette au gouvernement américain de consulter toutes les données hébergées par des entreprises américaines partout dans le monde ? Ben non. On a quand même des exemples malheureusement très récent avec l'Ukraine sur le fait de ne plus pouvoir payer parce que *Google* et *Apple* ont décidé de fermer la question du paiement par téléphone des données bancaires. Les gens ne peuvent plus payer parce qu'ils payaient tous avec leur application sur téléphone. C'est vraiment aussi tous ces enjeux, ce n'est pas un repliement nationaliste à la con mais c'est à un moment donné savoir les enjeux d'avoir tout délégué dans des entités qui ne sont elles absolument pas démocratiques. Ces entreprises elles n'ont aucune raison d'avoir une démocratie. Ce n'est pas Google en particulier mais leur truc c'est de faire de l'argent et des décisions sont prises par trois gaziers qui décident ce qu'ils veulent.

As-tu vu une évolution dans tes pratiques professionnelles avec l'arrivée de la donnée ?

Moi complètement mais parce que mon activité professionnelle aujourd'hui mon activité c'est former des acteurs du tourisme, des acteurs de collectivité, tout ça, contribuer à *Open Street Map* et à faire des cartes. Donc aujourd'hui, mon activité, c'est celle-là. Je ne pourrais pas avoir l'activité que j'ai aujourd'hui s'il n'y avait pas toute cette simplicité d'accès à Internet, tout cette capacité de gestion de gros volumes de données et ainsi de suite. Après j'ai commencé les premiers projets avec des gamins où on menait des projets pédagogiques dans les classes aux alentours de 2011-2012 sur la question d'*Open Street Map* et de contribuer à *Open Street Map*, donc il y a un peu plus de dix ans. C'est sûr qu'aujourd'hui il y a une prise en compte de ces questions qui est juste immense si je ne regarde que le biais d'*Open Street Map* aujourd'hui, si on considère que *Apple*, *Microsoft*, *Facebook*, *Amazon*, pour reprendre cet exemple-là, sont tous les quatre des utilisateurs des données *Open Street Map* pour faire leur fond de carte, ils utilisent les données *Open Street Map*, même ça c'est complètement transparent dans la plupart des milieux. Même *Google* commence à utiliser sur ses *maps*, sur ses *Google Maps*, des données venant d'*Open Street Map* qui sont des données *Open Street Map* sur la question des transports en commun. Il y a eu l'exemple en Pologne sur les *Google Maps* polonais. Et pourtant *Open Street Map* est un commun en licence ouverte. Et pourtant toutes ces boites s'en servent. Les *Apple plan* c'est *Tom Tom* et *Open street map* comme fond de cartes. *Amazon* et compagnie, dès que tu es hors des Etats-Unis et hors d'Europe, c'est *Open Street Map* la base de données de cartes. Donc oui c'est immense en fait la prise en compte avec des choses qui sont aussi à des endroits un peu particuliers entre les communs et les cinq plus gros. Si on prend les *GAFAM*, ils sont aussi des utilisateurs de ce qui est aujourd'hui pour nous le deuxième ou troisième plus gros commun numérique.

Si on prend *Wikipédia* d'un côté, *Wikidata* en termes de volume justement qui est quand même la base de données on va dire juste immense, à partir de là déjà *Open Street Map* doit être le troisième en termes de volume. Et tous ces volumes-là, on est sur des croissances exponentielles. Enfin, pas exponentielles parce que j'en sais rien, mais sur de fortes croissances. Sur *Open Street Map* par exemple, sur dix ans, en nombre de contributeurs, en nombre de volumes modifiés tous les jours, en nombre d'ajouts, on est sur des croissances immenses : ce n'est pas de la perception, c'est la réalité.

Aujourd'hui en l'espace de trois ans, on est passé de 200 contributeurs quotidiens en France à plus de 400/450 quotidiens en France de gens qui contribuent à *Open Street Map*. Sur des outils comme ça, il y a des stats très précises d'évolution de volumes de données traitées. La *data literacy* ne connait même pas l'existence de ces choses-là. Encore une fois c'est normal, mais aujourd'hui quand le gouvernement lance le site commun.gouv.fr, il met trois communs en avant : il met *Wikipédia* et toute sa panoplie, y compris *Wikidata*, donc on parle bien des données. Le deuxième commun qu'il met c'est *Open Food Facts*, la plus grande base de données des éléments alimentaires. Et le troisième c'est *Open Street Map*, la plus grosse base de données cartographiques libre. C'est ça qui est posé dans commun.gouv.fr, ce sont trois trucs dont l'activité principale sont des communs et des données. Alors aujourd'hui dans la *data literacy*, il y a un peu de *Wikipédia* à des endroits un peu quand même, le reste il n'y a pas grand-chose. Après oui beaucoup avec *Open Data* mais l'*open data* ce n'est pas des communs. *Open data*, c'est la publication de collectivités, de leurs données. C'est très bien et j'en suis un fervent défenseur. Mais ce n'est pas du commun numérique : c'est de la publication structurée de données. On peut en faire plein de choses. Mais au niveau des communs, je pense qu'on est en train de rater tout un truc sur la question des données mais ça va venir.

Par rapport à tout ça et toujours par rapport à ta pratique, est-ce que tu constates une évolution, en bien ou en mal, du cadre législatif. La loi Lemaire, le RGPD : est-ce que tu vois une différence dans tes pratiques de données depuis leur arrivée ?

Il y a quand même un certain nombre de choses mais après je vois les différences avec un biais de fait :c'est que je sais qui est derrière madada.fr parce que j'ai bossé avec eux, parce que je m'en sers. Les gens avec qui je bosse, avec qui je travaille, ou avec qui on est engagé sur des structures autour d'*Open Street Map* par exemple mais pas seulement sont des utilisateurs de *madada*. Oui, il y a des éléments qui bougent. Alors est-ce que c'est vite, pas assez vite, ça je n'en sais rien. C'est très difficile de se rendre compte mais parce que là on parle aussi de la question du changement et de comment est-ce qu'après avoir été embarqué dans tout ce discours ambiant french-tech compagnie, qui moi m'épuise, comment on a aussi une réflexion sur les questions un peu plus ouvertes, sur le service public, à quoi ça sert et ainsi de suite. Il y a des petits bouts dans des endroits de la loi qui amènent des choses sur ça.

Tu parlais de la loi pour une République Numérique avec effectivement le fait de publication par défaut ou par principe de bases de données ouvertes. La question de la loi *CADA*. La question de l'intégration des bases de données comme étant des documents administratifs, effectivement dans la loi *CADA*. Tous ces éléments-là sont quand même des choses aujourd'hui qui, quand on va voir une collectivité en lui disant « eh oh, vos données-là, tutututut », leur permettent de savoir de quoi il est question maintenant. Cela ne veut pas dire qu'ils le font. Il n'y a pas de sanctions aujourd'hui pour une collectivité qui ne libère pas ses données. Mais il n'empêche qu'aujourd'hui, il y a de moins en moins de collectivités qui disent « de quoi vous parlez ? ». Cela ne veut pas dire qu'ils sont encore dans la démarche. Mais il commence à y avoir des ressources. *Madada* c'est juste hyper bien.

Donc tu ne trouves pas que c'est trop restrictif d'un côté ou pas assez encadré de l'autre. Est-ce que tu ressens ce genre de choses ou pas du tout ?

Si on avait mis des sanctions, je pense qu'on aurait braqué beaucoup de choses. L'absence de sanctions aujourd'hui n'aide pas à ce que les gens s'alignent sur tout ça mais pour moi c'est forcément lent. J'ai des exemples : on a bossé sur la politique *open data* du [nom de département]. J'accompagnais quelqu'un que tu pourrais peut-être rencontrer [nom de personne] qui est plutôt assez affûté sur les questions d'*open data*. Qui a bossé longtemps au département. Qui n'est plus là aujourd'hui. Mais qui a beaucoup bossé sur les questions d'*open data*, de schémas. Je me suis perdu dans mon raisonnement.

Concrètement c'était surtout savoir si tu voyais une évolution positive ou pas en France avec l'instauration de de ce cadre législatif. Ou est-ce qu'il est trop restrictif.

Je ne sais pas quels sont les moyens qui sont vraiment donnés. Et l'absence de contraintes de fait n'aide pas. Il y a des politiques volontaristes de transparence de l'action publique mais comme ça je ne sais pas. C'est trop large pour avoir une idée sur ces questions-là. Après il y a tous les projets autour de la base adresse nationale, il y a plein d'endroits où il y a tellement de choses qui bougent que c'est difficile aussi de dire « il faudrait aller plus loin ». Chacun est dans son quotidien. Le changement prend du temps. Et je n'ai pas assez de recul pour savoir si c'est trop restrictif ou pas assez.

On va passer à la dernière partie de l'entretien. De ton point de vue, si on devait mettre en place une médiation aux données, déjà est-ce qu'elle va concerner tout le monde et comment on va la mettre en place dans un organisme ?

J'ai quand même un passif éducatif donc de fait à un moment donné, il faut rentrer dans la question de l'éducation et de la formation sur ce que sont les données, savoir de quoi on parle. Est-ce que c'est normal qu'un gamin de six ans ait un téléphone portable avec la géolocalisation activée ? Il peut y avoir aussi à ces endroits-là sans doute des choses à faire. Cela ne veut pas dire qu'il faut arrêter de lire, écrire, compter, bien évidemment, ce n'est pas à cet endroit-là. Mais je pense qu'il y a aussi des choses à faire sur ces questions d'identité numérique aujourd'hui en termes éducatifs et en termes d'enjeux de ce que ça peut signifier au niveau de l'Éducation Nationale. Il y a cet aspect-là mais là on parle pour dans trente ans. Il va falloir former les profs.

Pour donner un exemple : est arrivé en 2019 la compétence du programme Sciences numériques et techniques SNT. Le programme *SNT*, il y avait une compétence à l'intérieur qui est la compétence, donc obligatoire pour toutes les secondes de France, « contribuer de manière collaborative à Open Street Map ». Cela arrive dans le programme, c'est une obligation. Tout enfant de seconde doit faire, doit répondre à cette demande. Sauf que pour ça il faut qu'en face il y ait des enseignants qui sachent de quoi on parle. Alors je ne sais pas comment c'est arrivé dans le programme. Nous on ne le sait pas. On est assez intrigués sur la question. Comment le comité des programmes de l'éducation nationale a intégré, validé cette compétence. Mais par contre ce que l'on voit c'est qu'il n'y a aucune formation qui accompagne et donc là on a limite du vandalisme à des moments par des élèves de seconde à qui on donne les clés d'un commun numérique autour de la donnée cartographique, non encadrés, et ne comprenant rien.

Là par exemple il y a beaucoup de choses à faire mais ça prend du temps. Oui, il y a de la formation aux enseignants. Oui aujourd'hui il y a des enjeux immenses sur ça. Ce que je décris en disant qu'il faut le faire

au niveau de l'Éducation Nationale, pour moi de fait c'est plutôt une entrée universaliste et il faut le faire auprès de tout le monde. Autour d'un champ d'éducation populaire pour que tout le monde s'approprie ces questions-là. Mais il y a tellement de boulot sur tous les réseaux de médiation, tout ce qui est médiation numérique aujourd'hui. Il y a les grosses structures comme la *Mednum*, tous ces gros paquets que sont ces structures de la médiation numérique, de l'insertion numérique de la solidarité. Mais oui, il faut se poser les questions et auprès de tout le monde. Après, un changement de société ça prend une dizaine d'années.

Dans le cadre de la médiation à la donnée, qu'est-ce qu'il faudrait que les personnes qui ont suivi cet apprentissage, puissent développer en termes de compétences ? Des compétences techniques, de l'esprit critique, une culture générale...

Clairement, une question d'esprit critique sur la question des données et donc de positionnement par rapport à ça. C'est quoi une donnée ? De quoi on parle ? Qu'est-ce que j'en fais ? Et de le faire en partie en conscience. Y compris d'être conscient de se dire « je m'en fous ». Je peux décider mais en sachant d'avoir un compte *Google* et de l'avoir sur mon téléphone et d'être suivi tout le temps. Mais je le sais. Pour moi, il y a quelque chose à un moment donné de savoir. D'avoir conscience de. Pour moi, il y a un enjeu sans doute là-dessus. Mais il y a sans doute une nécessité de réappropriation par les services publics, par les collectivités. Il y a quelque temps c'était « pourquoi aujourd'hui on n'a pas dès que l'on nait une adresse mail donnée par l'État ? ». Cela ne veut pas dire que tu auras que cette adresse mail mais pourquoi aujourd'hui on n'a pas une adresse mail ? Pourquoi les collectivités et les Etats ne se disent pas « pour toutes les questions administratives, vous avez une adresse mail, elle vous est attribuée, elle est attribuée à votre carte d'identité ». S'il y a des modifications, on sait qu'avec cette adresse mail, on peut toujours vous joindre. Parce qu'aujourd'hui les adresses mail changent, les gens changent, des choses se perdent. Il y a un nomadisme qui s'instaure. D'une certaine manière, les gens commencent à avoir une fixité avec leur adresse *gmail*. Il pourrait y avoir des choses à cet endroit-là. Il y a une initiative dans les Landes qui se fait dans la mise à disposition d'adresses mail facilement par le département je crois. Il pourrait y avoir des choses à ces endroits-là parce qu'aujourd'hui l'adresse mail c'est quasiment l'adresse numérique. Pour créer un compte il faut une adresse mail. Tu perds ton numéro, ton mot de passe ? tu fais « mot de passe oublié » et ça te revient sur ton adresse mail. Et aujourd'hui c'est complètement approprié : 70% des adresses mail ça doit être *gmail*. Sur les 30% qui restent, 20% ça doit être *Outlook*, 5% *Yahoo*, etc.

Dans les enjeux aujourd'hui sur la médiation, c'est de se dire, ce qui se fait énormément en ce moment pour accéder aux services publics « pour accéder à votre compte en ligne, créez-vous une adresse *gmail* et après vous pourrez avoir accès, on vous accompagne sur les services publics ». Toute la dynamique d'accompagnement numérique sur les usages, les gens qui sont éloignés. Autour de la médiation aujourd'hui il y a aussi des enjeux à ce que les acteurs de la médiation comprennent. Et je râle régulièrement sur *Teams*. Et sur les gens qui font de l'accompagnement numérique à des gens qui sont éloignés du numérique en leur disant « on va se créer une adresse *gmail* ».

On va rebondir là-dessus : toi qui a vu plusieurs publics différents, qui a enseigné et qui continue à enseigner, comment vois-tu la situation en termes de niveaux de compétences et de culture au niveau de la médiation aux données. Est-ce que tu trouves que l'on est encore très bas ou vois-tu une évolution ?

Il y a une part d'évolution parce qu'aujourd'hui je rencontre peu de monde mais il y aurait des petits tests à faire avec des étudiants. J'en vois de moins en moins. Mais aujourd'hui avec les étudiants que je peux croiser, il y en a quand même peu qui ne sont pas conscients du fait que *Google*, *Apple*, ainsi de suite, ont une activité d'absorption de données qui pose question. Sans même aller jusqu'à « c'est pas bien », ce n'est même pas la question. Il y a au moins cette prise de conscience qui il y a quelques années était vraiment inconnue. Il y a quand même des enjeux autour de ces questions-là qui arrivent j'ai l'impression un peu au fur et à mesure. Il y a quelques affaires qui ont été marquantes pour ça et qui ont été aussi un peu dans les médias, c'était celle des militaires qui utilisaient *Strava* dans les bases de l'armée secrète, et dont on connaissait la localisation parce qu'ils avaient publié leurs traces sur Internet. Il y a des choses concrètes qui sont des éléments que les gens ont souvent entendus. Ce que je disais aussi, les *Cambridge Analytica*, des trucs tellement gros que j'ai l'impression qu'il y a une prise de conscience autour de ça. Après, encore une fois, on est encore bas parce que je pense qu'il y a une confrontation de modèles de société : on se bat avec des entreprises dont certaines ont le poids de plusieurs États. Et des capacités de communication et de maîtrise qui sont justes hallucinantes.

A l'inverse, quelles sont les difficultés, les freins, de la médiation aux données, de cette acculturation aux données ? Tu as cité les entreprises. Est-ce qu'il y a autre chose qui s'ajoute ?

Aujourd'hui, sur certains enjeux autour des données, ce qu'il peut effectivement manquer, ce sont des alternatives qui vont permettre de continuer à travailler en étant un peu plus *clean* sur les données. Je prenais l'exemple des mails mais beaucoup de services par exemple si on veut utiliser des services en ligne de médiation aux données, de *data literacy*, de dataviz, et ainsi de suite, il faut une adresse mail. Si on veut proposer à des gens de se créer des adresses mail, qu'est-ce qu'on leur propose aujourd'hui ? C'est pas une fixette sur le mail sauf que là on arrive à un endroit où on commence à être bloqué. Après il y a de super outils qui sont des outils propriétaires par exemple et qui fonctionnent super bien mais qui s'accaparent les données parce qu'en même temps en médiation des données il est parfois difficile de trouver l'alternative qui va permettre d'avancer et de faire des choses simplement aussi intuitives avec un design aussi intéressant et tout ça. Pour moi il y a un vrai investissement qui doit venir : j'aurais tendance à dire que cela devrait être de l'ordre du public, du service public, que d'avoir aussi une réflexion sur du soutien à des outils qui sont des outils partagés, mis en commun, qui ne sont pas appropriés par des entreprises au niveau des données. Cela ne veut pas dire que ce ne sont pas des entreprises qui montent leur activité autour de ça.

Nous notre activité tourne autour des données *open street map*. On est pas propriétaire des données *open street map*, elles n'appartiennent à personne. Donc c'est aussi tout ça qui sans doute a bougé. Mais il y a beaucoup de choses qui existent. Ce que fait *Etalab*. Tout la mission autour de Bastien guerry et du logiciel libre. Toutes ces questions-là ce sont des initiatives au niveau de l'Etat qui sont pas anodines. De fait, des initiatives de connexion unique des données assurées par l'Etat, *France Connect*, on est là sur des réflexions qui sont quand même intéressantes pour se dire qu'on ne doit pas avoir une adresse *gmail* pour avoir une connexion à impôts.gouv. Il y a une logique de réflexion sur la donnée et sur la conservation de la donnée. Il y a des choses qui bougent à ces endroits-là mais c'est complexe.

Est-ce qu'il y a un projet en médiation de données en France et que tu estimes avoir bien fonctionné et pourquoi ça a bien fonctionné justement.

Je trouve que les formations qu'on a pu faire par exemple au niveau du *CLEMI* sur la médiation aux données justement à travers soit de la dataviz, soit après mon entrée qui est beaucoup la carto, tout ce travail sur la question de la cartographie et des données cartographiques, des choses comme ça, cela matchait bien avec les gens que l'on avait. Après ça se sont des discussions qui ont eu lieu après avec des gens du *CLEMI* national et ainsi de suite pour bosser sur ces questions-là. Mais encore une fois, aujourd'hui toutes ces questions-là elles sont surtout portées par des gens dont ce n'est pas le métier et donc de manière bénévole, ou alors à la marge sur un certain nombre de petits projets.

En termes de gros projets, c'est clairement tout ce qui tourne autour de *Open Street Map* qui est vraiment sur la question des données. Alors est-ce que c'est une réussite en termes de datavisualisation, de *data literacy*, je ne sais pas, si ce n'est que à travers les échanges à toutes les échelles. On discute avec de gros interlocuteurs on va dire nationaux, à travers tout ce travail autour des données, autour de la question des données ouvertes, des communs, des communs numériques et des enjeux des données. Toutes ces questions-là avancent un peu. Est-ce que c'est une réussite je ne sais pas, mais ce sont des choses qui bougent dans ces milieux-là. On reçoit souvent des coups de fil de ministères qui se disent « ça serait bien qu'on comprenne quand même ». Effectivement, cela serait bien. Et encore une fois, c'est la question des données. Soit du fait des habitudes avec *Google Maps*, soit d'avoir des bases de données qui servent et qu'ils s'aperçoivent qu'ils ont plein de données qui pourraient être mises en commun avec plein de gens. Et peut-être qu'il faut réfléchir à ça aussi. A changer la manière de travailler les données. Donc là on revient sur des questions de médiation, de comment on fait aussi pour changer de posture, changer de manière de travailler, ce qui est important aussi. Et qui prend du temps.

En guise de conclusion, as-tu des recommandations pour améliorer cette médiation à la donnée ?

Je dirais : logiciel libre, données ouvertes, et licence de documentation en CC BY-SA. Le CC BY-SA j'insiste dans tout ce qui est ressources créatives libres, on trouve beaucoup de NC, de Non commercial, qui est une énorme connerie. Parce que ça empêche plein de gens de se servir de ça. Donc tout ce qui est association d'éducation populaire, entreprises, tout ça, ne peuvent pas se servir de ces ressources-là parce que souvent ils font des prestations donc ça pose problème. Si on doit bosser sur les questions aujourd'hui d'alignement sur la question de la médiation aux données pour des médiateurs et qu'on a envie que ça soit diffusé au plus grand nombre, il faut du logiciel libre, c'est de la donnée ouverte, c'est du CC BY-SA.

Associatif 6 (07/12/22)

Employez-vous le terme de *data literacy* et si non, lequel employez-vous ?

Oui, on utilise le terme de *data literacy*. On utilise aussi le terme de « culture générale de la donnée » de manière assez indistincte ou indifférente entre les deux, sans vraiment faire de distinction. Mais oui, on utilise le terme de *data literacy*.

Quelle définition lui en donnez-vous ?

Vaste sujet : je ne crois pas qu'on est très originaux de ce point de vue-là. En tout cas, on a pas de théorisation qui nous serait propre. En tout cas, pour ce qui me concerne personnellement. J'y suis venu essentiellement à travers les travaux de la *FING* et notamment la *Data Literacy Conference*, les multiples éditions de cette conférence, ont tenu un rôle important dans la création de [nom d'organisation] parce qu'on a lancé officiellement lors de la première *Data literacy conference*. Et ensuite, pour la deuxième édition, on avait fait notre fête d'anniversaire dans le cas de la *data literacy* et voilà c'est toujours quelque chose qui a été assez important pour nous. [nom de personne] a aussi joué un rôle je pense important dans la création de [nom d'organisation]. Donc ça c'est pour la petite histoire pour dire que c'est aussi par ce biais-là que j'ai découvert cette notion et l'univers des acteurs qui s'inscrivent dans ce champ-là.

La définition que j'en donnerais n'est pas très originale : c'est reprendre le parallèle avec la *literacy* tout court, ce qui est hyper bateau de dire « on a besoin dans notre société de savoir lire, écrire, compter » et on a aussi besoin d'avoir un certain degré de familiarité et de compréhension sur ce que sont les données et les usages qu'on peut en faire. Et ensuite avec les déclinaisons opérationnelles qu'on essaie d'en faire dans les univers dans lesquels on évolue.

On considère qu'on peut pas le dissocier soit des notions d'émancipation, soit d'*empowerment* selon le terme que l'on veut utiliser. Et que ça vient au service d'un projet d'émancipation individuelle et collective. C'est aussi une question intimement liée à des questions de rapports de force parce que finalement, si on analyse, on fait très régulièrement cela dans des cadres de formation qu'on appelle la culture générale de la donnée. Ça va depuis des contextes académiques à [nom d'établissement] par exemple ou dans d'autres établissements supérieurs à des contextes de formation continue dans les cadres professionnels publics / privés. Et ce qu'on explique toujours dans ces formations à la culture générale de la donnée, c'est que d'une certaine manière qu'on assiste bien à une forme de « révolution de la donnée », en tout cas de mise en donnée du monde, en s'appuyant sur les travaux de Kitchin. Et que derrière ces révolutions, il y a des questions de *shift*, de modification des rapports de force. Et que donc si on veut pouvoir rééquilibrer un tout petit peu ces rapports de force et redonner un petit peu d'agentivité, pour être un peu jargonnant, aux individus, ils ont besoin d'avoir un minimum de cette *data literacy*.

Après bien sûr cela se traduit de manière opérationnelle dans des spectres très variés : ça va de projets *self data* qu'on peut accompagner à des formations plutôt pour des cadres exécutifs ou cadres dirigeants et puis avec toute la palette intermédiaire. Par exemple, en ce moment, on est en train de faire un projet assez intéressant avec [nom d'organisation] : on fait un cycle de trois webinaires de formation continue pour ces agents-là sur justement les enjeux data et intelligence artificielle avec tous les guillemets qu'on peut mettre autour de ces termes-là, notamment en lien avec le monde du travail. C'est [nom de personne] qui fait l'introduction du webinaire et ensuite on regarde un petit peu toutes les transformations qui vont se produire dans le monde du travail sur un secteur d'activité qui est celui de ces agents-là et pour moi ça relève de cette *data literacy*.

Justement, par rapport à ce que vous venez de dire, si on devait donner certaines caractéristiques de la *data literacy* au sens d'une « nature », comment est-ce qu'on pourrait la définir ? Est-ce que ça serait plutôt une culture générale ? Un ensemble de pratiques et de techniques ? Un état d'esprit ? Comment voyez-vous tout ça ?

C'est une bonne question. Je ne sais pas s'il faut trancher entre les deux. C'est vrai que nous, on fait un peu des deux. On a des formations plutôt théoriques on va dire, plutôt culture générale. Le vernis un peu de l'honnête homme du XXIe siècle sur le sujet. Et puis aussi on a pas mal de formations, d'ateliers, beaucoup plus pratico-pratiques : on s'est beaucoup inspiré du travail de la *School of data* par exemple, d'*open knowledge* sur le sujet. Non seulement on s'en est inspiré mais en plus historiquement [nom de personne]y a pas contribué. Et pareil, la création de [nom d'organisation] entretient un lien étroit avec [nom d'organisation] donc on reprend aussi ces aspects-là.

Après pour nous l'objectif il n'est pas de former des techniciens de la donnée ou d'avoir une approche uniquement technique, ça peut être effectivement un outil d'*empowerment* et d'émancipation. Mais en tout cas, on ne réduit pas la *data literacy* à des aspects pratico-pratiques je dirais, ou de savoir-faire pratiques et techniques. Et la dimension « appréhender de manière critique » les enjeux de la « révolution des données » nous semble quand même importante. Mais c'est vrai que je pense qu'au sein même de notre organisation, on a un peu l'attention entre les deux pôles. On ne l'a peut-être pas forcément complètement explicité et théorisé. Après il y a aussi des appétences personnelles qui sont assez variables sur le sujet. Mais *de facto* on fait un peu les deux.

On a tout un ensemble de concepts liés aux données : on a la gouvernance des données, on a la data science. Est-ce que vous arrivez à faire, à défaut d'un distinguo, mais en tout cas de voir les interactions qui existent entre ces concepts-là et celui de la *data literacy* ?

On essaye oui. C'est une bonne question parce que je pense qu'on ne l'a pas explicité et cela mériterait d'être, sans doute, explicité. Mais ce qu'on dit souvent, c'est qu'en fait pour nous, en tout cas dans une organisation, la culture de la donnée, la *data literacy*, c'est un peu la condition nécessaire, pas forcément suffisante mais nécessaire pour le succès d'autres démarches data, de démarches de la gouvernance de la donnée, de conduites de projets de valorisation de la donnée, etc. Donc pour nous, c'est vraiment le socle un peu transversal sans lequel il est difficile de bâtir d'autres démarches.

La gouvernance de la donnée ensuite pour moi elle renvoie à une notion organisationnelle de process, d'organigrammes, de mise en place de la fonction data, etc.

Et le troisième terme que vous aviez évoqué. *Data science*, je crois que tout le monde revient un peu de ce terme-là peut être aujourd'hui mais pour le coup on est dans un champ de pratiques, à la fois un champ de savoirs et surtout de pratiques professionnelles à l'intersection entre trois notions. Là encore, j'invente rien, je reprends des choses assez bateau mais à l'intersection entre trois notions que sont : les statistiques, l'informatique et le métier. Et c'est ce que j'essaye d'expliquer souvent à nos clients qu'en fait un vrai *data scientist*, c'est celui qui est capable de mobiliser ces trois piliers et c'est pour ça que c'est difficile de trouver des vrais *data scientists*. Après, ce qu'on appelle *data scientist* en pratique sont souvent des gens qui ont un ou deux de ces trois piliers. Rarement les trois. Après à la limite ce n'est pas très grave parce qu'il y a plein d'autres métiers que celui de *data scientist*. Il y a plein d'autres approches que celles de la *data science*. Je pense que l'héroïsation du *data scientist* a duré un moment dont je crois qu'on revient quand même beaucoup aujourd'hui. Après il y a des effets de mode : aujourd'hui c'est le *data scientist*, puis c'est le *data engineer*, et autres. Mon expérience, parce

qu'on travaille avec des organisations publiques essentiellement, c'est qu'en fait les problèmes des organisations sont souvent beaucoup plus prosaïques que ceux qui sont en jeu dans la fiche de poste d'un *data scientist* par exemple.

Donc on essaye de les inciter souvent à faire les choses dans l'ordre, à se préoccuper d'abord de culture de la donnée et éventuellement de gouvernance de la donnée. Et ensuite de réfléchir jusqu'où il faut aller dans la valorisation de la donnée et c'est vrai que la tentation est souvent de faire l'inverse. On voit beaucoup de collectivités qui écrivent des fiches de poste de *data scientists* puis ensuite vaguement elles arrivent à recruter un *data scientist*, en tout cas quelqu'un qui est présenté comme tel. Et puis ensuite c'est compliqué parce qu'en fait personne n'a la culture de la donnée, pas de gouvernance, donc il n'y a pas de données en fait ou elles sont très difficiles à mobiliser. Il n'y a pas de demandes et c'est quand même le gros problème : on se retrouve avec des gens qui ont un titre de *data scientist* mais ils n'ont pas de répondant dans les métiers. Donc concrètement ils s'ennuient. Donc pour moi dans le cas de la *data science*, on est quand même dans quelque chose de beaucoup plus avancé pour lequel le recours à cette notion-là suppose déjà une certaine maturité.

Pour vous, quelle est l'importance d'être acculturé aux données ? Est-ce que c'est plus ou moins important que les compétences informatiques, informationnelles, ce qui est demandé aujourd'hui à un employé classique ?

C'est peut être un prisme, peut être un regard biaisé sur le sujet de ma part mais j'aurais tendance à dire que c'est classé au même rang que les autres éléments que vous avez cités parce qu'en fait je ne vois pas aujourd'hui un secteur d'activité qui n'est pas impacté directement ou indirectement et de différentes manières par la « révolution de la donnée ». Il y a des fois où c'est évident, des fois où ça l'est beaucoup moins, où parfois c'est assez ancien et parfois c'est peut être plutôt devant nous mais je ne vois pas aujourd'hui dans quel aspect de la vie professionnelle ou même la vie de citoyens où un certain degré de *data literacy* ne serait pas utile.

Alors encore une fois il ne s'agit pas de dire que tout le monde va devenir *data scientist*. Mais d'avoir un minimum de clé de compréhension. Après la question qui est peut être plus difficile à résoudre et je ne suis pas sûr que collectivement on sera très matures sur le sujet, c'est comment ça s'articule avec une littératie numérique plus large. Est-ce que c'est simplement un sous-ensemble ? Est-ce qu'une forme de littératie numérique emporte avec elle une forme de littératie de la donnée ? Est-ce que ce sont des notion distinctes ? On voit par exemple sur les sujets « données » aujourd'hui que l'approche de l'État, en tout cas de certains secteurs de l'État aujourd'hui c'est de dire « on met ça dans PIX » parce qu'on a ce truc-là qui marche bien donc on va mettre la donnée dedans.

Je pense que collectivement et en tout cas moins individuellement on n'est pas forcément complètement au clair là-dessus sur l'articulation du numérique, de la donnée, jusqu'à quel point l'un est un sous-ensemble de l'autre. Est-ce que ce sont des choses différentes ? Le risque étant, et on le voit aussi à certains égards, que si on dit : « la donnée c'est juste un sous-ensemble du numérique », on considère que sont légitimes à acculturer sur ce sujet-là les grands acteurs du numérique et on se retrouve avec *Pole Emploi* ou même l'Education nationale qui fait appel à *Google* pour nous expliquer le numérique. Je ne suis pas un spécialiste du numérique en général mais sur les sujets data en particulier c'est un vrai problème parce qu'ils ont un modèle économique vis-à-vis

de l'extractivisme de données qui est très spécifique et qui est assez orthogonal et contradictoire avec des visées émancipatrices quant à la culture de la donnée, donc là je pense qu'on a quand même un petit sujet. Mais bon, peut être que des gens plus spécialistes que moi de la culture numérique en général vous diraient « mais en fait, ce problème là il est vrai de manière générale avec *Google* et ce n'est pas que sur les données et c'est un problème que *Google* fasse de l'acculturation au numérique ».

Toujours en lien avec ça et je pense que vous êtes le mieux placé pour me le dire : depuis quand c'est devenu, en tout cas dans les discours, une nécessité cette acculturation à la donnée. Est-ce que temporellement vous arrivez à situer depuis quand c'est au centre des discours ?

Là encore j'ai des œillères sur le sujet parce que je suis rentré dans ce champ-là à peu près au moment où on a créé notre organisation. Cela ne veut pas dire que je ne m'y intéressais pas auparavant mais je m'y intéressais de manière beaucoup plus lointaine. Donc moi j'ai l'impression qu'en tout cas c'est sorti des catacombes je dirais à peu près en 2016, c'était aussi le moment de la première *Data literacy conference*, qui est le premier évènement de cette ampleur-là en France, avec des travaux de la *FING* sur le sujet, et la campagne *Infolab* de la *FING*, et ensuite de sa diffusion par capillarité. Il se trouve qu'en 2016, on a la loi pour une République Numérique, on a le RGPD : il y a quand même eu une conjonction d'éléments à ce moment-là qui me font penser que c'est un tournant mais encore une fois, c'est peut-être un biais de perception de ma part. Le pari qu'on fait c'est qu'on est peut être à l'orée d'un nouveau tournant avec la vague de textes européens en cours d'adoption : *Data act, Data governance act, AI act*, etc. C'est peut-être un biais de perception court-termiste mais il faut aussi reconnaître qu'un certain nombre de développements technologiques notamment autour de l'IA qui rendent la chose extrêmement perceptible par un public très large en fait. Les gros modèles textuels comme *Chat GPT* ces derniers jours ou *DALL-E* depuis quelques mois, etc., ça commence à être très tangible, très perceptible par un public globalement assez large là où sur certains sujets c'était des discussions un peu de spécialistes il y a encore quelques années.

On va passer à l'autre partie de l'entretien qui est plutôt axé sur les données en général et la première question que je veux vous poser et qui est elle aussi assez générale : qu'est-ce qu'une donnée pour vous ?

Alors là j'ai un peu tout appris de [nom de personne] parce que lui est vraiment un expert académique du sujet mais en fait peut-être un peu juste d'égo-histoire. Donc avec [nom de personne] on se connaît depuis très longtemps il a fait [nom de formation] à [nom de ville] des années après moi, on se connaît depuis cette époque-là. Après on s'est croisés régulièrement, on a vécu à [nom de ville] tous les deux à peu près à la même époque : lui était actif dans [nom d'organisation], à ce moment-là je poursuivais une thèse en sociologie que je n'ai jamais rattrapée mais je l'ai poursuivie un certain temps.

Et je faisais beaucoup de quanti, et je travaillais beaucoup sur des données publiques : des données de l'*INSEE*, des données de l'*IGN*, des données du ministère de l'Intérieur, etc., tout ce genre de choses. Et j'ai commencé ce travail-là en master 2 en 2005. Puis on créé notre organisation, j'ai vraiment vu l'évolution des modalités de diffusion des données publiques puisqu'en 2005 on envoyait un chèque à l'*INSEE* qui nous renvoyait par courrier un dvd avec les données du recensement et puis pareil, si on voulait les données de l'*IGN*, il fallait

payer, c'était assez cher, et ils nous renvoyaient des dvd, des téléchargements FTP, et pareil avec l'*INSEE*, c'était l'enfer.

Et puis progressivement, l'*open data* est passé par là, la diffusion en ligne est passée par là, et aujourd'hui c'est extrêmement facile d'avoir accès à ces données, elles sont toutes en *open data*, même celles de l'*IGN*, etc. Tout ça pour dire que du coup, moi en tant que doctorant, je militais aussi à mon petit niveau autour de l'*open data* ne serait-ce que pour des raisons assez pratiques aussi, pour des raisons philosophiques mais ne serait-ce que pour des raisons assez pratiques. Et donc j'ai aussi croisé [nom de personne] comme ça, il m'a invité à des évènements avec [nom d'organisation], autour de ces sujets-là. Je me rappelle avec débattu déjà sur les sujets d'ouverture des données électorales avec Henri Verdier et Laure Lucchesi en 2013, 2014 quelque chose comme ça. Et donc sur la définition de la donnée, et tout ça pour en dire que par ailleurs j'avais déjà lu déjà, je pense en début de ma thèse, Desrosières : j'étais totalement convaincu et enthousiasmé par cette approche là. Et puis je venais d'un *background* académique qui était plutôt compatible avec cette littérature là.

Et donc, on est tous les deux partis à [nom de ville] vers 2015, tous les deux en suivant nos compagnes respectives et en disant on verra bien ce qu'on va faire. Je faisais surtout du *freelance* à l'époque donc je travaillais à [nom de ville] c'était à peu près pareil, et on s'est retrouvé à faire des cours ensemble à [nom d'organisation] c'est comme ça que nous est venu l'idée de notre organisation en discutant autour de la thèse de [nom de personne] qu'il était en train de finir et en disant ben qu'est-ce qu'on pourrait faire ? Tiens il y a peut être un truc autour de ça et tout. Et donc tout ça pour dire que moi ma définition de la donnée, et plus généralement, mon approche théorique de la culture générale de la donnée, je l'ai entièrement reprise de [nom de personne] [coupure de son] mais moi ça reste Desrosières, ça reste la sociologie de la quantification et donc pour moi si je devais définir une donnée de manière peu originale, je dirais que c'est un enregistrement du réel qui obéit finalement à la dialectique « convenir et mesurer ». C'est vrai que c'est à chaque fois cette double dimension et double étape de l'aspect conventionnel, qu'est-ce qu'on saisit à travers la donnée et puis mesurer l'instrument de mesure et qu'est-ce qu'il fait à la donnée, donc une approche assez constructiviste de la donnée. Et c'est vraiment le genre de chose qu'on transmet aussi dans nos formations, de faire toucher du doigt cet aspect à la fois conventionnel et en même temps de construction de la réalité, d'éléments de construction de la réalité qu'est la donnée.

Selon vous, quels sont les grands enjeux autour des données ?

Vaste sujet, ça dépend du contexte et du public avec qui on en discute. J'aurais tendance à considérer que les deux principaux enjeux sont d'une part la question au sens très large de gouvernance. Je ne parle pas de gouvernance au sein d'une organisation mais vraiment de cadre à la fois réglementaire, éthique, organisationnel, etc., sur la donnée. Je pense qu'on a un problème de métaphore pour comprendre la donnée, c'est-à-dire qu'on parle beaucoup de propriétés de la donnée et moi je pense que la métaphore de la propriété n'est probablement pas une bonne métaphore pour réfléchir à la relation que des personnes physiques ou morales entretiennent avec des données.

Et d'ailleurs c'est pas vraiment la métaphore que retient le droit de fait, puisque e cadre juridique des données personnelles n'est pas un cadre qui repose sur la propriété mais plutôt sur la protection de la vie privée et sur la protection de libertés fondamentales et de droit, et le cadre que retient le droit de la propriété intellectuelle

des affaires n'est pas vraiment non plus un cadre de propriété puisqu'en fait il n'y a pas de propriétés *sui generis* sur les données. Il peut y en avoir à certaines conditions mais donc pour moi, une première grande famille d'enjeux tient des règles qu'on pose. Si vous me permettez la petite blague « qui a le droit de faire quoi avec les données », « à qui appartiennent les données » si tant est que ce terme soit approprié, etc. Et on voit bien qu'il y a des approches très différentes quand on est en Europe, aux États-Unis, en Asie, de ces questions-là. Il y a le cadre juridique et il y a la réalité : et c'est une réalité qui est quand même aujourd'hui, je ne vais pas reprendre toute la littérature sur l'extractivisme de données, le modèle capitaliste californien.

Le deuxième enjeu est très en lien avec le premier : c'est la question de la valeur au sens économique et au sens plus large du terme, c'est-à-dire qui s'approprie la valeur des données. Déjà, qui la créé et qui se l'approprie ? Et comment elle est partagée ? Et je suis un militant du modèle coopératif. Ce qui définit le modèle coopératif c'est une double conception, à la fois du modèle de gouvernance d'une entreprise, gouvernance égalitaire entre les salariés, en tout cas dans une coopérative de production, et du modèle de partage de la valeur qui se fait au profit des salariés et là aussi de manière égalitaire. Et je pense qu'on a besoin dans le monde de la donnée, de pouvoir inventer des modèles qui permettent aussi de garantir ces deux aspects et on voit bien en entreprise que l'aspect de gouvernance et l'aspect de partage de la valeur sont assez intimement liés et il en va de même sur le sujet des données.

Justement, à l'intérieur de ces enjeux que vous venez de citer, est-ce que pour vous il y a des types de données qui sont au cœur de ces enjeux-là ou s'il y en a pas de particuliers...

Est-ce qu'il y a des types de données ? C'est difficile à dire parce que même si les données en tant que *capta* pour reprendre ce que dit Kitchin ou d'autres, même si ces données répondent toujours à une intentionnalité dans leur capture justement pour autant cela ne préjuge pas des usages. Et donc des données qui peuvent paraître anodines dans un premier temps peuvent se révéler extrêmement stratégiques, avoir une très grande valeur pour telle et telle raison, etc. Donc je ne suis pas sûr que ce soit quelque chose qu'on puisse totalement préjuger. En tout cas dans le modèle européen. On a fait des données à caractère personnel un enjeu très fort, très spécifique parce qu'on a considéré que c'est un enjeu de protection des libertés mais on voit bien que le champ de ce qu'est ou non une donnée à caractère personnel est variable et n'est pas prédéfini à l'avance : des données qui, à un moment donné, sont définies comme non réidentifiantes peuvent le devenir dans un autre contexte, que la sensibilité collective aussi sur ce sujet-là évolue. Et évolue assez rapidement je dirais. Donc il y a sans doute un sujet un peu particulier là-dessus mais dans le même temps je pense qu'il y a des données qui ne sont pas nécessairement des données à caractère personnel en tant que telles mais qui sont extrêmement stratégiques : je crois que c'est aussi les enseignements des grands modèles d'intelligence artificielle qui sont sortis ces derniers temps : *Chat Gpt*, *Dall-E* ou autre. C'est qu'en fait, ces modèles mettent en valeur des corpus extrêmement vastes, extrêmement hétérogènes avec des données qui, prisent individuellement, ont donc peu ou pas de valeur, mais qui, assemblées dans ces dispositifs, prennent une valeur incroyable. Donc je me garderai de préjuger de savoir si telle ou telle catégorie de données est plus ou moins stratégique, concernée par tel ou tel enjeu.

Vous l'avez cité tout à l'heure mais j'aimerais qu'on aille plus dans le détail dans les réglementations autour des données. Depuis un moment, on a pas mal de législation en vigueur : la loi Lemaire, le

RGPD, le Data act et le Data governance act qui vont arriver. A quel point ces législations ont un impact sur les pratiques professionnelles, en tout cas peut être les vôtres, et si vous avez un avis sur ce cadre législatif.

Alors à quel point ça a eu un impact, ça je pense que c'est bien réel, en tout cas pour le *RGPD*. Je pense que l'impact du *RGPD* est indéniable partout, en tout cas dans de très nombreuses organisations, à la fois en termes d'évolutions parfois assez profondes des pratiques. Et puis en termes organisationnels il n'y a quasiment plus d'acteurs publics ou de grandes organisations privées qui n'ont pas un délégué à la protection des données, qui n'ont pas mise en place une organisation *ad-hoc*, etc., qui, de manière plus ou moins avancée n'ont pas mis en place un registre des traitements. Donc sur le *RGPD*, c'est assez indéniable.

Sur la loi pour une République Numérique, je pense que le bilan aujourd'hui est beaucoup plus mitigé. Disons que cela a été un outil de prise de conscience, d'acculturation, de pédagogie. Je ne pense pas que d'un point de vue juridique on puisse considérer, en tout cas sur le sujet de l'ouverture des données, que cela a été un outil contraignant et opérationnel. En tout cas, ça n'a pas été perçu comme ça par les acteurs. En 2022 aujourd'hui, on a encore l'essentiel des organisations publiques pour qui oui l'*open data* c'est vaguement inscrit dans la loi mais bon à la fin des fins on fait bien comme on veut. Et je ne parle même pas d'autres enjeux inscrits dans la loi pour une République numérique comme l'ouverture des algorithmes par exemple, la transparence algorithmique où là on est encore plus dans les limbes aujourd'hui de ces sujets-là. Et je crois que c'était plutôt un outil d'incitation et de pédagogie et une sorte d'affichage politique qu'une réglementation contraignante et opérationnelle parce que d'ailleurs il n'y a pas de mode d'emploi dans la loi pour une République numérique. Donc voilà sur l'impact. Après, mon avis sur ces réglementations, je ne sais pas si j'ai un avis très tranché et si ça a beaucoup d'intérêt. Je pense que ce qu'on peut dire c'est que la loi pour une République Numérique illustre le fait, et je pense que c'est largement partagé, qu'il ne suffit pas d'écrire les choses dans la loi pour qu'elles se traduisent en réalité et donc soit on est dans des approches qui sont purement incitatives où finalement on compte sur l'alignement de certains intérêts ou certaines incitations pour que les choses se passent.

Soit on va jusqu'au bout du truc et on met en place une police. Et on voit d'ailleurs que ce qui est intéressant, ce qui s'est passé après la loi Lemaire dans certains secteurs où on a des réglementations sectorielles sur la gouvernance des données qui là pour le coup sont beaucoup plus contraignantes et mettent en place à la fois des modes d'emploi et les gendarmes. Je pense à la loi *LOM*, la loi sur les mobilités, sur les données de transport et de mobilité aujourd'hui on a à la fois un mode d'emploi en disant « il y a telle donnée qui doit être ouverte par telle personne partagée avec telle autre, en suivant tel standard, etc ». Et puis un gendarme, l'autorité a changé de nom : c'est l'ancienne autorité de régulation ferroviaire qui a été étendue à l'ensemble des transports et qui a le pouvoir de sanction à l'égard des acteurs qui ne respecteraient pas les règles. Donc du coup, il y a une tension aujourd'hui je trouve, entre des approches très sectorielles et très précises, et des approches beaucoup plus transversales comme la loi Lemaire qui restent quand même un peu du droit nul comme dirait l'autre sans forcément que l'une ou l'autre des approches soit forcément complètement convaincante mais voilà je pense que dans les années qui viennent et notamment dans la transposition en France du cadre réglementaire européen qui est en train d'être posé dans le cadre d'une éventuelle Loi République II ou je sais pas comment elle s'appellera, sera sans doute quelque chose à bien réfléchir à cette articulation entre approche sectorielle et approche transversale.

On voit que même organisationnellement l'État est pas encore complètement au clair là-dessus. Un exemple intéressant c'est celui de transport.data.gouv, le point d'accès national sur les transports et la mobilité plus largement maintenant qui au départ était portée, très intégrée à *Etalab*. Et puis finalement, un peu au forceps, il y a eu une restitution de cet outil à la direction générale des transports, *DGITM* je crois, enfin la direction générale en charge des transports. Ce n'est pas évident en fait et on voit que c'est un truc un peu non identifié administrativement et politiquemente t qu'il n'y a pas vraiment de doctrine ou de mode d'emploi complètement stabilisé sur le sujet.

Comment pourrait-on mettre en place une acculturation à la donnée, une *data literacy* dans un organisme de manière générale et surtout est-ce que ça doit concerner tout le monde ?

Je vais peut-être répondre en commençant par la deuxième partie de la question : oui, j'ai la conviction que ça doit concerner tout le monde. J'ai cette conviction et elle est partagée je pense par mon organisation. On dit régulièrement que cela ne doit pas être un sujet réservé à des codirs, des conex, aux élus, etc. : c'est vraiment un sujet qui concerne tout le monde et ça se comprend aisément parce qu'en fait les données sont souvent saisies, manipulées par des agents de terrain, des agents de guichet, etc. Et si à eux on ne leur explique pas l'importance du sujet : pourquoi ce qu'ils font est important, comment ensuite ça a des répercussions ailleurs, etc, on a aucune chance d'avoir des données de qualité, exploitables à la fin. Donc même d'un point de vue très utilitariste il y a un intérêt à ce que ça concerne tout le monde.

Sur le comment en revanche, là je suis très humble. Je pense pas que nous on ait complètement craqué le problème aujourd'hui, résolu le problème. Nous on fait des choses assez classiques en fait : des sessions de formation, de sensibilisation, des ateliers, etc., qui sont des piqûres à un moment donné. Après le vrai sujet c'est comment on maintient et on entretient ça dans la durée, sur des sujets qui évoluent rapidement que ce soit techniquement ou d'un point de vue juridique. Donc j'avoue qu'on n'a pas construit de doctrine très forte sur le sujet. On voit bien aussi qu'autant pour une organisation c'est facile de former soit ses cadres supérieurs, soit tel service sur le sujet, autant former tout le monde c'est compliqué. Il y a peu d'organisations qui sont vraiment allées jusque-là. Il y en a quelques-unes, j'ai vu des départements par exemple où il y a vraiment une approche qui est offerte à tout le monde. Bon est-ce que ça va être des choses comme *PIX* qui vont prendre le relais ? est-ce que ça va être intégré à la formation initiale, finalement au collège, au lycée ? on sait que c'est déjà le cas avec les enseignements *NSI*, numérique et sciences de l'information, quelque chose comme ça. Et les sujets data sont déjà abordés, en tout cas par certains aspects.

Vous avez parlé de sensibilisation : qu'est-ce qu'une personne qui suit une formation d'acculturation à la donnée doit développer ? Est-ce que ce sont des compétences ? De certaines connaissances sur le sujet ?

Comme je vous le disais plus tôt en fait, nous on a deux types de formats : des formats plutôt culture générale de la donnée où là ce sont vraiment des clés de compréhension pour décrypter les enjeux essentiellement. Donc on explique ce que ce que sont les données, ce qu'est le big data, ce qu'est le *machine learning*, tout ce genre de choses. Donc là c'est vraiment de la culture générale : comprendre un certain nombre de notions, comprendre les enjeux, être capable de décrypter.

Et puis on a aussi des choses pratico-pratiques, alors un peu sur toutes les étapes du cycle de vie de la donnée depuis la collecte des données et parfois on fait des ateliers très pratico-pratiques. Par exemple, au début, avec [nom d'organisation] on avait fait des expéditions de données pour la CAF de [nom de ville] avec des bénéficiaires de la CAF sur des sujets comme la parentalitér. Où on faisait produire de la donnée par les participants qui étaient vraiment des bénéficiaires lambda de la CAF et qui avaient une culture de la donnée très faible au départ. On leur faisait par exemple mettre sur une carte les lieux en lien avec la parentalité qui pouvaient avoir dans leurs quartiers par exemple. Où on leur faisait produire des dataviz avec des gommettes, ce genre de choses. Ça peut aller de ça à des ateliers *airtable* : on m'a appelé dans une collectivité il y a pas longtemps pour utiliser *airtable* pour *scrapper* un bout de projet data de manière assez autonome, à de la dataviz. On fait pas mal d'ateliers dataviz depuis les niveaux très débutants jusqu'à des niveaux plutôt avancés. Cela peut être un peu tout ce genre de choses. On en a fait pas mal mais on n'en fait plus maintenant, on a fait pas mal de formation à R par exemple. Historiquement, on a des collègues qui continuent à faire des formations Python par exemple. Plutôt dans des cadres universitaires. Ça peut être aussi ce genre de choses. On peut montrer ce qui peut être fait avec *Open food facts* ou avec *OSM*. Souvent on fait des petites formations, comme utiliser *Overpass turbo* avec *OSM*, ce genre de chose un peu orienté outil mais sur des choses rapidement appropriables sans avoir à faire des formations de trois mois pour devenir expert en R.

Justement, avec votre vision, comment estimez-vous le niveau de culture générale des personnes que vous rencontrez ou même des personnes que vous employez. Est-ce qu'on est sur quelque chose qui augmente en compétences, en connaissances au fur et à mesure du temps ou est-ce qu'on est arrivé à un plateau et qu'on en bouge plus depuis ?

Alors ça je n'en sais rien du tout. J'ai forcément plein de biais de perception là-dedans. C'est une bonne question. Ce que je dirais, c'est qu'il y a quand même un truc : j'ai l'impression que la donnée était un peu l'apanage de gens qui étaient passés par des formations de statistiques. Soit à titre principal, soit à titre accessoire dans leur formation académique et qu'aujourd'hui on a de plus en plus de gens qui peuvent avoir des compétences techniques relativement avancées en données sans avoir de bases statistiques. Après est-ce que c'est bien ou non, c'est peut être un point de vue de vieux con de ma part de considérer que c'est compliqué de pas avoir de base de stats sur le sujet. Mais c'est vrai qu'aujourd'hui, je pense qu'il y a d'autres points d'entrée dans la data que les points d'entrée traditionnels on va dire. Ce qui à la fois est bien car cela ouvre des opportunités à plein de gens qui ne les auraient pas eu auparavant. De ce point de vue-là, ça monte plutôt.

Mais d'un autre point de vue, on se retrouve aussi avec des situations où les gens font un peu n'importe quoi mais en même temps je ne suis pas non plus favorable à ce qu'il y ait une sorte de *gate keeping* trop forte de la part du monde des statisticiens parce qu'ils n'ont pas le monopole du sujet de la data. Et même historiquement c'est un sujet qui ne les intéressait pas beaucoup en réalité. Donc je pense que globalement il y a plutôt un développement de cultures avec un s de la donnée qui peuvent être relativement hétérogènes. Et sans forcément une seule culture commune qui pendant longtemps était quand même la culture statistique.

A contrario, quels sont pour vous les freins, ou en tout cas les difficultés générales que vous avez recensées qui bloquent cette acculturation à la donnée ?

Je pense que le principal c'est sans doute l'idée que c'est un sujet technique, un sujet de spécialistes, qu'il faut être bon en maths pour comprendre ces sujets-là. Alors, dans nos formations de culture générale à la donnée, il n'y a pas de maths, nulle part : on n'a pas besoin de bases en maths pour comprendre. Donc je pense qu'il y a un peu ce côté très clivant en France du rapport aux maths et aux sciences et techniques, et aux ingénieurs. Donc il y a soit ceux qui considèrent qu'ils sont bons en maths, qu'ils sont ingénieurs et donc qu'ils sont légitimes sur le sujet. Soit tous les autres qui ont été un peu traumatisés par ces sujets-là à l'école et qui au contraire se disent « ouh lala, ce n'est pas pour moi, j'y comprends rien, je ne peux pas comprendre ». C'est peut-être très franco-français comme point de vue.

J'aimerais vous interroger par rapport à deux termes. Le premier c'est l'esprit critique, être critique vis-à-vis des données. Est-ce que vous ça vous parle et si oui qu'est-ce que vous mettez derrière ?

Oui, ça nous parle totalement. Ça fait complètement partie des choses que l'on essaie de transmettre dans le cadre de nos formations par exemple. Et qu'est-ce que je mets derrière ? Il n'y a rien qui m'énerve plus que la pub qu'a faite le ministère de la santé en disant « on peut discuter de tout sauf des chiffres » à propos du Covid. Je trouve ça vraiment contre-productif comme message, complètement faux évidemment. Donc oui, esprit critique, il est absolument nécessaire et ça consiste je pense précisément à s'interroger sur ce que je disais tout à l'heure : les deux dimensions « convenir et mesurer ». Qu'est-ce qu'on peut dire de la convention qui a été adoptée pour saisir telle et telle donnée et qu'est-ce qu'on peut dire d'un instrument de mesure qui la saisie ? Ce qui ne signifie pas pour moi que toutes les opinions se valent et que tout peut être mis au même plan. Je crois que ça c'est fort dans la sociologie de Desrosières où il montre bien qu'une fois qu'on a convenu et mesuré, cela construit la réalité. Et donc on ne peut pas s'en affranchir si facilement et considérer que cela n'a aucune valeur. Mais ça reste la base d'un débat démocratique et il construit quand même une conception de la quantification qui s'inscrit dans une société démocratique et donc dans un débat possible et dans une critique possible. Voilà pour moi, c'est ça que je mets derrière.

Même question par rapport à l'éthique des données. Est-ce que ça vous parle et qu'est-ce que vous mettez derrière ?

Je vous avoue que je suis moins à l'aise sur ce sujet. C'est vrai qu'il s'agit de termes qu'on voit beaucoup aujourd'hui : éthique des données, IA éthique, etc. Y compris chez nos clients qui écrivent parfois des chartes éthiques de la donnée. Je trouve que c'est un sujet plus compliqué qu'il n'y paraît. Dont le risque est de tomber dans des approches un peu simplistes, notamment très intentionnalistes considérant que si l'intention est bonne, le résultat sera bon et donc finalement que se sont les intentions qu'il faudra cadrer. Donc je ne suis pas certain que les choses soient si simples que ça. Et qu'une approche intentionnaliste, je suis pas du tout philosophe ou spécialiste de l'éthique mais le peu que j'en comprends me laisse me permet de savoir qu'il y a quand même d'autres approches de l'éthique que des approches intentionnalistes et que globalement c'est compliqué. Donc je ne sais pas si on ne risque pas par des approches intentionnalistes « on a fait une charte, on a fait un cadre réglementaire, on a fait ceci, cela » de finalement s'empêcher de voir toute la complexité des sujets qu'il peut y avoir malgré toute la bonne intention. Je reste très humble et prudent sur ce sujet-là pour l'instant.

En guise de conclusion, est-ce que vous avez en tête un projet en *data literacy* qui a bien fonctionné, si oui lequel et si possible pourquoi ?

C'est difficile à dire des projets qui ont bien fonctionné. On a mis en place à [nom de ville] depuis cinq ans maintenant, tout un cursus de culture générale à la donnée dans le cadre de [nom de formation] en partant de l'idée que les étudiants restent quatre ou cinq ans dans l'institut donc on a des formats pédagogiques tout au long du cursus depuis la première année jusqu'à la cinquième année avec des choses plutôt théoriques, plutôt descendantes et des choses très pratiques, notamment le *challenge data* qu'on fait en quatrième année où les étudiants accompagnent de manière très opérationnelle les collectivités territoriales dans l'ouverture de leurs données ; l'ouverture et la valorisation de leurs données. Et ça c'est un truc donc je suis assez fier parce qu'on touche 120 étudiants par an. On a une vraie progression pédagogique sur le sujet sur des étudiants qui vont ensuite occuper des positions de cadres intermédiaires ou supérieurs dans toute une série de secteurs assez large. Donc voilà, je me dis qu'à notre modeste échelle ça doit avoir un petit impact et probablement un peu plus dans la durée que les organisations ponctuelles qu'on peut faire dans telle ou telle organisation. Honnêtement c'est difficile de savoir un an, deux ans, trois ans après qu'est-ce qu'il en reste : on en sait rien en fait.

En guise d'ouverture, à défaut de recommandation, en tout cas des remarques pour améliorer l'acculturation aux données en France.

Vaste sujet. Des recommandations non. La question que je me pose, qu'on se pose collectivement dans notre organisation, c'est celle du financement de tout ça. Nous on adore créer des contenus libres. D'ailleurs tous nos supports de formation sont tous en ligne sous licence CC-BY SA. Mais c'est vrai qu'en fait on manque de temps essentiellement pour pouvoir en faire davantage. On aimerait faire ça et on est en train de réfléchir à comment financer ça. Mais voilà, je trouve qu'il y aurait peut-être une approche collective, soit des pouvoirs publics, soit même des employeurs privés mais en collectivisant un peu les choses, en mettant un peu en commun les choses pour créer davantage de ressources, de formation, de sensibilisation, d'acculturation sur le sujet parce que finalement les ressources publiques sur le sujet il y en a relativement peu quoi. Là je plaide un peu pour ma paroisse puisque nous c'est quelque chose qu'on aime bien faire. Mais ça bouge comme je le disais : il se passe des trucs du côté de *PIX*, il se passe des trucs du côté de la *Mednum* aussi. Par exemple, quand on voit les moyens qui ont été mis, et à juste titre, sur l'inclusion numérique, on se dit qu'il y a peut être une fraction de ces moyens qui pourrait être mise sur la question de littératie des données qui pourrait avoir pas mal d'impact.

Annexe 12 : les entretiens avec le public « entrepreneur »

Entrepreneur 1 (19/10/21)

Comment définissez-vous la *data literacy* et employez-vous ce terme ?

Alors la *data literacy* pour moi c'est une question de culture ou de *mindset*. Et de fait culture c'est la tradition française et *mindset* on ne sort pas du sujet de la *data literacy* en tant que tel, et notamment du fait qu'il n'a pas d'équivalent en tant que tel dans la langue française. Donc *data literacy*, *mindset*, j'emploie ces mots anglo-saxons parce qu'en fait ils contribuent, surtout parmi les gens qui ont des pré-formatage audit/conseil anglo-saxon, à leur fixer les idées dans la langue de Shakespeare, en anglais. En français, je le traduis pour essayer de rendre ce discours accessible à d'autres gens. Je parle de la « culture de la donnée ». Au-delà de ça en fait la notion de culture commence à s'envisager dans l'esprit des gens qu'à partir du moment où il y a des enjeux qui sont liés à la maîtrise de cette culture. Donc je navigue un peu entre tout ça : pour moi *data literacy* quand je le traduis c'est « culture de la donnée ». Mais ce que je mets avant tout en évidence, ce sont tous les enjeux qui vont se rattacher à l'acquisition de cette culture.

Il existe tout un écosystème de données : on parle de *data science*, de gouvernance des données. Où peut-on situer la *data literacy* dans cet écosystème ?

Pour moi la *data literacy* c'est le socle commun. En l'occurrence, c'est pour ça que je parlais de culture : c'est quelque chose qu'on partage parce qu'on a établi un maillage et tout un relationnel, et on s'interroge finalement assez rarement sur sa propre culture. En tout cas, on ne le fait généralement que quand on rencontre des gens d'autres cultures. En fait, si on fait le parallèle avec la donnée, cela veut dire que quand on est dans une entreprise, tout le monde est supposé avoir les mêmes bases, notamment par rapport à la donnée. Je pense que ça illustre bien le propos parce qu'il suffit d'avoir fait un tour dans une entreprise pour regarder un peu ces sujets-là pour se rendre compte que c'est très loin d'être le cas.

Si on devait décrire la « nature » de la *data literacy*, ses caractéristiques : parlerait-on d'une culture, d'un état d'esprit, d'un ensemble de pratiques et de techniques ? Tout ça à la fois ?

J'aime bien, comment dire, le parallèle avec cette notion de culture parce que finalement il y a beaucoup de choses qui ne sont pas réfléchies dans cette histoire de *data literacy*. En d'autres termes, il y a toute une approche théorique de ce qu'il faudrait savoir et après il y a toute une approche pratique de ce qu'on fait vraiment. Et du reste, je pense qu'on ne peut pas considérer cette discipline sans adresser les deux approches. Je vous donne un exemple. J'ai eu l'opportunité de tremper dans le projet [Nom de projet]. notre premier client ça a été [Nom du client], donc aérospatial-défense, théoriquement c'est un compte qui doit avoir adressé toutes les problématiques liées à la protection des données, pratiquement. Eh bien, ils ont eu l'intelligence de considérer qu'il y avait un décalage entre ce qu'ils souhaitaient voir, leurs employés et leurs parties prenantes faire, et ce qu'ils faisaient vraiment dans la réalité. Pour la *data literacy* ce principe-là, il continue d'être vrai sur d'autres sujets comme la valorisation des données, comme la gouvernance des données et comme tous les enjeux qui peuvent se rattacher au sens large, à la donnée.

Si on devait comparer aux autres nécessitées apparues au fil du temps comme la technicité informatique par exemple, où se situe la *data literacy* ? Est-ce central ou juste une compétence parmi tant d'autres ?

En fait, je pense que tout le monde a conscience de l'importance de cette approche. Sauf que je parlais tout à l'heure des cabinets de conseil anglo-saxons où il est plus facile de se concentrer sur des sujets qui ont des ROI faciles à dresser et à construire que sur ce type de sujets qui sont relatifs à la culture, au *mindset* et à de l'évolution de comportements et qui touchent à des dimensions beaucoup plus humaines. Donc pour répondre, je pense que les problèmes sont bien connus. D'ailleurs, dans la plupart des cas, quand on discute avec les gens, ils finissent par l'avouer parce que je ne vais pas rentrer dans les détails mais il y a toute une approche *corporate* qui fait qu'il y a des sujets dont on ne parle pas facilement. Et c'est exactement ce que je veux mettre en évidence. C'est-à-dire que la dimension humaine d'un point de vue *corporate*, on ne l'adresse pas facilement parce qu'on sait que c'est un sujet casse-gueule pour le dire avec des mots vulgaires.

Arrivez-vous à situer temporellement à quel moment l'acculturation aux données est devenue une nécessité ?

Alors en fait, je ne pense pas que ça soit encore perçu comme une réalité sur l'ensemble du spectre. En revanche, il y a certaines approches aujourd'hui sur lesquelles les entreprises se disent « WOW ! Vraiment, là je ne peux plus faire l'impasse ». Et notamment quand j'évoquais tout à l'heure la protection des données, là maintenant je pense qu'il y a eu suffisamment d'exemples d'entreprises qui avaient pris des gros tampons pour qu'elles se rendent compte justement de l'intérêt de combler ce vide entre la théorie de « vous voudrez bien faire tout cela, respecter tous les principes de la charte informatique, etc. », en clair de mettre des croix dans les cases et la théorie de « au quotidien, je reçois des mails qui sont vachement bien fait avec des liens pourris. Et comme je ne suis pas du tout habitué à ce genre de choses et qu'on ne m'a rien dit, qu'on ne m'a pas *challengé* dessus, je clique dessus, je vais mettre les coordonnées du RIB bancaire de mon entreprise » et puis du coup derrière il se passe des trucs définitifs parce qu'en l'occurrence quand on s'embarque là-dedans et qu'on est une petite boite, ça coûte très cher. C'est donc cette dimension un peu « pognon » qui arrive à draguer le reste dans la mesure où les exemples commencent à être relativement bien connus et du coup ça fait évoluer la considération qu'ont les décideurs de ce type de problématique.

On arrive sur ce point central qui est la donnée : quelle définition donnez-vous de la donnée ?

Alors pour moi la donnée c'est à prendre en compte dans une acceptation plus large que la donnée de la base des données. En fait, la donnée c'est une matière brute dont la nature a beaucoup changé avec l'informatique, c'est-à-dire que, et du reste on y reviendra sans doute là aussi beaucoup, ce qui change par rapport à toute cette histoire de données, et de *data culture*, de *data literacy*, etc., c'est que la donnée, elle est devenue numérique, donc extrêmement fluide et toute l'ubiquité dont on entend parler par ailleurs, eh bien c'est la donnée qui reflète ça parce qu'au bout du compte, comme ça peut transiter très vite, tout le monde peut en avoir des usages dont certains sont complètement innovants. Et il faut arriver à considérer cela comme une matière première dans laquelle vont se loger d'autres réalités. Exemple : il y a la donnée de la base de données qui en fait partie, ça c'est une évidence, mais tout ce qui est document pour moi, c'est une famille de données à part entière. Tout ce qui est brevet, c'est encore une autre famille de données. Tout ce qui est message, *mail*, tout ce qui est « donnée non structurée » comme on dit, c'est par essence pas forcément stocké dans une base de données, donc ça ouvre le spectre de toutes les natures de données qui peuvent être à considérer. En clair, dès

qu'on commence à pouvoir poser par écrit quelque chose de mental et une pensée humaine, eh bien, il faut savoir considérer ça comme une donnée.

Vous avez évoqué l'évolution numérique des données : est-ce que vous avez senti un vrai changement dans vos pratiques avec l'évolution de la donnée ?

Alors oui, très clairement. Je vais prendre un exemple. C'est-à-dire qu'en ce moment je suis en train de réfléchir au fait de mettre par écrit toutes les pensées, justement comme j'étais en train d'évoquer à l'instant, d'essayer d'ordonner un peu cette pensée-là à travers l'écriture d'un ouvrage. Et je me suis posé la question à moi-même : est-ce que je le fais avec un crayon et un papier, ou est-ce que je le fais avec un ordinateur et un clavier ? Et après avoir essayé les deux, je vais vraisemblablement le faire avec un ordinateur et un clavier. Je pense que ça résume bien le propos, c'est-à-dire en l'occurrence pendant la guerre froide, si on voit *Les Trois Jours du Condor*, on n'en était encore à une culture essentiellement papier et il fallait qu'il y ait des mecs qui lisent les notes, les bouquins, etc. Aujourd'hui tout cela a été transposé dans une sous-infrastructure numérique et, oui, définitivement, cela fait changer les pratiques là-dessus. Je précise juste le propos : on procède toujours de la même façon, c'est-à-dire que de toute manière cette donnée-là il faut l'appréhender sous une certaine forme et même sous un certain format. Mais au bout du compte, il faut toujours en prendre connaissance, il faut toujours l'analyser et il faut toujours essayer de la raffiner pour qu'elle nous apporte véritablement quelque chose. Donc les mécanismes sont restés identiques. Le format, le formalisme et l'infrastructure sont entrain de changer et du reste il suffit de voir les problèmes que ça pose par rapport à l'histoire, à l'archivage et à toutes ces choses-là.

Quels sont pour vous les enjeux des données, de quoi parle-t-on ? Et quels types de données sont au cœur de ces enjeux ?

Alors pour moi, les principaux enjeux, j'invente rien. En Europe, on est en train de remettre sur le haut du tas, *Gaïa-X*, à savoir toutes les problématiques qui sont liées à l'infrastructure. Je crois d'ailleurs qu'on a discuté ensemble de demander aux gens la différence qu'ils font entre informatique et numérique. Et juste pour remettre un peu les choses un peu au clair par rapport à ça : pour moi, l'informatique c'est une problématique d'infrastructure, donc de tuyaux. Et le numérique, c'est une problématique de ce qu'on fait transiter dans ces fameux tuyaux finalement : de liquides et de fluides qu'on fait transiter dans ces fameux tuyaux. Et je reviens à l'informatique : en fait, le problème s'est déjà posé. C'est-à-dire que, et c'est rigolo parce que moi je donne des cours et je fais réfléchir mes étudiants là-dessus, il y a un moment où, avec le plan *Calcul*, on s'est dit « il faut absolument qu'on maîtrise notre infrastructure de tuyaux pour être certains d'avoir la capacité de calcul qui va bien pour nos grands projets de bombes nucléaires ». En fait, on n'a pas résolu le problème à ce moment-là sur les histoires d'infrastructures mais pour autant ça nous dérange pas de le reposer exactement de la même façon aujourd'hui avec les histoires de numérique, à savoir, quand on est en train de se dire avec *Gaïa-X*, je vous remets en tête que *Cloudwatt* et *Numergy* c'était déjà le même sujet et on s'est déjà raté une fois. Mais voilà on en revient systématiquement à repenser les choses, exactement de la même chose de la même façon. Reposer les bases des stratégies exactement avec les mêmes ancrages. Et finalement, pour en revenir à votre question, je pense qu'à chaque fois qu'on échoue devant ce type de challenge, on perd un bout de ce qui est

important en matière de souveraineté, en matière de capacité à gérer son autonomie, en matière de capacité à prendre ses propres décisions.

Aujourd'hui on parle de la donnée : pour moi, on a déjà raté un épisode avant avec les tuyaux. Je reviens à votre question sur la donnée. Aujourd'hui, tout est important dans la mesure où finalement pour revenir à la définition qu'on donnait tout à l'heure, c'est le propre même de l'intelligence que d'être capable d'analyser les données qui sont soumises à notre sagacité, à nos sens, etc. Et c'est sur la base de ces données-là qu'on forme nos jugements. Si on n'a pas la maîtrise de nos données, on a un jugement qui est juste faussé. Mais voilà sur les fondamentaux c'est ça la base. Donc aujourd'hui, si on n'a pas la maîtrise de la donnée, si on n'a pas la maîtrise de l'infrastructure, quand on prend une décision, elle est ancrée et fondée sur quelle fondamentale solide ? Je vous laisse en juger par vous-même.

Et justement, est-ce que la *data literacy* se présente comme une solution ou comme un artifice supplémentaire pour tenter d'endiguer cela ?

Alors pour moi très clairement c'est une solution. En revanche, c'est une solution multiface et qui remet en cause énormément de choses et c'est pour ça que je n'aime pas tellement parler de solution parce que ça laisse à penser qu'il y a une unicité derrière qui va être valable pour tous les contextes. Il ne faut pas du tout le voir comme ça. Pour moi, justement, cette histoire de *data literacy*, il va falloir en implanter des bouts dans les entreprises, il va falloir en implanter des bouts dans l'administration et la chose publique. Par définition, il va falloir en planter des bouts dans l'Éducation nationale, dans la vie sociale, peut-être même dans les arts, puisque finalement quand on vient tester de nouvelles pratiques, on s'ouvre des horizons qui nous font réagir, on apprend avec nos émotions. Pour moi, le sujet est extrêmement vaste et parler d'une solution, je trouve ça extrêmement réducteur. C'est évident qu'on peut se servir de la *data literacy* et surtout de cette histoire de culture. C'est un peu comme si je vous disais, comme-ci vous me posiez la question : si on devait développer une culture particulière, dans quel domaine il faudra la faire s'exercer ? Dans tous les domaines !

Et donc si je vous suis, il n' y a pas une pratique de *data literacy*, ce n'est pas quelque chose d'homogène, ni d'unifié : ça s'adapte à tous les contextes.

Alors, au même titre qu'une culture, j'aime bien ce parallèle-là. C'est pour ça qu'en fait, une culture elle a son unicité. Pour autant de ce côté du monde, on a développé certaines façons de faire. Quand on se retrouve à l'autre bout du monde, on découvre qu'il y en a d'autres qui fonctionnent d'une façon complètement différente. Donc il y a bien une notion d'unicité. Mais en revanche, par rapport à tous les évènements d'une vie, ça touche à tellement de domaines différents que justement une des difficultés de la *data literacy* c'est de poser les fondamentaux pour être certain que par rapport à la notion de socle commun que j'évoquais tout à l'heure, on a au moins adressé les éléments les plus essentiels. Et après ça se construit au fil de l'eau, au bout du compte, en détaillant les pratiques, en détaillant les usages et, comme je le disais, pas forcément que sur des approches techniques, mais peut être en essayant des choses nouvelles sur des sujets plus artistiques, par exemple.

Si vous avez parlé de socle commun, vous mettriez quoi dans ce socle commun ?

Alors moi je viens de finir un bouquin sur l'école et sur l'éducation parce que finalement quand on regarde un peu autour de quoi c'est articulé, notamment aux États-Unis qui nous ont quand même assez largement précédé sur tout ça, ça touche à plein de sujets. Certains qui sont très basiques : comme par exemple comment on collecte de l'information. Certains qui sont un peu moins basiques : comment on exploite cette information, comment on l'enrichit, comment on l'analyse, etc. Et finalement, moi quand je range tout ça, j'ai un acronyme pour ça qui s'appelle *DIKW* et dans lequel j'adresse les points essentiels. Donc la collecte, l'analyse, la capitalisation et après l'influence, pour l'essentiel. Cela donne un truc super vaste. Parce que tous ces domaines si on veut vraiment partager des théories, des pratiques relatives à tout ça, eh bien oui, il faut se donner les moyens. Alors soit ce que j'essaye de faire en ce moment, de constituer une sorte de mini-socle qu'on va aller présenter à des gens différents. J'ai commencé dans les *business school* et aujourd'hui je vais dans les entreprises pour essayer de faire connaître cette façon-là, cette approche-là et essayer de susciter l'intérêt des gens sur cette notion de socle commun. On évoquera les difficultés que cela ça engendre derrière. Mais voilà, pour dire une démo de ce qui est ennuyeux à travers cette approche-là, c'est qu'au bout du compte, on se retrouve à venir proposer quelque chose d'horizontal à des organisations qui dans bien des cas ont des structures verticales. Donc, tout le monde parle de « désiloter » les organisations mais en fait, moi par exemple, j'ai énormément de mal à trouver des interlocuteurs qui savent penser horizontal. Voilà, ça nous ramène encore à cette histoire de culture et ça montre encore à quel point, finalement, ça va être délicat d'organiser toute cette transformation et de généraliser cette culture.

Si nous restons dans le niveau politique et législatif plus précisément, on a eu un certain nombre de lois et d'applications qui ont émergé. La loi CADA est très ancienne mais plus récemment on a eu la loi Lemaire, le RGPD. Est-ce que ça a changé les pratiques et les mentalités ?

Je serais tenté de répondre non. C'est un peu dur, mais en fait, le point positif relatif à tout ça, c'est qu'au moins ça pose le sujet : ça en fait un sujet. Donc de ce point de vue-là, c'est une très bonne chose. Maintenant après derrière, j'ai lu avec beaucoup d'attention le bouquin de David Graeber sur les *Bullshit jobs*. Je pense que quand on commence par la réglementation, ça veut dire qu'on *bypass*, on *bypass* toutes les étapes qui sont liées à l'éducation. Et finalement, c'est ça qui ne va pas dans cette approche-là. C'est-à-dire que quand on indique la loi sans avoir pris le temps de former les gens, qu'est-ce qui se passe ? On met des *statements* avec des croix avec des boites au bout et en fait les gens vont se borner à mettre des croix dans les boites. Et toute la partie éducative, elle se trouve évacuée pratiquement de fait. Donc comme la partie éducative est évacuée, les pratiques changent à la marge après un gros problème et une mise en défaut par la loi qui avait été édictée, on se retrouve à être obligé de faire les choses parce qu'on a pris conscience, que la croix dans les cases finalement ça nous n'avait pas permis de développer toute la partie éducative. Donc en fait, on fait les choses à rebours et c'est ça qui est très ennuyeux dans cette approche-là.

Comment peut-on s'y prendre pour mettre en place une culture de la donnée dans une organisation ?

Pour m'être déjà sérieusement posé la question, je pense qu'il ne faut pas commencer par le comment. Il faut commencer par le qui, à savoir qui décide de la mise en œuvre d'une culture de la donnée. Et là pour répondre très clairement à votre question, c'est une décision politique. Et du reste, c'est assez rigolo de voir

qu'aujourd'hui quand on remonte *Gaïa-X*, on est en train de se dire « oh ben dites donc, il y a beaucoup d'informations qui nous concernent qui ne sont pas chez nous ! ». Bah oui, c'est construit comme ça. Je te remets en tête les tuyaux, les liquides, etc. Si t'as rien qui coule vers chez toi, t'as rien chez toi. Donc après c'est une décision politique : est-ce qu'on accepte ça et de voir notre futur défini par des gens qui n'habitent pas sur notre territoire ? Ou est-ce qu'on décide de s'en donner les moyens ? Le qui, c'est la nation, c'est le pays. Et sur toute l'épaisseur du trait que j'ai évoqué tout à l'heure, avec tout ce que ça implique derrière. C'est-à-dire qu'après si on décide ça, il va falloir aussi s'en donner les moyens, il va falloir définir notre rapport aux autres. Il va falloir se reposer sur des grandes questions quand même. D'où cette dimension essentiellement politique finalement.

Vous parlez politique, donc à une échelle je pense assez nationale ?

Parce que aujourd'hui la plupart des initiatives dont vous parliez comme le *RGPD*, *DSA*, DMA, etc., tout ça ce sont des initiatives européennes. Après c'est vrai qu'en Europe, on transcrit du droit européen, on fait la loi Pacte, et du coup on essaye d'ouvrir les chakras par rapport à tout ça. Mais je réinsiste là-dessus : s'il n'y a pas une vraie volonté politique, c'est très compliqué d'arriver, au titre de la conduite du changement, à créer ce sentiment d'urgence qui est une absolue nécessité pour faire bouger les gens.

Toujours par rapport à cette histoire d'application de culture des données. Si on s'arrête à l'échelle d'un organisme, d'une entreprise, comment peut-on la mettre en place et est-ce que ça doit concerner tout le monde ?

Cela doit toucher tout le monde déjà, pour être clair. Et aujourd'hui, c'est un peu ce que je laissais entendre tout à l'heure, c'est-à-dire que si on veut marquer le coup, et si on veut rendre le sujet visible, il faut effectivement rentrer dans cette logique d'adaptation de l'organisation. C'est-à-dire qu'on ne peut pas diffuser une culture de la donnée dans des structures à une verticalité comparable à celles qui existent aujourd'hui. C'est juste pas possible. Ou alors si on peut le faire : mais que dans la verticale à ce moment-là. Et c'est ce qui fait qu'on n'en tirera pas un réel bénéfique. Donc pour moi, concomitamment à cette histoire de culture de la donnée, il va y avoir des adaptations organisationnelles qui seront juste nécessaires. Et en l'occurrence, ce que je préconise à travers les approches dont j'assure la promotion, c'est de créer des équipes transverses qui vont se charger elles d'agréger, au sein de chacune de leurs verticales la plupart du temps, les principaux usages. Et ensuite, on va pouvoir organiser les échanges, les partages, les relations et les mises en relations autour des différentes familles de données, des différentes valorisations qui les concernent. Et pour moi, c'est justement cette approche, cette création d'une approche transversale au sein d'une organisation. C'est un prérequis. Après oui on peut monter des projets mais de toute façon on ne les fera que dans un cadre préexistant, ce qui fait qu'on ne tirera pas réel bénéfice de ce que pourrait apporter la valorisation des données en tant que telle.

Et ces équipes transverses, comme vous le précisez, est-ce que les personnes qui sont dans ces équipes-là, est-ce qu'elles ont des postes bien définis ou est-ce que cela peut être n'importe qui dans l'entreprise ?

Elles ont des postes bien définis. Après, la définition de ce poste n'est pas corrélée à la notion de donnée en tant que telle. Généralement derrière le poste il y a une mission, même pas un rôle : une mission. Bon après je ne vais pas m'étendre sur les histoires d'adaptation de l'organisation, mais en l'occurrence, vous devinez peut-être dans mon propos que j'aurais tendance à définir plutôt des rôles plutôt que des missions, mais on ne va pas chipoter. Au bout du compte, finalement, les fonctions, elles existent déjà, elles ont été définies.

La vraie question, tout l'enjeu de cette histoire de culture, c'est : comment on fait pour des gens qui ont des *statements* ? Je reviens à mon histoire de croix dans les cases. On leur dit : « voilà ta mission, c'est ça avec une croix au bout ». Peut-être dans la plupart des cas, avec une mesure, avec un entretien annuel. Comment on fait pour introduire une dose de proportionnelle pour revenir à mes approches politiques en disant : « voilà tu dois faire tout ça et en plus tu dois être conscient de toute cette approche qui est liée à la donnée, et justement dans tout ce cadre-là, on va te donner le rôle d'être le mec qui est capable d'évaluer l'humeur de l'équipe qui est dans le bâtiment bidule, tu nous diras s'ils vont bien, s'ils vont pas bien ». C'est très con quoi, mais voilà juste un exemple.

Et du coup après, derrière, comment on fait pour l'intégrer dans l'entreprise ? Qui va le prendre en charge ? D'une entreprise à l'autre ça va être très différent. Il y en a qui vont considérer que c'est de la communication interne parce qu'il faut, si on se rend compte que les gens ne vont pas bien, il va falloir faire en sorte de leur soumettre des éléments qui vont leur permettre de gérer tout ça. Dans d'autres cas, ça va être une problématique de RH. Dans d'autres cas, ça va être une problématique de secrétariat général. Nous on a fait l'expérience sur les histoires de protection de données avec *Get them* et je dirais que nos interlocuteurs finalement, même s'ils avaient des intitulés de postes différents, il y en avait toute une partie qui venait de tout ce qui était sécurité, parce que sur la protection des données, c'est eux qui se tape toutes les galères qui vont derrière. Il y en a qui étaient justement sur la partie marketing-com, parce que c'est eux qui connaissent les codes, les valeurs de la boîte et qui savent comment communiquer. Et les derniers c'était tous ceux qui étaient sur la partie RH et formation, parce que tôt ou tard, même les non experts, il va falloir rentrer ça dans leur définition de poste. Et c'est gens-là qui ont ce type de mission-là au sein des organisations. Et après par extension on pourrait mettre dans la boucle les R&D pour les problématiques de brevets. On pourrait mettre des tas de gens dans la boucle pour adresser des approches particulières liées aux différents usages et pratiques qu'on peut construire autour de la donnée.

Donc en fait si je vous suis bien, tout ce qui est mise en place « culture des données », c'est avant tout une affaire de médiation et de communication plus que de compétences techniques et de savoirs.

Alors dans ce sens-là oui carrément. Et moi, c'est un truc sur lequel je suis très militant dans la mesure où finalement quand on regarde comment s'appréhende le problème dans les grandes entreprises, dans les grands groupes aujourd'hui, c'est vraiment une caricature. Quand je vois le nombre de grandes boîtes qui embauchent des *data scientists*, qui se plaignent de ne pas pouvoir les recruter, qui au bout du compte leur donnent des boulots sur lesquels moi j'aimerais avoir juste un peu de *feedback* sur la valeur de ce que ça a absolument généré. Comparer finalement à toute cette approche effectivement de communication auprès de l'ensemble des salariés sur « bah voilà, est-ce qu'on va réfléchir ensemble à vos boulots ? Qu'est-ce qu'ils vont devenir ? ». Est-ce qu'on va réfléchir ensemble à tout ce sur quoi vous vous trouvez embringués au titre des réseaux sociaux ? Ce type de

réflexion, comme vous le soulignez, est toujours appréhendée d'un point de vue purement technique et jamais sur cette approche humaine dont je parlais tout à l'heure. Donc c'est très dérangeant.

Je pense aussi que, à un moment, je ne sais pas si on en parlera mais il faut s'interroger sur ce qu'on met dans la valeur de la donnée. Moi je reste convaincu qu'économiquement parlant, on va avoir un léger problème à gérer, à savoir qu'une donnée elle est réplicable à l'envi donc que si on veut gérer sa propriété, on va se retrouver avec des tas de soucis qui seront très complexes à gérer d'un point de vue purement économique. Donc autant s'ouvrir les chakras maintenant et essayer de poser d'autres familles d'indicateurs pour mesurer cette valeur, pour évaluer les pratiques, etc. Moi en tout cas c'est vers ça que je milite et que je tends parce que je pense que la dimension politique dont je parlais tout à l'heure : ce n'est pas pour prendre une décision à un moment. C'est-à-dire que dans l'après il va y avoir toute cette implication. Quand on parle de communication aujourd'hui, je suis curieux de voir comment va se régler la présidentielle par exemple, notamment en termes d'influence. Je ne vous fais pas un tableau sur ce à quoi je fais référence par rapport à l'actualité très récente mais il y a des déséquilibres qui sautent aux yeux. Toute cette pensée critique, toute cette culture de la donnée et de l'information, elle touche aussi à ça. Et il y a des moments où, comme je le disais à l'instant, elle saute aux yeux. La nécessité, je veux dire, d'être définie et en tout cas au moins précisée.

J'ai plusieurs questions qui me viennent en tête. Tout d'abord sur la valeur de la donnée : est-ce qu'il y a selon vous une vision trop restreinte sur la valeur de la donnée ? Qu'est cette valeur et est-ce qu'on est pas bloqué dans une vision particulière actuellement ?

Carrément. Pour répondre en un mot : carrément. On est carrément bloqué parce que c'est exactement le terme sur ce que j'appelle négocier. Parce que si on reste sur une dimension économique et, je dirais même plus précisément, financière de la donnée, on va être bloqués dans ces modèles-là. Et ces modèles ce sont des modèles qui visent à identifier une famille de données et de regarder comment on va pouvoir les négocier, comment on va pouvoir organiser des transactions avec ces données-là entre deux tiers. Donc ça c'est un modèle purement économique. Et aujourd'hui, toutes les boites qui se montent ne le font pas, je veux dire par définition concrète, parce que de toute façon ce sont des sociétés à but commercial, donc elles n'adressent quasi exclusivement que ce type de données. Je dis quasi exclusivement car au-delà de ça, il y a quand même quelques initiatives vraiment intéressantes où les gars, par exemple ceux qui bossent en association trouvent des modèles qui ne sont pas ceux de la société à but commercial. Et du coup on voit émerger des choses très intéressantes avec des gens qui, dans leurs communautés, ou dans leurs régions, dans leurs territoires à rendre des services et du coup vont chercher des compensations qui ne sont pas toujours immédiatement financières. C'est assez rare mais ça existe. Et ça existe en tout cas suffisamment pour être noté. Parce que c'est sans doute finalement auprès de ceux-là qu'il faudra chercher d'autres familles d'indicateurs qui leur permettent de se regarder avancer. Parce que la vraie question à se poser par rapport au schéma monolithique dans lequel on est, c'est : si on ne mesure pas notre valeur ajoutée en argent, on la mesure en quoi ? Et je pense que c'est exactement vers là qu'il va falloir tendre par rapport à cette histoire de données. Il y a déjà des choses existantes : est-ce que c'est du pourcentage de CO_2 ? Est-ce que c'est du nombre de *followers* ? Est-ce que je veux dire là maintenant qu'il y a des gens qui commencent à mettre des billets parce que l'influence de bidule sur Youtube, il a un million de *followers* ? Quand il dit un truc normalement on doit pouvoir se faire un peu de blé du coup derrière. Donc, je ne veux pas revenir sur la dimension philosophique qui fait qu'en toile de fond on aura toujours cette histoire d'argent

qui traine d'une façon ou d'une autre, mais n'empêche qu'au bout du compte, et notamment par rapport à des jeunes initiatives et des montages de *start-up*, il y a plein de nouvelles données qui seront à considérer pour générer et regarder comment on valorise d'autres familles de données.

Et parmi ces données, vous citeriez lesquelles ?

Des données de valorisation ? J'en ai cité deux par exemple. C'est-à-dire qu'il ne viendrait à personne aujourd'hui d'acheter une voiture qui a des caractéristiques techniques proches de celles des années 40/50 parce qu'ils auraient peur du regard des voisins, que culturellement, il se le ferait reprocher. Ce ne serait pas bien. Donc il y a des choses qui existent déjà sur lesquelles s'appuie une certaine forme de publicité. Je pense qu'il faut encourager ça, c'est-à-dire qu'au bout du compte, il faut se donner les moyens d'aller les mesurer un peu partout. Donc politiquement, on se met d'accord sur le fait que c'est une donnée importante, qu'on va généraliser la mesure auprès de tous ceux qui ont des activités importantes pour la communauté. Et après on le fait. Donc aujourd'hui sur les frigos, sur les fours, sur un peu tout et n'importe quoi, sur les appartements, il y a une échelle de A à E si je me rappelle bien et on regarde quel est le pourcentage. Alors en fait, je ne suis pas trop rentré dans les détails, mais j'ai l'impression qu'il y a une petite subtilité dans les mesures qui sont proposées. N'empêche que ce qui est fondamental, c'est quand même quelque chose d'écologique et de vertueux par rapport à l'impact que peut produire la chose concernée sur l'environnement. Et cette mesure d'impact propre à cet usage-là me semble intéressante et à reproduire pour d'autres choses. J'ai évoqué aussi l'exemple des *followers* : aujourd'hui il y a plein de boites dans la com justement qui construisent des écuries d'influenceurs tout simplement parce qu'elles savent très bien qu'elles pourront monnayer l'audience de ces gens-là.

Vous avez évoqué l'idée d'être critique vis-à-vis de la donnée : qu'est-ce que vous mettez derrière cet esprit critique justement ?

Alors pour essayer de préciser et je pense que ça se devine en partie dans mon propos, je ne suis pas tellement critique envers la donnée en tant que telle parce que la donnée c'est juste un reflet de l'interprétation qu'on en fait et qui peut être sujette à critique. Mais au bout du compte, c'est cette histoire d'interprétations et même d'usage et même de, comment dire, de commerce de la donnée qui me dérange. C'est-à-dire qu'au bout du compte, je l'ai dit en une tournure de phrase tout à l'heure : je comprends très bien qu'on utilise de l'argent quand il s'agit de fabriquer un bien. Ce bien il est unique, il est posé à cet endroit-là. Quand je m'en sépare, quand je le donne à quelqu'un, le quelqu'un en question, il me donne de l'argent en contrepartie. Ça je comprends bien, ça me pose aucun problème. Après le service, je ne peux pas acheter le truc durablement mais je vais m'en servir pendant un temps. Mais je sais exactement où il est et comment il fonctionne, et ce qu'il va m'apporter, etc. Que je paye pour ça, ça je comprends aussi.

En revanche, quand on arrive à la donnée et que la donnée, je peux la répliquer douze millions de fois, que ça coûtera un pouième de temps machine, je veux dire le pouième est tellement négligeable qu'au bout du compte on devrait pouvoir gérer ça autrement. Et là coup, ça devient dérangeant d'arriver, de mettre la main sur certaines familles de données très particulières en disant, « voilà, je ne vous vends pas ça quelques centimes, mais je vous vends ça quelques dizaines de milliers ou quelques centaines de milliers d'euros parce que j'ai compris que vous en aviez vraiment besoin et que du coup je vais faire mon *business* là-dessus ». Là je trouve

que ça commence à être un peu critiquable parce que je pourrais le reproduire plein de fois. Et je ne suis pas à l'aise avec ça pour le dire clairement et je pense que ça posera problème tôt ou tard. Parce qu'en plus ça dans la propriété, dans le transfert même de propriété lié à quelque chose de réplicable, on voit déjà ce que ça donne avec l'IP, avec tout ce qui est propriété intellectuelle, il y a plein de trucs qui sont d'une grande complexité. Et qui s'arbitre différemment selon les lieux, selon les périodes, selon les *business*, etc. Donc c'est une notion qui n'est pas solide, pas suffisamment sérieuse pour avoir le niveau de pérennité qui politiquement est acceptable.

Est-ce que cette vision des données affecte les pratiques en entreprise ?

Oui je le disais tout à l'heure. En fait, je vais le dire avec mes mots mais finalement on a appliqué à la donnée ce qu'on avait appliqué aux services préalablement et ce qu'on avait appliqué aux produits encore avant, c'est-à-dire une grande vénalité, c'est « comment je vais faire du blé avec ça ? ». Et autant, comme je le disais, je n'ai pas de problème avec les deux items précédents, autant quand on en arrive à la donnée, ça créé des déséquilibres tellement flagrant que si politiquement on ne corrige pas un peu le tir, eh bien comme je le laissais entendre un peu tout à l'heure, on va tous en subir les conséquences. Et ça me dérange beaucoup. Ça ne veut pas dire que j'ai des solutions toutes faites à ça, parce que finalement je ne sais pas comment lutter à vrai dire. Mais je dirais que la monoculture économique et financière est pour moi le pire qui puisse arriver quand on cherche à s'intéresser à une valorisation pérenne des données, parce que du coup, on se retrouve sous l'influence déterminante, en tout cas aujourd'hui, de ces critères financiers.

Lors d'une application de la culture des données dans un organisme, que doivent développer les personnes qui reçoivent cette culture, cette formation aux données en termes, par exemple, de compétences, de savoir-être ?

La pensée critique. Clairement, la pensée critique. C'est-à-dire aujourd'hui moi je rencontre plein de gens qui sont des amis proches, qui me disent « mais moi, dans ma boîte je fais de la merde, je le sais très bien mais je suis bien payé pour en faire et du coup comme aujourd'hui toute ma vie repose sur le pognon que je gagne, eh bien je vois pas pourquoi je changerai ». Mais alors, il y a la pensée critique, parce que ça permet de prendre conscience de ça, et après il y a la capacité d'action. C'est-à-dire, au bout du compte, qu'est-ce qui peut faire que du jour au lendemain ces gens-là passent à l'action ? Alors là, on est encore sur d'autres sujets, mais qui sont tous liés à la transformation. C'est-à-dire qu'au bout du compte je reste convaincu que la recette que je cherche moi, c'est comment on fait bouger quelqu'un, comment je fais bouger mes amis qui gagnent plein de blé, qui offrent des tas de possibilités à leurs enfants, sachant que si j'en crois tout ce que je lis et tout ce que à quoi je m'intéresse, leurs enfants auront de toute évidence pas la même qualité de vie qu'eux. Quand on a tout ça c'est effectivement très vaste et on peut s'y perdre très facilement. Mais c'est un peu ça le jeu, si on ne décolle pas d'une approche purement « pognon » du truc, tous les problèmes dont on entend parler là, on s'en sortira pas, on trouvera pas de solutions. Parce que justement dans la pensée critique et toutes les idées qu'on peut être amené à faire naître au titre de cette pensée critique, il y a des solutions qu'on *bypassera* de fait parce qu'elles ne rapportent pas de blé. Manque de bol, justement, c'est vraisemblablement dans ces solutions là qu'il y a la solution, dans ces problèmes là qu'il y a la solution qu'il faudrait sans doute généraliser. Je vous laisse imaginer deux secondes aussi, si on en arrive à un moment au prix du litre d'air qu'on respire, y a pas

de CO_2 dedans. C'est juste même pas concevable. Il y a plein de trucs qu'il faudra savoir considérer comme apportant de la valeur même si on ne sait pas combien.

Est-ce que les formations actuelles forment des spécialistes de type *data scientist* ou est-ce qu'il y a quand même des profils assez divers ?

Alors en fait, je peux vous apporter une réponse circonstanciée à ça parce que j'ai eu l'opportunité de participer à une étude étant donné que je suis [nom de fonction]. J'ai participé à cette étude et cette étude a révélé, et c'est justement ce qu'on disait tout à l'heure, que les profils sont très déséquilibrés dans le sens où, je vais le dire avec mes mots, il y a beaucoup de « crânes d'œuf », c'est-à-dire beaucoup de gens qui pensent et qui intellectualisent toutes les problématiques de la donnée, de l'intelligence artificielle et tout ce qui s'ensuit. Mais finalement, il y en a très peu qui « font », c'est-à-dire que des gens qui sont capables d'aller trouver des applications concrètes à tout ça et de déployer ces applications, il y a très peu de formations qui les concernent. Je le dis en langage français, universitaire et éducatif : on forme du bac + 5, on forme très peu de bac +3, bac +2, les mains d'œuvre de la donnée.

Quels types de métiers mettez-vous derrière ces bacs + 2 et + 3, si ça existe d'ailleurs ?

C'est ce que j'allais vous dire. Pour moi aujourd'hui, toute la logique qui consiste à penser les métiers comme étant une succession de missions à faire, c'est bien parce que ça permet de poser le sujet avec les gens tel qu'ils le connaissent aujourd'hui. Mais moi aujourd'hui, je reste convaincu que ça reste assez peu adapté à ce qu'il faudra faire demain. Je vais le dire d'une autre façon : c'est-à-dire que je pense que demain, et d'ailleurs il y a une terminologie qui est en train de m'échapper sur la façon dont on nomme ça aujourd'hui. Mais en fait, par rapport aux métiers existant aujourd'hui, demain on aura des profils qui seront plutôt amenés à piocher parmi plusieurs métiers existants pour remplir des rôles particuliers et qui les obligeront quelque part à faire certaines opérations pour véritablement produire la valeur pour laquelle ils sont disposés comme ça dans l'organisation. Je ne suis pas super clair là-dessus parce qu'on n'est pas dans le futur. Mais en tout cas, je vous prends un exemple aujourd'hui : si on considère la qualité des données, si on raisonne telles que les entreprises sont structurées aujourd'hui, on va nommer un *data scientist* pour travailler sur telle famille de données. Il va commencer à pondre toute sa démarche, il va fixer ses outils, il va organiser ses traitements et puis jusqu'à l'étape où il va dire « maintenant, il me faut les volumes de données et il faut que je tape dedans et il faut que je vois ce que ça donne par rapport à tout ce que j'ai prévu ». Et généralement c'est à ce moment-là qu'il se rend compte qu'il a des données pourries et plein de trucs qui ne sont pas renseignés ou qui sont manquants ou qui sont incomplets, etc. Et ce qui fait qu'à ce moment-là, il perd un temps considérable parce qu'il va essayer de trouver le gars dont le job c'est de *cleaner* les données. Et je suis convaincu que dans le futur on ne pourra pas faire comme ça. On ne pourra plus faire comme ça. Et il faudra qu'il y ait des gens qui acceptent de *cleaner* les données. Alors bien évidemment après, quand il y a des gisements entiers de données qui sont pourris, là on confiera ça à des gens qui utiliseront des gros trucs vraiment *tankés* pour remettre ça d'équerre. Mais voilà, dans l'absolu, on sera obligé de faire plein de petites choses différentes pour avoir une vue d'ensemble sur la valeur qu'on génère avec la famille de données sur laquelle on bosse. Je reviens à ce que vous disiez tout à l'heure : en fait, je ne crois pas trop à cette histoire de métier mais je pense que pour revenir à cette notion de culture qu'il y aura un

socle de base et après on choisira des petites briques, si je puis dire ça comme ça, ou plutôt des petits pétales qui conviendront à tel ou tel usage plus marqué, lié à des familles de données spécifiques.

Si je vous comprends bien, aujourd'hui, on centralise trop de tâches sur un métier en particulier et on n'a pas pensé à développer, je prends à défaut le terme de métier que justement vous dénoncez, d'autres métiers pour faire ça...

Je suis convaincu que de ça. C'est-à-dire qu'aujourd'hui, on a une approche du sujet qui est de l'ordre photographique et il faut qu'on mette la dynamique en route parce qu'aujourd'hui elle ne l'est pas. Donc en clair, quand on aura une approche plus *movie*, plus vidéo, plus animée, eh bien effectivement on se rendra compte que la façon dont ça bouge montre à quel point l'organisation ne peut pas continuer de fonctionner sur cet angle. De cette façon-là en tout cas.

Est-ce que vous avez en tête un projet en *data literacy* ? Vous avez cité notamment Gaïa-X....

Pour s'en tenir à *Gaïa-X* et à la totalité du projet, j'ai nommé tout à l'heure *Cloudwatt* et *Numergy*. Comme quoi, on a déjà fait un premier épisode et j'espère qu'on a tiré le bénéfice de tout ce qu'on a décidé à ce moment-là. Et quand on voit le résultat que ça produit en l'occurrence c'est un truc qui me frappe tout le temps : je pense qu'il y a eu des gens de grande qualité qui se sont succédés dans l'open data au sein du gouvernement. Je pense à Henri Verdier et à l'impulsion qu'il a réussi à donner. Moi je l'ai vraiment suivi de près parce que j'ai trouvé ça absolument remarquable. C'est un garçon qui a une vraie vision, une vision politique, comme je le disais tout à l'heure, de la donnée et je trouve hallucinant qu'on l'ait collé au placard parce que s'il y en a un qui faisait le boulot, s'il y en a un qui avait la vision, s'il y en a un qui contribuait à essayer de faire évoluer les mentalités auprès d'une institution qui est quand même âgée et sclérosée. Bref je ne vous fais pas le débat, c'était vraiment lui.

Et du coup lui a eu plein d'initiatives en matière de projets. Les *start-up* publiques par exemple, c'est une super idée. Des choses comme *PIX*. Moi j'ai eu l'opportunité d'échanger avec le directeur de *PIX* et même de le faire intervenir dans les conférences, et je trouve ça absolument remarquable. « Dites-le nous une seule fois » ça aussi c'est une super initiative. Donc voilà, il y a plein de choses bien qui se font mais je dirais qu'on en reste, c'est-à-dire qu'on considère le numérique comme étant une branche, une verticale de plus. Et tant qu'on résonnera comme ça il y a un stade qu'on ne pourra pas franchir. Donc voilà, je ne peux pas le faire autrement. En fait, les projets il y a plein de trucs bien. Quand on les contingente à une verticale donnée et qu'on ne se donne pas les moyens de les diffuser dans les autres, il y a des choses qui sont impossibles. Donc je ne sais pas comment on peut tirer un enseignement de ça. J'ai vu qu'il y avait des initiatives, je suis tombé dessus très récemment, notamment dans le Ministère de la transformation publique qui visait à mettre de côté, et je trouve ça super intéressant. D'un autre côté l'appel à projet est en cours, moi ça m'intéresse d'aller taper pour voir la réalité qu'il y a derrière, ce que je suis en train d'évoquer, à savoir : le Ministère de la Transformation Publique se propose de suspendre les réglementations sur un territoire donné pour faire des vraies expérimentations relatives à la donnée. Et je trouve ça super intéressant dans la démarche. Je ne sais pas ce qu'il va y avoir derrière et je trouve ça super intéressant.

Si on dit suspension des réglementations, cela veut dire qu'il y a des réglementations qui bloquent ces initiatives ? Avec toutes les réglementations qui ont été produites, estimez-vous que c'est un frein ?

C'est absolument évident. Une réglementation c'est un cadre. Donc ça veut dire qu'on fige de toute façon ce qui va être à l'intérieur et à l'extérieur du cadre. Donc évidemment que c'est une contrainte. Après ça oblige les gens à restreindre toutes les explorations qu'ils pourront faire autour d'une idée pour être certain que ça rentre dans ce cadre-là. Alors comme je disais tout à l'heure, justement, la bonne nouvelle, c'est que, en traçant le cadre, on laisse à entendre que tout le vide qu'il y a au milieu c'est un truc important. Mais au bout du compte, les mêmes qui doivent se challenger là-dessus ils doivent systématiquement obligés de faire attention là où ils mettent les pieds ou s'ils prennent un tampon parce qu'ils regardent pas forcément, ils sont sortis du cadre, c'est stupide et ça n'encourage ni l'entreprise, ni l'innovation, ni l'initiative au sens large.

Je vais peut-être aller plus en profondeur : est-ce que vous avez une réglementation précise qui bloque justement cette créativité ?

Alors un exemple précis. Si moi demain je décide de construire mon propre programme de *data literacy*, je vais me prendre le Ministère de l'éducation nationale et que je décide que mes mômes ou ceux de ma rue entre huit et dix ans, je leur réserve trois mois où c'est moi qui gère, voilà ce que je leur propose en matière d'éducation. Je ne rentre même pas dans le sujet. Je sais déjà que je vais m'en prendre plein la tronche et que ça ne va pas être possible. Ou que, justement je viens de finir le bouquin de Ken Robinson, la *creative schools* et les gens qui ont monté ce type d'initiative là, il y'en a mais du coup, ils en ont fait un métier à part entière et ils font que ça de leur temps maintenant. Voilà, donc tout ça pour dire que quand on cherche à résoudre un problème ponctuel pour essayer de trouver des solutions qui vont permettre d'adresser d'autres problèmes ensuite, il y a toujours des réglementations et il faut aller chercher des spécialistes. Alors ça ne veut pas dire que ça ne va pas se faire. C'est pas ce que je dis. Ça veut dire qu'au bout du compte, entre le moment où on décide de faire avancer le projet et le moment où on arrive à ce qu'on voulait *checker*, on multiplie le temps par je ne sais pas combien quoi. Et en plus, il y a des sujets sur lesquels on n'ira pas. Et notamment sur l'humain. On n'ira pas parce qu'on sait pertinemment que comme il n'y a de gain pognon qui pourrait justifier qu'on déplace des montagnes, eh bien il va falloir qu'on se cogne les réglementations à lever à la place. Voilà, c'est court mais assez symptomatique à vrai dire.

Est-ce que selon vous ces projets sont connus de l'ensemble des personnes qui travaillent sur les données ou c'est vraiment des initiatives individuelles qui émergent sans de véritables concertation ?

Je pense qu'il y a plusieurs scenarios. Je pense que dans l'*open data*, il y a une fluidité qui est plutôt entretenue et plutôt bien foutue. Là il y a une conférence, ou plutôt une non conférence qui se monte courant novembre. Moi je vois qu'il y a une vraie communauté, que ça bouge. Alors peut-être pas autant qu'on souhaiterait mais en tout cas cette approche de diffusion, d'essayer de faire connaître ce qu'on fait, ça me semble tout à fait palpable et tout à fait concret. Donc dans l'*open data* pour moi, ça le fait, y a pas de soucis. Je pense à l'extrême dans tout ce qui est *Big Data* quand on évoquait les *data scientists* tout à l'heure, les mecs sont sursollicités et pareil ils jouent un peu les divas et donc sous-entendu ils ont un peu une visibilité des différents projets sur lesquels on vient les chercher, sur lesquels ils veulent aller ou pas. Donc pareil, ça a une autre façon de considérer la ca-

pacité à faire connaître les initiatives. Mais je pense que dans ce cadre-là, il y a des bons projets. Il y a des projets moins bons et grosso modo, autour de ces approches de *Big data*, on peut, dans ces rails là parce que c'est pas les mêmes que ceux de l'*open data*, il y a des gens qui peuvent avoir une vision des initiatives et être capables d'en parler. En plus, il y a un côté médiatique qui est assez sexy, donc il y a des trucs qui tombent, ça va aussi. Et entre les deux il y a ce que j'appelle moi le *smart data* où là on n'est ni dans des problématiques d'ouverture comme l'*open data*, ni dans des problématiques de volume comme le *big data*, mais on est dans le côté intelligence si je puis dire. Au double sens de cognitif et de renseignement venant de l'onglet quoi. Et là-dedans, je pense que les gens en revanche, et c'est pour moi le plus gros morceau, ne communiquent pas facilement sur leurs initiatives parce que tout simplement ils ne sont pas très clairs sur ce que ça leur apporte et sur ce que ça va leur donner comme avantages par rapport à leurs concurrents dans la suite et je dirais qu'ils baignent dans un milieu que j'ai pas mal côtoyé qui est celui de la compétition et de la concurrence qu'on adresse en France à l'outil de l'intelligence économique et stratégique. Et en fait, dans ces milieux-là, cette culture du renseignement, elle fait en sorte qu'on a tendance plutôt à cacher ses succès qu'à en parler. Donc voilà, dans les deux extrêmes que j'identifie *open data* et *big data*, oui je pense que les initiatives sont relativement connues dans les communautés. Et pour celles du *big data* ça tombe un peu dans les médias. Et au milieu, c'est un peu une sorte de trou noir, c'est-à-dire qu'il y a plein de petites boites qui prennent des super initiatives mais généralement elles les gardent jalousement parce que ça leur donne un avantage concurrentiel, parce que finalement ça va leur faire tourner leur business d'une façon qu'elles veulent pas forcément partager avec d'autres.

Rapidement, que mettez-vous derrière le smart data ?

Le *smart data* c'est en fait la façon dont j'ai articulé mon discours. C'est tout cette approche DIKW, c'est-à-dire tous les cycles de valeur qui vont être centrés sur la donnée. Et justement dans les petites boites, il y a plein de fondateurs qui ont des visions, qui savent plutôt bien d'ailleurs de quelles données ils ont besoin pour, donnée-informations-connaissances-etc., ils ont besoin pour accomplir cette vision et qui le font quoi. Parce qu'en France on a des entrepreneurs de qualité. Et que si on se donnait de temps en temps la capacité de les mettre en valeur, on aurait beaucoup à apprendre de la façon dont ils ont procédé et dont ils ont utilisé les données dont ils avaient besoin pour accomplir la vision qu'ils avaient en tête.

Pour terminer, avez-vous des recommandations pour une meilleure acculturation aux données ?

Eh ben déjà je pense qu'un gars comme [Nom de personne] en secrétariat d'État au numérique, je pense que, de par les fonctions qui sont les siennes, et j'ai rien contre la personne parce que je pense au contraire qu'elle essaie de faire beaucoup pour faire connaître les initiatives des entreprises. Mais c'est un peu ça qui est ennuyeux. C'est-à-dire que finalement elle développe une vision qui est quasi exclusivement économique de la donnée. Et ça pour moi quand je faisais un parallèle, quand je citais tout à l'heure Henri Verdier. Il faut savoir agréger toutes ces différentes visions qui, pour moi, sont complémentaires de la donnée. Donc j'ai nommé les trois que je me suis construites. Donc l'*open data* pour tout ce qui est ouverture, gratuité, mise à disposition, facilité d'utilisation, etc. Tout ce qui est *big data* pour approche un peu plus pointue, liée au volume, à l'exploitation de l'intelligence collective, etc. Et entre les deux, il y a quelque chose qui est beaucoup plus délicat à formaliser que j'appelle la *smart data* parce que ça va dépendre complètement des humains qui sont impliqués dans les

organisations et de leurs capacités. Pour le dire en une formule : de réfléchir intelligemment et de construire leurs propres recettes par rapport à tout ce qu'on a pu évoquer aujourd'hui.

Donc pour moi, les recommandations, ce serait déjà d'adresser la vraie dimension politique de la donnée, c'est-à-dire en renforçant non pas tant les pouvoirs mais que la capacité de communication et d'influence d'un secrétariat au numérique pour lui donner les moyens d'être entendu par les autres ministères. Et aujourd'hui je pense que c'est peu le cas. Je viens de lire le rapport de la Cour des comptes sur les grands projets de l'État français et notamment avec un gros focus sur la *DINSIC*, à savoir l'informatique du gouvernement. Aujourd'hui je fais toujours le parallèle entre les données et les tuyaux. Quand je vois au niveau des tuyaux, l'interopérabilité qui existe dans les ministères pour construire un vrai gouvernement français, je me dis qu'on a déjà un gros boulot à faire là avant de commencer à parler des données qui transitent et de la façon dont on va devoir gérer tout ça. Je ne parle même pas des problématiques de souveraineté, de sécurité et de tout ce qui peut aller autour de ces différentes familles de données. Mais au bout du compte, si on est une collectivité, cette culture du transversal, de l'horizontalité, etc., eh ben on est obligé de la prendre. Tant qu'on ne la prend pas en compte pour ce qu'elle est, je le répète encore une fois, il y a plein de *steps* qu'on ne peut pas franchir. Quand on ne peut pas franchir ces *steps* là, eh bien on reste prisonnier des anciens schémas.

Question supplémentaire posée après l'entretien :

Comment définir l'éthique des données et quels sont les enjeux qui y sont associés ?

Concernant l'éthique des données, ça devrait être une question clé. Et ça reste malheureusement une petite rubrique juridico-administrative où on se contente de traiter de ce qu'on a le droit ou pas le droit de faire (regardez le rôle de la plupart des DPO dans les boites…). Je pense que très peu de gens ont compris les impacts liés à l'usage des données (surconsommation, manipulation, propagande, désinformation, j'en passe…). Encore moins s'imaginent les dérives dont elles peuvent faire l'objet (encore qu'il y a un vrai travail en cours pour faire émerger ce sujet, notamment autour de la cybersécurité). Je suis toujours surpris de constater que parmi les plus jeunes que je croise en *business school*, bien peu ont entendu parler du *social credit program* chinois. Et une proportion tout aussi faible a regardé des docs accessibles sur Netflix comme *The great hack* ou *The social dilemma*. Du coup, vouloir parler d'éthique devient un délire de spécialistes où vont se détacher deux grandes familles : les partisans du « ce qui doit se faire » et ceux du « ce qui peut se vendre ». Je suis tombé sur un document sur Arte : *Instagram, la foire aux vanités*. Vers la fin, le groupe Meta est évoqué au sujet de ces efforts à tenir compte du développement des enfants (et à ne pas interférer dans ce processus). Notamment, une directrice du groupe est interrogée par des sénateurs au congrès. C'est juste confondant de cynisme : « laissez-nous faire du profit avec vos données, nous veillons sur vos chères têtes blondes ! »

Bref, vouloir mettre de la morale dans l'usage de la donnée, c'est déjà commencer par demander un droit de regard sur tous ceux qui prétendent s'intéresser à sa valorisation. Et tant qu'ils gagneront autant d'argent (*Apple*, *Google* et *Microsoft* sont dans le top 5 des entreprises qui gagnent le plus d'argent dans le monde), rien ne sera possible sans leur assentiment (*Google* est des plus dépensiers en matière de *lobbying* à Bruxelles alors que se décident DSA, DMA, DGA DA,etc…) Pour me mouiller vraiment, je suis carrément du côté des progressistes : je pense que la donnée n'est pas bonne à vendre pour la bonne et simple raison qu'elle est facilement réplicable :

on ne se dépossède pas quand on partage une donnée. Dans le contexte actuel, je pense que derrière la notion d'éthique, c'est surtout la remise en cause d'un système économique a bout de souffle qui pose problème. On ne peut plus durablement considérer de la même façon l'argent créé (ou investi) par *Meta* (déjà plusieurs milliards dans le métaverse !) autour de la donnée et celui péniblement gagné par un paysan de la Creuse avec son champ de soja. S'il doit y avoir une éthique de la donnée, c'est pour dépasser les approches purement réglementaires induites par une économie au bord de la chute et reprendre l'acculturation par là où elle aurait du commencer : apprendre à reconnaître / analyser / utiliser / partager la donnée à tous les acteurs de notre communauté / société pour organiser une concertation autour d'un vivre ensemble durablement. En un mot, (re)mettre chacun en maîtrise de ce qu'il génère.

Entrepreneur 2 (09/02/22)

En première question, j'aimerais connaître votre définition de la définition de la *data literacy* et si vous employez ce terme.

Alors dans notre milieu de data journalistes, on est très « frenglish » en permanence parce qu'on est dans un milieu qui baigne énormément dans le monde anglo-saxon que ce soit pour les outils, pour les inspirations, parce que c'est vrai que le data journalisme est plutôt né aux États-Unis et Angleterre donc c'est vrai que je continue de dire *data literacy* et j'ai bien du mal à sortir de ce terme. Après plus généralement quand j'en parle autour de moi je vais plus dire la culture des données. Mais je vois bien qu'en France, tout d'un coup, il y a un problème de compréhension. Culture des données, il y a un côté où tout d'un coup on va parler des maths, on va parler de la science. Quand on revient au terme « données » je vois bien qu'en France c'est un peu plus mou en fait, et que c'est pour ça que je m'accroche un peu à « data literacy » sans avoir trop d'équivalents français qui ne me satisfassent parce que je sens bien que quand je dis « data literacy », je projette autre chose que « culture des données » ou la façon dont « culture des données » est réceptionnée en France. Je ne sais pas trop comment m'en sortir mais du coup je m'accroche à ce terme anglais parce que je me sens plus à l'aise. Je travaille dans un milieu qui est plus anglophone mais ce n'est pas très satisfaisant.

Si on devait définir en une phrase la *data literacy*, comment le feriez-vous ?

J'en ai une vision assez citoyenne. Je dirais que c'est une éducation civique à la donnée. A la donnée qui nous entoure, à la donnée que l'on produit en tant qu'être humain dans nos interactions sociales, dans nos interactions numériques, dans nos interactions avec la ville, avec notre environnement. Donc c'est tout ce monde invisible de la donnée qui est né il y a maintenant bien plus de dix ans [passage retiré] qui forme une couche pour le citoyen qui a peu ou mal connaissance. La *data literacy* c'est l'accès à cette information qui est aujourd'hui pour moi une lacune forte en France : c'est-à-dire qu'en tant que citoyen, on va me parler beaucoup de juridique, de politique, de législatif mais on ne va pas forcément me parler de la data. C'est un objet virtuel, c'est difficile à conceptualiser, il n'est pas très concret, et puis en plus c'est du domaine que souvent je vais mettre en-dehors de la politique parce qu'en fait pour beaucoup c'est plus de l'ordre du *marketing*, c'est plus quelque chose de l'ordre de l'expert, de l'ingénieur, mais pas de moi en tant que citoyen. Alors que pour moi c'est le cœur de nos

civilisations à venir. Enfin : actuelles ET à venir. Et qu'en tant que citoyen c'est une culture au même titre que connaître le fonctionnement politique de mon pays, que connaître mes grands classiques de littérature, que connaître l'histoire de mon pays, la géopolitique, je le mets au même titre. Mais aujourd'hui c'est vu comme l'objet mathématique, l'objet scientifique, l'objet des experts, l'objet dont d'ailleurs on fait tout pour m'exclure et pour ne pas m'expliquer : en m'expliquant que c'est une boîte noire, c'est très compliqué ! Je ne vais vraiment pas comprendre donc il ne faut vraiment pas m'expliquer. C'est un lieu de connaissances dont sont exclus les citoyens et dont il faut que l'on se ré-empare très vite.

Juste pour revenir sur la « nature » de la *data literacy*: vous avez parlé de culture. Est-ce que pour vous, si l'on tentait de la définir, il s'agit plus d'une culture, un ensemble de pratiques et de techniques, un état d'esprit, tout cela à la fois ? Est-ce que l'on a une manière globale de la définir ?

Oui, je serai très normand(e) et je dirais que c'est un peu tout ça à la fois parce qu'on ne peut pas se contenter de la data uniquement dans son aspect technique. L'aspect technique ce sont des outils, je peux les maîtriser : finalement, c'est une compétence technique. Mais si je me contente de ça, je ne comprends pas l'aspect politique, tout simplement. Ne serait-ce que des outils : qui a créé ces outils pour enregistrer les données ? Comment sont fait les calculs ? Si on ne comprend pas que la donnée est un objet politique au sens où elle est faite par des humains, enregistrée par des humains, calculée par des humains, et donc biaisée par des humains en fait, on ne comprend pas que la data n'est pas neutre. A partir du moment où l'on a compris ça, on ne la voit plus comme cet objet qui est juste en face de nous avec notre thermomètre : tout d'un coup, elle prend une autre dimension, donc on ne peut pas se contenter de l'aspect juste technique. Il est important. Il n'est peut être pas premier, en tout cas dans ma conception très citoyenne des choses. L'idée n'est pas d'outiller techniquement les gens. C'est d'ailleurs de les outiller intellectuellement en fait : de les sensibiliser, de savoir qu'il y en partout, qu'on est émetteur de données. Je reviens à cette histoire de couches : j'ai l'impression de voir des humains dans la rue et qu'ils sont entourés de données et qu'ils ne les voient pas, alors qu'elles sont structurantes de toute leur vie aujourd'hui quand même. Donc j'aurais du mal à distinguer. C'est culture / connaissances, c'est apprentissage, c'est outil, c'est état d'esprit. C'est pour ça que je parle d'éducation civique parce que quand j'apprends à l'école comment fonctionne mon pays, après on ne me demande pas de devenir législateur ou spécialiste juridique. En fait, c'est de comprendre le contexte dans lequel je m'inscris moi en tant que citoyen. C'est une compréhension un peu macro pour dézoomer de sa vie perso, de se dire qu'en fait on est dans une société, dans une collectivité. Et dans les données, il y a un peu de ça aussi parce que si les opérateurs ne récupéraient que ma donnée, ils n'en feraient pas grand chose. Tout l'intérêt c'est de récupérer ma donnée, celle de mon voisin, de mon immeuble, de ma ville, de mon pays et ce sont ces millions de données qui génèrent ce qui permet de classer dans des catégories de consommateurs. En fait, c'est ma petite émission de données collectivisée qui va créer ce truc macro. Donc c'est une culture un peu globale.

Justement, la question qui se pose par rapport à la *data literacy*, c'est celle de sa place dans « l'écosystème » des données : on parle de gouvernance de données, de *data science*... la question que je me pose est : où situe-t-on la *data literacy* ? Est-ce la même chose ? Un engrenage supplémentaire ?

Alors *data science* et gouvernance des données je ne les mets pas dans les mêmes cases. Déjà pour gouvernance des données, il y a quelque chose de l'ordre de se réapproprier les données au niveau individuel. Et après, aussi, au niveau, plus géopolitique et international : comment les entités publiques, les États, régulent des données qui sont aujourd'hui aux mains ultra majoritairement d'opérateurs privés. Donc pour moi la gouvernance, il y a quelque chose de l'ordre de l'ordonnancement un peu du monde des données qui soit au niveau individuel, d'avoir ma maîtrise de mes données à moi, de comprendre de ce que j'émets, de ce que je peux maîtriser, de ce que je peux récupérer, de ce que je peux limiter, de ce que je peux bloquer, etc. C'est un peu ma gouvernance à moi. Et la gouvernance plus géopolitique au niveau, d'une ville par exemple, donc bien sûr ville puis pays puis niveau international. Alors que la *data science*, je la mettrais plus dans cette case des outils dont on parlait, c'est-à-dire des compétences techniques qui permettent d'explorer des énormes masses de données via des outils, via des langages informatiques, des langages de calcul donc je ne les mets pas aux mêmes endroits. D'où coup votre question c'était de comment ?

De connaître les rapports qu'entretient la *data literacy* avec cet écosystème.

Pour moi, c'est encore une méta enveloppe autour de ces deux choses-là mais aussi d'autres choses. Je trouve qu'un citoyen du XXIe siècle, il ne devrait pas être *data scientist* : il devrait juste savoir ce que c'est ce métier, à quoi ça sert, etc. Il devrait être sensibilisé à sa gouvernance des données et à comment dans le monde on gère des données. Mais effectivement il devrait être aussi un producteur averti et engagé de données. On a l'impression que la donnée je la vois comme quelque chose, pas de négatif mais en tout cas de quelque chose qui oppresse, quelque chose qu'on nous vole, etc. Alors que la donnée, à partir du moment où on s'en est réemparé pour moi c'est un axe de liberté. Donc la *data literacy*, c'est une culture qui rend plus libre parce que quand je suis conscient de ce que je fais, ce que j'émets, on peut aussi être producteur citoyen de données, c'est-à-dire qu'on peut être acteur de données aussi. Donc c'est vrai que la *data literacy* c'est plus toute cette grande méta-culture qui permet d'être plus libre. Parce qu'un citoyen averti en vaut deux ! C'est plus un outil de liberté.

Vous répondez déjà aux questions que je comptais vous poser. La première est : quelle est l'importance et la nécessité aujourd'hui de la *data literacy* comparée à d'autres domaines comme la technicité informationnelle, la technicité informatique, etc ?

Je dirais qu'à chaque niveau, c'est d'offrir cette transparence pour qu'on puisse, entre guillemets « travailler, » le mot n'est pas juste, collaborer entre citoyens et politiques. Si je prends au niveau d'une ville : pour moi, une ville, si elle avait un engagement de *data literacy*, cela serait bien sûr dans ses actions qu'elle fait aujourd'hui. C'est assez remarquable en France. En tout cas de plus en plus remarquable ce travail sur l'*Open Data* : je suis une collectivité, une entité publique, une commune, une région, peu importe, je produis de la donnée. Avant je la gardais. Maintenant je la mets à disposition de tous sur des portails *open data* : c'est une forme de *data literacy*. Mais la mise à disposition ne vaut pas littératie. C'est juste que je mets à disposition mais le citoyen, ma mère, ne se lève pas en allant télécharger des tableurs *Excel* sur l'*Open data* de sa ville. Donc justement la *data literacy* c'est avec ce travail de narration, ce travail de rendre ces données plus servicielles, ce travail pour expliquer comment j'ai produit ces données, pourquoi je produis ces données. C'est en fait cette couche d'information pour comprendre le contexte en fait, pour utiliser ces données en conscience. Donc au niveau d'une vie ça pourrait être ça.

Au niveau d'une entreprise, c'est que chacun dans l'entreprise puisse comprendre les données qui sont enregistrées sur chacun des salariés : qu'est-ce qu'il peut faire avec ? est-ce qu'il peut s'opposer à certaines des données qui sont enregistrées ou pas ? Et puis, qu'est-ce l'entreprise dans laquelle il travaille et enregistre par ailleurs. Que sont toutes ces activités avec une vraie transparence ? C'est permettre l'appropriation par tous des méthodes d'enregistrement, des méthodes de calcul, et puis de ce que l'on fait de ces données. Mais c'est peut être utopique. Mais c'est se dire que les données ne peuvent être une boîte noire. Du moins, cela ne sera qu'un élément de servitude, et potentiellement, le mot est un peu fort, d'oppression. Si on n'a pas connaissance de ces « data » qui sont enregistrées sur nous, qui sont utilisées, et je dirais à toutes les strates, qu'elles soient politiques, entrepreneuriales, même associatives. On ne parle pas du monde associatif mais il enregistre aussi plein de choses sur nous. Si on donne à une ONG, on a un certain nombre d'informations qui sont enregistrées sur nous : des fichiers qui sont mutualisés entre les ONG pour cibler, entre guillemets, les « bons » donateurs, etc. Cela serait une position un peu globale que tout opérateur qui utilise les données en parle de façon explicite sur son site Internet, en interne, etc. Qu'on ne soit pas juste avec de la livraison de fichiers ou de la totale boîte noire. Donc je dirais que c'est un état d'esprit global, on revient sur votre notion d'état d'esprit en fait.

Vous y avez un peu répondu en début d'entretien : pour vous cette nécessité de s'approprier les données, la *data literacy*, quand est-ce que vous la situeriez ? Vous avez dit « peut-être dix ans » mais est-ce plus ou moins que cela ?

J'ai dit plus de dix ans car c'est mon point de vue du data journalisme qui, du coup, n'est qu'une partie de cet écosystème de la data, qui n'est qu'une pratique. Après je pense que je le situe là pour le data journalisme parce dans le fait d'enquêter à partir de données, de raconter des histoires à partir de données, le data journalisme a eu besoin de contraindre certains opérateurs de données à mieux raconter d'où venaient ces données. Parce que nous, en tant que data journalistes, moi, éthiquement, je ne peux pas utiliser des données dont je ne peux attester l'origine. Je ne comprends pas comment elles ont été calculées, je ne comprends pas quels outils ont été utilisées. Donc le data journalisme a contraint un certain nombre d'opérateurs, publics et privés, à mieux raconter leurs pratiques de données parce que nous on en avait besoin pour les utiliser. Et donc, peut être, que l'on a contribué à une certaine *data literacy* ne serait-ce que pour ces opérateurs. Et on a très vite vu après ceux qui acceptaient de communiquer ou non sur leurs méthodologies, leurs façons de faire, etc. Donc c'est vrai que je le situe d'un point de vue très data journalisme mais je n'ai pas forcément connaissance finalement d'un mouvement qui serait peut être plus profond et plus ancien et qui serait nourri par d'autres choses. C'est vrai que là c'est que de ma petite lucarne de data journalisme qui a une dizaine d'années si on regarde les premières enquêtes *New York Times* et *Guardian* qui étaient un peu les deux journaux pionniers. C'est à peu près ça : leur rubrique data c'était il y a à peu près dix ans. Mais effectivement, ça ne témoigne pas forcément d'un mouvement qui est peut être beaucoup plus ample et beaucoup plus ancien. C'est que de mon point de vue professionnel. C'est pour ça que je dis dix ans par notre pratique et puis par le début de l'*Open data* : le portail data.gouv.fr a fêté ses dix ans. Donc, du côté français, il y a aussi une pratique de l'*Open data* qui a contribué à cette *data literacy*, ou en tout cas qui est un des éléments de cette *data literacy* d'il y a dix ans aussi. Donc je mets ça là mais vous aurez certainement peut-être d'autres sons de cloche qui le positionneront à un autre moment de l'histoire.

Sur la donnée elle-même, il existe de nombreuses difficultés d'un point de vue conceptuel, du point de vue de l'appropriation, et je souhaiterais revenir à une question centrale : comment définissez-vous la donnée ? Qu'est-ce qu'une donnée ?

Pour moi une donnée, c'est un ensemble structuré d'informations. Sa structuration est souvent positionnée dans un outil qu'on appelle le tableur. Et cette structuration en lignes et en colonnes d'une information, entre guillemets, « propre », c'est-à-dire : dans une cellule, j'ai une information, qu'elle soit textuelle ou chiffrée, qui me permet d'appliquer un certain nombre de formules de calcul et de sortir de l'information une information autre, une information synthétique, de cet ensemble de lignes et de colonnes. Donc pour moi c'est basiquement ça : une information structurée et utilisable immédiatement pour produire une nouvelle couche d'informations synthétiques.

On a beaucoup parlé des enjeux actuels de la donnée : la question que je me pose est : comment la *data literacy* permet de répondre à ces enjeux et est-ce qu'elle permet justement d'y répondre ?

Je ne pourrais pas dire qu'elle permet d'y répondre car je trouve, en tout cas en France, qu'on en est hyper loin. J'ai l'impression, enfin dans mon petit milieu j'entends, que les gens qui utilisent cette expression, je vois peu de choses concrètes. Donc, est-ce qu'elle permettra de répondre à tous les enjeux, je n'en sais rien. Je pense que ce n'est pas un outil justement, j'ai bien expliqué que ce n'était pas un outil. C'est un état d'esprit, des compétences, quelque chose de méta qui doit se répandre dans la société, se répandre dans nos pratiques, dans l'école, dans le monde associatif, dans les entreprises, partout. Maintenant, j'ai une difficulté à dire si dans un pays c'est quelque chose qui est mis un peu plus en avant. Je sais que dans les pays anglo-saxons il y en a beaucoup plus à l'école notamment. Maintenant, est-ce que ça produit des effets ? Sur les enjeux qui m'intéressent, notamment l'appropriation de ces données personnelles, je ne vois pas de pays qui se distinguent en termes de citoyens qui ont une connaissance et une action suffisante pour se réapproprier leurs données sinon il y aurait des manifs dans la rue pour obtenir que les GAFA nous rendent plus accessibles nos données. Je n'en vois pas. Je ne vois pas de manifs de gens qui veulent se réapproprier les données donc visiblement on en est soit au début, soit ça n'existera jamais. Mais bon, je ne vois pas d'autres moyens pour une massification de cette sensibilisation. C'est-à-dire que moi je vis dans un milieu, chacun en a fait l'expérience : on a fait nos demandes de données auprès des GAFA pour les récupérer. Mais on est une ultra-minorité à essayer de comprendre ce que l'on fait avec nos données parce que nous c'est notre quotidien. Maintenant, comment massifier cette prise de conscience, c'est tout l'enjeu de mon métier. C'est d'essayer de créer notamment des expériences interactives ou de la sensibilisation visuelle pour que les gens se rendent compte que ces données sont partout mais qu'en même temps s'ils ne sont plus là et qu'ils ne donnent plus leurs data, il n' y a plus de GAFA non plus. Donc en fait c'est tout un système. Je dis ça dans les écoles de journalisme tout le temps : quand vous utilisez une appli, quand c'est gratuit c'est vous le produit. Souvent cette phrase les marque en disant « ah bon? ». Et alors, bien sûr, on a l'impression qu'ils sont benêts mais ce n'est pas du tout ça : ils n'ont pas fait tous les liens en fait ! Il y a une utilisation, une gratuité, un service génial et donc j'y vais. Et bien sûr je ne me rends pas compte de tout ce qu'il y a innervé derrière tout un système de récupération de données. Donc, je pense que c'est un travail de très longue haleine. Je pense effectivement que cette massification doit passer par l'école parce que je ne vois pas d'autres endroits où nous passons tous à un moment de notre vie. Et sans en faire « un cours de plus », je le mettrais au même titre que l'éducation civique. A un moment, je ne comprends pas comment je peux

apprendre comment marche mon Assemblée nationale et ne pas avoir cette sensibilisation à tout ce que l'on va faire de ma vie via les données. Et là je parle beaucoup des GAFA, il y a un côté un peu France - États-Unis qui est un peu ridicule.

Mais ce sont aussi les données qui font qu' aujourd'hui dans *Parcoursup* ma vie va se passer comme ça, que bientôt je vais avoir le dossier médical personnalisé et donc mes données médicales. On voit bien que de toute façon on va vers une numérisation même de notre vie de citoyen, notre relation aujourd'hui avec tous les services publics elle se fait de plus en plus via le numérique, via des données. Ne parlons pas du Covid où l'ensemble des informations sont allées chez *Doctolib*. Bien sûr on va avoir un QR code sur notre prochaine carte d'identité : quand je dis tout ça, je ne suis pas contre, c'est juste que je vois un mouvement d'accélération dans tout ça qui apportent mille choses magnifiques : beaucoup plus de facilités pour des services publics qui ne pouvaient pas être physiquement présents. Il y a plein de positif dans tout ça. Mais il y a ce mouvement que l'on nous présente comme naturel et comme positif. A un moment il faudrait me décrypter ce qu'il va se passer !

Alors on peut me dire que les journalistes ils font aussi leur job : il y a des articles sur ce qu'est le dossier médical personnalisé mais pour moi on ne peut pas se contenter d'informations de presse à un moment où les médias, où le niveau de confiance est extrêmement bas. Il y a une crise de la confiance terrible actuellement donc je me dis qu'il faut que cela passe ailleurs. Et en même temps à l'école on m'explique ce que l'on va faire de mes données de santé, de mes données de vote, de mes données médicales, de mes données de mobilité quand je me balade dans la vie. Cela serait me raconter le monde dans lequel je vis tout simplement. Et sans un discours blanc ou noir. Il n'y a pas de méchants et de gentils dans cette histoire : ce que j'aime dans le mot "littératie", c'est ce côté d'une culture que l'on doit partager. Et une culture doit être éveillée, éveillée aux data. Si après être éveillé, j'ai envie de tout donner aux GAFA, c'est ma liberté. Mais c'est juste que je le sais ! En fait, c'est de faire des choses en conscience. Je pense qu'il y a une forme d'inconscience qui est voulue par beaucoup d'opérateurs. C'est de se réveiller à ce monde des données pour mieux agir en conscience.

Justement, pour rebondir sur ce point, vu que l'on est au cœur de ces enjeux par rapport aux données : existe-t-il des « types » de données en particulier qui sont au cœur de ces enjeux ? Vous avez cité l'open data, les données personnelles...

Je ne fais pas vraiment de hiérarchie parce qu'en fait on se rend compte qu'une donnée qui peut être récupérée d'une façon qui a l'air hyper ouverte, transparente, peut servir à plein d'autres choses. Donc je ne ferais pas de classification. Je pense qu'il y a un grand nombre de données que l'on ne perçoit pas. Je pense notamment à tout ce que l'on lit aujourd'hui avec par exemple ce terme de la *smart city*. Je pense que dans la ville aujourd'hui...enfin, la ville ou la campagne mais majoritairement dans la ville parce que c'est là qu'il y a plus de points de contact avec des panneaux publicitaires ou autres. Aujourd'hui je pense qu'il y a peu de gens qui savent que les panneaux publicitaires, lorsque l'on passe devant, on est enregistré. Cela leur permet de savoir le nombre de gens qui passent dans cette rue, à cette heure, etc. Alors eux c'est pour dire combien de personnes ont vu le panneau, ou du moins ont potentiellement vu le panneau parce qu'ils sont passés devant, pour vendre des opérations marketing en disant « voilà, on a eu un taux de passage de tant » mais ça enregistre des mouvements dans la ville qui sont au-delà de la simple publicité. Donc je pense que la *smart city*, qu'il y

a peu de conscience de ça. J'ai du mal à faire de hiérarchie entre tout ce monde numérique. Je parle beaucoup des GAFA, mais en fait toutes nos interactions avec des applis quelles qu'elles soient, tout le fait qu'on accepte effectivement de laisser des enregistrements que se soit ma musique préférée sur *Spotify*, tous les services sur lesquels on accepte d'être traqué.

Après, les données, il y a aussi tout ce beau mouvement de la France participative car c'est assez incroyable. Il y a bien sûr je compte les papillons et les coquelicots pendant un week-end en cliquant sur l'appli donc ça effectivement ce sont des choses qui se sont développées, où il y a vraiment des passionnés. Ce sont des milliers et des milliers de gens qui participent à ça et c'est là qu'ils sont producteurs de données mais volontaires [passage manquant]. J'en vois chez les anglo-saxons mais pas encore en France où aujourd'hui la communauté scientifique s'appuie énormément sur les citoyens pour les aider à faire des sortes d'enquêtes scientifiques, en tout cas à faire des actions qui, parce qu'elles sont et seront massifiées par les milliers de gens qui vont participer, vont permettre à la science d'aller dix fois plus vite. Et donc là c'est de la production de données qui rejoint un peu plus votre ancien métier de documentaliste.

L'autre jour, j'ai vu une opération vraiment géniale, au pôle Nord je crois, ils ont un centre scientifique et ils ont mis une caméra qui fait des photos toutes les trente secondes de ce qui est en face. C'est vraiment pour savoir s'il y a des ours qui passent. Donc, à raison de trente secondes pendant des mois, ce sont des millions de photos où on voit juste du blanc. Les gens participent et dès qu'ils voient un objet ou du moins un être qui n'est pas de la neige, ils disent « là, il y a quelque chose » et cela permet de gagner des milliers d'heures aux scientifiques, de repérer les moments où il y a un animal. Donc c'est une action qui est très proche de ce que l'on appelle un peu « les esclaves du web » qui font des choses un peu débiles, parce que là on est vraiment dans de la micro-action de web, mais à la fin on produit une base de données qui fait accélérer la science. Donc il y a aussi cette donnée dont je suis producteur conscient et participant à la science et on pourrait imaginer cela dans d'autres domaines.

Je suis étonné(e) qu'il n'y ait pas d'autres domaines qui s'appuient sur cette intelligence collective que sont les citoyens où on pourrait produire de l'intelligence à partir de photos, de documents anciens, même si ça se fait aussi effectivement chez les anglo-saxons avec la numérisation de documents très anciens qui ne peuvent pas être OCRisés. On a des citoyens qui lisent des documents hyper anciens. L'autre jour j'ai vu des relevés météo : la météo anglaise a maintenant en tables de données tous ses enregistrements météo depuis 1800 à aujourd'hui. Et toute la période 1800 - 1940/1950 cela a été entièrement fait par des citoyens qui ont relu les chiffres et qui les ont retapé car c'étaient des chiffres à la plume. Donc ça c'est beau aussi, ça c'est une forme de data dont on parle moins parce qu'on est obsédés par la data qu'on nous aspire mais il y a aussi la data que je produis. Je trouve que c'est pour ça qu'on ne peut pas être binaire dans cette *data literacy*. Je pense intimement que c'est aussi un objet de libération, c'est aussi un objet de connaissances, un objet d'amélioration de la connaissance du monde. Lais pour ça il faut massifier. Je pense que si on comprend mieux la donnée, on est plus apte ou enclin à participer à ce genre de choses car en fait on comprend à quoi ça sert aussi. Tout d'un coup si on me dit que pendant des heures tu vas remplir des tableurs avec des chiffres à la plume de paon, puis qu'on se dit qu'avec ces bases de données on va mieux comprendre le changement climatique, tout d'un coup tout prend du sens. Enfin, un autre sens.

Par rapport à la donnée elle-même : qu'est-ce que l'arrivée des données, cet usage des données, a changé dans vos pratiques professionnelles ?

Vous voulez dire la donnée dans mon métier ? [j'approuve de la tête]. Alors moi je peux dire que ça a tout changé ! C'est comme-si j'avais eu une première vie professionnelle : je suis journaliste depuis vingt ans. Quand je suis sorti(e) du *CFJ*, l'école de journalisme, en 2000, ce n'est pas si loin mais c'est loin quand même, je n'avais pas ouvert un tableur de ma vie. Jamais au lycée, jamais en école de journalisme : la data était inexistante. Le journaliste français jusqu'à il y a huit / neuf ans, n'avait aucune acculturation à la donnée. Aucune.

Si on aimait les maths, on allait en école d'ingénieurs mais on ne faisait pas du journalisme. On avait un journalisme très « Albert Londres », la plume dans la plaie, chroniqueur, reportages au long cours, portraits, interviews, etc. Donc j'ai appris tous les modes d'expressions journalistiques et dès qu'il y avait une donnée, c'était un non-sujet. Ce qui est quand même hallucinant. On peut se dire « oui,il y a vingt ans, il y avait moins de données ». Mais néanmoins, et c'est qui m'a vraiment horrifié en fait, dans ma pratique j'ai travaillé à [nom de journal]. Pendant assez longtemps, je me suis retrouvé(e) aux services Sport et Économie pour des raisons diverses et variées : ce sont deux choses que j'aime et ce sont deux choses où il n' y a que des chiffres. Quand je me retrouvais avec un communiqué de presse avec plein de chiffres et que j'étais là à faire un copier-coller parce que j'étais incapable d'interroger un chiffre, tout d'un coup je me suis dit « il y a un truc qui va pas ! ». Pourquoi on ne m'a pas appris à interroger un chiffre ? et c'est là que j'ai commencé à creuser ça de mon côté mais on me prenait un peu pour un(e) dingo. Il y avait un côté « non mais les chiffres c'est la vérité ". C'est une vision hyper restrictive que je trouve horrifiante du métier de journaliste. Les chiffres « cela ne s'interroge pas ». Et bien, c'est un vrai problème en fait.

Pendant dix ans j'ai eu une pratique, entre guillemets, assez « classique », même si je me posais plein de questions et que du coup je me suis formé(e) tout(e) seul(e) parce qu'il y avait pas grand chose. J'en parlais avec d'autres mais c'est un truc dont on ne parlait pas du tout. Et puis il y a dix ans j'ai créé l'agence pour ne faire que ça car je me suis rendu(e) compte que je ne pouvais pas, enfin après c'est moi et ma façon d'être, je suis entier(e), c'est tout ou rien, mais je me suis dit « bon je ne peux pas faire juste ma propre acculturation à la donnée ». A un moment je me suis rendu(e) compte qu'il y avait des milliers de choses à raconter. Et puis il y a eu l'*open data*. Je voyais que c'était un mouvement grandissant et que si j'avais envie d'en faire partie, j'avais envie de m'y consacrer complètement.

Donc j'ai créé l'agence et c'est vrai depuis j'ai été très actif(ve) dans les écoles de journalisme avec d'autres qui sont venus petit à petit. Maintenant, il n' y a aucun étudiant sortant des quatorze écoles reconnues de journalisme qui n'a pas eu au minimum quinze jours de data dans sa scolarité. Data c'est : tableurs, scrappeurs de données. Vraiment, il a tous les basiques. Après il en fait ou pas mais aujourd'hui c'est considéré comme une compétence de tout journaliste qui sort d'école. Au même titre que savoir enregistrer, faire une interview, faire du son, faire de la télé : c'est une compétence qui est mise au même niveau. Mais ça c'est une évolution de ces huit dernières années.

Donc ce que ça a changé, c'est que j'ai eu l'impression d'avoir deux vies professionnelles. J'ai eu une vie avant et j'ai eu une vie après. Je fais partie des « vieux/vieilles » du data journalisme. On a un groupe de data journalistes

francophones, on est une centaine, je pense que je suis la/le plus vieille/vieux. Mais aujourd'hui ce que je trouve très bon pour notre métier de journaliste, c'est que cela a beaucoup renouvelé les profils parce qu'avant peut être que quelqu'un qui aimait les maths ou qui était très scientifique, il n'osait pas aller dans le journalisme parce qu'il se disait « ah si je peux être journaliste à la rubrique science » mais ce n'était pas forcément ça qui l'intéressait.

Et aujourd'hui on a une génération de jeunes journalistes qui arrivent de filières d'ingénieurs. On a aussi des *data scientists* qui se reconvertissent dans le journalisme donc on a une vraie diversité de cerveaux. On n'a pas juste des sciences po, histoire, droit, les trois filières qui monopolisaient le monde journalistique et ça c'est très positif pour notre métier. C'est d'avoir effectivement tout cette génération maintenant qui sait coder et qui sait écrire, qui sait faire de la *data science* et qui sait enquêter. Donc oui, pour moi, pour mon métier, c'est un vrai bouleversement. J'ai même créé une boîte alors que jamais dans ma vie je comptais le faire. Je me destinais à être journaliste dans une rédaction, jamais je pensais créer une boîte dans ma vie. Donc du coup je me retrouve chef(fe) d'entreprise ! Donc pour moi cela a tout changé. Mais pour le journalisme en général, je peux parler du data journalisme francophone car je le connais bien, c'est de la diversité d'esprit, c'est de la collaboration aussi (je pense que je finirais là-dessus car je parle beaucoup).

Peut être ce que je n'ai pas connu dans d'autres rubriques de journaux, les journalistes judiciaires, les journalistes politiques, c'est un aspect un petit peu sectaire de notre métier, et entre journalistes judiciaires, entre journalistes politiques... quand je dis entre, c'est entre journaux. On ne peut pas dire qu'il y a toujours une entraide. C'est un métier où il y a une hyper concurrence entre les titres. Et ce que j'apprécie énormément, et pour moi ça fait partie de ce côté de valeurs que je trouve que la data porte quand on y est éveillé, c'est la collaboration.

Aujourd'hui je pense qu'il y a peu de journaux qui le savent mais les data journalistes français s'entraident tous les jours. C'est-à-dire quand on a des galères sur un fichier (parce que les data journalistes travaillent sur les mêmes sujets), y a l'actualité, y a le covid, y a en ce moment la présidentielle : tout le monde galère sur des fichiers plus ou moins propres, plus ou moins disponibles, on a un groupe sur *Signal* et tout le monde s'entraide en permanence. Et personne ne sait que le *Figaro* travaille avec *France Info* qui travaille avec d'autres, et ça je trouve qu'il y a une sorte de solidarité très forte. Il y a une collaboration. Peut être que moi je l'associe à la data et aux valeurs positives qu'elle peut avoir quand elle est conscientisée, mais ça je ne l'ai jamais vu dans d'autres secteurs. Donc voilà, je trouve que le monde des data journalistes est un beau milieu. Malgré la concurrence entre les journaux, on s'entraide en permanence et ça je trouve que je ne l'aurais peut être pas connu si j'avais continué dans un journalisme plus classique.

Toujours en lien avec vos pratiques : est-ce que le cadre législatif, je pense à la loi Lemaire, le RGPD... Est-ce que toutes ces réglementations ont un impact sur vos pratiques depuis leur arrivée ?

Alors oui, dans des domaines très différents parce que la loi Lemaire est pour nous un point d'appui important pour porter un message auprès d'institutions qui ne libèrent pas complètement leurs données. Dire que la loi n'impose pas l'open data, que la loi le prône et qu'on aurait aimé que dans les décrets d'application il y ait une obligation d'open data, on n'a pas pu aller jusque-là, mais c'est quand même un point d'appui important sur

plein d'aspects. La loi Lemaire c'est un texte et des articles et un esprit aussi sur lesquels on peut s'appuyer dans nos négociations pour obtenir les données. Nous les journalistes on veut toujours les données qui ne sont pas en *open data*.

Après je dois dire que le plus efficace et c'est étonnant car c'est une loi beaucoup plus ancienne, je pense que c'est 1974, c'est la création de la *CADA*. Nous aujourd'hui pour obtenir des documents, de data notamment, on fait des demandes à la *CADA* toutes les semaines. Il y a même un site qui a été créé par des data journalistes pour faire de la demande un peu plus automatisée parce qu'autrement c'est hyper pénible. Ça s'appelle *Madada*, *Macada*, je ne sais plus...et en fait on a un site qui a été créé par des data journalistes et on fait des demandes récurrentes. On s'appuie sur une commission qui n'est pas si récente mais qui nous permet d'accéder en tant que citoyen aux documents administratifs. Donc on fait beaucoup de demandes avec la *CADA*. Pour le *RGPD*, en tout cas pour ma pratique, ça n'a pas vraiment de changement car on utilise très peu, voire pas de données personnelles dans nos enquêtes. Donc à la limite je le vois plus comme quelque chose de protecteur et qui a permis d'avoir un certain nombre d'articles et d'explications. Parce qu'en fait, via le RGPD, on découvre qu'on essaye de réguler, de protéger les citoyens qui ne se rendaient pas compte que leurs données étaient utilisées un peu n'importe comment. C'est plus un cadre régulateur et contraignant pour un certain nombre d'opérateurs.

On revient plus pour moi à cette gouvernance des données dont on parlait tout à l'heure, c'est-à-dire un moment où les entités publiques essayent de reprendre la main sur ce qui peut être fait, pas fait, etc., ce qui est très drôle car hier, il me semble que c'était hier, Mark Zuckerberg disait qu'il va stopper *Facebook* et *Instagram* en Europe parce qu'il ne peut pas faire ce qu'il veut avec les données. Pauvre Mark, c'est vraiment très triste. Mais voilà, ces sujets-là, quand ils émergent dans l'actualité, j'espère qu'un certain nombre de gens qui découvrent les choses, qu'ils s'informent ou qu'ils s'alimentent. Après j'aimerais que, par exemple s'il y avait un cours de data ou de *data literacy* à l'école, il y ait des actus toutes les semaines. Ça serait génial : pendant une heure « pourquoi Mark ne veut plus » on peut faire un cours génial. Pourquoi Mark veut fermer *Insta* et *Facebook* de vos téléphones. Expliquer ça aux gamins. Et expliquer par un peu d'absurde : « il veut les fermer car il peut pas faire grand chose avec vos données » « ah bon il fait des choses avec nos données ? ». En fait, l'actualité, le RGPD, cela devrait être expliqué à l'école à l'image d'une loi. C'est pour ça que je trouve que l'école peut être un terreau vraiment génial parce qu'en fait toutes les semaines on apprend des trucs de dingue et que ça peut nourrir des moments de débats, des moments de compréhension. Justement ce ne sont pas des moments de cours. C'est plus de débats, d'acculturation et on revient à cette notion de culture. De sensibiliser que cette info-là ce n'est pas une info technique, ce n'est pas une info géopolitique, c'est une info qui te concerne toi. Donc de raccrocher toujours ce lien entre soi, producteur et émetteur de données. Et puis des décisions qui ont l'air très politiques, macro et qui ne sont nourries que par nous. Donc oui, chaque loi est une occasion d'acculturation. Après dans nos pratiques, elles influent à différents niveaux mais très peu pour les données personnelles.

De manière plutôt négative ? Positive ? Ni l'une ni l'autre ? Ou cela dépend des contextes ?

Oui, cela dépend vraiment des contextes. La loi Lemaire c'est très positif, même si on aimerait que ça aille plus loin et qu'il y ait une loi qui oblige à l'*open data* en France, qui contraigne à l'*open data*. Le *RGPD*, c'est plutôt

neutre en fait pour nous. Ça dépend vraiment des contextes. Aujourd'hui, nous ce qui est le plus attendu c'est effectivement qu'il y ait un fonctionnement de la *CADA* plus rapide. C'est une vieille commission d'une loi des années 70 : alors je ne vais pas dire que les gens de la commission ils sont hyper vieux, on a déjà le bonheur qu'elle existe, mais c'est vrai que cela mériterait d'être remis au goût du jour et qu'il y ait une suite à la loi Lemaire. On espérait que dans ce quinquennat, je veux dire la loi Lemaire c'est quand même l'ancienne mandature, là ça veut dire, depuis cinq ans en tout cas, on n'a pas eu en termes de données de grandes lois qui ont accéléré les choses pour la transparence, c'est ça qui est dommage peut être pour ce mandat.

Pour aborder la dernière partie de l'entretien : cette culture des données, cette *data literacy,* comment peut-on la mettre en place pour qu'elle puisse être partagée par un certain nombre de personnes et doit-elle concerner tout le monde ou certains profils en particulier ?

Pour la mise en place, c'est vrai que je mise beaucoup sur l'école comme je vous l'ai dit parce que pour moi c'est le lieu de la massification et qu'il faut qu'on arrête d'en faire un sujet d'expert ou d'un microcosme de data. Il faut que ça redevienne un sujet citoyen. Pour moi, les citoyens c'est l'école. Après bien sûr, il y a le terreau associatif, l'éducation populaire, il y a plein de ramifications au-delà de l'école. Mais l'école, cela reste aujourd'hui le creuset le plus massif. Pour moi oui, par l'école, cela doit concerner tout le monde. Je pense que cela peut permettre aussi d'avoir cette acculturation, peut être en fait de pas en faire, au contraire même, de l'éducation civique. Je suis allé(e) à l'école il y a longtemps, donc il y avait ce côté un peu très cours. Si je ne dis pas de bêtises, je crois que je n'en ai pas eu en primaire, peut être au collège et au lycée. Pour moi cette acculturation aux données je la vois dès le primaire en fait. C'est un accompagnement au cours de ma vie. Il y a largement de quoi raconter la donnée depuis la maternelle jusqu'à la terminale parce que ça brasse tous les sujets de ma vie, ça brasse des enjeux géopolitiques, économiques, etc.

Donc après il faut aller crescendo en fonction de son éveil au monde, mais dès la maternelle. Enfin, à la maternelle ils en font déjà un peu plus maintenant : on peut apprendre à faire du code avec des légos, il y a beaucoup de choses qui sont possibles. Et là j'englobe un côté un peu plus large de la data et ce monde du code. Parce que le code, c'est pareil : tout le monde ne va pas devenir codeur ou développeur. Pour moi cela rejoint le même enjeu : c'est comprendre ce qu'il y a derrière la machine. Aujourd'hui, on est tous des consommateurs de ces objets technologiques. J'ai de la chance, car je suis très vieille/vieux, je fais partie de cette génération qui pendant deux ans en France, dans les années 80, ont connu un grand plan informatique partout en France. En CM1, CM2, je m'en rappelle encore ça m'avait hyper marqué, on m'a appris avec un *DOS*, on apprenait à faire fonctionner un ordinateur où il fallait lancer des lignes de codes et on faisait des jeux très basiques. Mais on faisait partie pendant deux ans d'un plan qui visait à ce que tous les jeunes de France aient cette initiation au code, à un nouveau langage. Parce que vous êtes spécialiste de la littératie, c'est une langue, en fait, le code. Et moi ça m'avait hyper marqué. Et puis tout d'un coup ça s'est arrêté. Il n' y avait plus de fonds, on avait changé de gouvernement.

Et je pense que c'est un apprentissage hyper important de comprendre que là, derrière mon interface, là je vois par exemple de belles fenêtres, je vois des couleurs, etc., que derrière c'est un langage qui permet de produire ça. Et pour moi ça fait partie de cette réappropriation, un peu comme les données. Aujourd'hui je ris beaucoup avec cette génération Z, Y, enfin je ne sais pas comment la nommer, parce que moi je les vois en cours : mais je

n'ai jamais vu des gens aussi inaptes. Inaptes à l'informatique. Ils ne savent pas ouvrir un traitement de texte. Alors ils font plein de trucs sur leurs petits doigts sur leur petit écran. Ils montrent des vidéos sur 8cm² et ça c'est assez fort, mais ils ne comprennent plus du tout la logique de l'informatique. Et ça c'est un danger car du coup on est vraiment dans un truc de consommateur non averti. On produit de la donnée sans se poser de questions : on consomme de l'informatique, du numérique, du digital, mais sans plus rien comprendre en fait. Et voilà, un citoyen qui est en interaction constante avec ces objets numériques, pour moi il devrait comprendre ce qu'il y a derrière la machine. Qu'est-ce qui se passe quand je fais « Enter »? Et donc en cela je les rejoins, cela doit concerner tout le monde, à l'école principalement parce que c'est là où je vois vraiment le côté le plus. Et tout au long de la scolarité. A la fin, on aurait peut-être des personnes qui font des choix professionnels plus éclairés en sachant ce que l'on peut faire ou pas faire. C'est un espoir un peu utopique. Mais oui, je vois ça de façon assez massive.

Vous avez anticipé ma question, à savoir que doivent développer les personnes à l'issue d'un apprentissage de la donnée en termes de compétences, de techniques, d'état d'esprit...

Oui, vous avez vu que je suis assez normand(eà dans toutes mes réponses. En fait, je n'arrive pas à dissocier complètement : pour moi la technique, elle est au service d'un état d'esprit. Je n'ai pas l'état d'esprit, je ne vois pas pourquoi j'apprendrais la technique [coupure de son] faire ça, parce que ça va servir à me libérer de ça, parce que ça va m'éveiller à ça, etc. Donc je ne peux pas dissocier les deux. Si on avait quelque chose, une acculturation tout au long de l'école on aurait le temps de poser toutes ces briques et pas de faire, allez, pendant un an, une éducation à la data, c'est complètement débile. C'est comme-si on imposait tout d'un coup un cours dans un programme ! je pense que c'est vraiment au cours de la vie, ça devrait même ne pas s'arrêter après. Mais en tout cas, si on avait un terreau vous imaginez, je ne sais pas on entre à l'école à quatre ans maintenant, de mes quatre à mes dix-huit ans. Mais quatorze ans à essayer de comprendre ces données dans les différents aspects, philosophiques, politiques, techniques. Dans le monde qui vient, je ne comprends même pas que cela n'existe pas.

Par rapport à cela, quelle est selon vous la situation actuelle en termes de compétences des personnes : quel est le niveau général en culture des données ?

Alors, j'ai une vision très mal située parce que j'ai des data journalistes et des *data designers* donc ce sont des gens qui ont un très bon niveau. J'ai bien conscience qu'ils sont rares. Après, du coup, si je dézoome un peu et que je regarde peut être un peu plus les promotions de journalistes, puisque ce sont les jeunes journalistes que moi je forme le plus, qui n'ont pas tous vocation à devenir data journalistes, là oui je suis un peu surpris(e) parce qu'on est quand même sur du bac + 3, en tout cas dans les écoles où je suis. Quand je vous disais tout à l'heure « quand c'est gratuit, c'est vous le produit » et que j'en vois qui ont l'air de découvrir le truc, je suis un peu inquiet(e). On parle de gens qui ont fait des études supérieures et qui ont l'air de découvrir la vie. Oui, ça me fait un peu stresser. Donc oui c'est un peu ma seule référence. Après je ne vais pas parler de ma famille. Dans ma famille qui est très loin de la data, ils me prennent pour un(e) folle/fou en fait. Ils ne comprennent même pas ce que je leur raconte. Donc si je les prends comme référence de la société française, on est mal. Et si je vois les jeunes journalistes actuels avec un bac + 3, bac + 4, qui ont l'air parfois, pas tous, mais quand même,

sur des choses qui moi me paraissent évidentes, ils ont l'air de découvrir la Lune, je pense qu'on est très bas. En France, on est très très bas.

Et comment expliquez-vous ces difficultés actuelles ? Vous avez précisé que la donnée était un outil conceptuel difficile à définir par exemple.

Je pense qu'on présente mal les choses. Parce qu'en fait, quand on va parler data, les gens vont croire que c'est cours de maths. Je pense que tant qu'on arrivera pas à dissocier, à expliquer beaucoup plus globalement qu' on ne parle pas de maths, ce n'est pas un cours de maths ! Pour moi, c'est un cours de vie citoyenne. Et tant qu'on arrivera pas à présenter ça, à le propulser comme une culture générale, tout simplement car c'est de la culture générale, je pense que le côté matheux, le côté projeté dans le côté maths, ça bloque énormément de gens. Je pense que c'est un problème d'approche du sujet. Il me semble.

Nous terminons maintenant sur deux points. Le premier : avez-vous en tête un projet en *data literacy* qui a bien fonctionné et en quoi il peut être un modèle pour d'autres projets à venir ?

Alors je ne sais pas comment le classer : à chaque fois que je fais des interventions auprès d'élèves, je fais aussi de l'éducation aux médias dans des cours pour les plus petits, j'en parle beaucoup car c'est un peu mon prisme, ma spécialité. Après cela se fait plus sous forme : je vais prendre un sujet d'actualité comme on l'a fait tout à l'heure. Je vais prendre l'actualité de la semaine et on va en parler sous l'angle de la data. C'est plus sous forme de débats donc je dirais que ce n'est pas très formalisé. C'est plus de la sensibilisation. Cela mériterait d'être un peu plus formalisé peut être dans des cours d'éducation aux médias. Et après je vois que ça s'adresse à un public plus large de que ce que l'on a fait. Là on a lancé il y a un mois un site qui s'appelle datajournos.fr, ce collectif de data journalistes français, qui se réunit sur des plateformes un peu virtuelles, on a décidé de lui donner une existence pour repartager ce que nous on partage en interne. Donc c'est un site de ressources pour devenir plutôt data journalistes, donc je restreins pas mal. On essaye de mettre pas mal de ressources qui sensibilisent à la donnée pour ceux qui se posent des questions du type : est-ce que c'est technique ? on essaye de doser un peu. Vraiment des ressources qui sont spécifiques aux data journalistes pour apprendre un langage par xemple. Et puis on essaye de mettre des ressources un tout petit peu plus globales ou un peu plus grand public, on va dire, pour sensibiliser à la data et à notre métier.

Donc aujourd'hui j'ai cette incarnation là avec ce site et c'est plus effectivement après, dès que je peux, dans mes interventions en école, d'aborder ce sujet mais il n' y a pas de formalisation. Mais j'aimerais, maintenant que *datajournos* existe, ouvrir dessus un espace justement un peu plus de *data literacy*. C'est sur ce site que je vais essayer de développer un peu plus ça dès que le temps me le permettra. Aujourd'hui je n'ai pas tant d'exemples à proposer parce que j'ai l'impression de le vivre un peu chaque jour dans tous mes échanges.

Dernier point : quelles seraient vos recommandations pour une meilleure appropriation de la culture des données en France ?

Au-delà de l'école, qui est pour moi majeure, je sais que plus personne ne regarde la télé mais ça m'avait toujours marqué : sur la BBC, ils avaient, via Hans Rosling qui est un statisticien décédé maintenant et qui était

vraiment un maître de la statistique, il avait quand même une émission tous les samedis. En gros, sur la BBC, il y avait une heure de stats tous les samedis. Alors, des stats à la Hans Rosling : hyper drôles. Et en fait cet homme arrivait à faire de l'acculturation grand public à la donnée sur une chaîne nationale tous les weekends, je crois que c'était une heure, si vous voulez je vous retrouverai le nom de l'émission, et je me disais « imagine ça sur France Télévision ». Que la télé française, en plus maintenant depuis le Covid, on a un canal qui essaye de se focaliser plus sur les ressources, les contenus un peu plus pour les étudiants, les jeunes, etc. Mais justement on pourrait faire pour dédramatiser la stat, montrer toutes les implications, montrer tous les biais algorithmiques : il y a mille et un sujets en fait. On pourrait raconter, avec une émission, la stat dans tous ses états. Qu'on nous raconte tout un tas d'histoires. Je me dis que oui que des canaux comme la télévision, les sites bien sûr des grandes chaînes, des grands médias, qui pourraient aussi participer à ça, certainement.

Entrepreneur 3 (10/02/22)

Comment définis-tu la *data literacy* et emploies-tu ce terme ?

Bonne question. Qu'est-ce que c'est que la culture de la donnée ? C'est une bonne question. En fait, il y a la culture et l'acculturation à la donnée. L'action vis-à-vis de la culture. Qu'est-ce que c'est la culture à la base ? Toi, quand tu parles de *data literacy*, et il faudrait que tu m'en dises plus sur ce mot, je crois qu'il y a le côté vocabulaire derrière. Le point de départ de la culture de la donnée je pensais que c'était de comprendre le vocabulaire qu'il y avait autour de ce qu'était la donnée. C'était ce que je croyais. Mais vu les échanges que j'ai eus, je me suis rendu(e) compte que ce n'était pas le sujet du tout en fait le vocabulaire. En fait, la culture de la donnée elle va dépendre de ta cible finale que tu vas devoir acculturer. Au départ, je pensais que ça devait passer par le vocabulaire et je me suis rendu(e) compte que c'était plutôt comprendre le fonctionnement global, les enjeux et les impacts que la cible va avoir ensuite dans son quotidien. Donc en fait la culture de la donnée, elle dépend finalement du profil de la cible finale que tu vas devoir acculturer. Ça peut prendre plein de formes et ça agrège plein de choses. Tu as la partie vocabulaire. Tu as la partie « les cas d'usage ». Qu'est-ce que tu vas pouvoir faire de la donnée ? Le fait de comprendre tous les usages qu'il va y avoir derrière, ça fait partie de la culture. Connaître les outils que tu vas pouvoir utiliser par rapport à la donnée, ça fait aussi partie de la culture. C'est très large ! Ça regroupe plein de choses. Mais je pense que ce qu'on met derrière dépend de la cible qu'on va vouloir acculturer.

Et de ce fait, en rebondissant là-dessus, si on devait définir une « essence » de la *data literacy*, que dirait-on : une culture, un ensemble de pratiques et de techniques, un état d'esprit, tout ça à la fois ?

C'est une bonne question. C'est tout ce qui se trouve avant la formation. L'acculturation, la culture de la donnée, c'est tout ce qui se trouve avant que les gens que tu vas acculturer rentrent dans l'étape de la formation où là ils font une démarche volontaire de leur part de...enfin « volontaire », parfois on leur impose de se former. Mais où ils sont obligés personnellement de se former, en disant « je vais me former à la donnée », à un sujet particulier autour de la donnée. Donc tout ce qui est culture, c'est tout ce qui est avant. C'est pour ça que ce n'est pas facile à définir parce qu'on est dans l'ordre de la sensibilisation. Donc c'est comprendre schématique-

ment, ressentir le potentiel, dépasser les mythes qu'on peut avoir. Donc, est-ce que c'est technique ? c'est un petit mix de tout ça.

A partir de quand est-ce que quelqu'un peut dire « j'ai profité, je me suis approprié une culture de la donnée », je pense que c'est le moment où il commence à se dire « j'ai envie de me former ». Je te le dis avec mon point de vue. Mon objectif, c'est d'accompagner les gens. C'est de les amener à cette étape là où ils vont se dire « oui, c'est pour moi la donnée et oui je décide de m'y engager » quels que soient leurs métiers. Pas les *data scientists*, la culture de la donnée elle est pour tous les métiers. Puisque la donnée va toucher absolument tous les métiers, quel que soit le taff que tu fais il y a de la donnée derrière et c'est toute cette phase de sensibilisation de s'en rendre compte et jusqu'au moment où les gens sont ensuite dans la démarche volontaire de vouloir se former.

Actuellement on parle de gouvernance de la donnée, de data science, de tout un écosystème de données. Où peut-on situer la culture de la donnée dedans ? Est-ce la même chose que la *data science* ? Est-ce que ça a un côté culturel ? Voilà, où peut-on situer la culture de la donnée ?

La culture de la donnée, c'est le socle obligatoire, nécessaire, pour réussir à embarquer. Moi je parle toujours dans le cadre d'une entreprise. On est d'accord, c'est toujours dans le cadre d'une entreprise. Mais en gros, c'est le socle, le minimum vital pour pouvoir mettre l'ensemble de l'entreprise dans une logique *data-driven*. Si tu n'as pas tous tes collaborateurs qui ont été acculturés à la donnée, tu ne peux pas passer à l'étape d'après. Ta gouvernance de la data, elle n'a de sens que si tes collaborateurs, ils savent de quoi tu parles. Même par gouvernance de la donnée j'ai mis du temps à comprendre ce que ça voulait dire. Le côté *data science*, eux ce sont déjà des experts. Eux, ils ont dépassé la problématique de culture. C'est une expertise de la donnée, c'est leur cœur de métier. Alors que la culture, elle touche tous les gens dont ce n'est pas du tout le cœur de métier à la base. Tous les collaborateurs de l'entreprise quels qu'ils soient.

Et au-delà de l'essence, je parlerai de l'importance de la culture de la donnée aujourd'hui. Quelle est son importance aujourd'hui et depuis quand ?

Y a deux types de problématiques : ce qui est de la sphère professionnelle et ce qui est de la sphère privée. La compréhension de la culture de la donnée à titre personnel, moi c'est ce qui m'a en premier le plus titillé parce que je travaillais beaucoup sur la partie marketing digital qui brasse des données de partout. Et il n'y avait, je trouve, aucune transparence sur l'usage des données. Alors que nous en Europe même si on a le *RGPD*, on essaye vraiment de protéger l'individu, le citoyen, d'obliger les entreprises à expliciter ce qu'elles font des données, c'est tout nouveau ! Les trucs qui te demandent si tu acceptes les *cookies* ou pas, avant c'était même pas le sujet : on te traquait de tous les côtés. Et en fait le citoyen n'avait tellement pas de notions et pourtant moi je travaillais dedans. Je découvre des trucs tous les jours et ça me hérissait le poil. Mais vraiment je trouvais ça hallucinant.

Et pour moi c'était vital parce que ça permet ensuite au citoyen, c'est ce qui se met tout doucement en place avec le *RGPD*, de pouvoir dire oui et de pouvoir dire non. Je pense que la base d'un bon *business*, c'est la relation de confiance avec les entreprises, entre un consommateur client et entreprise. Et je trouve qu'avec tout ce

qui a été fait des données, il y avait un vrai risque que le jour où le citoyen, où le client allait se rendre compte de ce qui était fait de ses données, qu'il perde totalement confiance dans l'entreprise et coupe la relation. Pour moi, je mettais ça en parallèle avec la même crise que l'on a connue, dans la bouffe, etc., où on a découvert il y a quelques années de toutes les merdes qu'on bouffait qui étaient dedans et sur lesquelles il n'y avait aucune transparence et de fait, il y avait une vraie perte de confiance des consommateurs dans les marques. Et moi j'étais très convaincu(e)... enfin, je le suis encore d'ailleurs, que s'il n'y a pas plus de transparence de la part des entreprises sur ce qu'elles font des données et qu'elles ne le font pas de manière volontaire. Pas qu'on leur impose de le faire comme c'est le cas dans le *RGPD* mais que ce soit une vraie démarche de fond, c'est extrêmement dangereux car tu risques vraiment de perdre la confiance de tes consommateurs. Ça c'était la sphère privée.

Et dans la sphère professionnelle, je pense que c'est fondamental parce que c'est une responsabilité je pense de la part des entreprises d'acculturer leurs collaborateurs parce qu'elles le savent : tous les métiers vont être impactés par la donnée. Et si elles n'acculturent pas leurs collaborateurs, les besoins qu'elles vont avoir les entreprises dans les années à venir, les profils des collaborateurs, cela ne sera plus raccord. Et donc il y a des gens qui vont se retrouver sur le bord de la route, clairement, si du moins ils ne sont pas acculturés. Je ne te dis même pas de se reconvertir, c'est juste qu'ils comprennent et qu'ils fassent partie de l'aventure dans laquelle toutes les entreprises sont parties. Il n'y a pas un plan stratégique d'entreprise que j'ai entre les mains, et j'en ai de plus en plus maintenant, dans lequel il n'y a pas le mot *data-driven*. Ne serait-ce que la phrase clé qui est en haut de chaque truc, mais vraiment ! Donc si on n'embarque pas les collaborateurs, je vois vraiment ça aussi comme une opportunité. C'est-à-dire que c'est une opportunité pour les entreprises aussi d'ajouter ça dans leurs démarches RSE, de le voir comme un engagement et pouvoir communiquer dessus : moi je suis là pour faire du *business*, ce que j'aime c'est faire du *business* donc je vois aussi l'intérêt qu'elles peuvent y trouver. Mais un intérêt partagé parce qu'au final ça embarque aussi les collaborateurs et ça évite aussi qu'ils ne se retrouvent après plus utiles à l'entreprise parce que leurs jobs sont en danger aussi.

Ce mouvement de fond sur les données, arrives-tu à le situer temporellement ?

Je l'ai identifié par un détail, à mon avis ça va t'intéresser : l'évolution du nom des directions ! Il y a quatre, cinq ans, on parlait de la direction du digital et [nom d'organisation] est un bon exemple pour ça. Il y a quatre, cinq ans quand ils ont débuté la transformation digitale, la direction s'appelait la direction du digital. C'est devenu il y a trois ans la direction de la data et du digital. Et maintenant c'est devenu, je ne sais plus dans quel ordre je t'avoue, mais je me souviens que l'ordre était stratégique : data, digital et innovation. Il y a quatre, cinq ans c'est passé du digital à la data. En fait, on a d'abord fait le vernis extérieur du digital et puis ensuite on s'est rendu compte que pour vraiment aller jusqu'au bout du digital il fallait avoir des données de qualité, il fallait avoir de la donnée, donc il fallait que les collaborateurs saisissent les 06 de leurs clients qu'ils voyaient passer, etc. Et de fait, on a plongé dans la data en démarrant par le digital je pense.

Pour rentrer dans le cœur du sujet, je veux te demander : qu'est-ce qu'une donnée ? Comment définis-tu une donnée ?

Alors moi je définis rien parce que je n'ai pas de compétences techniques. Par contre, ce que j'ai entendu souvent c'est la différence entre la donnée et l'information, toujours la fameuse définition : je n'ai pas du tout de compétences techniques donc j'écoute la manière dont les gens définissent les choses et j'essaye de jouer avec. Autant la culture de la donnée c'est plus mon sujet à moi, autant comme je n'ai pas la compétence, je ne suis pas ingénieur(e), je n'ai pas une approche scientifique du sujet, mais la donnée c'est l'élément brut qui devient une information à partir du moment où elle est mise en contexte, nettoyée, relativisée, etc. Donc voilà pour moi c'est l'élément brut de base. D'ailleurs, est-ce qu'on dit donnée, est-ce qu'on dit data, c'est un grand sujet.

Quels sont les enjeux des données et comment la *data literacy* peut y répondre ?

Ce sont les problématiques autour des données qu'ont les entreprises. La première, c'est avoir des données. Et pour avoir des données, il faut des gens qui saisissent les données : c'est ce qu'on disait tout à l'heure. Et donc le truc de base quand j'ai bossé pour [nom d'organisation], ils ont des caissières et en fait les toutes premières données qu'ils ont eues. C'est qu'en fonction des entreprises, la culture de la donnée est venue par un endroit différent. Et donc par exemple chez [nom d'organisation], c'est venu par les cartes de fidélité. Cartes de fidélité qui étaient scannées par les caissières et donc les toutes premières données qu'ils ont commencé à avoir c'étaient celles de leurs clients. C'était du marketing, c'était l'angle marketing. Et donc la problématique c'est que si les caissières ne scannent pas la carte, eh bien il n'y a pas de données. Donc si elle n'a pas compris, la caissière, l'impact que pouvait avoir le fait qu'elle scanne la carte, eh ben il n'y a pas de données derrière. Et s'il n'y a pas de données derrière. Quand ils ont la carte de fidélité, ils ont aussi tout le ticket de caisse donc ils savent exactement tout ce que t'as acheté, ils ont l'historique donc ils savent si t'as des enfants à la maison, ils savent absolument tout. Donc la première chose c'est réussir à avoir de la donnée. Et pour le petit clin d'œil si ça t'intéresse : justement, on a parlé de la problématique des caissières qui ne scannaient pas assez les cartes. Je pose la question au *chief data officer* de chez [nom d'organisation] et je lui dis « mais, le président directeur général monde, celui qui est tout en haut de [nom d'organisation], dans son tableau de bord, s'il suit 50 indicateurs, y a quoi comme indicateurs qu'il regarde tous les matins, genre le truc qui l'empêche de dormir la nuit ? ». Il me dit « ça, ça, ça » et il me dit le nombre d'encartés. Donc les trucs qu'il regarde et qui sont essentiels pour lui, le grand directeur général monde et ses 200 000 collaborateurs, c'est le nombre de gens de personnes qui sont encartés dont les cartes de fidélité sont passées. Donc en fait, la caissière au Chili elle a un impact direct sur le tableau de bord du président directeur général et je lui dis « mais est-ce qu'elle le sait ça notre caissière au Chili qu'elle scanne un truc, ça rajoute un +1 sur le tableau du gars tout en haut ? ». Et en fait on est parti sur cette idée-là de le communiquer ça aux caissières parce que si elles se rendent compte de l'impact qu'elles ont, petites fourmis au milieu de l'énorme chemin de la donnée, eh bien en fait ça a une valeur inestimable leur action. Quand elles voient passer la carte elles risquent, et c'est ça l'objectif final, de changer les usages dans la culture de la donnée, c'est la manipulation dans le bon sens du terme, enfin j'espère, qu'elles finissent par y penser à pas zapper la carte. Voilà c'était le petit clin d'œil sur la caissière chilienne.

Toujours sur ces enjeux des données, est-ce qu'il y a des types de données qui sont au cœur de ces enjeux ?

Ça dépend vraiment des entreprises. En fait, tu as le côté historique : donc là chez [nom d'organisation], c'était la donnée marketing, c'étaient les premières qui étaient rentrées. Mais maintenant il y a des données à absolu-

ment tous les niveaux. Les métiers de la logistique qui utilisent des données et des algorithmes pour réussir à optimiser comment les camions sont remplis, c'est ouf. Les équipes magasin qui utilisent des algorithmes et des données par rapport à la météo locale de chaque magasin pour savoir quels produits il faut mettre en avant : c'est hallucinant ! Je ne sais pas si toi tu savais mais moi je ne savais pas. Il y a des données à absolument tous les niveaux donc je ne pense pas qu'il y a une gradation, en tout cas qu'il y a de priorisation. Tu as toujours la donnée historique, la première qui est rentrée, qu'on chouchoute le plus. Je donne un autre exemple : il y avait une autre boite qui est dans la gestion d'actifs, eux leurs données de base, c'est les données financières du marché parce qu'historiquement, le mec me racontait qu'historiquement c'était dans le journal. Tous les matins, on recevait les données du jour. Et ils feuilletaient, enfin pas les *traders*, ceux qui faisaient les choix de stratégie, ils recevaient les informations comme ça alors que maintenant ils sont connectés et reçoivent des données de partout et je pense plutôt qu'il n' y a pas de priorisation. Par contre, il y a le côté historique de par où les premières données ont été créées, enfin à quel endroit de l'entreprise les premières données ont été créées. Ça veut dire aussi que c'est, en général, une population qui est acculturée depuis plus longtemps à la donnée. Je n'ai pas encore eu l'occasion de le faire mais en tout cas c'est une réflexion que j'ai eue, je pense, des bons ambassadeurs vis-à-vis des autres populations. Parce qu'ils ont eu plus de contacts depuis plus longtemps avec la donnée en fait.

Est-ce que l'arrivée de la donnée a changé ta pratique professionnelle ?

Je suis indépendant(e) donc je n'ai pas des tonnes d'équipes et des indicateurs à suivre de tous les côtés. Moi où je suis la donnée, c'est la partie un petit peu marketing : les envois de newsletters, les taux d'ouvertures, tout ça. C'est quand même très minimaliste. Ce n'est pas du tout le cœur de mon activité. Mais après, il y a vraiment la partie émergée de l'iceberg, la partie émergée énormissime où tous traitent des données dans tous les sens. C'est juste qu'on ne le perçoit pas et qu'on ne le visualise pas comme ça, c'est tout. Sinon il y en a des tonnes derrière mais voilà, c'est pas forcément intentionnel on va dire.

Toujours en lien avec ta pratique professionnelle, ressens-tu au quotidien le cadre législatif, je pense à la loi Lemaire, au RGPD ? Ressens-tu une différence dans tes pratiques ?

Perso, j'ai du mal à le voir comme un citoyen lambda. Comme ça fait partie de mes sujets qui me préoccupent, qui sont importants pour moi, forcément quand je vois passer une question sur le *RGPD*, sur les cookies, etc., je l'analyse plus que n'importe qui puisque c'est un sujet qui m'intéresse, tu vois ce que je veux dire : je vais avoir tendance à être plus vigilant(e) pour voir comment chaque site internet joue le jeu, ne joue pas le jeu, lequel m'a demandé un abonnement payant, pourquoi est-ce qu'il me l'a demandé, est-ce que j'aurai cliqué à sa place.

Mais après je le vois comme faisant partie de l'écosystème data, je le vois avec un œil un peu critique. J'essaye de comprendre comment les entreprises elles se positionnent par rapport à ça. J'avais écrit un truc, je l'ai un peu en tête mais je le développe beaucoup moins c'est *data confidence* parce que j'étais vraiment sur ce sujet de garder la confiance entre les marques et les consommateurs en jouant la transparence, en explicitant les risques pour que la démarche des clients soit vraiment volontaire. C'est un peu poil à gratter par rapport aux entreprises mais je pense que c'est stratégique si elles veulent pas que ça parte en sucette leur lien de confiance.

Si l'on voulait mettre en place une culture des données dans un organisme, que faudrait-il faire en termes de moyens, de durée ?

Je peux te dire ce qu'il ne faut pas faire. Ce que j'ai vu et qui n'a pas fonctionné pour l'instant. Le réflexe qu'on a souvent et je le vois vraiment parce que, avec les clients avec lesquels j'interagis, c'est que le réflexe que tu vas avoir c'est de tout de suite produire des contenus parce que tu vas te dire que culture de la donnée c'est du vocabulaire donc vite vite on va créer un lexique des mots de la donnée pour essayer de l'expliciter aux gens. Vite, on va tourner des conférences vidéo où on va faire parler des *data scientists* pour que les gens comprennent ce qu'ils font. Et très souvent le premier réflexe ça va être de partir dans un planning de production de contenus mais ce n'est pas une bonne idée parce qu'en fait je me suis vraiment rendu compte que ce qu'il faut faire avant tout ça, parce qu'en fait les gens qui produisent les contenus ce sont des gens qui sont déjà acculturés, c'est comme-ci toi tu parles anglais et tu essayes de créer un support. Alors oui, la grande différence, la grande particularité de l'acculturation, c'est que tu vas toucher des gens qui n'en ont rien à foutre. Quand tu proposes à des gens de leur apprendre l'anglais, ils voient tout de suite à quoi ça va pouvoir leur servir et ils ont envie de partir en vacances en Angleterre donc ça tombe bien, et donc ils vont avoir une démarche d'action. Alors que quand tu vas voir des collaborateurs et tu vas leur dire « Coucou, je vais t'acculturer à la data », le gars il se dit « mais ça ne me concerne pas ! En quoi ça me concerne ? Attends, moi je suis caissière. Ce n'est pas ma vie, je ne connais même pas ce mot, c'est un truc d'ingénieur, c'est un truc de tecos, rien à voir avec moi ». Les gens ne se sentent pas du tout concernés. Et donc tu vas avoir tout un travail d'abord. Si tu leur fais directement des contenus où tu expliques à ta caissière c'est quoi un *data scientist*, mais elle n'en a rien à carrer ! Et savoir ce que ça veut dire un algorithme, elle n'en a rien à carrer. Donc en fait, le truc dont je me suis rendu compte : la première étape qu'il faut absolument c'est d'abord définir tes grandes populations que tu vas devoir adresser en fonction de dénominateurs communs qui seront propres à ton entreprise. Pour certains ça sera le niveau d'acculturation des différentes populations, certains ça sera en fonction de leur côté spatial, on va la jouer par pays, d'autres on va la jouer par métiers, tu vas rassembler plusieurs métiers... il y a une vraie réflexion à avoir sur les différentes typologies de populations auxquelles tu vas t'adresser.

Et pour chaque population comprendre : c'est quoi son enjeu métier ? Hormis, la data, tu oublies la data parce que la data on s'en tamponne, c'est juste un outil qui va permettre quelque chose mais au final l'enjeu de la caissière qui arrive le matin au taf, elle a un sujet. C'est quoi son sujet ? Son sujet c'est « à quelle heure je termine aujourd'hui ? Comment est-ce que je peux m'organiser parce qu'on me prévient toujours la veille ? ». Tu sais elles ont des plannings hyper mouvants. Et ça les stresse. Ça fait partie des problématiques qu'elles peuvent avoir quand elles commencent leur taf. Et donc si à elles, tu viens leur expliquer que grâce au petit algorithme qui a été mis en place, on arrive à optimiser les plannings de tout le monde. De fait elle va pouvoir savoir une semaine avant qu'est-ce qu'elle a comme planning : ça ça l'intéresse ! Et qu'on ne lui a pas du tout expliqué qu'il y a un algorithme, ce n'est pas le sujet. Par contre, tu lui fais se rendre compte de l'impact que ça peut avoir sur elle. Mais ça ça marche parce que tu as bien compris d'abord pour elle quel est son enjeu personnel à elle en dehors de la data ! Pour moi, il faut zapper la data au départ. Ce n'est pas le sujet. C'est à toi de réfléchir en quoi la data aujourd'hui lui apporte quelque chose par rapport à son enjeu.

Et ça c'est un peu biaisé parce qu'en général, tu parles à un *chief data officer* et tu lui dis « ok, on a les grandes populations, alors c'est quoi la problématique de ces grandes populations ? » et là tout de suite il va parler data

parce qu'on est tous biaisés. Et alors je suis obligé(e) de lui dire « non mais zappe la data : le matin, au taf, c'est quoi son problème ? C'est quoi son sujet ? ». De fait, le gars avec ses camions, c'est de remplir ses camions et d'optimiser le chargement dedans parce que sinon ça lui prend plus de temps. Alors est-ce que tu as en face de toi un outil data qu'il peut utiliser ou un *use case* qui a démontré qu'en utilisant tel truc, et il n'y a pas besoin de savoir ce qu'il y a derrière le tel truc, il réussissait à gagner du temps par exemple. Parce que lui ce qui lui importe c'est de gagner du temps.

Donc il faut revenir aux grandes populations et définir pour chacune leurs enjeux hors data, leurs freins par rapport à la data, leurs mythes, les mythes qu'il va falloir déconstruire pour les rassurer. La vendeuse elle est persuadée qu'on va la sortir parce qu'on va la remplacer par un ordinateur, tu vois, mais chacun a ses propres mythes. Les leviers psychologiques sur lesquels tu vas pouvoir travailler, là c'est le côté manipulation un peu horrible. Le CODIR il kiffe les *benchmarks*, tu vas lui dire « comment ça se passe chez les concurrents ? », c'est le meilleur moyen de faire bouger un CODIR. Il est obligé d'y aller. Pour certaines populations, ça va être la gamification qui fait le côté défi, le jeu. Chacun va trouver un levier psychologique. Et puis après c'est de trouver les bons canaux sur lesquels tu vas pouvoir interagir avec les populations : est-ce que c'est par mail ? Est-ce que ça sera du papier ? On s'est rendu compte que les gens dans les magasins [nom d'organisation], ils n'ont pas de mails donc t'arrives à des bornes, c'est des A4 qu'on a imprimé. Et ensuite, le format, parce qu'on sait pas quel format ça va être : de vidéo ?

Et si je te suis bien, cela peut s'adresser à tout le monde la culture des données ?

Ça doit ! Aujourd'hui la donnée est dans tous les secteurs et à tous les étages de l'entreprise.

Et à la sortie d'une formation à la culture des données, que doivent développer les personnes ? Est-ce que ce sont des techniques, un esprit critique, etc. ?

Je pense que tu es un cran plus loin. Je n'aurais pas mis l'esprit critique qui pour moi est encore un cran après. Peut-être dans la sphère de la vie perso. Parce que s'ils comprennent, ils ont un esprit critique et de fait, en conscience, ils vont cliquer sur les cookies. Mais je pense que dans le cadre de l'entreprise, surtout la première étape c'est réussir à comprendre quel est, parmi tout le parcours de la donnée au sein de leur entreprise, à quel endroit ils sont en fait, quel maillon de la chaîne ils sont. Je pense que déjà si déjà tes collaborateurs, chacun a compris par rapport à ton job où est-ce qu'il est, je pense que c'est déjà énorme. C'est déjà quelque chose de très important.

Après en fonction des métiers tu as des outils qui sont utilisés, tu en as des tonnes. Maintenant chaque métier a ses outils qui utilisent de la donnée différemment. Ça peut être plus ou moins la maîtrise ou en tout cas l'appropriation de l'outil. C'est marrant, il y a un *chief data officer*, celui que j'ai interrogé la semaine dernière, il me disait « moi je pense que j'ai réussi à acculturer mes collaborateurs quand ils pensent à me contacter quand ils ont un problème ». C'est vachement intéressant je trouve. Ça veut dire que pour les collaborateurs, la data est devenue une solution potentielle, ils ne savent pas comment, et à la limite tant mieux on ne leur demande pas de savoir comment ça fonctionne la donnée derrière, de vraiment comprendre l'algorithme, le truc technique, mais ils ont compris que ça pouvait être source de solutions. Donc, quand ils appellent le *chief*

data officer, « j'ai un problème, je pense qu'on pourrait faire un truc pour industrialiser, nous faire gagner du temps » mais ils ne savent pas vraiment comment. Eh ben pour lui c'est ça la réussite de l'acculturation. Tu vois, on est à l'étage de la sensibilisation, pas à l'étage de la maîtrise et de la formation qui là pour moi sera pris en charge par les RH et c'est complètement une autre histoire. Mais nous, le socle de l'acculturation c'est les préparer dans les premières strates jusqu'au moment où les RH vont prendre le relai.

Je reviens rapidement sur l'esprit critique : qu'est-ce que tu mettrais derrière un esprit critique par rapport à la donnée ? Comment cela peut se traduire en actes ?

C'est marrant ce que tu dis, j'avais parlé avec justement des gens qui étaient *data scientists*, et ils donnaient des cours dans des écoles d'ingénieurs. Et sur des cours d'éthique. Et je leur disais « c'est quoi ton plus grand sujet sur l'éthique par rapport à la donnée ? Qu'est-ce que tu transmets à tes étudiants ? » et il m'avait dit « c'est de pas être des robots en fait ». C'est d'avoir un esprit critique sur l'usage de la donnée, de se questionner à chaque fois en disant « est-ce que c'est bien de le faire ? ». On est quand même pas très loin de potentielles manipulations. En fait, je trouve que ça se joue beaucoup dans la transparence des actions qu'on fait avec la donnée. Je pense que, et c'est un vrai sujet, si tu le fais en transparence, que tu n'es pas en train de tromper les personnes. Je pense que la data n'est pas l'outil et au final c'est toujours un homme, un être humain qui va faire la démarche. Et utiliser la donnée, une intention peut être bonne, peut être mauvaise, peut être dangereuse pour l'autre. Et s'il n'y avait pas le *RGPD*, il y aura plus d'abus. Il y a déjà de l'abus dans tous les sens, pour moi ça devient un abus à partir du moment où la personne, que chaque fois tu vas sur un site internet direct il y a 30 traqueurs qui arrivent sur ton ordi, c'est horrible, qui le sait ça ? Il faut leur dire ! C'est flippant parce qu'on ne te le dit pas.

La situation actuelle en termes de compétences sur la culture des données des personnes qui sont recrutées. Est-ce qu'on est à un plateau, est-ce que ça a évolué ?

C'est une question qui est intéressante. Ça fait deux ans que j'interviens pour un groupe d'écoles de l'enseignement supérieur. Il y a douze écoles dedans, 30 000 étudiants, quand je vois tous les parcours qu'ils proposent. Et je fais pas mal de veille là-dessus. Et donc on va dire depuis un an, j'ai vu apparaître des couches de données, de formation à la donnée sur le domaine *business*, sur des domaines qui n'étaient pas ingénieurs à la base. Donc, j'imagine que vont arriver dans les entreprises une nouvelle génération qui aura eu cette couche de sensibilisation à la donnée. Mais c'est tout neuf, ça doit faire un an. Tous les autres, ils n'ont clairement pas eu cette couche de sensibilisation dès leur formation, dès l'enseignement supérieur. Je pense que d'ici quelques années, une fois que tous ces jeunes qui sont en train d'être formés, qui vont entrer dans les entreprises, l'acculturation à la donnée sera moins nécessaire. Parce que là on est à un changement d'étape. Il y aura le *gap* entre ceux qui auront été sensibilisés qui vont arriver en entreprise et qui n'auront pas besoin de ça et les autres, et il en a un paquet qui n'ont eu aucune sensibilisation au sujet. Mais je pense que ça arrive. En tout cas, tout ce que j'ai vu sortir, c'est vraiment une couche transverse qui s'est ajoutée dans toutes les formations.

Est-ce qu'on va donner la même formation à tous les employés ou scinder les formations en fonction des publics ?

Ça dépend de ton besoin par rapport à chaque population. Et tu utilises le mot formation, pour moi tu es sorti de l'acculturation. Tu es déjà dans la partie formation. En fait, tout ça, c'est en fonction des besoins de l'entreprise. J'ai découvert le mot *gpec* il y a trois jours, c'est une *chief data officer* qui m'a parlé de ça, je pense que ça dépend de leur plan prévisionnel des futurs besoins qu'ils vont avoir dans leur entreprise. La donnée, tout ça, c'est au service de quelque chose : tu ne formes pas juste pour former. Donc ce n'est pas nécessaire pour tout le monde de rentrer dans une phase de formation je pense.

Est-ce que tu vois des difficultés sur cette acculturation ?

Ah ben il n'y a que ça. C'est un sujet qui est extrêmement compliqué. Super méga compliqué. Justement parce que, comme on le disait tout à l'heure, tu dois toucher des gens qui n'en ont rien à faire. Et tu dois d'abord réussir à aller les toucher : c'est la première étape, ce n'est pas toujours simple. Tu dois réussir à les sensibiliser. A les faire sortir de leur zone de confort. De leur faire réaliser des choses. T'as un accompagnement au changement qui est extrêmement important derrière. Ça le petit truc, la petite réflexion que je te partage aussi : tu connais le *nudge* ? Le *nudge*, que je connais sous l'angle marketing, il peut être utile dans le cadre de l'acculturation. Parce que c'est vraiment un biais qui va jouer sur les leviers psychologiques pour changer le comportement des personnes. Tu ne peux pas être dans le « il faut », ça ne marche pas. On ne peut pas être dans le « on doit », ça ne marche pas. Je trouve qu'il y a un côté un peu malin, il faut hacker la situation sous un autre angle.

Donc si je te suis, il y a surtout à la base un problème d'intérêt plus que de compréhension ?

C'est l'étape d'après la compréhension. Alors l'intérêt tu as de la chance quand tu arrives à susciter l'intérêt, ce n'est pas facile. Je pense que le plus difficile c'est de susciter leur intérêt. Avant de susciter l'intérêt, tu peux susciter la curiosité. C'est peut-être le premier cran. Tu vas passer de la curiosité tout doucement à l'intérêt. Il y a plein d'étapes entre deux. Mais si t'arrives à susciter l'intérêt t'es fort déjà.

Est-ce que tu as en tête un projet en culture des données qui a bien fonctionné et si oui pourquoi ?

Aujourd'hui comme tu le disais c'est le *Far West*. On essaye tous des trucs. T'as pas aujourd'hui, de tous les projets, de tout ce que j'ai pu entendre, le truc idéal, il y a une recette spécifique à trouver en fonction de chaque entreprise. Et donc la recette idéale est souvent un agrégat des retours d'expérience de plein d'autres remélangés à ta propre sauce à toi. Il n'y a pas de recette qui peut dire « tu fais ça, ça fonctionne à tous les coups ». Il y a un truc, un dénominateur commun que j'ai trouvé, qui se retrouve absolument dans tous les gens avec qui j'ai parlés et qui a fonctionné, c'est le fait d'avoir un réseau d'ambassadeurs au sein des entreprises qui va te servir de levier de support pour aller acculturer. Ça me parait impossible aujourd'hui, en tout cas c'est comme ça que j'accompagne mes clients, ça fait partie des incontournables, d'avoir un réseau d'ambassadeurs.

Quelles seraient tes recommandations pour une meilleure acculturation à la donnée ?

C'est une bonne question. Il ne faut pas faire de la haute couture mais sur mesure quand même à chaque fois. Moi le vrai retour d'expérience, et je l'ai dit tout à l'heure, c'est ne foncez pas direct dans la production

de contenus ! Il y a 15 000 questions à se poser avant. Donc il y avait les questions sur les populations, il y a la question sur la gouvernance de l'acculturation dans le cadre de ton entreprise, donc qui tu vas embarquer, auprès de qui tu vas remonter ce que t'as mis en place, quels acteurs tu vas embarquer dans l'histoire. Les RH ? Tu sais qu'il y a plein de boites avec lesquelles j'échange, ils n'ont même pas pensé du tout à mettre les RH dans l'histoire parce que eux en général...ah et d'ailleurs il y a un truc intéressant pour toi : en général, l'acculturation aujourd'hui est menée par les *chief data officer*, pas par les RH. Donc les *chief data officer* font leurs trucs dans leur coin. Et je leur dis « mais t'as pris contact avec les RH ? » « bah non pourquoi ? » « parce que nous on fait la première base avant d'apporter la formation, il faut absolument les intégrer dans l'histoire ! » ils n'y ont pas pensé.

Ou la com interne. « Mais pourquoi la com interne ? C'est super, t'as fait de supers contenus mais comment tu vas faire pour que ces contenus ils aillent toucher la cible ? Qui aujourd'hui dans l'entreprise a l'habitude d'aller toucher les cibles que tu veux toucher ? Les RH, parce que un bulletin de paie ça arrive toujours à celui qui doit le recevoir, donc eux ils savent. Et la com interne. » Donc si tu mets ces deux personnes là dans la boucle, déjà tu t'assures d'un bon socle pour t'assurer que ce que toi tu vas produire comme contenus géniaux vont bien toucher leurs cibles au final. Tu vois ce genre de choses.

Donc il y a un milliard de questions à se poser avant de se lancer dans la production de contenus. Aujourd'hui, pour moi, quand on parle de *data literacy* ou de culture de la donnée, on imagine tout de suite des contenus alors que le contenu ça va être l'outil final. Mais que t'as plein de questions à te poser avant. Il faut que tu marketes ta réflexion. Marketer au sens dans la logique marketing : bien définir ta cible, bien définir le canal sur le quel tu vas partir, et le format pour qu'il s'adapte bien à ta cible. Il y a une sorte de marketing de ton offre d'acculturation pour qu'au final elle soit vraiment efficace sinon tu fais de la com. Juste : tu as produit des contenus super chouettes, canons, mais est-ce qu'ils vont vraiment toucher, est-ce qu'ils vont vraiment atteindre l'objectif de faire bouger les lignes, ça c'est moins évident. Et si tu fais de la com, tu ne pilotes pas. Tu n'as pas d'indicateurs. Si tu fais ça avec une approche marketing, t'as les indicateurs, tu impactes, tu pilotes.

Question supplémentaire posée après l'entretien

Comment définis-tu l'éthique des données et qu'est-ce qu'elle recouvre ?

L'éthique des données, selon moi, regroupe tous les risques associés à l'utilisation des données. A la fois tous les enjeux de subjectivité, les biais dans les algos, faire dire ce que l'on veut à des données etc. Et puis les enjeux liés au manque de transparence de l'utilisation des données (usages des données personnelles par des tiers : entreprises pour le marketing ou tous autres organismes). Tout ce qui devrait requérir un cadre « éthique » pour éviter des abus.

Entrepreneur 7 (24/06/22)

Comment définis-tu la *data literacy* et emploies-tu ce terme ?

Comment définir la *data literacy* : je vais le définir comme une compétence avec une vision RH. C'est la capacité à utiliser les données dans le quotidien. Et dans le fait d'utiliser les données, il y a aussi cette notion d'avoir du recul dessus. De se dire que quand il y a un indicateur qui est produit, c'est le résultat d'un certain nombre de décisions, d'un certain nombre de calculs et, parfois, de biais. C'est donc être capable d'utiliser la donnée mais aussi de se dire que ce n'est pas forcément la seule information dont on a besoin pour décider.

Tu emploies donc bien ce terme et pas, par exemple, culture des données...

Le terme est quand même pas mal rentré dans le langage courant, poussé bien entendu par les anglo-saxons qui à chaque fois parlent des sujets chauds avant tout le monde. J'ai entendu « datalphabétisation » aussi que je trouve intéressant mais dur à dire. Donc ça marche bien à l'écrit. « Culture des données » ça me semble assez représentatif en fait du véritable enjeu. Parce qu'avec « data literacy » on a une vision assez individuelle de la capacité à utiliser alors que la culture des données a une dimension plus collective, organisationnelle et donc qui met des enjeux différents. Je dirais que la culture des données intègre mieux cela.

Dans l'écosystème des données, comme la gouvernance des données et la *data science*, quelle est la place de la *data literacy* ?

Ce n'est pas évident bien entendu. Je pense que la *data literacy* est partout. Quelque part c'est facile à dire mais je la place au niveau de la stratégie parce que si on n'a pas une réflexion stratégique sur la culture des données et sur les actions pour atteindre cette culture des données, on n'aura pas de cohérence dans l'organisation. C'est une réponse de consultant mais on a les mêmes pratiques. Et après à chaque étape de l'organisation si on parle de l'architecture, si on arrive à faciliter l'accès aux données, on va bien entendu développer une *data literacy* parce que les gens vont accéder plus facilement aux données et donc plus facilement s'en servir. Si on parle de gouvernance, on a des outils comme les catalogues de données qui facilitent la *data literacy*. Quand on clarifie les processus de modification de données, d'évolution, etc., qui sont traités dans le cadre de la gouvernance des données, on facilite la *data literacy*. Si on a des *data scientists* qui sont capables de parler de quoi ils font, de vulgariser, alors c'est de la *data literacy*. A tous les niveaux de l'organisation, dans toute la chaîne de la data, il y a de la *data literacy*.

Si on devait définir la « nature » de la *data literacy*, que dirais-tu ? Plutôt une culture, un état d'esprit, un socle de pratiques techniques ?

Je le verrais comme un état d'esprit car cela a le mérite d'englober un petit peu une culture, la *data literacy* c'est forcément un petit peu une culture, mais aussi de curiosité. Cette curiosité c'est elle qui permet de développer les pratiques parce qu'il y a ce côté « ok j'ai de la donnée, qu'est-ce que je vais en faire pour m'aider dans mon quotidien. Et le fondement de la *data literacy* c'est ça : « est-ce que je peux me servir des données ? » en réalité.

Les données sont devenues centrales : peux-tu définir l'importance aujourd'hui d'être acculturé aux données ? Et à quel niveau d'importance cela se situe comparé à d'autres compétences comme la technicité informationnelle et informatique ?

Aujourd'hui, on est dans l'étape où cela donne un coup d'avance. C'est-à-dire que le fait de bien utiliser les données, cela permet de comprendre des mécanismes qui sont sous-jacents à cette utilisation et cela permet de découvrir de nouveaux enjeux, de découvrir de nouveaux services, découvrir de nouvelles opportunités. Et véritablement les gens qui maîtrisent la donnée, ils ont des startups, ils offrent des services qui sont hyper innovant, qui sont à très forte valeur. On n'est pas encore dans le stade, comme peut l'être l'informatique aujourd'hui et depuis une dizaine d'années, où c'est discriminant, où si tu ne maîtrises pas, tu es sur le banc de touche. Par contre, c'est certain que l'on va y arriver. C'est une conviction. Et la *data literacy*, c'est aussi travailler pour éviter cette fracture numérique quelque part.

Arrives-tu à situer temporellement l'émergence de cette nécessité ?

Je suis un peu jeune mais avec les observations que je fais en RH, ce n'est pas encore essentiel. C'est important mais pas encore essentiel. Cela le devient en réalité. Cela le devient aujourd'hui sur la valorisation de la donnée et cela le devient de plus en plus avec le contexte du covid, du travail hybride, etc. Pour piloter les équipes et les ressources humaines, on a besoin de quantifier un peu plus parce qu'on ressent un peu moins, parce qu'on a mis de la distance entre les gens.

Je vais maintenant passer à la seconde partie de cet entretien qui sera basée sur la donnée. Ma première question est « comment définis-tu la donnée ? »

Tu as la définition académique : c'est le résultat d'une mesure, d'un fait. On a mis un thermomètre dans l'eau chaude : quel est le résultat ? Après la question c'est quelle interprétation et quel contexte on y place pour savoir si c'est bien, si c'est mal, si c'est meilleur, si c'est moins bon. Et c'est tout ce contexte métier qui fait la valeur des données.

Quels sont les plus grands enjeux des données aujourd'hui selon toi ? Et est-ce que la *data literacy* peut y répondre ?

Alors pour la deuxième question la réponse est oui. Et à la première question, qui est beaucoup plus ouverte, je dirais que c'est un enjeu de confiance. C'est une réponse hyper bateau car ça marche souvent l'enjeu de confiance mais aujourd'hui, les utilisateurs de *Google*, *Facebook*, etc., sont de moins en moins certains que leurs données sont utilisées à bon escient. Ils ont compris que c'est du *retargeting*, *marketing*, etc. Mais ils sont de moins en moins confiants dans le fait d'utiliser leurs données pour avoir des services personnalisés parce qu'on a eu *Cambridge Analytica*, on a eu plein de choses qui fait que la confiance est érodée. En RH, c'est la même chose. Sauf qu'il y a un lien assez particulier entre l'employeur et le salarié : un lien de subordination qui fait que le salarié, de fait, est une personne vulnérable vis-à-vis de l'employeur dans l'utilisation des données au sens du *RGPD*. Ok je donne des données pour la gestion administrative, pour la paye, pour la formation, etc., et je veux m'assurer que ces données seront utilisées correctement. Ensuite, il y a une confiance aussi dans la direction des ressources humaines de dire « j'ai des données j'ai le droit de m'en servir ». Et là c'est un sujet que je suis en train de traiter aujourd'hui, je fais une formation à [nom d'organisation] sur le *big data*, *machine learning*, etc., et je dois produire un mémoire. J'ai proposé à mon employeur d'analyser les données RH et c'est très compliqué d'avoir accès aux données RH alors que je suis consultant RH, que j'ai accès aux données

de mes clients, qui font 50000, 100 000 salariés, mais ceux de mon employeur ce n'est pas évident. Donc dans les RH on a des données, c'est ok de s'en servir mais il faut poser un cadre dans l'utilisation de ces données. Et pour recouper avec le dernier bout de la question : la *data literacy*, c'est la capacité à dire « ok mes données, elles peuvent servir à ça et je sais dans quel cadre elles sont utilisées, je connais les mécanismes sous-jacents qui font que l'on m'a recommandé telle formation, que l'on m'a recommandé telle publicité dans *Google* ».

Arrives-tu à identifier des types de données particuliers qui sont précisément au cœurs de ces enjeux ?

J'en vois plusieurs, il y a un sujet et ça on en parle beaucoup côté [nom d'organisation], on a des gens qui sont intéressants là-dessus : la frugalité de la donnée. C'est s'assurer que l'on ne va pas utiliser toute la donnée à disposition parce qu'elle est à disposition. Avec des enjeux éthiques mais aussi environnementaux. Une manière de répondre à ça c'est l'*open data* parce que si on a la *RATP* qui communique ses données, on n'a pas besoin de faire du *web scrapping*, etc., par ailleurs. On a une approche frugale de l'utilisation de la donnée et ça c'est important. En RH cela nous intéresse beaucoup car quand on a des données de chômage, des données de dynamisme de bassins d'emplois, etc., on arrive à contextualiser un peu plus nos formations. Et en RH le sujet chaud aujourd'hui, c'est plus un domaine RH qu'un type de données, ce sont les parcours : les parcours des individus dans l'organisation. Et on arrive à voir, à partir de l'entrée, où est-ce qu'ils arrivent dans l'organisation, quels sont leurs déplacements, de carrière, de mobilité, etc. Et c'est être capable de modéliser un environnement extrêmement complexe avec assez peu de données.

Tu as évoqué le RGPD : ressens-tu l'influence du cadre législatif dans ta pratique professionnelle et quel est ton avis dessus ?

Je travaille à mettre en conformité mon client sur le *RGPD* donc je le sens beaucoup. Mon avis là-dessus c'est qu'il y a une intention assez intéressante de poser un cadre de travail et de collaboration vis-à-vis de l'exploitation des données. Un avis qui est plus personnel c'est que parfois, à vouloir avoir une bonne intention de protection, en réalité on va dégrader l'expérience des législateurs, et là je prends un exemple un peu à part avec la gestion des cookies : je trouve que c'est infernal que dès que l'on va sur un site de devoir accepter ou refuser les cookies. On a des normes européennes qui imposent des pratiques, très bien, ce n'est pas pensé UIC, c'est pensé protection. Et c'est ce que je regrette beaucoup c'est que finalement on ne protège pas l'individu s'il accepte systématiquement pour être tranquille. On n'a pas éduqué finalement sur ce sujet. Et encore une fois il n'y a pas de *data literacy*, il y a un bouton : t'acceptes ou tu refuses. Je ne sais pas pourquoi tu acceptes ou pourquoi tu refuses. Et donc je pense qu'on légifère beaucoup et on ne rend pas très intelligible.

Comment pourrait-on mettre en place la *data literacy* dans un organisme et est-ce que cela doit s'adresser à tout le monde ?

C'est une vraie question encore une fois. Comment mettre en place ? Je pense que ça passe par une première action. Je vais te parler de : nous on a une offre pour les RH qui a pour visée de les aider à exploiter leurs données en partant un peu de zéro. Et la première action que l'on fait c'est qu'on les réunit, on fait une vraie action de formation de « voici ce qu'est la donnée ». Comme ça on replace tout à plat. Et l'action suivante c'est

de dire « okay, vous avez des irritants, dites-nous quelles sont vos difficultés, qu'est-ce qui est difficile pour vous, et on va réfléchir ensemble à comment la donnée peut y répondre ». C'est vraiment cette démarche de dire : ce n'est pas la donnée, c'est le service qui peut être rendu par la donnée. Et c'est comme ça que l'on engage ces actions de *data literacy*.

Après la question c'est : est-ce que cela doit être réalisé pour tout le monde ? Pas forcément pour tout le monde, en tout cas pas à la même échelle. Et là je suis vraiment dans les entreprises, je ne me place pas au niveau citoyen : je ne ne suis pas certain qu'un cariste ait besoin du même niveau d'information sur la donnée qu'un gestionnaire de paie qui saisit dans un applicatif et qui a un rôle très important dans la qualité de la donnée. Que le cariste sache ce que l'on fait de ces données là on est plus dans l'ordre de l'information versus, dans un autre cas, plutôt de la formation. Différentes échelles mais une action de *data literacy* qui peut être intéressante dans les deux cas. Après au niveau citoyen je pense qu'on doit tous avoir un socle de connaissances au niveau citoyen tout simplement parce que l'on a des données qui ont une valeur que l'on sous-estime radicalement, et aujourd'hui très peu de gens ont conscience de ce qui est fait de leurs données par les *GAFAM* au quotidien.

A l'issue d'une formation en *data literacy*, pour toi, qu'est-ce qui le plus important, pour les personnes, à développer ?

En priorité une acculturation à la donnée, d'être capable d'avoir ce schéma de pensée, de dire comment fonctionne la donnée, c'est cette pensée de croisement de données pour extraire de l'information, c'est vraiment une démarche d'analyse quantitative qui permet de dire « ok j'ai une donnée qui donne une information. J'ai d'autres données qui donnent d'autres informations. Si je les mets ensemble, je produis de la valeur. Je produis quelque chose d'intéressant. Les compétences techniques, il y aura toujours des gens pour développer à leur place mais le métier, dans un contexte d'organisation, c'est lui qui sait ce dont il a besoin en réalité. Si on arrive à faire exprimer de façon claire de quoi a besoin le métier (les développeurs y en a à la pelle c'est pas le problème).

Il y a une expression qui revient souvent, c'est celle d'esprit critique, est-ce que cette expression te parle et si oui qu'est-ce que tu mets derrière ?

Cette expression elle me parle. Ce que je mets derrière c'est, j'en parlais un petit peu lors de la définition de la *data literacy*, de ne pas prendre la data comme d'une vérité absolue. Il y a plusieurs cas pour ça : déjà, il y a des enjeux de qualité, ça c'est tout con, c'est mal renseigné = c'est mal renseigné, tu ne peux pas considérer que c'est fiable si c'est mal renseigné, ça se monitore. Ensuite, il y a des enjeux d'arbitrage. Une donnée c'est le résultat d'une mesure et suivant si l'on fait du degré Celsius ou du degré Fahrenheit, on ne va pas avoir le même résultat à proprement parler et donc on a fait des arbitrages à un moment pour parler en degré Celsius et c'est comme ça. Et puis il y a des affaires de biais : nous on travaille beaucoup dans le recrutement sur de la décision humaine, et la data rend compte d'une décision humaine. Elle ne peut pas se porter garante de la décision humaine, elle ne fait que constater. Et si, ça arrive, les femmes sont moins recrutées que les hommes, si les gens qui habitent dans le nord de Paris sont moins recrutés que ceux qui habitent dans le sud, ce sont des choses que la data pourrait reproduire. La data, en tout cas, ne va pas alerter dessus. C'est donc en avoir

conscience pour utiliser la data pour monitorer cela. En avoir conscience c'est pourquoi pas rééquilibrer un peu les pratiques.

As-tu un avis sur le niveau de compétences et de connaissances générales sur les données de l'ensemble des publics auxquels tu as affaire ?

En RH, c'est assez disparate. On a des cas où la data c'est celui qui maîtrisait mieux *Excel* que les autres, à qui on a demandé de faire du quanti depuis des années, on lui a collé un *ETL* dans les mains, et en fait personne comprend ce qu'il fait mais il arrive à sortir des tableaux de bord. En gros. T'as des organisations qui ont pris le sujet un petit peu plus niveau groupe avec une vraie vision stratégique et qui développent une culture de la data. Toute l'organisation n'est pas forcément dedans, mais il y a des personnes qui sont un peu plus moteurs et il y a des moyens qui sont mis en œuvre et là on se rend compte que vraiment il y a un sujet. Globalement je dirais que le niveau moyen de la data n'est pas évident. J'ai travaillé avec des RH qui avaient une approche plus qualitative avec des gestionnaires de paye qui sont des matheux à jour sur la réglementation sociale mais pour les données à proprement parler ce n'est pas toujours très clair, c'est hyper hétérogène en réalité.

Et par rapport à cela, as-tu pu identifier les difficultés à cette acculturation ?

Une des difficultés c'est de ne pas utiliser la donnée RH du tout. C'est de dire simplement « non c'est une donnée qui est confidentielle ». Parfois on va dire qu'elle est sensible à tort, donc c'est se mettre des freins soi-même. Cela peut être parfois pour des sujets éthiques mais tant que l'on n'a pas essayé… l'éthique c'est l'intention qu'il y a derrière et l'usage que l'on va en faire, c'est pas le cas à proprement parler. Donc on sous-utilise la donnée, c'est le premier frein.

Et ensuite un sujet dans les RH, c'est pas mal un sujet de compétences aussi, parce que la compétence data n'est pas vraiment détenue par les RH (on a un peu de tableaux de bords, un peu de dataviz, etc. mais ça ne va pas très très loin dans l'exploitation des données) ; et aussi un problème d'architecture des données, avec des applicatifs qui sont assez sillotés entre gestion administrative, paye, formation, etc. Et en fait, des modèles des données qui sont difficiles à croiser et il faudra le faire manuellement parce qu'il n'y a pas de *data lake*, y a rien de tel, donc aussi un problème d'infrastructure.

C'est aussi un problème de représentation de la donnée, de ce que cela signifie au quotidien ?

Pas tant que ça en réalité pour la donnée en tant que mesure parce qu'aujourd'hui c'est la gestion de la paye, la gestion des temps, il y a beaucoup de flux de données en RH qui transitent tous les mois. Et la gestion de la paye c'est un excellent exemple car il y a la déclaration sociale numérique envoyée à l'État pour valider les déclarations. Souvent le corps RH c'est le cœur de l'habilitation au système d'information de toute l'entreprise. On sait ce qu'est la donnée mais on a peur de s'en servir.

As-tu en tête un projet en *data literacy* qui a fonctionné et si oui pourquoi ?

Non je n'en ai pas en tête parce que je n'ai pas eu l'occasion d'en mener, d'en réaliser. J'en ai observé un, ce n'était pas terrible. C'était chez un client, c'était sa DSI qui faisait ça. C'est pas un projet de *data literacy* en soi mais les parcours de formation qu'on propose aux consultants RH ça peut rentrer dedans. On a trois parcours de formation différents selon l'appétence et l'expertise des consultants sur le sujet. Cela marche plutôt bien. Bon c'est plus intense qu'une simple action de *data literacy* : on fait du *data management*, de l'introduction à de la *BI*, cela va assez loin quand même. Et cela a le mérite de faire monter en compétences, de créer de la curiosité. Si ça marche bien c'est parce déjà qu'on a un enjeu *business* : on forme les consultants et on leur dit qu'on les forme parce qu'on a envie de faire ça. C'est plus qu'une approche support, c'est une approche *business*. Ça marche aussi parce que c'est cadré et qu'on a beaucoup de formations, beaucoup d'experts en interne qui savent former sur le sujet, ça marche parce qu'on a proposé des parcours, qu'on a promis aux gens, de se placer, etc. Et ça marche parce qu'en fait ils sont ici, ils sont dans ce groupe d'intérêt parce qu'ils étaient intéressés par la data. Ils bénéficient de la formation par choix.

En guise de conclusion, as-tu des remarques pour aider à l'amélioration de l'acculturation aux données en France ?

A tous les RH je dirais « allez regarder mon site web ». Et sinon en France, à une échelle plus citoyenne, je pense que c'est quelque chose qui doit passer sans doute par l'éducation nationale. Je crois savoir qu'en technologie on apprend de plus en plus à coder, ce genre de choses, mais avoir une approche vraiment citoyenne de ce qu'est la donnée, de ce qu'est sa valeur, on pourrait se retrouver avec peut être un pays tiers qui fait de l'ingérence dans nos élections en faisant des pubs sur *Facebook*, ce genre de choses, c'est véritablement les enjeux qu'il y a derrière. On parle beaucoup de business « oui on va préparer des *dashboards* » mais il y a des enjeux qui sont plus souverains derrière la donnée et dans un monde qui se polarise de plus en plus on a des voisins qui sont beaucoup moins bien intentionnés et qui n'hésitent pas à utiliser la donnée comme une arme alors que cela peut servir à des choses beaucoup plus positives.

Question supplémentaire posée après l'entretien

Comment définirais-tu l'éthique des données ?

L'éthique des données doit être vue dans l'ensemble de la chaîne de traitement. Comment les données sont-elles acquises ? Quels croisement sont réalisés ? Quelles questions sont posées dans les analyses et quelles décisions sont prises ? Toutes ces questions ont des réponses personnelles, individuelles, mais nous rappellent que la donnée est puissante. Les impacts seront positifs ou négatifs selon les intentions de leurs utilisateurs. La transparence imposée par le *RGPD* nous oblige à plus d'éthique car un traitement compris par les personnes concernées peut plus facilement être *challengé*.

Entrepreneur 9 (05/10/22)

Comment définissez-vous la *data literacy* et employez-vous ce terme ?

La *data literacy* : oui, j'emploie ce terme. J'aime bien l'employer justement parce que je trouve qu'il raccroche cet univers des données à quelque chose qui relève plus des sciences humaines, en tout cas dans le côté littératie je pense à littérature, donc il y a moins ce côté « on est dans une discipline d'experts techniques ».

Et je l'utilise pour parler de l'aptitude à comprendre, lire, s'y retrouver dans les données, les bases de données. Un peu comme la maitrise d'une langue étrangère : on a cette formule sur les CV des fois « maîtrise parfaite » ou « lu, écrit, parlé ». C'est un peu pour moi l'équivalence de ce que regroupe le terme de *data literacy*. C'est le niveau d'aisance avec les données qui va de « je ne sais pas de quoi on parle et je panique face à une base de données » à « j'ai des rudiments, je sais me débrouiller avec *Excel*, je sais globalement comment se présente une base de données, comment structurer des données, quel format ça prend » à « je sais faire des requêtes dans des langages de programmation et je sais manipuler de gros volumes de données ».

Il existe divers concepts imbriqués dans un écosystème de données, je pense à la *data science*, à la gouvernance des données... Est-ce que vous arrivez à situer la *data literacy* dans cet écosystème ? Est-ce équivalent à certains concepts ? Comment voyez-vous les choses ?

Je le vois comme un concept assez à part dans le sens où ce n'est pas une discipline, c'est pas un domaine de recherche. C'est compliqué de vous dire ça mais c'est pas une discipline académique, ce n'est pas un métier comme *data scientist*, on est pas data litera... « practiseur » ? Je sais pas quoi, ça n'existe pas. Donc je le vois plutôt comme un concept qui décrit des attitudes, des savoir-faire, parce que ça recoupe cette capacité à s'y retrouver face à un jeu de données, à savoir faire des requêtes, à savoir l'ouvrir, des choses plutôt techniques, mais ça recoupe aussi des aptitudes plus cognitives. Je vous l'ai dit tout à l'heure : ne pas paniquer face à une base de données dans le sens se sentir à l'aise avec, c'est-à-dire savoir concrètement qu'est-ce que c'est cette information ? Comment ça se présente ? A quoi ça correspond ? Comment on met de l'information là-dedans ? C'est un fichier ? C'est pas un fichier ? Où est-ce que c'est ? Où il est situé ? Où on peut y accéder ? des éléments un peu concrets où j'arrive à me faire une image mentale de cette base de données ou de ce fichier.

Mais aussi de la culture un peu générale qui m'aide à prendre des distances par exemple par rapport à l'auteur d'une source, à avoir quelques réflexes d'analyse de type « quelle est la source des données », « est-elle fiable ? », « d'où proviennent les enregistrements ? », « est-ce qu'il peut y avoir des doublons ? », « est-ce que ça peut porter atteinte au respect de la vie personnelle ? Des données personnelles ? », il y a un peu un réflexe comme ça qui relève de la culture générale des données mais qui font que face à un jeu de données, on est *literate* ou pas, on est éduqué ou pas. Eduqué ça fait un peu « éducation des bons sauvages », ce n'est pas très adapté mais on a un minimum de culture de cette matière qui fait que j'ai de bons réflexes de traitement de cette information, de lecture de cette information, j'ai quelques bases techniques qui permettent d'en faire quelque chose.

Vous avez anticipé ma prochaine question qui était, et j'emploie le terme avec des guillemets, si on devait définir la « nature » de la *data literacy*, qu'est-ce que cela serait ? Serait-ce plutôt une culture, un ensemble de pratiques et de techniques, un état d'esprit, tout ça à la fois ?

Ouais je pense que c'est un ensemble d'aptitudes et de connaissances et de savoir-faire. Et donc je le mettrais plutôt du côté de la culture.

On a eu un certain panel de compétences qui se sont ajoutées au fur et à mesure. J'entends par-là tout ce qui est compétence informatique, technicité informationnelle, et maintenant on parle de compétences en données, est-ce que pour vous la *data literacy* est au même niveau, plus importe, moins importante que les autres compétences ?

Ça dépend un peu de votre métier. Si vous êtes aide-soignant je ne sais pas si c'est aussi important que si vous travaillez dans le marketing ou dans le journalisme. Je pense que ça serait bien que ça rentre dans le périmètre de la culture générale, qu'on l'enseigne aux enfants très tôt. Il en va de notre capacité de savoir réagir sur les réseaux sociaux, face à l'information qu'on peut lire sur Internet ou dans les médias plus traditionnels, il en va d'un certain nombre de choses qui nous aident à bien appréhender notre environnement, à bien comprendre le monde qui nous entoure, je pense que c'est très important d'avoir un minimum de *background* sur ce que sont les bases de données, leur compréhension et comment on en tire de l'information. Après il y a beaucoup de métiers où je pense que c'est quelque chose de très important car aujourd'hui une grande partie de l'information manipulée dans beaucoup de métiers. On parle des métiers très manuels qui manipulent peu de sources d'informations écrites ou informatiques. Si on est dans milieu où il n'y a pas du tout d'informatique cela ne va pas être très important, mais dès qu'on travaille avec un ordinateur et de l'information connectée, je pense que c'est une compétence vraiment importante parce que la massification des données elle a vraiment envahi tous les métiers.

Si on regarde des gens, par exemple, du marketing, peut-être qu'avant, même dans les industries culturelles, on recevait des listings de vente de la Fnac, de Culture ou de Leclerc. Maintenant on va requêter ces chiffres de vente, ces chiffres d'écoute sur une API. Ce ne sont plus du tout les mêmes compétences qui sont demandées. On n'est plus du tout dans le même écosystème des données, des informations. Le monde des RH a énormément changé avec l'entrée d'outils de suivi, d'outils de management des compétences, d'outils de recrutement, qui intègrent beaucoup d'algorithmie. Avec d'ailleurs des potentiels biais sur les algorithmes de recommandation. Donc c'est un métier où typiquement les RH sont assez mal formées selon moi. Ce sont des métiers qui sont souvent...les personnes qui font ces métiers ont rarement des backgrounds d'études scientifiques. Plus rarement que des cursus plus économiques ou littéraires et donc ils n'ont pas forcément le *background* théorique pour comprendre ces algorithmes, les biais qui peuvent en résulter, etc.

Donc c'est vraiment quelque chose qui envahit tous les métiers qui travaillent avec un ordinateur et comme c'est une compétence qui est vraiment importante, une culture qui est vraiment importante parce que c'est l'accès à notre information, c'est l'information avec laquelle on travaille au quotidien, qui nous permet de comprendre notre métier, de comprendre le monde professionnel qui nous entoure, de prendre des décisions, d'agir dans notre métier. Notre paysage informationnel, professionnel, il est de plus en plus déterminé par des données et des données de plus en plus massives donc pour moi c'est une compétence, une culture, qui est vraiment très importante. Parce que de là que dépend notre capacité à faire tout le reste. La bonne compréhension des informations avec lesquelles on travaille c'est quand même indispensable si on veut travailler, avoir des bons résultats ou comprendre ce que l'on fait sur les autres étapes.

Si on devait dater ce moment où l'acculturation des données est devenue une véritable nécessité, à quelle période la situerez-vous ?

Je dirais depuis les années 2000 mais c'est peut-être un biais parce que j'ai commencé à travailler dans ces années-là. Donc je ne saurais pas forcément. Disons que mon impression c'est que ça date des années 2000 mais ça peut être biaisé par le fait que moi-même j'ai commencé à travailler à ce moment-là donc je suis plus susceptible de repérer ces phénomènes là à partir de ce moment-là.

Est-ce que vous arrivez à déterminer la raison de pourquoi on parle des données aujourd'hui ?

Pour moi ça vient de leur massification. Le fait qu'elles occupent une place de plus en plus importante dans le monde de l'entreprise là où avant elles étaient cantonnées à l'univers du *reporting*, de la *business intelligence*, donc un rôle très silloté, très cloisonné d'analyse des résultats de l'entreprise, d'analyse des performances de l'entreprise, de suivi des résultats de l'entreprise, éventuellement aussi de suivi de gestion comptable. La massification des données, le fait qu'on collecte et qu'on ait accès à des bases de données de plus en plus importantes, qu'on collecte des données de plus en plus importantes sur les réseaux professionnels, leur a fait passer la barrière de ce simple usage du reporting et de la *business intelligence* pour envahir beaucoup d'autres sphères professionnelles et pour devenir une source d'informations et une matière première des outils de travail pour plein de métiers.

Donc on pourrait dire, pour parler simplement, la massification des données, donc des nouveaux usages, donc ça envahit les sphères professionnelles. Je pense que ça va avec des évolutions techniques bien sûr, les deux vont ensemble : il y a massification des données parce qu'on a des progrès techniques sur le stockage des données, la capacité de stocker des données beaucoup plus volumineuses dans des espaces plus petits, d'accéder à des données plus volumineuses en moins de temps, la réduction du coût de stockage qui fait que ça devient possible de stocker des gros volumes de données alors que cela ne l'était pas avant. Donc les performances techniques sur le stockage nourrissent ça mais aussi l'évolution des programmes informatiques qui fait qu'on collecte plus de données. Les deux s'entretiennent forcément. Et puis, l'amélioration des performances des navigateurs, en tout cas moi je le vois en visualisation de données, le fait que les navigateurs d'aujourd'hui (*Chrome, Safari, Firefox*), les versions d'aujourd'hui ne sont pas du tout les mêmes que celles des années 2000 en termes de performance, de capacité à charger dans une page web des choses lourdes, des éléments, que ce soit du visuel, de l'image, de la vidéo et des données. On a des performances de navigateur beaucoup plus grandes qu'avant parce que nos ordinateurs deviennent plus puissants, parce qu'une carte graphique a une puissance plus de 100 fois supérieur à un ordinateur des années 2000… du coup tout ça est très tiré aussi par le progrès des appareils informatiques, des programmes informatiques. Les deux sont liés, les usages et la technologie.

Quelle définition donneriez-vous de la donnée ?

C'est une façon d'encoder l'information qui est particulièrement efficace pour transmettre des grands volumes d'informations et qui est à la base un format fait plutôt pour la communication entre machines ou entre ordinateurs. C'est une façon d'encoder l'information qui permet son traitement par des machines, par des algorithmes, par des ordinateurs, de façon programmatique, qui permettent des échanges d'informations entre des machines, mais qui n'est pas un format d'échange qui est fait pour des humains à la différence du langage ou de la photo qui est une façon d'encapsuler de l'information, de transmettre de l'information qui est plutôt fait pour des humains. Pour les machines, il n'y a qu'à voir les galères de la reconnaissances automatique d'image

pour comprendre que les ordinateurs ont du mal à faire ce que nous on fait très bien les humains, à savoir reconnaître un truc sur une image. Et justement à l'inverse les données c'est un format qui n'est pas facile à comprendre pour les humains et qui est vraiment fait pour être compris par les machines.

Quels sont pour vous les grands enjeux actuels autour des données ? Et en question subsidiaire, est-ce que la *data literacy* peut y répondre ?

Il y en a plein. Il y a des enjeux importants, éthiques sur la manipulation de l'information via des programmes informatiques de type *deepfake* ou de type *deep learning*. Donc avec des données, comme par exemple des tas de vidéos sur une personnalité : on peut reproduire la façon de parler, de s'exprimer, de se déplacer, les mouvements de visage de cette personnalité et lui faire dire ce qu'on veut. Pareil sur n'importe quelle déformation d'image.

J'inclue donnée au sens très large car pour moi cela relève de l'intelligence artificielle et donc derrière, des données. Si on prend les données au sens très large, il y a beaucoup d'enjeux en ce moment éthiques autour de l'intelligence artificielle, que ce soit sur la capacité de déformer de l'information, pour propager des fausses informations via les intelligences artificielles ou la façon dont on améliore la propagation des fausses informations par des algorithmes de recommandation qui sont faits pour valoriser les contenus qui font réagir. Donc il y a des contenus éthiques sur tout ce qui concerne la propagation de fausses informations par exemple, que ce soit sur comment se créent ces fausses informations ou comment on les propage. Il y a des contenus éthiques sur les biais dans l'usage des intelligences artificielles dans le cadre de la justice, dans le cadre de l'attribution des prêts, dans le cadre de tout un tas d'opérations qu'il va falloir choisir ou *scholer*, attribuer une note, faire passer ou pas un dossier... les algorithmes qui font ce score, quand ils reposent sur du *machine learning* ou l'intelligence artificielle reposent sur des bases d'entraînement qui leur permettent de reproduire des résultats tels qu'on les a déjà observés.

Donc des très grandes bases d'entraînement, de l'intelligence artificielle. Si j'ai un cas similaire qui arrive, je vais pouvoir lui donner tel résultat ou tel score. Et en fait nos bases d'entraînement peuvent inclure des biais mais aussi la façon dont sont codés les programmes d'intelligence artificielle peuvent favoriser certains biais. Si on prend un programme d'entraînement basé sur un truc de justice, est-ce que je prends un casier judiciaire sur un cas où il y a un litige et je demande à une intelligence artificielle « est-ce que cette personne elle a plus ou moins de chance d'être coupable ? ». Si ma base d'entraînement, elle est faite de tous les cas judiciaires pendant la ségrégation raciale aux États-Unis ou pendant la période coloniale en France, je risque d'avoir un biais qui fait que les personnes de couleur sont plus facilement accusées ou plus facilement considérées comme criminelles. On connaît la raison historique de ce biais-là. Donc si je me sers de ça comme base d'entraînement, je risque de reproduire ce biais. Sur des cas plus proches de nous, par exemple, pour des algorithmes de recommandations pour du recrutement sur des postes de manager, les algorithmes vont avoir tendance à moins souvent faire ressortir des CV de femmes parce que dans leurs bases d'entraînement elles ont peu de postes de top management. Et donc le fait d'être une femme sur un cv va plutôt dégrader sa note. Ce qui sont des choses que l'on peut corriger bien sûr : les biais les plus évidents comme ça ont été mis à jour. Ils sont corrigés après par les personnes qui programment ces algorithmes mais il y en a des tas d'autres de biais comme ça. Donc il peut y avoir un sujet éthique de compréhension de ces biais, de détection de ces biais, de compensation de ces biais et surtout

un sujet éthique sur dans quel cas c'est justifié d'utiliser une notation qui vient de l'intelligence artificielle ou qui vient d'un programme d'entraînement basé sur des bases de données. Donc il y a des cas comme ça où il y a des tas de choses positives sur l'amélioration de traitement de certains dossiers, la capacité à la reconnaissance d'images ça permet de faire beaucoup de progrès dans la médecine, d'avoir des premières analyses qui sont faites automatiquement sur des lectures de radio par exemple qui mobilisent moins le radiologue qui peut se concentrer sur des résultats plus litigieux, donc plein de cas d'usage dans lesquels c'est positif. Il y a des cas d'usages dans lesquels ça fait gagner du temps et des cas d'usages dans lesquels il peut être justifié de se poser la question « est-ce qu'on en a vraiment besoin ? Si on en a besoin, comment on se sert de cette première information ? Et comment on prend la décision avec ? ». Donc voilà, il y a des cas éthiques qui me semblent être un gros enjeu et c'est là pour moi où justement c'est important que ça rejoigne la culture générale parce que ce ne sont pas des choses réservées… disons que le problème serait que ça soit réservé, que la compréhension de ces problématiques-là soit réservée à quelques experts de l'analyse des données ou de la *data science*, que du coup ça soit un débat entre experts qui échappe au citoyen et à l'ensemble des citoyens et leurs représentants parce que ce n'est pas que une problématique technique en fait. On ne va pas dire « on va faire un meilleur algorithme pour résoudre ça ». C'est vraiment une question de choix.

Il y a un très bon exemple là-dessus par exemple sur les voitures autonomes. Les voitures autonomes il y aura forcément des cas où quelqu'un va traverser pile devant la voiture et la voiture va avoir le choix entre piler pour épargner la personne qui traverse au risque de projeter les personnes qui sont dans la voiture contre la vitre et les blesser, voire de les tuer. Ou bien de pas piler pour épargner les personnes qui sont dans l'habitacle mais prendre le risque de tuer la personne qui vient de traverser. Et il y a un test qui a été fait par un laboratoire de recherches américain qui s'appelle le « moral test » où on va présenter différents cas. Par exemple, si c'est une personne âgée qui traverse la route et une personne jeune dans l'habitacle, est-ce que je pile ou pas ? Si c'est un animal et il traverse la route et il y a des personnes humaines dans l'habitacle, est-ce que je pile ou pas ? Si c'est l'inverse ? Si ce sont des personnes qui ont un diplôme élevé qui sont dans l'habitacle et que quelqu'un a un poste de balayeur qui traverse, est-ce que je pile ou pas ? Et c'est intéressant : le test a été fait sur des personnes d'origines culturelles différentes et montrent que les choix sont différents. Alors ce ne sont pas des choix aussi grossiers que je viens de vous présenter mais vraiment ça permet de prendre conscience que si d'une culture à l'autre on ne fait pas les mêmes choix, il n'y a aucune raison qu'une intelligence artificielle soit la mieux placée pour prendre cette décision. Il n'y a pas de rationalité absolue dans ce genre d'arbitrage. Et donc ce sont des choix culturels, des choix politiques dans le sens la chose publique sur lesquelles des communautés d'individus doivent se mettre d'accord et donc c'est vraiment très important pour l'avenir des systèmes informatiques qui nous entourent parce que ça va prendre plus de place qu'à présent. Donc c'est pour ça que pour moi c'est important que ça devienne quelque chose intégrée à la culture générale, intégrée à des débats politiques et sociétaux, et pas réservée aux experts. Donc ça c'est un des gros enjeux. C'est même probablement le plus gros.

Il y a de gros enjeux de la protection de la vie privée et des données personnelles parce que, par le biais du téléphone ou de nos comportements en ligne, une grosse partie de notre activité quotidienne se fait sur des supports numériques et donc laisse des traces numériques, et donc des données personnelles, des données qui concernent notre activité présente. Et ces données-là sont le fruit d'interactions entre nous et tous les services avec lesquels on interagit, que ce soit les outils techniques ou les recherches sur *Google Map*. *Google* a fourni

la *search bar*, le moteur de recherche, et nous on rentre quelque chose dans cette *search bar*. A l'interaction de *Google* et de nous, il y a une donnée qui est « j'ai fait une recherche à telle date, telle heure » ce terme-là. Et donc le statut de ces données-là est pas évident. Parce que laisser à chaque individu le choix de communiquer ou pas ses données personnelles c'est un peu ce qu'on fait aujourd'hui avec la multitude de cookies, bannières... chaque fois qu'on entre sur un site, on est censé regarder qu'est-ce qu'on donne comme autorisation de ce qui est ou pas une donnée personnelle... j'ai l'impression de passer je sais pas combien de temps par jour à *dismiss*, à désactiver, à accepter des bannières cookies et bien sûr je ne relis pas les trucs auxquels je consens parce que c'est juste pas gérable vu le nombre d'interactions qu'on a aujourd'hui avec des programmes informatiques. Ça va forcément changer, ce n'est pas tenable. Mais aussi parce qu'une grosse partie de l'économie numérique repose sur la valorisation de ces données personnelles. Pas pour connaître chaque individu personnellement mais parce que dans l'ensemble elles produisent de l'information quand on les agrège, elles produisent des informations qui sont importantes pour faire fonctionner tout un pan de l'économie, le marketing, le ciblage publicitaire, d'améliorer les ventes, etc. Mais aussi parce que cette masse de données personnelles prises pour connaître un ensemble de comportements des individus sont aussi une matière première extrêmement utile pour améliorer la gestion des catastrophes naturelles, pour l'organisation des villes, pour l'urbanisme, le plan de circulation des voitures et donc pour tout un tas d'améliorations du monde qui nous entoure.

Donc encore une fois c'est pareil on ne peut pas juste dire qu'on va arrêter de collecter nos données personnelles puis personne le fait parce qu'on se prive d'autres outils et d'autres usages qui sont extrêmement intéressants. Et là encore une fois c'est difficile de faire la part des choses entre des usages qui voilà s'il y a juste « moi, mes données, je regarde juste mes données » c'est intéressant mais le fait de faire ça à grande échelle crée d'autres problèmes. Et donc là aussi, il y a des enjeux, des enjeux de débats politiques, sociétaux... on ne peut pas laisser chaque individu faire ses petits arbitrages personnels. Il y a quelque chose de plus grand qui entre en jeu. Et donc ça aussi me semble être un enjeu important : comment on fixe la réglementation sur l'usage des données personnelles.

Donc éthique sur l'intelligence artificielle, réglementation sur l'usage des données personnelles, ce sont quand même les deux gros cas dont on parle beaucoup sur les enjeux. Après il y a beaucoup d'enjeux d'améliorations techniques dans le domaine de la médecine, dans le domaine de l'aérospatial, dans le domaine du développement durable, dans la mobilité propre... il y a énormément de progrès techniques qui dépendent de notre capacité à bien traiter ces données, à bien traiter ou collecter des données, dans chaque industrie, dans chaque métier, dans chaque secteur académique. Il y a une partie des progrès à venir qui vient de la capacité à traiter des données donc après chaque industrie va y voir ses propres enjeux. C'est vraiment une façon de traduire de l'information. A terme c'est bien quelque chose comme de parler de l'écriture, de parler du langage ou bien de parler de l'image, ou bien du dessin... ça devient un medium. Il peut y avoir des enjeux spécifiques au langage, il y a des gens qui étudient le langage toute leur vie. Et il y a des gens qui étudieront les données en tant que façon d'encapsuler l'information, oui. Ça devient tellement présent partout que c'est difficile d'en faire un sujet à part entière. L'intelligence artificielle en ce moment c'est relativement nouveau pour qu'on se dise « bon ben là il y a un domaine de l'intelligence artificielle ». Peut-être que dans cinq ans ça sera aussi diffus, ça sera tellement partout aussi que cela ne sera plus un sujet en soi. On pourra parler du sujet de l'intelligence artificielle dans tel domaine ou l'intelligence artificielle appliquée à ceci ou à cela. Les données j'ai un peu l'impression

que c'est ça maintenant. C'est comme parler du langage, c'est comme parler du texte ou c'est comme parler de l'image, ouais c'est juste une façon d'encapsuler l'information.

Est-ce que vous avez en tête des types de données qui sont au cœur de ces enjeux ? Vous avez cité les données personnelles en priorité. Est-ce que vous en voyez d'autre ou est-ce que les données personnelles c'est ce qu'il y a de plus central actuellement ?

C'est très central mais il y a plein de choses dans les données personnelles. Il y a les données médicales, les données de déplacement, il y a les données d'interaction sur les services en ligne. Est-ce qu'il y a d'autres sujets ? là comme ça non, je n'ai pas de cas qui me reviennent, il y en a sans doute plein.

J'aimerais avoir votre avis sur l'influence du cadre législatif. Ce que j'entends par-là c'est qu'on a eu, bon la loi CADA, la loi Lemaire, le RGPD... y a le Data Act et le Data Governance Act qui sont en cours d'élaboration. Est-ce que vous vous constatez l'influence sur vos pratiques et si c'est en bien ou en mal ? Enfin en bien ou en mal c'est caricatural mais quel est votre avis sur ce système législatif ?

Je ne ressens pas de changement dans mon activité lié à ça. Je ne suis pas confronté(e) à ça. Ça n'impacte pas directement mon activité. C'est un vrai sujet. C'est une très bonne chose que l'Union Européenne travaille sur la réglementation de ces sujets parce que c'est un sujet qui mérite d'être encadré réglementairement, ce que je vous disais c'est qu'il doit y avoir des choix de société, des choix politiques qui encadrent comment on traite ces données, comment on les utilise, comment les valoriser économiquement. Et c'est un choix qui peut être fait d'une certaine façon ou d'une autre, c'est vraiment à une communauté d'individus de se mettre d'accord sur comment on veut que ça se passe. Donc c'est vraiment le rôle des politiques selon moi de faire ce travail-là d'encadrement par la réglementation. Donc c'est très bien que cela soit fait. Je trouve très bien que cela soit fait à l'échelle européenne parce que c'est assez dur à maintenir à une petite échelle locale. Mettre des barrières artificielles sur ce qui se passe... les services sont tellement globalisés que c'est compliqué de le faire à l'échelle nationale donc très bien à l'échelle européenne.

Le RGPD est un énorme progrès je pense dans l'état d'esprit de cette réglementation sur le fait d'encadrer l'utilisation commerciale de données personnelles, de façon à protéger la vie privée des individus. De façon aussi à protéger les individus d'un usage abusif des données personnelles. Donc c'est une réglementation qui va dans le bon sens et qui est importante. Maintenant, je pense que dans la façon dont elle est mise en œuvre il va falloir que ça change parce que dans les faits l'objectif de cette loi était d'obliger les gens qui collectent des données personnelles à recueillir un consentement manifeste explicite des utilisateurs sur les données qui sont collectées. La façon dont c'est fait aujourd'hui, à savoir je sais pas 30, 50, 100 bannières *cookies* à disposition par jour, personne ne lit. Et donc le consentement me parait collecté de façon un peu partielle et on pourrait dire « c'est la faute des gens, ils avaient qu'à lire les petites lignes » cela ne me semble pas très tenable. Je pense qu'il y a des choses à faire là-dessus. Ce n'est pas forcément l'objet de la réglementation d'organiser derrière comment c'est mis en œuvre et il y a des solutions qui peuvent venir de la société civile, des entrepreneurs, pour améliorer ça. Mais voilà la réglementation me semble bonne, utile, nécessaire et plutôt un progrès.

Dans sa mise en œuvre, il y a quand même pas mal de frictions que ce soit pour les utilisateurs mais aussi pour les gens qui développent des sites ou tout simplement, on a le cas parfois, pour stocker, sur quoi par exemple... imaginons que par défaut vous faites une recherche sur un trajet dans votre moteur de recherche d'itinéraire parce que régulièrement vous avez besoin de faire ce trajet. Imaginons qu'à chaque fois que vous deviez rouvrir ce site et refaire ce trajet le site, si vous n'avez pas donné votre consentement explicite, ne peut pas vous réafficher, par exemple, la date de la dernière recherche que vous avez faite, sachant que l'information n'est pas stockée sur un serveur, elle est stockée en local sur votre ordinateur, elle ne transite même pas vers la boite qui édite cette solution logiciel, donc y a des cas où c'est quand même beaucoup de frictions et pour les gens qui développent les applications et surtout pour les utilisateurs à la fin parce que c'est l'utilisateur qui doit ressaisir encore une fois sa recherche, que toutes les semaines il est obligé de la ressaisir. Donc c'est pas vraiment pour les développeurs d'applications, le problème c'est plutôt pour l'utilisateur à la fin où ça créé même des frictions d'usage qu'on pourra peut être éviter et donc il y a des progrès à faire là-dessus. Mais dans l'ensemble c'est une bonne réglementation.

Il y a des choses intéressantes à suivre sur l'intelligence artificielle. Alors bien sûr, même sujet, même combat : il faut encadrer de façon réglementaire les usages de l'IA. Et de la même façon que les données personnelles, c'est un choix politique, c'est un choix de société, ce serait bien que ces débats-là. Ça serait bien que ça soit porté au débat européen. J'aimerais bien que la pédagogie soit faite auprès du grand public peut être un petit peu mieux pour que ces sujets ne soient pas forcément des sujets très lointains, traités par l'union européenne par nos représentants, nos députés, nos parlementaires européens... les gens ont l'impression que c'est un peu loin, ils ne se sentent pas concernés alors que c'est un sujet qui mériterait d'être repris, d'être animé à une échelle locale beaucoup plus et qu'il y ait une culture générale de ces sujets qui soit mieux partagée. L'échelle européenne est indispensable mais en même temps notre capacité d'attention à ce qu'il se passe au Parlement européen est quasiment faible. Donc c'est le bon échelon pour cette réglementation mais cela l'éloigne aussi beaucoup des sujets quotidiens et des sujets très traités dans notre périmètre proche. Du coup, cela n'améliore pas la compréhension du grand public de ces sujets. Donc cela serait pas mal que cela soit des choses qui soient relayées et beaucoup plus proches des gens.

Et donc sur l'intelligence artificielle, je ne connais pas encore le texte là en préparation sur l'*IA ACT*. Je ne vais pas me positionner dessus mais c'est important qu'il y ait une réglementation. Après, il faut faire attention à bien différencier, à ce que cela ne prenne pas la recherche fondamentale ou le travail des chercheurs. Il faut différencier les usages économiques et les usages pour les besoins de la recherche parce qu'il ne faut pas laisser les autres pays (les États, la Chine) être les seuls à travailler sur certains algorithmes, sur certains modèles d'intelligence artificielle, sur certaines technologies, pour qu'on soit en mesure de connaître ces technologies qui peuvent nous arriver d'une façon ou d'une autre. Je sais pas, en reconnaissance d'image, je reprends l'exemple d'une personne avec qui je discutais récemment : on va pas s'interdire la reconnaissance de visages sur les images de caméras de surveillance par exemple. On s'interdit tous les programmes de reconnaissance faciale pour pas qu'on puisse traquer un individu dans un supermarché, une gare ou simplement le traquer avec du ciblage publicitaire, de le reconnaître et de lui envoyer de la pub contextuelle dans le lieu dans lequel il sera passé parce que ça instaure un *tracking* des individus auxquels ils ne vont pas consentir, aucun acteur économique n'a besoin de savoir exactement quel était le déplacement de quelqu'un. Donc ça bien sûr mais par contre on peut avoir besoin de reconnaissance faciale dans le cas d'un enlèvement d'un enfant et on veut le retrouver vite.

Et donc si on s'interdit de faire de la recherche sur la reconnaissance, la détection de visages automatique, si on s'interdit de faire de la recherche on ne peut pas bénéficier de cette technologie pour des usages qui sont plus légitimes. Et encore il faudrait débattre est-ce que c'est légitime d'utiliser cette reconnaissance d'image dans le cadre d'un enlèvement d'enfant. La question peut être posée. Mais surtout sur des sujets comme les *fake news*, les *deep fake* (reproduire le visage d'un individu et lui faire dire des choses), ces technologies là c'est aussi le fait comment sont créées ces *fake news* et ces faux programmes et c'est aussi ça qui permet de mieux les reconnaître et de mieux lutter. Si on ne s'intéresse pas du tout à ces algorithmes là en disant « nous on interdit tout ça, on interdit toute recherche là-dessus », ben on sera pas non plus en mesure de s'en protéger, de s'en défendre, de les reconnaître quand ils viendront des États-Unis, de la Chine ou d'ailleurs.

Donc c'est important d'encadrer mais il faut veiller à ce que néanmoins notre connaissance, notre capacité à faire de la recherche sur ces sujets ne soient pas limitées mais je ne pense pas du tout, je ne pense pas du tout que la réglementation européenne limite nos capacités à faire ce genre de projets.

Comment selon vous on peut mettre en place une acculturation aux données dans un organisme et surtout est-ce que ça doit concerner tout le monde ou certaines personnes en particulier ?

Très intéressant. Je vous invite à voir ce que fait AirBnb sur ce sujet qui est bien documenté, ses programmes pour améliorer la *data literacy* de tous ses salariés. C'est un sujet très important et je défends l'idée que c'est un sujet qui doit concerner tous les membres d'une organisation, tous les salariés d'une entreprise parce qu'il en va de la capacité de cette entreprise à s'adapter aux évolutions de son environnement, à améliorer sa façon de travailler, à améliorer sa capacité à comprendre l'information. Donc ce n'est pas un sujet qui doit être réservé à certains départements ou certains métiers. Et du coup ça pose une question qui est vraiment importante : comment on met tout le monde dans une organisation en capacité de faire lui-même ses requêtes dans une base de données pour recueillir l'information dont il a besoin, de comprendre un *dashboard*, d'avoir les bons réflexes quand il reçoit une information ou quand il manipule des données essentielles. Cela passe par une série de choses : on pourrait se dire, on va recruter que des gens qui ont eu un background qui les a exposés à faire un peu de *data science* ou d'analyse de données… cela nous prive d'une partie importante de la population pour le recrutement. Peut-être à terme ça sera le cas quand les programmes scolaires auront intégré ça comme socle de base mais c'est pas le cas aujourd'hui. Donc on ne peut pas tellement se dire qu'on va recruter que des gens qui sont bons avec les données, donc il faut former les gens. Il y a pas d'autres solutions que de la formation. La formation, la formation interne… la propagation de bonnes pratiques, faire des témoignages, des petites réunions pour présenter comment les gens font, des manuels, des webinaires, tout ce qu'on peut inventer comme formats de transmission de connaissances, de partage de bonnes pratiques et de partage de connaissances.

Ça passe aussi par les outils et ça c'est pour le coup mon métier. C'est aussi des outils de requêtage qui ne nécessitent pas de savoir coder, c'est aussi des outils de visualisation qui permettent de comprendre de l'information qui vient des données sans avoir forcément un background en data science important. Qui permettent de traduire des grandes quantités d'informations dans un langage qui est plus familier pour les humains, dans un langage visuel, interactif. Et donc c'est aussi un gros travail sur les outils qui permettent à des gens qui ne sont pas experts de l'analyse des données de travailler avec une information qui vient des données.

Et enfin, il y a un travail de gouvernance, mot très à la mode dans le monde de l'entreprise. Il y a un vrai travail de gouvernance parce que le cas classique c'est : il y a plusieurs départements qui vont utiliser les mêmes indicateurs. Je ne sais pas, par exemple, des indicateurs comme le chiffre d'affaires, il va y avoir le service comptable, la gestion compta qui a sa propre définition du chiffre d'affaires. Et du côté marketing ils ont aussi leur propre définition du chiffre d'affaires. Et puis côté ventes, ils ont encore une autre définition. Et puis à la stratégie ils ont encore une autre définition. Soit parce qu'ils ont vraiment des façons différentes de le calculer en intégrant le prévisionnel ou pas... soit parce qu'ils ont tous la même façon de le calculer mais ils le font tous, certains le font trois mois plus tôt, d'autres plus tard. Certains prennent une base qui est dressée, d'autre pas, et à la fin ils n'ont pas tous le même résultat. Et pourtant ils ont tous l'impression d'utiliser le même indicateur. Et donc chaque fois qu'on parle du chiffre d'affaires, personne n'a le même chiffre et donc pour le coup ça créé de la perte de confiance dans les données et à la fin les gens disent « de toutes façons, les résultats, tout ça c'est biaisé ». Donc pour le coup je prends un cas caricatural mais ça arrive très souvent parce que justement donc comme tous les départements d'une entreprise sont concernés par cette massification des données, tous s'adaptent, font leurs indicateurs, font leurs petites tambouilles, mettent en place des systèmes de récupération d'information, etc. Et en fait, il n'y a pas eu forcément d'homogénéisation dès le début avec « voilà, on définit chaque indicateur un par un, y a qu'une seule façon de le calculer, y a qu'une seule personne responsable de sa définition... ». Et donc ça fait que dans les grosses entreprises surtout, y a une espèce de sédimentation, de strates, de KPI définis à droite à gauche, de façons d'extraire de l'information, de bases qui cohabitent, etc, qui fait que globalement c'est un peu le bordel et tout le monde a l'impression que c'est pas très fiable, c'est pas très maintenu, etc. Et ça c'est justement un problème de gouvernance des données : il faut mettre en place des mécanismes qui permettent d'avoir confiance dans les indicateurs qui sont extraits des données, dans l'information qui est tirée des données, et ça passe par une organisation qui va définir voilà quelle est la définition officielle dans notre entreprise de cet indicateur et il y en a qu'une. Si on veut faire une définition différente, si ça doit porter un nom différent, il y a une méthodologie : la définition de chaque indicateur, de chaque KPI, il y a une personne qui est responsable de maintenir cette méthodologie. Éventuellement la faire évoluer pour les besoins de l'entreprise. Il y a une « source de vérité unique » entre guillemets, une base de données de référence où il y a : ça doit être possible de savoir si l'indicateur est à jour ou pas donc il y a des statuts associés à chaque indicateur, de quand est-ce qu'il a été mis à jour, où a été prise l'information, il y a tout un tas de documentation, de méta informations autour des données qui sont présentées qui va permettre d'avoir un ensemble de faisceaux d'informations qui nous permettent d'avoir confiance. Donc ça enlève beaucoup d'activités mais il faut définir, il faut poser à plat les méthodologies, il faut nommer les personnes responsables, il faut d'un point de vue technique harmoniser les bases et les façons d'enquêter. Donc ça relève à la fois de la tuyauterie des systèmes techniques et de la méthodologie, des définitions, des process qu'il faut harmoniser et sur lesquels ils faut qu'une personne soit en charge, soit garante de maintenir cette organisation propre. Et ça c'est la gouvernance. Et aussi, qui est responsable si cet indicateur n'est pas à jour, qui je dois aller voir si j'ai une question, etc. Et derrière aussi, qui est responsable en cas de problème avec les données personnelles en cas de fuite, qui s'assure bien qu'on respecte la réglementation RGPD dans notre usage des données personnelles. Et donc ça c'est de l'organisation au sens, voilà, des postes, des responsabilités, des tâches, des attributions, des systèmes techniques qui vont derrière et c'est un autre enjeu important pour assurer que les gens ont cette *data literacy*, c'est-à-dire ne se désintéressent pas de ces sujets, ne considèrent pas que les données ne sont pas fausses ou ne sont pas à jour, et aussi se sentent en sécurité quand ils travaillent avec cette matière. En sécurité dans le sens où

ils n'ont pas l'impression de toujours manipuler un truc biaisé, faussé, de pas savoir lequel prendre, ça enlève aussi des frictions.

Donc voilà les trois étapes : formation, et je pense que c'est le plus important. Outil. Et gouvernance.

Qu'est-ce qu'il est important pour vous que les personnes développent à l'issue de la formation comme des connaissances de base sur la donnée et ses enjeux, des compétences techniques... que doivent-ils développer en priorité ?

En priorité une compréhension de base de ce que sont les enjeux. Et pour aller jusqu'au bout, pour faire le job, il faut quand même aller jusqu'à une compréhension un peu technique de ce qu'est une base de données, de comment on requête dedans. Même basique : l'objectif est pas forcément que tout le monde sache coder en SQL ou faire une requête SQL à la fin de la formation. Si on y arrive c'est pas mal. Mais au moins quelques notions de base là-dessus, comment ça marche.

Il y a une expression qui revient souvent qui est d'être critique vis-à-vis de la donnée, l'esprit critique vis-à-vis de la donnée. Est-ce que cela vous parle et si oui, que mettez-vous derrière ?

Ni plus, ni moins, que l'esprit critique en général. C'est pareil par rapport à une image, elle peut être totalement manipulée. On n'a pas attendu les intelligences artificielles pour avoir des photomontages et de l'information manipulée quand elle est présentée sous formes d'images. L'esprit critique on en parle beaucoup par rapport justement à la propagation des *fake news* mais y a rien de nouveau. Et les *fake news*, c'est pas non plus nouveau mais elles ne venaient pas forcément d'une intelligence artificielle avant. Je ne vois rien de neuf sur l'esprit critique par rapport aux données. Ça peut prendre une ampleur plus importante mais il faut toujours s'intéresser à la source, à l'émetteur, à ce qu'on nous présente et ce qu'on ne nous présente pas, et ça c'est dramatique mais ça devrait tellement être enseigné à l'école. C'est pas assez souvent le cas ou ce n'est pas assez bien fait. C'est important de l'appliquer aux données, c'est sûr, mais le même esprit critique par rapport à n'importe quelle source d'information pour moi.

Quel est le niveau d'acculturation aux données des publics que vous rencontrez (vos clients, vos collègues) ?

Vu que je travaille dans ce domaine, mes collègues ils sont plutôt sur un bon niveau de *data literacy*. Dans le monde des entreprises, dans le monde de nos clients, ça a progressé, certes, mais parce que leurs métiers ont évolué. Ils n'ont pas eu d'autres choix que de regarder. Je trouve quand même que ça pourrait être mieux. Et pour le grand public, je trouve que c'est catastrophique. Autant dans les métiers qui ont été confrontés à ça, de par l'évolution de leurs métiers, les gens ont été forcés de s'y intéresser et donc peuvent avoir quelques réflexes, quelques compréhensions, ont acquis certaines connaissances sur des outils, certains réflexes liés à leur métier. Mais par contre je ne suis pas sûr(e) qu'ils soient capables de le transférer à d'autres métiers ou à des choses plus quotidiennes, pragmatiques. Soit tu peux avoir quelqu'un qui est ultra formé et qui travaille sur des données sensibles, qui est ultra formé à ne pas laisser échapper des données personnelles, ou les manipuler de façon à ce qu'elles conservent leur anonymisation et néanmoins sur les réseaux sociaux poster des choses qui sont

potentiellement problématiques par rapport à la protection de sa vie privée. Je pense que ça arrive souvent. Je pense que ce sont des savoir-faire métier avant d'être une compréhension vraiment culturelle, générale de ces phénomènes mais bon je suis peut-être un poil pessimiste.

Mais par contre, sur le grand public, en-dehors des métiers qui travaillent avec des données, je suis quand même très pessimiste. Je trouve que cette compréhension elle est très faible, elle s'accompagne de beaucoup d'appréhension et de crainte, justement par rapport à quelque chose que l'on ne comprend pas. Et elle est très dépendante de quelques cas qui sont parus dans les médias, qui ont fait un peu de buzz, elle est très biaisée par ce prisme médiatique. Et elle est très peu technique. Des fois je suis abasourdi(e) de voir que des gens ne savent comment fonctionne l'envoi d'un *e-mail*, même pas le fait de ne pas savoir comment appuyer sur un bouton parce que tout ça, l'expérience utilisateur fait que c'est guidé. Mais de savoir comment ça marche, par où ça circule, des basiques de compréhension des réseaux. C'est un peu… tout à l'air magique. Les gens du grand public, beaucoup de personnes vont avoir une approche de ces questions techniques comme quelque chose de magique : ça marche mais on ne sait pas comment. Et donc du coup, ils ne peuvent pas du tout anticiper les risques, les problèmes. Ils peuvent éventuellement avoir acquis des réflexes du type « tu n'envoies pas des photos de toi sur un réseau social » parce que, par l'expérience, tout comme on a appris à ne pas mettre sa main dans le feu parce que c'est chaud quand on est petit et que ça brûle, on a appris à ne pas envoyer de photos de soi nues sur un réseau social parce que ça peut se retourner contre soi. Il n'y a pas forcément une grosse compréhension derrière de comment ça marche et de qu'est-ce qu'on fait de ces données, où est-ce qu'elles sont stockées, pour combien de temps elles sont stockées, quelle est ma capacité d'y accéder, est-ce que ça va être public, qui y a accès, etc. Et donc du coup les gens ont une petite connaissance par empirisme de ces sujets mais un tel manque de culture générale technique que ça reste difficile pour eux, pour beaucoup de gens, d'y voir autre chose qu'un truc magique et ça me semble un énorme problème pour comprendre le monde qui nous entoure. Je suis très dépité(e) de voir que ça pénètre la compréhension technique. Quand j'étais à l'école primaire au collège, c'était le début des ordinateurs à l'école. Je n'en avais pas à l'école primaire, au début du collège ça commençait à rentrer mais tout le monde n'avait pas un ordinateur à la maison. Donc ok à cette époque-là ce n'était pas encore ça mais maintenant c'est très différent ! Les enfants ont un téléphone dans les mains avant même d'aller à l'école donc à aucun moment se poser la question de la formation aux usages de ces technologies me semble quand même être un gros sujet et je n'ai pas l'impression que les programmes scolaires aient évolué et ça me choque beaucoup. Qu'on peut toujours chercher à rattraper après avec de la formation professionnelle, avec de la formation spécialisée… c'est vraiment le monde qui nous entoure, c'est comme apprendre que si tu mets la main dans le feu ça brûle. Donc ça pourrait être pas mal qu'on se préoccupe plus de la formation des enfants à ça.

Par rapport à cette nécessité de formation, quels sont les freins à la non acculturation aux données ?

Cela prend du temps de faire de la formation. Il faut trouver des gens qui font ça bien. On me demande souvent de faire des interventions sur la visualisation de données. Je fais des interventions quand elles ne sont pas longues. Et toutes les fois où on m'a proposé de faire un cours dans des cursus en journalisme, en marketing, pour des étudiants à l'université, à chaque fois qu'on m'a proposé un cours de 15/20h, dans mon métier je ne peux pas libérer 15/20h dans le mois dans mon emploi du temps, même si vous me rémunérez pour ça, j'ai un

autre métier. Donc c'est un peu le problème de ces sujets qui sont très récents : c'est que les gens qui connaissent le mieux ces sujets sont ceux qui travaillent sur ces technologies là et par définition ne sont pas professeurs. Et donc avant qu'il y ait des formateurs spécialisés, il faut qu'il y ait un petit temps pour qu'il y ait de la recherche, pour qu'il y ait des gens qui aient une activité professionnelle sur ce secteur et qui aient envie de changer et de faire autre chose donc ça prend toujours un peu de temps. Il y a un petit décalage. Et je ne pense pas qu'il y ait beaucoup de boites qui fassent de la formation à la *data literacy*. Je n'en connais pas en tout cas. Je pense que c'est quelque chose qui relève beaucoup de la formation interne parce que ça dépend beaucoup des enjeux de l'entreprise, de ce que fait l'entreprise. Ce n'est pas mal si c'est fait en interne. C'est aussi peut-être souvent fait en interne que c'est fait parce qu'il n'y a pas beaucoup de formateurs externes. Et donc un des freins, ça demande du temps, des moyens, de faire une formation correcte. Et souvent les personnes en charge des formations dans les entreprises c'est pas des gens qui sont très techniques et donc c'est aussi un travail d'évolution des métiers RH, de développement des carrières, etc. Donc la formation c'est pas si facile que ça à mettre en place : ça demande du temps, des ressources dédiées.

Tout la partie sur la gouvernance, c'est pareil, ça demande du temps et des ressources. Ça c'est un bon frein.

La partie sur les outils, soit ce sont des outils internes qui vont être développés, ça va demander du temps et des ressources, soit c'est encore une fois des outils qu'on achète et ça demande aussi de l'argent.

En fait, c'est un investissement. Et on peut se dire que c'est un investissement qui est rentable parce qu'une entreprise elle va être mieux adaptée à son environnement, être plus performante et avoir des salariés qui sont mieux formés. Qui peuvent se dire si ça tourne comme ça, on peut continuer comme ça.

Avez-vous en tête un projet de culture des données et si oui pourquoi ?

Faut vraiment que vous regardiez, faites s'y moi penser après notre entretien, je vois envoie une série de liens sur comment *AirBnb* a adressé cette question-là. Je n'ai pas travaillé chez *AirBnb* donc dans les faits je ne sais pas si ça c'est passé comme ils le décrivent dans leurs articles, mais si c'est le cas je trouve que c'est vraiment brillant. Ça me semble combiner tous ces trois approches dont je vous ai parlées. La formation, les outils, la gouvernance. Et c'est raconté par rapport à l'histoire de l'entreprise et à ses enjeux, et ça décrit quel programme de formation ils ont mis en place, quels outils ils ont mis en place et c'est vraiment intéressant. Je vous enverrai, il y a une série d'articles dessus.

Auriez-vous des recommandations ou des remarques pour améliorer l'acculturation aux données en France ?

Si je dois résumer et en garder qu'une c'est d'améliorer l'éducation des enfants sur ces sujets. Et bien sûr après je veux faire ma pub : utilisez des outils de visualisation de données !

Annexe 13 : les entretiens avec le public « institutionnel »

Institutionnel 1 (20/12/21)

Comment définissez-vous la culture des données puisque vous employez ce terme ?

Alors petit préambule : c'est vrai que ces textes que vous avez pu lire, que j'ai pu moi-même écrire, voire co-écrire pour certains avec d'autres collègues de réseau, ce ne sont pas forcément des textes qui ont vocation à fonder le concept. Ce qu'on a essayé de faire nous, ou moi en tout cas, c'est de voir comment les données pouvaient renseigner le travail et les tâches métier des enseignants. Donc il y a un élément culturel autour duquel on pouvait pivoter ces professionnels qui renseignent la pratique, les pratiques éducatives, et que c'était intéressant d'essayer de voir en quoi ça pouvait consister. Donc c'est un peu comme ça qu'on a travaillé sur cette notion de « culture des données » et on l'a défini en termes de « littératie des données », c'est-à-dire le « lire, écrire, compter ». Et effectivement, quand on s'est penché sur le comment lire et pourquoi lire, écrire, compter avec les données, à un moment donné on s'est posé la question de la « culture des données » parce que derrière, et d'ailleurs on s'en rend compte encore de manière assez pratique dans mon quotidien, que la notion de données elle n'était pas clairement comprise par le milieu enseignant et qu'il y avait un travail d'acculturation à faire, de construire une culture, et c'est pour ça qu'on a parlé de culture des données. Je ne sais pas si ça vous renseigne mais en tout cas c'était un peu la démarche qui était la nôtre. Et du coup, c'était une démarche assez pragmatique avant même l'idée même d'une culture des données.

Justement, vous employez le terme de culture des données, vous avez parlé d'acculturation : si on devez définir une « nature » de la *data literacy*, quelle serait-elle ? Une culture ? Un ensemble de pratiques et de techniques ? Un état d'esprit ? Tout ça à la fois ?

C'est pour ça qu'on a employé le terme de culture : la culture c'est à la fois une compréhension on va dire globale, conceptuelle, une entrée intellectuelle sur un domaine, et c'est aussi une pratique, parce que la culture elle se vit et elle se fait et elle se concrétise, donc c'est pour ça qu'on a parlé de culture. De comprendre, quand on parle de données, de quoi il s'agit, à quoi on fait référence, d'avoir en fait un peu tout le bagage définitionnel derrière. De savoir un petit peu comment ça impacte la vie de chacun : c'est pour ça qu'on parlait aussi de culture, en fait ça nous définit dans notre être à la société et de notre rapport à l'autre, c'est pour ça qu'on parlait de culture. Et avec la culture, l'idée de pouvoir créer et d'apporter les contenus nouveaux. Voilà pourquoi on a parlé de culture des données, c'était de global en fait.

Est-ce que d'un point de vue conceptuel, vous arrivez à replacer la culture des données vis-à-vis d'autres concepts comme la gouvernance des données et la *data science* ? Est-ce équivalent, en relation ?

Quand on parle de culture des données, on parle aussi de repères qui permet d'englober les champs de la donnée sans être enfermé dans l'un ou dans l'autre, d'avoir la vue globale générale. Pourquoi je dis ça ? Parce que moi,

mon public principal c'est le monde de l'éducation, essentiellement des cadres de l'éducation, des enseignants. Et quand on leur parle de données par exemple, la première idée qui leur vient en tête, il y en a plusieurs, on parle en tout cas aux praticiens que sont les enseignants, c'est la question de la donnée personnelle et sa nécessaire protection. Il y a ce champ-là. Quand on parle parfois avec les cadres, ils en sont aussi à cette approche de la donnée qui est la donnée en tant que source et processus de récolement pour comprendre les processus cognitifs des élèves, la donnée qui a été collectée dans les environnements numériques notamment. Après chez les enseignants, il y a aussi cette notion de donnée qui est lié aux concepts de … que l'on trouve très tôt à l'école. L'idée c'était d'englober l'ensemble des champs et puis après il y a aussi un certain nombre de conclusions chez eux pour des termes relatifs à la donnée : le *big data* par exemple. L'*open access*, ils sont un peu perdus, ils ont du mal à faire le distinguo. C'est donc leur apporter des éléments pour savoir de quoi on parle.

Si on devait définir la donnée, ce grand concept qui peut être encore flou aujourd'hui, comment le feriez-vous ?

Quand je parle de données avec mes collègues, si on est dans la granularité de l'information, c'est l'élément le plus simple qui, en lui-même, n'a aucun sens. C'est juste un enregistrement qui, oui, en relation avec d'autres enregistrements, d'autres données, va finir par produire ou permettre de produire une information. Cela n'a pas, elle n'a pas de sens toute seule.

Vous l'avez cité tout à l'heure quand vous avez évoqué les données personnelles : actuellement, quels sont les grands enjeux auxquels on doit faire face par rapport aux données et comment la culture des données permet d'y répondre ?

Alors dans le domaine dans quel j'interviens, c'est-à-dire plutôt dans l'éducation, il y a ce concept de données personnelles qui est vu comme un enjeu très fort. Permettre à chacun de garder la main c'est beaucoup dire. En fait, de pouvoir arbitrer sur la manière dont les données, qu'ils vont générer dans leurs comportements lors de leurs épisodes de connexion, comment ils vont pouvoir arbitrer afin de garder ou non la main dessus et de continuer à évoluer dans leurs environnements numériques. Donc il y a ce concept-là qui est vraiment très fort, qui est incarné dans l'éducation, par l'éducation aux médias et à l'information notamment, il y a beaucoup d'interventions dans ce champ-là. Donc nous c'est pas forcément le domaine, le champ qui nous a le plus intéressé d'emblée parce que je pense qu'on n'aurait pas eu grand-chose à apporter dessus, même si je pense qu'il y a encore des éléments d'éclairage à apporter : moi, je le vois quand je rencontre des jeunes, des enseignants, ou même des parents. La manière dont sont traitées les données ou les traces qu'ils déposent lors de leur épisode de connexion, ils ne sont pas du tout conscients donc là je pense qu'il y a un effort à faire, non pas dans la manière, pas que sur la manière dont les données sont récoltées mais sur la manière dont elles sont traitées ; je pense que là il y a un truc à faire. Donc il y avait ce champ-là, ce champ de l'EMI avec les questions de traitements algorithmiques qui vont derrière par exemple, et de mise au travail ou non d'intelligences artificielles. Ça, ce n'est pas le champ qui nous a le plus intéressé.

Nous, le champ qui nous a le plus intéressé, c'est le « lire, écrire, compter » avec les données. C'est pour ça qu'en fait, on a mis en œuvre ces dispositifs de datasprint, tout ça dans l'éducation parce qu'on trouvait que là on était vraiment dans autre chose. Et surtout, on était dans l'idée du, alors je sais pas si vous avez vu, j'avais

produit un graphique en forme de flèche qui monte vers le haut, inspiré d'un travail des Québécois sur les littératies numériques, et nous on l'avait repris pour donner une lecture sur ce que peuvent être les littératies, la littératie des données en tout cas. Et c'est vraiment sur ce champ-là qu'on avait travaillé. Parce que l'idée, c'était vraiment de se dire que les données, aujourd'hui, on vit dans un milieu qui est numérique : la donnée elle est partout. Il y a la donnée que l'on produit de manière consciente ou inconsciente, mais il n'y a pas que ça. Il y a aussi la manière dont on peut produire du savoir ou de l'information avec les données. Ça, ça nous intéressait plus. On est plus allé dans ce champ-là.

Du fait de l'importance prise par la donnée, quelle est l'importance de la culture des données par rapport à d'autres nécessités comme la culture de l'information, la culture numérique ?

C'est un peu la logique qu'on a voulu... si je dis on, c'est parce qu'il y a des choses que j'ai faites et après je les remets à l'ouvrage ou sur l'ouvrage avec mon collègue d'Avignon parce qu'on travaille beaucoup ensemble. Le schéma je sais que je l'ai produit un matin comme ça : j'étais à mon bureau et je me suis dit « il faut vraiment qu'on le formalise » donc ça a pris cette forme-là et à la fin on l'a validé avec lui. Et l'idée qu'on avait derrière, c'était de celui de rendre un peu du pouvoir en fait dans cette capacité qu'on a d'agir dans un monde qui parfois nous échappe ou qui est enfermé et c'est là qu'on a vu l'enjeu principal. C'était de se dire aujourd'hui : les gisements de données sont partout. Il y en a qui sont ouverts : on peut s'en saisir pour produire de l'information et de la connaissance. Dans l'éducation, c'est important de le savoir et de voir comment on peut faire, ce qu'on peut faire. Ce qui nous a semblé aussi important, c'est parce qu'on avait lu forcément la loi pour une République Numérique de 2016, et on s'est dit c'est aussi permettre aux enseignants et aux élèves de construire des habiletés numériques, d'agir en tant que citoyens en étant coproducteurs ou producteurs de contenus d'information à partir de gisements de données et que, si on faisait ça, en plus on pouvait, un petit comme ça existe dans les dispositifs d'apprentissage par la recherche mais pour l'école, faire des élèves des petits chercheurs avec les données, faire en sorte que, tout en travaillant des points au programme, ils arrivent aussi en contexte à développer des habiletés et des compétences numériques.

De ce que je comprends, l'arrivée de la démocratisation, à défaut d'autres termes, de la donnée, qu'elle ait maintenant imprégner l'ensemble de la société, j'imagine qu'en tant que professionnel cela a dû changer vos pratiques : manière d'enseigner, manière de transmettre...

Je suis professeur documentaliste de formation et cette histoire de données je ne l'ai pas choisie, elle me poursuit. Quand je suis arrivé dans l'éducation nationale, très vite on a demandé d'intervenir dans le plan de formation des professeurs-documentalistes et d'accompagner les entrants dans le métier. J'ai accepté et on m'a donné très vite, parce que j'y étais sensible, à accompagner ces collègues autour de la question de la donnée, parce qu'on commençait à parler du web sémantique et tout ça et en fait c'était jamais traité et ça m'est tombé dessus. Donc en fait, ça n'a pas changé grand-chose dans mes pratiques professionnelles, ça les a orienté en quelque sorte. Donc très tôt j'ai commencé à travailler avec eux sur la constitution de bases de données, sur l'architecture des bases de données, et après ça m'a accompagné tout le temps. Donc la donnée dure, la métadonnée aussi bien sûr, cela n'a pas trop changé mes approches professionnelles, cela les a plus accompagnées, orientées.

Et est-ce que vous pouvez dater d'à partir de quand cela a commencé à être prégnant l'arrivée des données ?

Dans ma pratique ?

La vôtre et même dans la profession en général.

Moi j'ai commencé à en parler à mes collègues vraiment, ça devait être en 2005. Et cela a pris une place vraiment plus importante un peu après. Cela a pris une place plus importante, je sais pas, on est en 2013-2014.

Vous avez évoqué la loi Lemaire, on a maintenant le RGPD : est-ce que ce dispositif législatif a modifié les programmes, les pratiques quotidiennes, est-ce que ça a impulsé quelque chose ou au contraire réduit le champ d'action ?

Je trouve que le *RGPD* a phagocyté un peu les approches. Les approches réglementaires et sécuritaires du truc, c'est-à-dire que d'un seul coup, on a compris que c'était un peu la seule entrée qui comptait. J'ai trouvé l'impact du *RGPD* dans ma pratique plus fort que le cadre de confiance qui était donné la loi de 2016.

Il y a vraiment une différence entre les deux que vous pouvez citer ?

C'est-à-dire qu'au moment où est sorti le *RGPD*, beaucoup de demandes qu'on nous faisait rentraient dans le souci du *RGPD*. Est-ce que ce que l'on faisait c'était garant du *RGPD* ? Il fallait garantir la protection des données personnelles des élèves, ce qui pour moi est une vraie question. Aujourd'hui, quand on parle de données, là où je suis un peu parfois en décalage c'est qu'aujourd'hui il y a la, comment dire, cette espèce de mirage du solutionnisme de la donnée. C'est-à-dire qu'on va avoir des gisements de données un peu partout, on va mettre de grosses machines autour pour les traiter et puis à la fin ça va tout régler. Je schématise volontairement. Cette approche-là me parait parfois… on maîtrise enfin le sujet et ce que j'aime bien justement sur l'espèce d'être dans une logique d'encapacitation, d'*empowerment* autour de travailler avec les données, ce que j'aime bien c'est que c'est une approche centrée sur l'individu et le fait de se dire qu'on va lui rendre, lui donner, une capacité d'action. Ce qui est moins le cas dans le cas que j'ai évoqué avant : tu prends plus le sujet comme un objet que comme un sujet avec une capacité d'action. On voit ça avec les neurosciences. C'est-à-dire que plus on va récolter de données sur les processus cognitifs d'un individu, plus on va être capable de s'adresser à lui pour qu'il apprenne. J'ai rien contre ça mais du coup [inaudible] plutôt que de se servir de ses traces pour faire en sorte qu'il soit pleinement opérationnel dans un univers social en quelque sorte.

C'est dépasser les outils, la version techniciste, pour passer à quelque chose de plus humain. Puisqu'on en parle, si on veut mettre en place une culture des données dans un organisme, dans l'éducation nationale par exemple, par quoi faudrait-il commencer ? Quel budget ? Quels types de moyens ? Et est-ce que ça doit concerner tout le monde ?

Alors je ne sais pas si c'est qu'une question de budget. Je pense que c'est une question d'approche. Je vais ancrer mon propos dans le réel de ce que l'on vit aujourd'hui dans l'éducation nationale. Vous avez dans l'éducation

nationale, vous en avez peut-être entendu parler, ce que l'on appelle les territoires numériques éducatifs. Dedans, il y a un axe intitulé « agir et apprendre dans un monde de données ». Je suis un peu responsable de cette appellation parce qu'avant ça s'appelait « les datas et leurs enjeux », ce qui en termes de lecture pour moi n'est pas du tout la même chose. Et du coup, là il y a du budget si l'on veut, avec de la formation derrière qui s'adresse aux enseignants pour les sensibiliser à ces questions-là. Ce dont on se rend compte, c'est que quand on alimente cet axe de formation, le mot de donnée fait peur, du coup il n'y a pas une appétence très forte autour de cet objet ou de ce sujet-là. Donc je pense pas que ce soit juste une question de moyens qu'on met, c'est une question d'approche que l'on développe et la manière dont on va s'adresser au public.

Donc qui ça concerne pour répondre à votre deuxième question, eh bien je pense que ça concerne beaucoup les cadres déjà parce qu'il faut qu'ils comprennent de quoi on parle. Et puis ensuite, forcément le public cible, c'est les enseignants. Pour que ces approches et ces méthodes soient lui [coupure de son] jusqu'à la classe. Donc maintenant comment on fait pour s'adresser à eux ? Je pense qu'il faut prendre des chemins détournés pour leur montrer que la donnée est partout. On peut travailler avec les données sans forcément travailler sous une forme numérique, je m'explique (et là je vois que ça marche) : vous proposez un programme de formation à des enseignants autour d'une question quelle qu'elle soit. Prenons par exemple les objectifs de développement durable qui sont une entrée transverse. Vous proposez à des enseignants, vous leur dites « eh ben voilà, en fait, on va faire de la récolte des données pour voir un petit peu comment on peut traiter des questions de gaspillage ou de bien-être alimentaire ». Vous leur faites réaliser des grands tableaux que vous mettez en fond de classe avec des symboles à cocher, et puis comme ça vous faites de la quantification. Puis vous transformez cette quantification en visualisation en utilisant par exemple des briques de *Lego*. Et vous êtes capable de voir combien de kilos de vêtements ont été ramassés dans la cour, donc ils sont potentiellement perdus ou gaspillés. Vous faites une autre couleur, une autre catégorie et vous avez les déchets alimentaires. Vous faites une autre couleur, une autre catégorie pour les quantités d'énergie dépensées dans l'établissement et puis là vous travaillez avec les données puisque vous avez mis en place un processus de récolement des données. Vous travaillez vos compétences mathématiques parce que vous faites additions, des comparaisons, etc. Vous faites de la visualisation de données de manière analogique comme ça on comprend tout de suite et ça donne à voir. Et puis à la fin ce que vous avez fait, c'est pas travailler les données, vous avez travaillé les objectifs de développement durable dans votre établissement et tout ça c'était fait de manière intégrée. Quand on fait ça, on est sur une approche très concrète : on désacralise un petit peu, on décomplexifie le sujet, mais à la fin, en fait, on a travaillé avec des données et on a mis un petit pied à l'étrier. Quand on travaille avec les datasprints, c'est exactement la même chose qu'on fait, sauf que là on travaille avec des données qui sont des données ouvertes qu'on est allé récolter à l'extérieur. On prend les lieux de naissance, les dates de mort, etc., des poilus. On les traite, on fait de la cartographie, on arrive à dessiner les lignes de front, on arrive à dessiner le portrait robot du poilu… en fait, on a fait de l'histoire en pratiquant les mathématiques et la recherche documentaire. Et qu'est-ce qu'on a fait pour faire ça ? On a traité des données. Et quand on arrive à faire faire ça à des enseignants, je pense qu'on leur fait toucher du doigt ce qu'est l'environnement de la donnée, comment on la traite, où est-ce qu'on la trouve, comment on l'enrichit et comment on la transforme pour éclairer un programme. Et si on s'adresse à des enseignants comme ça, je pense qu'on y aller. Il faut rentrer par les pratiques disciplinaires, par le développement de compétences numériques en les mettant en relation avec des référentiels type *PIX*. Et c'est bien en s'inscrivant dans les compétences, dans les programmes qu'on y arrive. L'idée c'est juste de traiter ou d'utiliser la donnée comme un matériau qui permet d'alimenter le cours, exactement comme on le ferait

avec un texte, une vidéo, etc., sauf que là on a un matériau d'une forme nouvelle, enfin nouvelle, pas nouvelle ! Mais pour l'enseignant, nouvelle en tout cas.

C'est enrichissant de voir comment tout cela se met en place. On parle beaucoup de la donnée maintenant mais il y a encore beaucoup de mal à l'appréhender. Et vous y avez un peu répondu : la culture des données, qu'est-ce que les personnes doivent développer à l'issue de cet apprentissage ? Une culture, des pratiques, un esprit critique, tout ça à la fois ?

Quand on m'a demandé de faire un parcours de formation sur les données, je l'ai divisé en six axes. Il y avait un premier axe sur l'aspect, on va dire, général, comprendre de quoi on parle quand on parle de données. Tout ce qui est aspect définitionnel, contexte social, essentiellement ça. Après, il y avait un deuxième axe sur comprendre le traitement des données et notamment le traitement algorithmique des données parce que ça, ça nous était demandé. On a parlé de recul critique. Il y a un autre axe qui était cette fois un axe manipulatoire, donc là qui passait par tout ce qui est lecture, manipulation de données, que ce soit de données tabulaires, y compris de données liées pour aborder la question de la métadonnée. Le quatrième axe qui s'intitulait « lire des représentations de données » pour qu'on voit un petit peu, une fois que le traitement a été fait, ben comment on peut lire et comment développer un point de vue critique sur le traitement de la donnée parce que … alors ça aussi, on avait mis quelques formations en ligne pour le faire, on faisait des corrections de graphes, c'était assez éloquent en corrigeant les échelles par exemple. Après il y avait un autre aspect, c'était de les produire ces représentations de données donc là on était dans la création. Et puis le sixième c'était le « collaborer autour des données ». On retrouve ces entrées qu'il y avait dans ce schéma lié à l'*empowerment* : donc il y a comprendre, traiter, lire, produire, créer, collaborer autour de la donnée.

Donc il y a un cœur de compétences.

C'est ce cœur-là qu'on avait identifié. Et puis c'est vrai que quand on a mis en place concrètement les datasprint avec les élèves, on avait également, là c'est d'autres compétences, qui étaient plus liées aux compétences dites du XXIe siècle qu'on avait associées à la donnée [coupure de son] gérer une équipe, des choses comme ça. C'était plus pour emmener les élèves, les engager encore davantage. En fait, ce dont on se rendait compte, c'est qu'il n'y avait pas besoin de les engager par des artifices dès lors qu'on les mettait dans un travail collectif en leur disant « ce que vous allez produire, cela va être utile à chacun », ça c'était un moteur suffisamment engageant pour les garder en activité et faire en sorte qu'ils soient impliqués.

Sur les formations en général, on a cité les cadres de l'éducation nationale et les professeurs. Est-ce que les compétences actuelles, si on voulait se mettre dans l'idée qu'on enseigne une culture des données efficace, sont atteintes ou est-ce qu'on en est encore très loin ?

On est hyper loin. On est à ce point loin que quand je discute avec des représentants de la *DGESCO*, ils disent « ah mais vous n'imaginez même pas faire ça avec des élèves d'élémentaire ! ça n'a pas de sens ! » alors qu'on le fait très bien. Le premier datasprint qu'on a lancé, il y avait 25 classes qui y ont participé, ils ont tous produit des visualisations de données.

Est-ce du coup une méconnaissance du public ? Et deuxième question, est-ce que dans l'éducation nationale tout ce qui est mis à disposition : les manuels scolaires, le CLEMI, Canopé et autres, est adapté pour pouvoir proposer une pédagogie sur les données ? En fait, est-ce un manque de ressources ou un manque de connaissances général de la part du monde éducatif ?

Je pense que c'est une vision assez peu claire. Je rentre pas dans la question avec le même angle de vue que mes collègues dans l'éducation nationale donc il y a un hiatus là, mais en tout cas je pense que c'est un manque de vision d'ensemble du sujet et du coup on a, par exemple, un ministère et une partie du réseau Canopé qui est très dans un travail de proximité avec les [inaudible], du coup cette partie là de collègues entre un peu par le côté du solutionnisme dont on parlait tout à l'heure. On va développer une technologie basée sur le récolement des données pour accompagner le professeur par exemple dans sa pratique de différenciation pédagogique. Et du coup on rate un peu, quand on est là-dedans, le côté peut être éducation autour de la protection, de la préservation de ses propres données, à travers la gestion de ses traces notamment. Et ça le *CLEMI* va s'en occuper. Et puis enfin, l'utilisation de données on va en faire en sciences, en SVT, en Histoire-Géographie où on va faire des cartes, mais on va perdre de vue les deux autres pôles. Donc je pense, c'est l'ensemble, la vision intégrée, qui marche. Et puis, peut-être, la façon de se dire de comment cette question du traitement de la donnée aujourd'hui peut renseigner l'ensemble des disciplines.

Et en même temps c'est pas très nouveau, parce qu'avant les humanités numériques, c'était très compliqué de faire comprendre à des enseignants, au monde de la recherche par exemple en lettres ou histoire de l'art, que le numérique et l'exploitation des données pouvait renseigner la discipline en tant que tel. Dernière difficulté, quand on est dans une approche d'éducation à la donnée, sous cet angle-là du travail pédagogique avec un matériau qui serait des données, la grosse difficulté qu'on a, c'est qu'on est jamais dans une proposition pédagogique clé en main, qui soit basée sur un manuel où on va dérouler quelque chose qui fait qu'à la fin l'élève aura appris ce qu'on attend de lui. On est dans un acte pédagogique qui est beaucoup plus complexe et on passe dans une pédagogie qui interroge ce que l'on appelle la forme scolaire parce qu'il faut changer son point de vue d'enseignant : on est plus celui qui transmet mais celui qui fait avec, donc déjà ça ce n'est pas facile.

Ensuite on fait avec mais avec un matériau qu'on connaît mal, auquel il faut se confronter, donc déjà il y a un pas à faire. Et en plus, non seulement on le fait avec un matériau que l'on connaît mal, mais on va le faire dans un environnement qu'on maîtrise peu parce que c'est un environnement, c'est une technologie de l'écrit, qui est complexe, le numérique, et c'est vrai que souvent les enseignants ils sont aussi pressés et il faut que ça marche. Ils ont des programmes et en fait, soit ils ont pas le temps, soit ils ne peuvent pas prendre le temps, ou soit ils leur manquent des outils pour avancer en confiance avec leurs classes. Donc je pense qu'on se heurte à tout ça notamment. Et la solution, ça serait le rapport Jules Ferry 3.0 : vous nous mettez une épreuve au bac où il y ait un mobilier des habiletés numériques qui soit, pourquoi pas, outillé, alimenté par le traitement des données.

Vous anticipez ma dernière question qui est : quelles recommandations pouvez-vous faire pour améliorer l'acculturation aux données en général ? Vous me dites si j'extrapole mais la non visibilité de la donnée dans l'éducation nationale provient du fait qu'il n'y a pas une matière particulière entièrement dédiée à cela avec des épreuves dédiées à cela.

Ben non, il n'y en a pas. Et c'est vrai qu'en plus, moi je suis convaincu du fait que travailler avec les données, comme ça peut se faire dans toutes les disciplines, il n'y a pas ce dispositif interdisciplinaire qui serait appuyé sur la donnée. Et puis, on est dans un système encore, il faut se l'avouer, très cloisonné entre les disciplines. Donc vous voulez mettre de l'interdisciplinarité, c'est compliqué. Moi je l'ai vu dans le dispositif qu'on a proposé : il a fallu aller les chercher les enseignants ! Parce qu'ils nous disent « ah ben oui, moi, est-ce que ça va venir au bon moment dans mon programme ? », alors qu'ils sont maîtres de la temporalité de leurs programmes. Ça veut dire que si je suis dans le secondaire, il faut que j'aille travailler avec mon voisin. Mon voisin de technologie, mon voisin de mathématiques, mon voisin d'histoire-géographie ou de lettres : « ah ben non jpeux pas c'est compliqué ». Donc il y a ça. Et puis, par contre, la vision elle, elle existe parce qu'elle est formalisée dans le rapport Ferry 3.0 avec ce fameux bac « humanités numériques », et il y a un an, c'est l'*INRIA*, je crois, qui sortait ce document dont j'oublie le titre, « numérique et construction des savoirs » un truc comme ça, et tout est dit dans ces deux documents. Nous c'est dans ces approches-là qu'on a pu, que moi j'ai puisé avec d'autres, du moins ceux qui m'accompagnent, c'est là où on a construit notre vision. C'était avec le rapport Jules Ferry 3.0, tout ce que dit Bernard Stiegler autour de l'organologie, tout ça, on n'invente rien. Nous ce qu'on essaie c'est de trouver une solution pour le terrain, c'est tout.

Justement, mon but c'est de vous interroger sur votre vision, par rapport au terrain, faire émerger des solutions à ces problématiques-là autour de la donnée. Et en parlant de projets, vous aviez cité les datasprints, vous avez aussi parlé de la manière détournée de parler des données à travers les lego. Est-ce que vous avez en tête d'autres projets en tête autour de la culture des données, qui ont fonctionné et que vous pouvez citer ?

Alors, ceux qui ont le mieux fonctionné sont les datasprint. Ça c'est évident. Alors, les confinements nous ont mis un point d'arrêt parce qu'il y avait quand même beaucoup de travail, on va dire en classe. Nous on formait les enseignants mais après on les accompagnait quand même, du moins pour les premiers, donc il y avait une dimension physique qui ne nous a plus été permise par la suite. On en avait un dans les cartons justement qui était autour de la, qui était un datasprint autour… en fait, on passait à la moulinette justement du texte parce qu'on voulait s'adresser aux profs de lettres aussi, et on avait pris les discours de Victor Hugo à l'Assemblée, y a un corpus dans *Wikipédia* : on les retrouve et on peut ensuite les traiter. Mais celui-là n'a pas pu avoir lieu. Donc au moins les deux datasprints c'est sûr. Après l'an dernier, au mois de juin, en ligne dans le cadre de la classe dehors, on a proposé un exercice de cartographie de l'environnement culturel de l'école. L'idée c'était de permettre aux enseignants de produire de la cartographie de leur environnement scolaire pour identifier les monuments historiques répertoriés au ministère de la culture. Les accès aux bibliothèques, théâtre, etc., de manière à pouvoir, à l'époque c'était l'enjeu, faire sortir l'élève de sa classe. L'idée c'était de les faire sortir et de faire en sorte qu'ils s'approprient l'environnement extérieur. Ça a eu lieu. Est-ce que ça eu un gros succès ? Ça a eu un petit succès mais ça n'a pas eu un gros succès, même si on a eu du monde. Après on a aussi ce qui peut fonctionner et là on a des collègues du sud de la France qui s'engagent là-dedans sur tout ce qui est programme de sciences participatives : comptage, par exemple, autour de l'ornithologie, comptage des oiseaux, les ruches connectées, les choses comme ça, ça peut fonctionner. Et puis ce dont je parlais, cette approche un peu débranchée de la donnée : ce sont des approches de formation qu'on a porté, que moi j'ai portées. Et mes collègues de l'institution qui n'étaient pas sensibles à cette question de la donnée là sont rentrés dedans en fait mais je pense qu'ils n'ont pas vu là où ils mettaient les pieds. D'où l'approche un peu détournée.

On arrive à la fin : en plus de tout ce que vous avez dit, vous avez des remarques, des recommandations, ou en tout cas un mot supplémentaire sur cette thématique de la culture des données.

Non, je n'ai pas grand-chose à apporter, mon point de vue il est modeste. Les approches qu'on a pu avoir avec mes collègues et la manière dont j'ai pu le documenter, tout est accessible et tout est ouvert. Après, je vois que le sujet d'année en année... alors, ce n'est pas toujours très flagrant mais en tout cas ça finit par s'ancrer quand même, ça fait un peu, pas boule de neige parce qu'on est pas jusqu'à la boule, mais ça se cristallise un peu j'ai l'impression. Je vois des énergies ... : votre appel en est un témoignage et il y a un peu moins d'un an je discutais avec une de vos collègues de Bordeaux qui s'appelait [nom de personne], j'ai eu le même genre d'entretien que j'ai avec vous aujourd'hui. Et oui, tout ça ce sont des signes qui me font dire que le sujet des littératies de la donnée arrive, est le signal aussi envoyé dans la commande autour des TNH sur débloquer un axe de formation autour des données, c'est un signe positif. Et ce qui est très positif c'est que nous avec mes collègues on a eu une approche un peu plus large de la question et tout de suite il n'y a pas eu d'opposition à ce qu'on relise un peu l'intitulé. Donc ça, ça m'a semblé aussi être un signe positif. Maintenant, je pense qu'on est loin du compte pour plein de raisons que je vous ai un peu exposées.

[Discussion hors entretien sur les objectifs de la thèse]

C'est l'écueil auquel on se heurte souvent, c'est-à-dire que quand on parle autour de nous de la manière dont on approche la question de la donnée et que quand on dit « ce qui nous intéresse, c'est la littératie de la donnée », bon déjà là on a perdu tout le monde. Ceux qui s'accrochent on leur expose notre approche : lire, écrire, compter avec la donnée et faire société avec la donnée. Pour à la fin faire société avec la donnée. En fait, ce qui manque, c'est la *punchline*. Et ça personne ne l'a trouvé. Je n'ai pas encore trouvé le truc qui me permette de l'aborder de manière hypersimple. Je pense que parfois on est victime de la pensée simple, on est aussi volontairement dans une démarche complexe : on postule que les gens sont intelligents et qu'il vaut mieux s'adresser à leur intelligence mais c'est vrai que du coup après c'est pas vendeur. On ne propose qu'une vision pratique. On est dans un monde francophone, on ne fonctionne pas en mode *bottom up*.

Institutionnel 2 (24/03/22)

Comment définissez-vous la *data literacy* et employez-vous ce terme ?

Je vais répondre à la deuxième question : est-ce que j'emploie ce terme ? Je l'emploie moins que ce que je ne l'employais il y a trois ou quatre ans. C'est plus conjoncturel : c'est plus en fonction des gens avec qui je travaille aujourd'hui que ceux avec qui je travaillais il y a trois, quatre ans. Ce n'est pas un désintérêt ou autre. Ce n'est pas un concept qui est toujours bien compris. Ce n'est pas un terme ou une expression que j'utilise très couramment avec les gens avec qui j'échange.

Si j'essaye de la définir,je pense qu'il y a deux grandes idées derrière *data literacy* : il y a bien sûr (je dis bien sûr mais cela n'a rien d'évident) quelque chose autour des compétences, de « capacités à ». Compétences, capacités

à traiter des données, à les collecter, les comprendre, les manipuler, les analyser. Toute une série de choses plus ou moins bien définies, plus ou moins bien standardisées, quelque chose comme ça. Mais il y avait un autre aspect qui me semblait tout aussi essentiel et qui me semble toujours aussi essentiel qui est un peu la question de l'état d'esprit. Dans ce que j'ai pu produire [passage retiré] avec mon co-auteur [nom du co-auteur], on avait identifié quelque chose, qui n'a rien de révolutionnaire encore une fois, qui était autour du *data mindset* : c'est-à-dire la capacité des organisations à avoir un état d'esprit qui soit orienté autour de la donnée. Et donc ce n'est pas simplement, pour moi, une organisation qui a cet état d'esprit « donnée », qui est orienté « donnée ». Ce n'est pas simplement parce que on peut dire, par exemple j'ai 60% des gens qui savent faire un tri croisé-dynamique : ce n'est pas ça en fait. Il y a vraiment les deux composantes pour moi : à la fois une composante de « qu'est-ce que je sais faire et qu'est-ce que je ne sais pas faire ou qu'est-ce que je devrais savoir faire », et puis une composante un peu plus organisationnelle de « c'est quoi l'état d'esprit » et la manière dont on voit les données pour s'en emparer.

Vous avez anticipé ma question sur la définition de la « nature » de la *data literacy* , à savoir est-ce que c'est un état d'esprit, un ensemble de pratiques, une culture... mais pour vous c'est avant tout un état d'esprit ?

Dans la définition que vous avez donnée, « état d'esprit » c'est sûr. Ensemble de pratiques c'est sûr aussi. Après la culture c'est quelque chose que l'on avait pas mal travaillé avec les organisateurs de la *Data literacy conference* effectivement. Déjà un état d'esprit c'est compliqué à saisir. Mais alors une culture, cela me semble.... En tout cas j'aurais du mal de dire de quelqu'un : « voilà il a la culture de la donnée ou alors il ne l'a pas ». Ce n'est pas binaire en fait.

C'est un peu toute la problématique quand on emploie le terme de *literacy* : on a tendance à le traduire par culture mais de ce que j'entends cela a l'air d'être quelque chose de compliqué à définir de manière générale.

Oui, et puis très honnêtement, pour travailler sur le sujet de la donnée, ça fait quinze ans que l'on ne sait pas définir de manière précise ce qu'est une donnée ouverte avec des débats sans fin si elle est, si elle l'est pas. J'allais dire, cela ne me perturbe pas plus que ça.

Au sein de cet ensemble de concepts de données, comme la gouvernance des données et la *data science*, quelle est la place de la *data literacy* selon vous?

Je pense que cela dépend de l'angle que l'on prend dans les trois que vous avez citées. Moi j'étais moins sur la partie culture que sur un ensemble de pratiques et d'état d'esprit. Quand on est sur un ensemble de pratiques : très clairement aujourd'hui dans la littératie des données, à mon avis on n'est pas dans la *data science* parce qu'on est beaucoup plus sur des objectifs de massification des connaissances. Justement, une manière dont moi je vais parler de la *data literacy* c'est de dire que l'on ne va pas transformer tout le monde en *data scientist*. C'est même en contraire de ! On ne va pas chercher à ! Mais pour autant, chacun d'entre nous, et moi le premier, ce n'est pas parce que je ne maîtrise pas les statistiques, les méthodes statistiques avancées, le *machine learning* ou je ne sais quoi, que pour autant il n'y a rien à faire. Bien sûr qu'il y a à faire, bien sûr il y a plein de capacités

mais qui ne sont pas liées au traitement de la donnée, qui sont aussi liées à la collecte de la donnée, qui sont aussi liées à identifier des sources, à les qualifier, à les comprendre, tout cet univers-là. Donc par rapport à la *data science* non, ou alors ça serait la *data science* pour tous mais cela ne serait plus la *data science*.

Après sur la question de l'état d'esprit, comment dire. Cela ne répond pas à votre question mais il y avait quelque chose qui m'avait beaucoup frappé et je m'étais beaucoup interrogé sur le fait qu'il y a beaucoup d'influence du côté de la *data literacy* autour de l'éducation aux médias. Et sur un champ très anglo-saxon encore une fois. C'est une certaine vision des choses et je me suis souvent posé la question : est-ce que la *data literacy* c'était la même chose ? Mais est-ce que le matériau entre guillemets que l'on traite n'est pas le même ? Ou bien justement ce n'était qu'un des champs de la *media literacy*, et puis voilà il y a eu de l'éducation aux médias. Sauf que, un débat que l'on a eu souvent, c'était de dire « est-ce qu'il faut dans les écoles former à la littératie des données ? » et qu'est-ce que cela voudrait dire former à la littératie des données. Est-ce que c'est la même chose que former à la littératie aux médias ? Est-ce que la littératie aux médias c'est la même chose ? Vous voyez, l'un des enjeux c'est de se demander qu'est-ce qui est commun à ce qui se fait déjà et est-ce qu'il y a des spécificités ou pas.

Les données sont devenues centrales : pouvez-vous définir l'importance aujourd'hui d'être « acculturé » aux données ? Et à quel niveau d'importance cela se situe comparé à d'autres compétences comme la technicité informationnelle et informatique ?

J'ai du mal à répondre clairement à cette question : pour moi il y a un parallèle, par exemple, dans les débats avec « est-ce qu'il faut apprendre à coder ? ». A quoi ça sert d'apprendre à coder ? Et dans les champs de l'éducation il y a eu des débats sans fin pour que tout le monde devienne codeur ou alors que l'objectif ne soit pas que tout le monde devienne codeur mais d'avoir vu une ligne de code, en apprenant à coder, tu deviens un acteur plus critique des systèmes, etc.

C'est un peu la même histoire, ce qui me frappe quand même c'est que tout cet univers depuis dix ans, quinze ans, il a été caractérisé quand même par plusieurs tendances assez lourdes. La première tendance c'est qu'on a de plus en plus accès à des sources de données qui sont plus riches ou massives ou complexes qu'auparavant. Donc il y a une vraie disponibilité : c'est certainement les résultats de l'*Open Data* mais on a beaucoup plus de choses accessibles que ce que l'on pouvait en avoir il y a encore dix ans. Le deuxième truc c'est que l'on a de plus en plus d'outils qui, quelque part, prétendent à rendre accessibles au plus grand nombre des compétences ou des capacités de traitement des données qui étaient réservées à un petit nombre d'acteurs avant. Il y a quand même cette idée-là de dire que de plus en plus quand on prend dans un environnement d'entreprise traditionnellement ce que l'on appelle de la business intelligence, avec des outils de business intelligence où chacun dans l'entreprise va être capable d'aller prendre des sources de données, de les interconnecter, de les visualiser avec un logiciel comme tableau, de faire de la préparation de données, etc. Donc il y a à la fois tout et à la fois ce qui me frappe c'est de dire c'est quoi les compétences qui sont requises ? Et il y a une grande partie de ces compétences qui pour le coup ne sont pas nouvelles. Je suis sorti des rangs de l'école il y a vingt, vingt-deux ans peut être, l'un des enseignements les plus utiles que j'ai reçu c'est de faire de la recherche d'information en ligne. La capacité à faire des requêtes complexes, à rebondir sur des recherches, à identifier des sources avant d'identifier des documents... tout un savoir-faire qui fait partie des sciences de la documentation mais appli-

qué à des moteurs de recherche. C'est un truc que l'on retrouve dans la *data literacy* finalement, vous voyez, on n'appelait pas ça de la *data literacy* mais tous ces savoir-faire là on les mobilise au quotidien. La manière dont on recherche l'information, cela n'est qu'une extension de tous ces savoirs-là.

Justement, comment définissez-vous la donnée ?

Il y a eu plein d'éminents chercheurs qui ont travaillé là-dessus mais la manière dont je définis la donnée c'est par rapport à l'information. C'est la plus petite unité, élémentaire, d'information. Mais c'est aussi inspiré d'une définition qui viendrait de l'informatique. Quand on s'est intéressé à l'ouverture des données il y a un peu plus de dix ans, le document c'est la maison. Nous ce qui nous intéresse, ce n'est pas la maison, ce n'est pas le mur, c'est la brique qui permet de construire le mur, qui permet de construire la maison. Mais après si je prends d'autres travaux : j'ai beaucoup lu Rob Kitchin. D'ailleurs je ne pense avoir dit que l'on a du mal à définir la donnée : ce qu'on avait du mal à définir c'était la donnée ouverte, la donnée brute. Et dans les travaux de Rob Kitchin il montrait très clairement que la donnée brute, on dit *raw data*, un des mantras au début de l'*open data* : on veut des *raw data*. Qu'est-ce que ça veut dire *raw data* ? Et bien en fait, les data elles sont plus ou moins brutes, c'est-à-dire à partir de quand ça commence. Je déploie des capteurs qui me renvoient, je sais pas, des données de consommation énergétique : est-ce que la donnée brute c'est la donnée que je réceptionne au plus bas du système de télécom par la trame brute, est-ce que c'est ça la donnée brute ? Est-ce que c'est une fois que j'ai l'encapsulation de la donnée et que je regarde plus que le message lui-même ? Est-ce que c'est autre chose ? Et on a appris, et c'est ce que je trouve assez rigolo, dans cet univers de la donnée à travailler avec des trucs pas très très bien définis quand même.

Quels sont les plus grands enjeux des données et quels sont les types de données particuliers (comme l'*open data*, le *big data*...) qui sont précisément au cœurs de ces enjeux ?

Il y a pas mal de choses intéressantes. Sur ces champs-là, très clairement les enjeux que l'on a aujourd'hui dans le secteur privé, c'est l'intégration des flux. Traditionnellement, on est passé d'un modèle de données plutôt sous forme tabulaire, qui sont finalement des extensions de la statistique classique. On arrive de plus en plus à traiter des flux. Et ça, traiter des flux de données, c'est des problèmes à la fois technique mais aussi des problèmes d'organisation et des problèmes de comment on les interprète. C'est un gros enjeu. Donc les flux de données, notamment avec l'Internet des objets, les capteurs et tout, disons qu'on a une montée en résolution au sens où là où avant on pouvait se contenter d'avoir des photos tous les mois. Maintenant la consommation énergétique est mesurée toutes les dix minutes, à chacun des points de livraison. Donc on a vraiment « comment j'intègre ça ». C'est ce champ-là.

Le deuxième champ que vous avez cité, la partie *open data*, on a par exemple aujourd'hui de plus en plus de jeux de données qui sont complexes, pas forcément parce qu'ils sont massifs, au sens où il y a des millions d'enregistrements, parce qu'ils reflètent une complexité métier. Je prends juste un exemple et là pour moi c'est un champ intéressant un peu nouveau la *data literacy.* Je ne sais pas si vous connaissez ce que fait, c'est une partie de l'*ADEME*, l'agence de la transition écologique, où tout le monde a un projet qui s'appelle *datagir*. Ils font un truc assez génial en fait. Ils sont de gros jeux de données produits par l'*ADEME*, et dans ces jeux de données, il y en avait sur les transports, la performance énergétique des bâtiments et l'impact carbone des pro-

duits alimentaires. Sur le diagnostic de performance énergétique, c'est des données qui sont ultra complexes. La complexité n'est pas qu'une complexité de traitement : je vais l'ouvrir dans *Excel* et cela ne va pas marcher parce qu'il y a trop de lignes, trop de colonnes. Il y a une très grosse complexité là-dessus. Et ce que je trouve marrant parce qu'on n'appelle pas ça la *data literacy* mais ça pourrait en être aussi une définition, cette petite équipe de *datagir* là elle fait vraiment ce rôle de ce que les anglo-saxons appellent *data stewarship*, c'est-à-dire de capacité à prendre la donnée, la documenter, et offrir un accès, en pur *open data*, aux données brutes à ceux qui le veulent. Offrir aussi des *API*, aussi des *Widgets*, qui permettent d'intégrer une partie de la donnée autour de ça. Bref, tout un travail qui vise à rendre la donnée plus facilement intelligible et réutilisable. Pour moi, ces gens font œuvre de *data literacy*. Ils ne travaillent pas sur les gens mais sur la donnée elle-même. Et le dernier enjeu, puisqu'on a parlé de données personnelles, je crois qu'il y a un gros enjeu autour de la capacité.. le *RGPD* avait introduit un droit à la portabilité qui très objectivement aujourd'hui est resté lettre morte. Le truc est très peu déployé, très peu utilisé ou autre. Mais il y avait quand même des initiatives depuis longtemps autour de l'idée : « est-ce que si je récupère tout un tas de données que des entreprises ou des organisations ont sur moi ou sur mes comportements ou sur mes achats dans leurs bases de données, est-ce que je suis capable d'en faire quelque chose ? ». Pour l'instant c'est à peu près de mon point de vue la même promesse depuis dix ans et sur laquelle on a encore beaucoup de mal à avancer.

Il y a eu une impulsion législative autour des données depuis plusieurs années : ressentez-vous l'influence du cadre législatif dans les pratiques professionnelles et quel est votre avis dessus ?

Je suis un peu juge et parti pour cette affaire. J'ai travaillé sur la loi pour une République Numérique. La loi pour une République Numérique elle est en 2016, ça fait 6 ans maintenant donc on a un peu de recul. Et quand je travaillais chez [nom d'organisation], j'ai assisté à sa préparation et à sa rédaction. Plusieurs choses : je pense tout d'abord qu'il y a un élément important à comprendre c'est qu'une grande loi numérique en France, c'est un train qui passe tous les dix ans. Donc, est-ce qu'on saisit ou pas cette opportunité-là ? Dans les années où la loi a commencé à être réfléchie, c'est-à-dire fin 2014, 2015 et début 2016, sur les questions de données, la question qui prédominait c'était la question de l'open data, la question de dire « il faut les ouvrir les données » et cet enjeu de dire « l'ouverture par défaut ». Et a posteriori, on s'est posé la question de se dire si c'est mature, pas mature, est-ce qu'on attend le train suivant ? Mais le train suivant, on serait encore en train de l'attendre. Je pense que ce cadre juridique-là, il a eu le mérite de poser la question de la donnée. Il a eu le défaut de le poser que sous l'angle de la gouvernance en mode open data, mais au moins il a posé la question de la donnée. Et je suis convaincu qu'on n'aurait jamais autant avancé sur ces questions, notamment dans le secteur public autour de la donnée, de son exploitation, des limites de ce que l'on collecte, si l'on n'avait pas mis des dispositions autour de la loi pour une République Numérique. C'est un déclencheur en fait.

On le voit notamment dans les ministères, avec l'impulsion d'Henri Verdier, même si cela ne montre rien de la praticité....

Oui et cela dépend en fait. Si on se concentre sur le secteur public, et l'Etat qui est encore un sous-morceau du secteur public, qui n'est pas négligeable mais qui est quand même un sous-morceau, ce qui est sûr c'est qu'il y a des cultures j'allais dire administratives de chaque ministère qui font que certains sont plus *data literate* que d'autres. Historiquement, dans la sphère socio-fiscale, on gère des informations depuis très longtemps, on gère

énormément de données : ce sont des gens qui font beaucoup plus autour de ces sujets-là que, spontanément, la culture ou la justice. Il y a des grandes cultures comme ça : on voit qu'elles sont différentes. Non, ce que je pense, et qui est aussi une question de littératie… je donne un exemple : y a beaucoup d'enjeux d'utilisation des données autour du changement climatique et pour le coup le ministère de la transition écologique est l'un de ceux qui, ces dernières années, s'est, à mon humble avis, le plus mobilisé sur ces questions-là. Vraiment très opérationnellement. Typiquement si aujourd'hui je veux dire : « le changement climatique c'est très large, si je prends un sous-sujet à l'intérieur de ça, le sous-sujet particulier de la rénovation énergétique des bâtiments publics ». Même ce truc-là est énorme. Et je vais d'abord me poser la question de dire : si je veux pouvoir cibler les bâtiments qui sont des passoires thermiques, qu'est-ce qu'il faut que je récupère ? Il faut que je récupère deux types de données : les données sur les bâtiments publics et des données sur la consommation énergétique. Les données sur les bâtiments publics aujourd'hui, si je redescends d'un niveau de détail, qu'est-ce qu'on utilise ? On utilise une base qui s'appelle *Majic*, Mise A Jour des Informations Cadastrales, qui est produit par la direction générale des finances publiques. On ne peut pas leur reprocher, c'est normal qu'il l'ait construite comme ça, mais la *DGFIP* l'ont construit comme ça pour collecter les taxes foncières. En règle générale, les bâtiments publics ne sont pas soumis aux taxes foncières, donc dans *Majic* je n'ai aucune certitude que j'ai tous les bâtiments publics à l'intérieur. Et ça pour moi c'est une question de littératie des données, très loin de dire « est-ce que je sais faire un tableau croisé-dynamique ? ». C'est une littératie au sens de dire « ah je comprends les conditions, les contraintes dans lesquelles ces données ont été produites » qui me permet d'être plus prudent ou plus responsable dans les utilisations ultérieures que j'en fait.

Est-ce que vous avez senti un changement de vos pratiques du fait de la généralisation de la donnée ?

Oui il y a eu pas mal de choses. Il y a quelques années, je ne sais pas si ça existe encore *Open Refine*…

Oui !

Il y a eu tout un mouvement autour d'outils de préparation des données, de nettoyage des données et qui continue encore aujourd'hui. J'ai quitté *Etalab* l'été dernier mais ma dernière mission c'était une mission avec l'*INSEE* où l'on a évalué les compétences data au sein de l'Etat, en sortant un chiffre avec le nom de *data scientists*, *data engineer*, etc. Ce qui ressortait c'était que le plus gros besoin n'était pas sur les data scientists, on avait pas besoin prioritairement de gens qui étaient capables de déployer du *machine learning* ou du *deep learning*. On avait besoin de gens qui savaient préparer des données. Et ces outils-là, que ce soit *Open Refine* dans la partie open source ou dans la partie propriétaire, il y avait vraiment ce travail autour de ça. Le deuxième champ c'était tous les outils de visualisation. Et là comme toujours il y avait classiquement les offres propriétaires et open source. Moi j'ai beaucoup utilisé *Tableau* qui avait l'inconvénient d'être très cher pour des usages individuels, et à la fois extrêmement intéressant car ça introduisait, comment dire, la plupart des outils que moi j'utilise en visualisation de données étaient plutôt des outils de représentation visuelle au sens que un *data wrapper* ou d'autres outils comme ça c'est « j'ai déjà analysé mes données, j'ai déjà une histoire à raconter, et je vais utiliser la visualisation pour raconter cette histoire ou appuyer ou soutenir une partie de cette histoire ». Et ce que me proposaient des outils comme Tableau, et ils le faisaient très bien, c'est pas la visualisation de données comme un moyen de représentation, de narration, c'est comme moyen d'explorer les données. Et c'est hyper différent.

C'est-à-dire que je ne pars pas d'une histoire, je me dis, tiens « si je croise ça avec ça, ou si je les regarde sous tel angle, ou si je fais tel truc, c'est quoi les outliers ? C'est quoi les tendances ? Est-ce qu'il ya des trucs ? Est-ce qu'il y a des valeurs aberrantes ? Pourquoi il y a ces valeurs aberrantes ? Est-ce que derrière ces valeurs aberrantes il y a une histoire ? » ce qui est très différent en fait.

Totalement, surtout que la datavisualisation, quand on lit ce qui est écrit dessus, la première mise en valeur c'est le *storytelling*, beaucoup plus que l'exploration elle-même. Ce qui est intéressant car d'avoir une autre vision plus ancrée dans la_ data literacy_, puisqu'en l'occurrence on peut vraiment dissocier les deux. Et si on va justement sur cette mise en place / pratique de la *data literacy* dans un organisme, doit-elle concerner tout le monde ? Et comment peut-on le faire aujourd'hui ?

Je ne sais pas dans quelle mesure c'est une question nouvelle en fait. Je serai curieux de savoir si au début de la micro-informatique personnelle, on ne s'est pas posé la même question pour savoir s'il fallait former d'abord les assistants et assistantes de direction avant de former tout le monde. Ou à quel moment on a dit que tout le monde soit capable d'utiliser *Word* niveau basique. Et je sais pas, c'est juste une hypothèse, dans quelle mesure ce sont des questions qui sont si nouvelles que ça. Par exemple, ce que moi j'ai vu dans les organisations publiques, c'est l'impact de *PIX*. Et qui est très intéressante parce qu'il y a certaines visions de ce qu'est la *data literacy*. Nous cela nous faisait marrer. Quand j'étais chez [nom d'organisation], on avait des stats de data.gouv, ou d'autres portails *open data*, et ils nous disaient « ah ben on a vu que PIX a changé leurs questions » et on a des pics de fréquentation sur un jeu de données en particulier. Et l'un des jeux de données contenaient les réponses d'un jeu de données de *PIX*. On est vraiment là-dedans : est-ce qu'on doit former tout le monde ? Est-ce que si on doit former tout le monde, c'est à quoi ? C'est quoi le niveau requis de base ? Ce que je sais c'est qu'on a fait beaucoup de travaux avec la petite équipe de la *FING* qui gérait le programme *Infolabs*, autour de sensibilisation un petit peu. Qu'est-ce qu'on peut faire pour que les gens soient sensibilisés aux enjeux et commencent à mettre les mains dans le cambouis. Il y a vraiment un truc qui était hyper important c'était mettre les mains dans le cambouis.

Et justement en parlant de « mettre les mains dans le cambouis », à l'issue d'une formation en *data literacy*, qu'est-ce qui est le plus important, pour les personnes, à développer ?

Je pense qu'il y a, comment dire… c'est une opinion personnelle : je pense que le truc le plus critique c'est l'esprit critique. Pourquoi je dis ça ? Parce que je pense qu'il y a encore aujourd'hui un peu de pensée magique autour de la capacité des données, de ce que cela représente, de la magie du *Big Data rebrandé* sous terme d'intelligence artificielle, de *machine learning*, mais c'est vraiment l'esprit critique.

Après dans les formats que j'ai déployés, j'ai le souvenir d'un bel évènement que l'on avait fait dans la campagne de l'infolab à Rennes : on avait fait une fresque sur la courbe de distribution des prénoms dans la ville. On avait récupéré sur le portail *open data* de la ville de Rennes, juste un fichier qui disait en 2008 ou 2010, les prénoms masculins donnés avec le premier prénom, le nombre de fois où il a été donné, etc. Et on avait pris une grande nappe de papier et on avait commencé à tracer : Nathan cela a été donné 43 fois, le deuxième tant de fois, etc. Et finalement à la fin de l'après-midi, on avait plusieurs groupes, c'étaient des adultes qui avaient travaillé sur ce truc-là, on était arrivé à un truc qui faisait 12 mètres de long. Et pour le coup, le soir on avait

une petite expo, les gens étaient venus. Et je me suis toujours posé la question de ce qu'on avait appris. Les participants qui ont fait ça, et on était à fond avec les mains dans le cambouis, parce que statistiquement on n'a pas mobilisé de compétences. La compétence qu'on a mobilisée c'est savoir lire un tableau à deux colonnes. Et puis savoir tracer une abscisse et une ordonnée, cela n'a rien d'exceptionnel. Par contre, ce qui est fondamental, c'est que tous ceux qui ont participé m'en reparlent encore parce qu'ils se sont fait super mal au dos, ce qu'ils ont retendu c'est que la courbe de distribution des prénoms était une super longue traîne. C'est-à-dire qu'à chaque fois le marronnier de début d'année, à Rennes comme ailleurs, dans la presse quotidienne régionale, on vous dit « les trois prénoms les plus données c'est ça, ça, ça », que c'est juste l'arbre qui cache la forêt. La plus grande probabilité pour un nouveau-né en France, c'est de porter un prénom rare. Et pas de s'appeler Léo, Théo, ou Nathan. Ce n'est pas ça les prénoms qui sont donnés aux Français, ce n'est pas vrai. Bon j'étais un peu long mais c'était pour vous donner un exemple pour vous dire « est-ce que la littératie des données, est-ce qu'elle s'apprend indépendamment d'apprendre quelque chose sur les phénomènes qu'on a cherché à mesurer à travers les données ? » Je ne sais pas en fait.

Si je suis bien : donc, l'esprit critique, c'est de savoir comment les données sont employées, produites, de comprendre qu'il n'y a pas de pensée magique ?

Oui, et c'est peut-être parce que je viens du monde de l'*open data* où on avait parfois la naïveté de croire qu'il y avait forcément des données, qu'elles étaient forcément ouvertes et qu'elles trouveraient forcément un public. Mais à un moment donné on a cassé un lien, c'est-à-dire que les données sont produites dans des contextes de production, et si elles arrivent comme par magie sur un portail *open data*, il n'y a pas de magie, il y a plein de travail et de gens qui sont impliqués, mais à un moment donné j'ai toujours eu la crainte que l'on perde le contexte de production. Parce qu'il ne se document pas facilement. Et le premier truc devrait être ça : je travaille en entreprise, je travaille dans le secteur public, je travaille dans le secteur de l'éducation…à un moment donné, même si ce n'est pas moi qui fait les traitements, je m'appuie sur d'autres pour les faire, je devrais être capable d'aller interroger. Ces données-là elles sont collectées par qui, comment ? Qu'est-ce qu'il y a dedans, qu'est-ce qu'il n'y a pas dedans ? Qu'est-ce qu'on a comme degrés de certitude ? Il y a plein de fois où on ne sait rien de ce qu'il y a dedans. Mais si on loupe tout ça, on loupe le sujet à mon avis.

Est-ce qu'il y a eu une évolution en termes d'esprit critique ?

J'aurais du mal à répondre. Ce que je vois c'est qu'il y a eu beaucoup d'efforts de sensibilisation dans beaucoup d'organisations. C'est très générique ce que je raconte mais il y a eu des efforts, c'est clair. C'est rare de trouver honnêtement, ce n'est pas toute la France mais si je me concentre sur les villes… si je prends la liste des 15 premières villes de France, cela serait super étonnant, je pense que cela n'a jamais existé, une ville qui n'a pas essayé de faire des démarches de sensibilisation de ses agents à l'*open data*, sur les données, etc. Il y en a qui ont déployé des programmes de formation à plus grande échelle aussi. Donc là c'est vrai qu'il y a eu quelque chose. Après, je ne sais pas véritablement. La sensibilisation c'est une chose et passer à la pratique en est une autre. Je ne sais pas encore vraiment mesurer si les pratiques ont complètement évolué.

Quels sont les freins à cette acculturation ?

Je crois que malgré les discours sur le potentiel des données, il y a un vrai frein : on a encore pas mal de gens en position de décision qui ne sont pas vraiment convaincus. Il y a un côté un peu « retour de hype ». « Vous avez parlé de la donnée pendant des années, que la donnée c'est un truc essentiel mais voilà on est passé à autre chose, maintenant c'est l'intelligence artificielle ».

Pas tant sur les définitions elles-mêmes du coup, ou est-ce que cela fait partie du problème ?

Je vais nuancer mon propos. Quand je dis que c'est un retour de *hype*, on a aussi manqué cruellement d'impact c'est-à-dire de dire d'être capable de dire qu'utiliser les données ça a tel impact positif. Faute d'avoir documenté les impacts, on n'est pas capable non plus de durablement engager des organisations sur ces questions-là. Et ça je pense que c'est un constat partagé. En ce moment je travaille avec l'*Open Data Institute* à Londres. Et très clairement eux-mêmes l'ODI ils le disent eux-mêmes « nous on a fait beaucoup de workshop, on continue à en faire mais quand on dit aux organisations si vous voulez de l'impact c'est pas juste de faire venir 40 personnes pendant deux jours sur un atelier ».

Si vous deviez citer un ou plusieurs projets en *data literacy* qui a fonctionné, cela serait lequel et pourquoi ?

Je suis très fan de ce que fait « fréquence écoles » à Lyon. Il y a pas mal de choses autour de ça. Alors depuis quelques années je travaille autour de la notion des algorithmes, et la littératie des algorithmes c'est encore un sujet qui est très émergent. Mais il y a des choses intéressantes qui se créent, notamment sur tout ce qui relève de la jouabilité, c'est-à-dire des systèmes avec lesquels je vais pouvoir interagir, essayer de changer les choses. Il y a plusieurs journaux qui ont tenté des trucs : le *Financial Times* avait créé un petit jeu où il fallait se mettre dans la peau d'un chauffeur Uber. *The Uber Gates* ça s'appelle. C'est accessible en ligne, c'est gratuit, il n'y a pas besoin d'être abonné au journal. Et tu vas pouvoir répondre à tout un ensemble de situations comme « je suis à Portland, est-ce que je vais à côté pour prendre des shifts ou pas ? J'y vais, combien ça coûte, etc. ». Et tout ce qui relève de la jouabilité je trouve ça intéressant.

Si vous aviez des recommandations pour améliorer l'acculturation aux données en France, quelles seraient-elles ?

Franchement je ne sais pas. Je me demande si une des recommandations, cela ne serait pas de savoir à quel point on doit ou pas arrimer ça aux enjeux de médiation numérique. Ce n'est pas très clair : ça revient à un point qu'on a discuté tout au début. J'ai été frappé de voir, et même dans la *data literacy conference*, les gens qui venaient MIT, du medialab qui avaient travaillé sur les questions d'inégalité, de discrimination, sur tous ces enjeux très sociaux, très sociétaux. Et la médiation numérique en France elle est restée très inspirée de ces approches : des droits, des égalités, toutes ces questions-là. Et je me demande dans quelle mesure c'est une piste de développement qui doit se poursuivre ou pas. Ce que je veux dire par là c'est qu'il y a sans doute des organisations si on leur présente la *data literacy* sous l'angle de la médiation numérique, il y en a plein qui iront, mais il y en aura encore plus qui n'iront pas parce que c'est un angle de médiation numérique.

Et qui est très prégnant d'un point de vue social, ce qui ne va pas empêcher des controverses d'apparaître.

Exactement et pour le coup je crois aussi que les anglo-saxons, toujours très décomplexés sur ces questions, certains d'entre eux ont réussi à en faire un *business*. Dans les sous qu'on avait fait venir dans les intervenants, il y avait le fondateur d'une boite qui s'appelait *Tuva*. Je continue à recevoir leur newsletter tous les mois. C'était y a six ou sept ans, il y est encore. Son truc a l'air de tourner, ça roule, c'était pas un feu de paille son affaire.

Institutionnel 3 (06/05/22)

Avez-vous une définition de la *data literacy* ? Et employez-vous ce terme ?

Nous, on est plutôt sur culture des données. Dans les collectivités, les anglicismes notamment ne sont pas toujours bien vécus.

Et en termes définitionnels ?

Ma définition ça serait la capacité de chacun à comprendre, exploiter tout le potentiel d'une ou des données de manière générale.

On a beaucoup de termes qui sont apparus et qui constituent cet écosystème des données : gouvernance de la donnée, *data science*... est-ce que vous arrivez à situer la culture de la donnée dans cet écosystème ?

Moi je suis chargée de la gouvernance des données pour le département [nom du département], c'est l'intitulé de mon poste. Dans les missions qui sont les miennes, on va dire qu'il y a des chantiers en lien étroit les uns avec les autres et qui sont des chantiers liés à la gouvernance : soit des chantiers sur lesquels je vais m'appuyer en interne sur des collègues qui ont déjà une forte compétence « donnée » comme les gens qui sont administrateurs des bases de données parce qu'ils maîtrisent tout ce qui est entrée et sortie dans le système de bases de données. Et après il va y avoir une sphère plus large de collègues qui vont être des gens qui sont capables de collecter, saisir de la donnée ou valoriser de la donnée dans certains outils comme le SIG par exemple. Et après j'ai d'autres chantiers encore plus satellites à tout ça qui touchent tout le monde, dont un des chantiers chez nous qui est la culture des données, donc vraiment amorcé pour la collectivité [nom de la collectivité] : c'est 1500 agents qui forment la collectivité aujourd'hui, c'est vraiment de développer, pour chacun, cette culture des données. D'être en mesure de dire demain, de se dire dans deux ans, dans les compétences numériques que les agents ont, l'exploitation des données devrait faire partie d'une des compétences numériques.

Donc c'est un cadre bien précis dans la gouvernance des données.

C'est ça. C'est un des chantiers de la gouvernance. Au-delà de tous les autres, ça fait partie de mes chantiers.

Et si on devait, je vais prendre le terme de nature avec de grosses pincettes, donner une « nature » à la culture des données, est-ce qu'on parlerait plutôt d'une culture, d'un ensemble de techniques et de pratiques, d'un état d'esprit ? Comment le voyez-vous ?

C'est un peu un tout pour moi. C'est une culture dans le sens où effectivement il faut permettre à chacun d'acquérir des notions de base, des notions très générales (définition d'une donnée versus information). Ce sont des choses toutes bêtes mais qui ne me semblent pas si claires que ça dans la tête des gens. Donc il y a cette partie un peu culturelle où on va apporter des grands principes, des grands préceptes de base à tout le monde. Et puis après, en fonction des missions et en fonction des appétences de chacun, mais aussi des pratiques professionnelles : c'est aussi commencer à comprendre ce que c'est qu'une donnée, une donnée structurée, une donnée fraîche, une donnée de qualité. Tous ces principes qui ne sont plus trop des principes généraux mais des principes assez techniques de la donnée en elle-même ; principes dont on a besoin pour mettre en place un certain nombre de projets. Mais encore une fois, c'est la cible de certaines professions ou de certaines appétences informatiques, donc numériques.

Est-ce qu'aujourd'hui cette culture des données est devenue aussi importante que la manipulation informatique, que les compétences informationnelles... est-ce que ça a pris une plus grande ampleur ?

Nous, on n'est pas rendus à la grande ampleur. On est plutôt rendus à la phase zéro : c'est qu'on a identifié le besoin. De par ma prise de poste, je sais que si je veux que ça prenne et si je veux que ça marche, il faut que je ne sois pas tout(e) seul(e) et que justement tous les collègues puissent être approchés de près ou de loin par quelque chose qui leur parle de la donnée à différents niveaux, différentes strates, etc. Pour une administration comme la nôtre, notre point de vue est plutôt de se dire : ça doit faire partie d'un package de compétences numériques qu'on est capable d'avoir : faire du tri dans les informations, trouver l'info qui m'intéresse à l'instant T et ne pas passer trois heures à chercher, des choses comme ça, avant d'être en mode « on doit piloter, on doit se piloter et piloter par la donnée » : on en est pas du tout rendus là.

Sauriez-vous situer dans le temps où la nécessité d'une maîtrise des données a pris de l'ampleur ?

Nous, ce sont des projets visibles qui ont fait que les choses ont pris de l'ampleur. Je crois que quand vous m'aviez vu en webinaire c'était avec [nom de personne] sur la gouvernance des données, etc. Moi j'expliquais dans cette démarche que j'avais lancé un portail *close data* pour le département qui est un outil d'évaluation de nos politiques publiques, ce qui permet d'avoir un baromètre sur l'état de santé de la collectivité. C'est ce sujet-là qui a permis de passer l'étape d'après et qui a permis à chacun de se dore : « Ok, voilà tout ce qu'on peut faire avec la donnée ». En fait, j'ai donné à voir et là on s'est dit « ah ouais ! On peut faire ça ! ». Et si on peut faire ça, on peut faire beaucoup plus. Et c'est ça qui a amorcé cette volonté de lancer cette grande phase d'acculturation chez nous.

Qu'est-ce qu'une donnée pour vous ?

Alors si je devais définir la donnée, je dirais : un élément brut qui contient des valeurs quelles qu'elles soient, sous forme de champs, sous forme d'enregistrements, qui sont non interprétés. Donc c'est brut.

Pouvez-vous à définir les enjeux actuels autour des données et comment la culture des données peut y répondre ?

Les enjeux autour des données, dans ma collectivité, dans les pratiques métiers qui sont les nôtres, ils sont de plusieurs ordres. Ils sont déjà d'avoir des données de référence pour mieux rendre service aux publics. Il s'avère que quand on commence à fouiller un peu dans les pratiques métiers, eh bien on se rend compte que… au département, il y a 170 métiers. Sur ces 170, il y a peut-être une centaine de métiers qui, en fonction de leurs missions professionnelles, accèdent à une donnée que l'on devrait appeler donnée de référence comme, par exemple, la base des agents. Mais en fait, chacun a son outil à droite à gauche, dans lequel il a déjà sa base agents mais qui ne fait pas référence à la vraie base agents de la DRH qui a les infos les plus à jour. Donc il y a des déjà des questions de données de référence. Après on va avoir aussi des questions sur la gestion des données et l'accès à ces données ; des questions de *RGPD* mais pas que.

Par exemple, si demain j'ai un audit sur mon système d'informations, je dois être en capacité de dire : la donnée qui concerne les bénéficiaires d'aides autour de la solidarité, par exemple, les ménages allocataires du RSA, les gens qui bénéficient de la protection maternelle et infantile, l'aide sociale à l'enfance, qui sont là des sujets *RGPD* sensibles ; je dois pouvoir dire : « ces données-là ne sont accessibles qu'à telle et telle personne parce qu'elles sont habilitées. La personne responsable de la donnée, c'est untel. Je dois pouvoir mettre des noms sur tout ça ». Et ça, c'est indispensable. Et c'est ce qui me permet de protéger mon système d'information et de protéger les données et l'accès à mes données. Après chez nous, il y a plein d'exploitations possibles, ou en tout cas de choses névralgiques sur lesquelles il faut qu'on soit très au fait. On intègre, par exemple, dans nos systèmes d'information, des données venant d'autres partenaires : la *CAF*, *Pôle Emploi*… ça c'est pareil, il y a à la fois le fait qu'en intégrant ces données, ça nous permette de calculer des aides à apporter aux bénéficiaires, comme pour les bénéficiaires du *RSA* par exemple. Nous n'avons donc pas le droit à l'erreur concernant certaines données, notamment celles qui permettent de faire des calculs d'aides, etc.

Donc en fait ce n'est pas nécessairement des types de données particuliers, comme le big data, l'open data, etc., qui sont au cœur des enjeux mais vraiment la pratique qu'il y a autour qu'il faut harmoniser.

C'est ça. Bien que la « data » puisse être rebutante pour certaines personnes, il est essentiel que chacun comprenne que c'est de sa pratique métier que la donnée est exploitée. C'est ce qui va me permettre de sensibiliser quand même : « ah regardez, si on avait des données bien propres, si on est certain de nos données, si on est sûr que c'est une donnée de référence, voilà ce qu'on arriverait à faire en close data ou en open data demain ». Mais cela ne va pas être, comment dire, le cœur de mes problématiques. Le cœur de mes problématiques ça va être de réussir à avoir des données fiables, de référence, qui transitent dans mon système d'information, sur lesquelles je peux m'appuyer, et quand je fais une extraction de telle ou telle donnée, je veux être certaine que c'est la bonne. Et non pas : si je fais trois extractions différentes, j'obtiens trois chiffres différents. Par exemple, le président du département [nom du département] est responsable légal des enfants qui sont confiés à l'aide sociale à l'enfance (*ASE*). Si je lance une requête sur mon système d'information, quelle que soit la requête, je dois systématiquement obtenir le même nombre d'enfants.

Et est-ce que les réglementations (loi Lemaire, RGPD) ont une influence sur vos pratiques ?

Non, ça a été plutôt perçu comme encore une réglementation de plus. Les services à qui s'imposaient ces réglementations, notamment les données obligatoires en open data par exemple, se sont pliés à cette réglementation. Ça n'a pas impulsé de changements majeurs. Par contre, ça devient un des arguments qui me permet, en l'occurrence, d'essayer d'impulser le changement et de mobiliser les troupes sur ces sujets-là en disant « il faut qu'on le fasse ». Alors, oui vous le faisiez déjà avant mais on va essayer de le faire encore mieux. Avec un peu moins de manipulation humaine, essayer de... ce sont donc des arguments sur lesquels je peux m'appuyer mais ça n'a pas amorcé un changement profond chez nous. En tout cas, pas nous département de la Mayenne.

Est-ce dû à votre structure ou au fait d'un fonctionnement interne ?

Non, c'est dû à notre structure parce que pour d'autres collectivités, cela a été vraiment un levier.

Comment peut-on mettre en place dans un organisme tel que le vôtre, quels moyens, quelle durée...et est-ce que ça doit concerner tout le monde ?

Alors c'est bien qu'on en parle parce qu'on est en plein dedans nous. Est-ce que ça doit concerner tout le monde ? Oui. Est-ce que ça doit concerner tout le monde de la même manière ? Non. Je pense qu'il y a différents niveaux en termes de strates à toucher et dans ces strates, différents messages à faire passer. Typiquement : je vais vous donner le cadre dans lequel on se trouve. Nous, on est là en train de préparer la quinzaine de la data au mois de juin qui est 15 jours de grosse com' autour de la donnée et on est dans le chantier de l'acculturation pure et dure. Comment on envisage les choses ? On envisage les choses principalement de deux manières.

La première c'est, pour essayer de toucher tout le monde : alors, ça va être des vidéos. Là l'idée c'est de faire à la fois : ludique, un peu décalé mais avec des messages-clés. Justement : qu'est-ce qu'une donnée ? Quelle est la différence entre une donnée et une information ? Et dans mon métier, moi, mes collègues dans leurs métiers, à quel moment je croise de la donnée ? Et pourquoi on parle de données aujourd'hui ? Et donc concrètement, je spoile un petit peu ce qu'on a prévu : par exemple, sur les vidéos qu'on envisage de diffuser sur l'Intranet, intranet qui est accessible à toute la collectivité, à tous les agents de la collectivité, on part d'un usage métier. Alors, le cas concret c'est : on a une jeune femme, qui est maman, qui vient d'accoucher, qui a un nouveau-né, et qui prend contact avec la protection maternelle et infantile pour avoir un peu d'aide dans son organisation, dans la gestion de ce nouveau-né dans sa famille, etc. Donc ça c'est le point de départ. Ça, ça va parler à 450/500 agents qui sont déjà dans la direction de la solidarité parce que c'est vraiment le cœur de leurs métiers. C'est aussi une des compétences majeures des départements, c'est la solidarité, donc là on met le doigt dessus. Et à partir de ça en fait on va dérouler : tiens, madame X, elle vient, elle donne son nom, son prénom, son adresse, etc., donc elle donne certaines de ses données à la collectivité. Elle fait offrande de ses données à la collectivité. Tiens, là on va pouvoir commencer à taper sur la définition originelle de la donnée. Et dans ces petites vidéos, on en prévoit a priori 3 ou 4, ça va être le moment de faire un zoom sur le cycle de vie de la donnée dans notre système d'information.

Sur le rôle de la gouvernance, donc mon rôle, dans la gestion de cette donnée dans notre système d'information, et si cette donnée elle est propre, voilà ce qu'on peut faire avec vos données, en gros, avec des tableaux de bord, d'évaluation, de la prospective, et si demain on croisait ces données avec des données présentes en *big data* ou si demain on avait une compétence interne pour faire un peu d'IA, on pourrait imaginer bien plus de choses. Mais pour toucher ces gens-là il faut qu'on soit super ludiques, il faut qu'on soit très accessibles, il ne faut pas qu'on soit en mode maître Capello qui donne un cours sur la data pendant deux heures chaque semaine : c'est vraiment des vidéos d'une minute trente qui font des zooms, donc on va parler de données géographiques, on va parler de données sensibles, on va parler de données *RGPD*, on va apporter tout un tas de connaissances générales sur ce sujet-là. A terme, on peut imaginer au bout de ces 3,4 vidéos à destination des professionnels de la collectivité, c'est faire un quiz par exemple, c'est mettre en place un petit quiz, un petit challenge interdirections qui dirait « alors, est-ce que vous avez compris ? Vous avez une semaine pour répondre au quiz. Si vous ne vous souvenez plus, vous pourrez regarder les vidéos » et d'être encore une fois en mode très ludique. Ça c'est la première phase. Donc là j'essaie de toucher le grand public de la collectivité, donc tous les agents quels qu'ils soient.

Deuxième phase, on est plutôt dans une attaque de type séminaire, séminaire ludique, un peu décalé, qui durerait deux heures, où on essaierait de toucher tous les managers de la collectivité. Chez nous les managers ce sont les directeurs qui sont rattachés aux DGS, il y a aussi des chefs de service, il y a des responsables de pôles, etc (ce sont 160 managers intermédiaires au CD53). L'idée c'est si on peut en toucher 50, déjà ça sera bien pour qu'après ça diffuse un petit peu. C'est aussi de filmer tous ces temps-là pour pouvoir les utiliser sous forme de replays, refaire d'autres actions dans l'année pour re-sensibiliser sur ces sujets-là mais en tout cas c'est de lancer ces séminaires. Dans ces séminaires, ça va être à la fois de partager ce qu'on évoque aussi dans les vidéos : la règlementation, l'*open data*, la démarche de gouvernance du comité départemental de la pétanque. Mais l'idée c'est aussi de leur dire : « on peut aller beaucoup plus loin ».

Et voilà ce qu'on peut faire aujourd'hui avec les données qui seraient propre, qui seraient correctement exploitables, qu'on aurait clairement identifié dans notre système d'information, voilà ce qu'on pourrait faire demain au travers d'un projet d'*open data* ou la création d'un nouveau service à destination des usagers, etc. Mais là l'objectif est de ne surtout pas être descendant, le but c'est vraiment de faire que pendant ces séminaires, maxi 12 personnes présentes, en petit comité, c'est de les faire jouer, de les faire manipuler en mode intelligence collective. Pourquoi pas des sortes de micro *serious games*, des choses qui les rendent acteurs et qui fassent que ce sujet, de la « donnée » il devient palpable. Et de cette partie séminaire en fait, on prévoit aussi d'ouvrir ces séminaires à une démarche territoriale donc aux communautés de commune, aux EPCI, de les inviter à participer à ces séminaires pour créer une sorte de culture territoriale, pas qu'une culture départementale, afin de réfléchir ensemble sur des nouveaux services à proposer à partir de nos données ou avec des données qu'on peut avoir en commun.

Et puis de ces séminaires découlent un troisième type d'action qui là va être proposé à des gens qui ont déjà une casquette un peu data dans la collectivité : les archives, le SIG, les collègues administrateurs de bases de données, ainsi que des référents « données » dans les directions métier. Nous allons leur proposer une journée entière d'acculturation où là on va plutôt être dans un mode hackathon je pense. Sur une journée : le matin, on va être plutôt de type quiz, challenge entre tables, etc., pour évaluer leur niveau de connaissances sur les

données parce que bien que j'ai des gens qui fassent du SIG, si ça se trouve, la compréhension qu'ils ont de la gouvernance elle est de zéro. La compréhension qu'ils ont de la donnée, à part une donnée géographique structurée dans une base de données dans un système de type postgre, etc., bah il n'y en a pas. Cela va être le moment et d'évaluer les connaissances, et confronter les points de vue, et apporter un peu de matière grise à tout ça.

Donc en fait c'est surtout une acculturation, un développement culturel que de pratiques vraiment techniques à développer.

Hum, complètement il n'y a pas de technique pour l'instant.

Est-ce que c'est quelque chose que vous visez par la suite ou pas du tout ?

On y viendra par la suite, mais pour la partie technique on va plutôt être en mode, à la méthode des petits pats en fait. Je pense qu'il va y avoir plusieurs manières d'attaquer ce sujet-là. Il y a justement dans un de mes chantiers de gouvernance, il va y avoir la création de référentiels, il y a la question de modélisation de certaines données ou bases de données qui n'étaient pas modélisées jusqu'à maintenant donc là on va devoir travailler ce genre de projets. C'est par ces projets-là qu'on va commencer l'acculturation technique. Et on va le faire projet par projet. Donc si j'ai une direction, je pense par exemple conseil départemental, il y a ce qu'on appelle une direction de « l'enseignement » qui gère tout ce qui est collège privé-public, et aussi une compétence « enseignement supérieur », donc les facs, les IUT, etc. Aujourd'hui, cette direction ne travaille qu'avec des fichiers *Excel*, il n'y a pas de bases de données. Alors pour la partie finance, ce sont des fichiers *Excel*. Pour la partie restauration dans les collèges, ce sont des fichiers *Excel*. Là clairement on aurait un outil à leur mettre en place pour qu'ils puissent faire des requêtes dans l'outil et sortir des stats sur les collèges, faire du publipostage, tout ce qu'on peut imaginer avec une base de données correctement préparée. Eh bien là clairement si on arrivait avec un projet de ce type pour la direction de l'enseignement, toute l'acculturation technique à la question des données, à la structuration aux données, l'accès aux données, la responsabilité sur les données, les questions de *RGPD*, etc., se ferait grâce à ce projet. Et uniquement grâce à ce projet. Donc nous on va être dans la méthode des petits pas : il y a un besoin à un moment, on profite de ce besoin pour placer un certain nombre de notions pour apporter un certain nombre de connaissances, etc. Clairement, dans le close data qu'il y a chez nous, pour chaque tableau de bord, c'est un accompagnement que j'ai fait, auprès de la direction métier qui voulait ce tableau de bord dans le *close data*. Avec des questions de type : « où est stockée la données ? Est-ce qu'on est sûre qu'elle est fraiche ? Est-ce qu'elle est fiable ? Est-ce qu'on peut automatiser le traitement de cette donnée-là ? quelle est la métadonnée ? Et donc là, c'est des temps de travail très itératifs, très réguliers. Par contre, j'apporte énormément d'informations dans ces temps de travail qui font que derrière, quand je parle à un collègue qui a vécu l'expérience close data, il y a beaucoup de notions qui sont déjà là et pour lesquelles la compréhension elle est là maintenant. Par exemple, le terme de données pivot : pour les gens qui ont travaillé de concert avec moi sur une production en close data , il y a plus de questions à se poser.

En lien avec cela justement, en termes de niveau général d'acculturation aux données chez les personnes que vous fréquentez dans votre milieu professionnel, est-ce que vous estimez que c'est une situation qui évolue ?

Pour nous, côté collectivité de [nom de département], on est plutôt sur un niveau bas. Je ne pense pas que c'est le cas partout quand même, au niveau des collectivités notamment. Il y a des collectivités qui ont atteint un niveau déjà plus que correct mais nous on est encore trop silotés. Il y a aussi des gens qui œuvrent pour l'intérêt personnel avant l'intérêt collectif, sur le service public, etc., et on a du mal à voir cette vision et cette culture de la donnée qui soit large et complète.

Et vous, quelles difficulté rencontrez-vous pour la mise en place d'une culture des données ?

Ah ça je peux en parler longtemps parce que j'étais encore en réunion où j'étais en difficulté ce matin : on est clairement sur des questions de platebandes. Si on regarde un peu l'historique, avant l'*open data*, qu'est-ce qu'il y avait ? Il y avait le *SIG*. Le *SIG* avec la directive *INSPIRE*, qui est quelque part, la base de l'*open data*. *INSPIRE*, c'est la fondation d'une démarche *open data* sauf qu'on était sur des données géographiques. Structuration de données, métadonnées, fraîcheur. Inspire a posé les bases d'une démarche aujourd'hui qu'on connaît sous le nom d'*open data* parce que l'open data ne se limite pas à des données géographiques. Dans les collectivités, du fait d'*INSPIRE*, il a été mis en place des systèmes d'information géographiques.

Donc il y a dix ans de ça, et c'était en 2012 au conseil départemental de [nom de département] : embarqué dans une démarche régionale, le conseil départemental a mis en place une plateforme qui s'appelle geo|nom de département, qui est dépendante de la plateforme régionale qui s'appelle geo|nom de région, donc Géo|nom de région, et qui est donc une infrastructure de données géographiques. Dans cette infrastructure de données géographiques, plusieurs fonctionnalités principales : le catalogue de données, une cartothèque, un outil de visualisation de données géographiques. Grâce au SIG sont arrivées d'autres gestions du type gestion de la route, gestion de la rivière, gestion du cadastre… voilà, des éléments dont on a besoin et dont les métiers ont besoin dans les collectivités comme dans les communes. Mais là, la question de la data, qui ne vient pas, éradiquer tout ce qu'a fait le *SIG*, mais la data (et l'*open data*) se placer à un niveau qui est juste au-dessus puisque l'*open data* va sortir du cadre strict de la donnée géographique ; on n'est plus dans une question de données géographiques, on est dans une question de données au sens large, quelles qu'elles soient. Et de valorisation de ces données. Et là moi la difficulté que je rencontre c'est quand initialement on a pas mis en place ce projet avec les équipes *SIG*, eh bien là il y a un point de friction qui se fait et d'incompréhension, entre ce que recouvre l'open data et ce que recouvre le *SIG*. Comment on peut faire parler les deux. Et la data et le *SIG* justement. Ce n'est pas qu'une question d'open data d'ailleurs. La donnée géographique soit d'inscrire dans la démarche de gouvernance des données d'une collectivité.

Donc concrètement, ce n'est pas vraiment un problème théorique sur ce qu'est la donnée qui pose le problème actuel.

Non. Non, la partie théorique, ça s'acquiert et ça s'acquiert en accompagnant les directions. Moi, je suis chargée de gouvernance des données mais le poste que j'occupe c'est chargée de transformation numérique, c'est-à-dire qu'effectivement, une de mes missions, c'est qu'à chaque projet data pour lequel je peux être saisie, j'apporte aussi de la transformation aux métiers : la transfo numérique mais pas que. Typiquement, je dois pouvoir apporter de l'expertise data en leur disant « non mais attendez vous voulez faire ça, c'est super intéressant mais regardez : vous avez telles données en interne, y a des données qui existent en open data, vous voulez en savoir

plus, eh bien on va croiser nos données avec des données open data. Là ça va vous apporter de la connaissance, ça va vous permettre d'élargir vos savoirs, et puis de commencer à prendre des décisions. Donc, en fait, cette acculturation méthodique des services, elle se fait au fur et à mesure des projets (par le faire, par la pratique). La difficulté majeure que l'on rencontre, c'est plutôt la compréhension fine de ce que recouvre la data, et des enjeux derrière un porte de gouvernance des données, et donc de mon positionnement au sein de la collectivité.

On entend souvent le terme d'esprit critique sur la donnée, est-ce que ça vous parle ?

Ce n'est pas un terme que j'utilise beaucoup. En tout cas sur la partie, dans mes missions, etc., on va très peu l'utiliser. On sait par contre quand même effectivement que quand on va notamment chercher à exploiter des données déjà présentes en open data, etc., qu'il faut une forme d'esprit critique et qu'il faut une forme de vigilance par rapport à ce que l'on va exploiter et au message que l'on va sortir de ces données-là. Mais sinon, non, on n'utilise pas beaucoup ce terme.

Est-ce que vous avez en tête un projet en culture des données qui a bien fonctionné et si oui pourquoi ?

Clairement, si je prends juste mon cas, c'est tout le chantier que j'ai lancé sur le *close data* qui a eu un vrai effet levier sur les directions qui sont impactées aujourd'hui par ce *close data*. J'ai des données RH, des données finances, des données autour de la solidarité, donc ce que je disais tout à l'heure : la PMI, l'aide sociale à l'enfance, les mineurs non accompagnés. J'ai des données autour des personnes âgées et des personnes handicapées, ce que l'on appelle des PA/PH, enfants et adultes. J'ai des données autour du Covid, suivi de l'épidémie. J'ai des données emploi-chômage. J'ai des données autour du RSA. Eh bien ça, toutes les directions qui ont participé à ça aujourd'hui, ça a été un vrai effet levier : il y a eu un accompagnement qui s'est fait direction par direction, donnée par donnée, jeux de données par jeux de données, avec « comment on va chercher la donnée dans l'outil métier », comment fait-on des requêtes. Tous ces gens-là aujourd'hui, l'acculturation elle est faite. Et ça fait l'effet waouh. C'est « ah ouais on peut faire ça ». Et ça fait un effet boule de neige… après avoir exploité nos propres données, les directions se disent : « on pourrait se comparer avec des départements de même strate parce que ça c'est quelque chose d'intéressant qui revient régulièrement dans l'évaluation des politiques publiques ». Et après j'ai le cas spécifique de l'observatoire du RSA par exemple qui est un tableau de bord qui permet de suivre le nombre de ménages allocataires du RSA dans le département et par zones d'action médico sociales (découpage géographique de [nom de département]) Et donc, cet observatoire du RSA (tableau de bord) a été conçu de telle façons qu'on compare systématiquement le phénomène départemental à une commission RSA (ex : pour les élus qui sont membres des commissions RSA du territoire). Historiquement, les données qui sont dans ce tableau de bord étaient transmises aux décideurs, via un fichier PDF, dont la préparation demandait environ 8h de travail par mois. A l'heure actuelle, la mise à jour des données du RSA dans cet outil, c'est 30 min. Donc au-delà de cette réussite technologique, ou numérique…au-delà d'avoir accompagné les métiers dans la définition de leurs besoins, les métadonnées, etc., il y a ce gain pour les services et là on est sur quelque chose qui est hyper visible parce qu'effectivement on est passé d'une petite journée de boulot par mois avant à moins de 30 minutes et on laisse le système tourner tout seul dans la nuit avec un système d'API. Toutes les données départementales sont à jour le lendemain matin et le tableau de bord est à jour pour le président à J+1.

En conclusion, si vous aviez des recommandations ou des remarques pour améliorer l'acculturation aux données en France ?

SI j'avais une remarque ou une recommandation, ça serait que déjà il faut que ça commence très tôt cette histoire d'acculturation aux données. C'est qu'on soit clairement sur des choses évoquées en collège par exemple. J'en reviens à une expérience que j'ai eue avant d'être chargée de la gouvernance des données, j'ai été pendant 3 ans, cheffe de projet *SIG*. A cette époque-là, je n'arrêtais pas de dire puisqu'on a les compétences numériques dans les collèges, c'est nous qui installons les ordinateurs dans les collèges, c'est nous qui dépannons et qui assistons les établissements scolaires en termes d'installations informatique… « pourquoi on ne va pas dans les collèges leur expliquer qu'ils ont une infrastructure de données géographiques à disposition, pourquoi est-ce qu'on explique pas aux profs de géo que demain leurs élèves, ils pourraient faire des cartes parce que l'outil le permet. » Voilà. Pourquoi on n'acculture pas les enfants dès le départ à apprendre à faire des cartes parce que ça va leur ouvrir des horizons ?

Eh ben clairement, sur la partie « donnée », la question est la même : les hackhatons, les serious games, tous ces outils ludiques, devraient être proposés en collège. Parce qu'en plus les enfants, ils ont des idées lumineuses dans certains cas. Dans des manifestations de type hackhatons, on utilise l'intelligence collective, le *design thinking*,etc., clairement aujourd'hui, même nous au niveau de notre expérience professionnelle, d'un âge déjà certain, on nous dit « allez-y » - exploiter les idées les plus farfelues, tentez tout ce que vous pouvez. Mais je pense vraiment, que les enfants là-dessus ont plein d'imagination qui méritent qu'on s'y atèle. Mais pour qu'on s'y atèle, il faut qu'ils sachent que ça existe et moi je pense que cette histoire d'acculturation, ces histoires de data, d'open data, de comprendre la donnée, de savoir ce que l'on peut faire avec la donnée. De justement, avoir un esprit critique sur la donnée, c'est à cet âge-là que ça s'attaque.

Nous aujourd'hui, département [nom du département], j'ai une collègue qui vient d'arriver, qui est chargée des usages numériques, qui est censée identifier tous les métiers qu'il y a dans la collectivité, et pour tous ces métiers de construire une sorte de package de base de connaissances numériques. C'est-à-dire que quand un nouveau professionnel arrive au CD, il aura un outil de type LMS pour apprendre, des vidéos, des tutoriels, pour mieux maitriser les outils numériques de notre collectivité. Donc la data s'inscrira dans ce parcours du nouvel arrivant. Nécessairement il va y avoir un petit quiz, des vidéos à regarder, des replays, des choses comme ça sur comment gagner en culture data. Mais ça devrait être fait bien avant. On a quitté l'ère d'*Excel*, c'est terminé. Il y a 5 ans, il y a 10 ans on était encore full *Excel*, les tableaux croisés dynamiques c'était génial. Ça permettait de faire des trucs super. Mais là il faut qu'on passe le cap d'après et le cap d'après c'est la donnée au sens large.

Institutionnel 7 (10/02/23)

La première question que je voudrais vous poser est : déjà est-ce que vous employez le terme de *data literacy* ? Et si vous ne l'employez pas, est-ce que vous en employez un autre comme culture des données ou acculturation aux données par exemple.

Le terme data fait peur ! L'utilisation de la data fait technique, *literacy* est un terme anglais, donc oui entre techniciens on peut parler *data literacy* mais quand on parle à des collègues ou quand on parle à des élus, moi je parle de culture de la donnée. On utilise trop de mots anglais dans le numérique ce que l'on reproche souvent, et l'enjeu est de rendre accessible le sujet et non d'utiliser du jargon technique.

Je voulais me présenter aussi pour spécifier quelle est la nature de mon poste : je ne suis pas intégrée à la DSI. Je ne suis pas tech. Je suis directement rattachée à une DGA ressource dans laquelle il y a la DSI mais également les assemblées, le juridique, la commande publique, etc. Ce positionnement est volontaire, je n' ai pas une formation dans l' informatique. J' 'ai un profil d'urbaniste au départ et ensuite plutôt autour de la communication et des usages numériques. Je pense que c'est important eu égard au sujet qu'on aborde.

J'essaye toujours d'aborder les questions sous un angle usages, services, et pas du tout technique. Et du fait de ma formation cela me permet naturellement d'avoir cette approche-là, sans être obligée de me forcer vu que ce n'est pas ma formation initiale. Dans la feuille de route transformation numérique, il y a une ambition générale qui est d'innover dans la production et la diffusion de services publics en prenant appui sur la data, l'IA, et dans des processus de design thinking permanents. On ne produit pas un service, on ne réfléchit pas à un outil si on n'a pas associé les futurs usagers. Donc on partage d'abord autour du besoin avant de proposer des services et des outils.

On a un deuxième pilier : cette société numérique est porteuse et tout le monde peut y accéder. Le défi est donc de ne laisser personne au bord du chemin avec une ambition forte autour de l'e-inclusion interne / externe, agents, élus, citoyens, professionnels. L'acculturation à la data est intégrée à notre démarche. Et on a un troisième pilier qui est l'accompagnement de la filière numérique locale mais avec un angle « numérique responsable ». La direction de l'économie aborde le sujet sous l'angle de l'attractivité économique. Ce que je viens de décrire est notre feuille de route de transformation numérique. La data y est extrêmement présente comme levier de transformation. On coopère aussi beaucoup avec Ekitia que vous connaissez sans doute avec cette volonté d'une data citoyenne, éthique, souveraine, sobre, etc. Voilà pour donner un cadrage général pour notre entretien.

Vous travaillez à la communauté d'agglomération [nom de l'organisation], une collectivité très spécifique...

Alors oui, une communauté d'agglomération est un EPCI, un établissement public de coopération intercommunale. On a 83 000 habitants et 36 communes et on déploie 18 compétences. On a quasiment toutes les compétences dans la relation à l'usager.

On a 700 agents sur 1200 qui travaillent dans le domaine social : de la crèche jusqu'au portage de repas à domicile. Donc on a vraiment une forte activité autour de l'humain. Nos autres compétences tournent autour de l'eau, l'assainissement, les déchets, l'économie, l'urbanisme, la voirie... ce qui fait que naturellement, on a énormément de data issues de toutes ces activités de gestion. On a également des délégations de service public pour, par exemple, la collecte des déchets, ou pour l'eau.

Quand je suis arrivée à mon poste en 2019, j'avais depuis 2016 ouvert nos données sur un portail *open data*. On est en [nom de région], c'est un territoire pionnier sur ces questions-là. Toulouse notamment a versé ses données depuis 2010. L'environnement était donc favorable. Déjà en 2012 je voulais qu'on ouvre les données : j'étais au département [nom de département] et j'avais proposé à mon collègue SIG de faire de l'*open data*. Ma proposition n'avait pas été retenue à l'époque. Quand je suis arrivé(e) au [nom d'organisation] j'avais cette préoccupation et me suis rapproché(e) de la Préfecture de région et du département parce que je savais qu'ils avaient des capacités à accompagner une démarche d'ouverture.

Il y a eu un projet qui s'appelait *Open Data Lab* qui a été développé par Fabien Moguen en 2015 et donc je me suis greffé(e) tout de suite sur ce projet-là. A l'époque j'étais directrice adjointe/directeur adjoint à la communication, ce qui n'a pas été neutre dans la manière dont j'ai approché ce projet. J'avais une feuille de route à dérouler : site internet, site intranet, réseaux sociaux, applications en ligne, etc. Une fois que tout était en place, ou presque, je me suis dit « l'open data c'est un moyen aussi de faire de la transparence publique, d'être un peu pionnier sur le sujet, du coup d'être visible, de faire de la dataviz pour partager une vision de nos politiques publiques mais aussi des données brutes.

A l'époque, le sujet était peu porteur dans les collectivités, on a donc eu un soutien très fort de l'Etat avec Fabien Moguen qui nous a accompagnés dans l'ouverture des données pendant un an gracieusement. Cette aide nous a vraiment permis d'accélérer et en 2016 on a ouvert notre portail. Un portail qui est hébergé par le département dans le cadre d'une démarche mutualisée. Ce fut le début d'un accompagnement fort des collectivités à l'*open data*. On était co-porteurs avec la cocréation de kits pédagogiques, de services types etc.

D'emblée j'ai porté le projet en interne alors que la Loi Lemaire disait « open data par défaut », j'ai préféré déployer des démonstrateurs opérationnels.

Six démonstrateurs qui permettaient de dire « à quoi sert l'open data ¿ ». On est donc parti sur open agenda. *Open agenda* qui a été à la source de *D-line* développé au sein d'*Open data France.*

D-line qui a été réalisé par Thierry Martin parce qu'il s'est dit « on ne va pas faire un outil chaque fois qu'il faudra sortir des données standardisées en faisant appel à des prestataires donc on va le faire nous-mêmes ». En tout cas, on est parti sur open agenda et vis-à-vis des élus, je leur disais « vous voulez rendre visibles vos évènements, vous n'avez pas de ressources humaines pour pouvoir réaliser cette mise en avant ? Nous vous proposons un service pour valoriser vos évènements. La moitié de nos communes ont moins de 500 habitants et n'ont pas de ressources dédiées à la communication. Open agenda c'est un moyen de pousser les données à un seul endroit et de les publier ensuite partout sans effort supplémentaire. Et en présentant ainsi le projet j'ai toujours facilité l'open data sans parler d'*open data*, en ne parlant que services et utilité. On a toujours pris cette entrée-là, usage et s'est relativement bien passé par rapport aux collègues. Ce fut le premier démonstrateur.

Ensuite il y a eu *Waze*. On a eu un partenariat avec *Waze* dès 2015 et donc mes collègues de la voirie publiaient sur un standard les données sur les chantiers pour proposer les contournements, donc là aussi c'était visuel parce qu'on voyait sur l'application tout de suite les données publiées par le Sicoval. Puis il a eu les menus de la restauration collective : là on a travaillé en coopération à l'échelle locale avec le Cotech data locale pour

proposer un standard qui est devenu le standard national. Il est né de nos réflexions avec l'idée d'avoir sur un même outil un service unifié, des tout petits jusqu'au lycée.

On a aussi souhaité proposer une application de service pratiques à nos habitants avec l'intégration de chatbots. Il s'agit à partir de la donnée, qu'elle soit "scrappée" sur des sites web ou à de données en *open data*, d'offrir un service sans couture pour les petites communes. On a développé une dizaine de *chatbots*. Un *chatbot* qui est donc celui du [nom de l'organisation] et les autres propres aux communes.

Les petites communes, même celles qui n'ont pas de site web peuvent offrir un service numérique. Ce type d'outil permet de montrer aussi que s' il n y a pas de données ou si la donnée n'est pas de qualité, c'est tout de suite visible. On se rend compte que si on n'a pas de donnée de qualité, le service n' est pas de qualité. Puis évidemment de l'open data on a glissé vers l'IA, le *big data* au sens le plus large, mais on a toujours mis un peu au cœur de notre transformation numérique, la donnée.

Quand on a mis en place la démarche *open data*, il a fallu accompagner à l'interne la sortie des données. J'avais un collègue qui avait de l'appétence et une fibre pédagogique, qui à partir d'outils que je lui avais partagé de l'Infolab, des jeux, de toute cette matière qu'il y avait au niveau national, dans le *Tubà* à Lyon dans les collectivités, s'est inspiré et a fait des jeux à destination des agents. Dans cette période est sorti le *RGPD* donc il y a eu particulièrement focus sur le *RGPD* mais pas que : il y a un atelier sur l'*open data* justement, sur la qualité de la donnée, le numérique responsable, les indicateurs, la mise à jour, etc. Tous les agents sont invités à ces ateliers, quel que soit leur métier, il n'y a pas de métier exclusivement dédié à la data.

Il y a aussi une quinzaine d'ambassadeurs de la donnée qui sont dans chacune des directions. Là encore le ludique prime : ils ont droit à des challenges, à des défis, à des trophées, voilà, pour créer un peu de stimulation et d'envie ; je les ai vu passer hier avec un chevalet et une roue pour identifier les données dans les directions pour alimenter un catalogue de la donnée. Nous menons donc une démarche lourde à l'interne sur cette « culture de la donnée » en l'adressant à tout le monde.

La diffusion du numérique s' est fait aussi à travers l'évaluation *PIX*. 600 agents sur 1200 ont joué le jeu avec la promesse de fournir des formations adaptées en suivant et selon son niveau de *PIX* obtenus. Dans ce cadre-là, on a construit avec le département [nom de département] des « communs numériques » de sensibilisation à destination des agents publics. Il y aura une partie destinée aux agents de [nom de département] et un autre lot déposé sur La Base, la ressourcerie pédagogique de l'*Anct*. C'est pour cela qu'on a proposé à Jean-Marie Bourgogne de faire un webinaire pour expliquer ce que c'est que la donnée, à quoi ça sert, flécher vers toutes les ressources d'*Open data France* qui existent parce qu'on se dit que, oui, la *data literacy*, la culture de la donnée, c'est indispensable pour un agent public, quel qu'il soit, où qu'il soit.

Et le pendant de toute cette démarche qui permet de sensibiliser à de la donnée de qualité est comment l'exploite-t'on?, comment l'analyse t'on quand on n'a pas de ressources humaines dédiées, que l'on n'a pas de *data scientists*, qu'on commence à avoir du *big data*...Par exemple on a mis fin à notre partenariat avec *Waze* parce qu'on avait des millions de données dont on ne faisait rien parce qu'on n' avait pas de *data scientist* capable de l'analyser, d'en faire quelque chose d'intelligent. On pense à l'externalisation, on a mutualisé aussi

avec le département un alternant de l'*École Simplon* qui nous a fait une petite application pour mieux gérer les flux d'accès aux déchetteries.

Cette problématique du manque de ressources pour analyser les données produites créé du « déceptif » par rapport aux collègues.

On a aussi un partenariat avec *Enedis* qui nous permet de récupérer à l'échelle de dix PDL et sur un pas de temps mensuel toutes les données *Linky* de tout le territoire, de tous les habitants, de toutes les entreprises, Potentiellement, on peut le faire. On a pris deux cas d'usage. Un cas d'usage sur la corrélation entre précarité énergétique et copropriétés dégradées et un autre sur le décret tertiaire. Ma collègue me dit « c'est positif d'avoir obtenu toutes ces données *Linky* mais qu'est-ce que j'en fais ? Je ne sais pas quoi faire de ça ». Et donc *Enedis* a été soutenant ils nous ont produit une petite analyse rapide. Mais le constat est là, on n'a pas les outils, on n'a pas les ressources humaines et donc on s'est retrouvé avec beaucoup de données mais pas de capacité à les exploiter pour en faire quelque chose d'intelligent.

Justement, par rapport à tout ça, puisque tu l'évoques : quelle est l'importance que vous donnez à cette acculturation aux données par rapport à d'autres nécessités. Parce qu'évidemment, depuis le temps, depuis plusieurs décennies, il y a la nécessité informatique, en tout cas de savoir utiliser certains outils de bureautique, savoir un peu décoder l'information. Là, le fait qu'on parle des données, est-ce que c'est devenu une nécessité pour vous, en tout cas pour les publics de ton organisation et ceux qui sont en-dehors. Donc, quelle est l'importance de l'acculturation aux données et depuis quand c'est devenu une nécessité en fait ?

Je ne sais pas si c'est l'or, si c'est le pétrole, si c'est le carburant de la transformation numérique mais oui c'est le cœur. C'est-à-dire que sans données, on ne fait pas de transformation numérique. On s'est rendu compte qu'il y avait cette question de la qualité de la donnée, de sa standardisation et de son partage.

Toutes ces questions-là, on les aborde depuis maintenant presque dix ans, et l'enjeu aujourd'hui, si on veut développer du service, c'est qu'on ait de la donnée de qualité. Et qu'on en dispose dans un cadre éthique, souverain. Ces valeurs sont très importantes pour nous. C'est pour cette raison qu'on est membre fondateur d'Ekitia parce que les élus portent vraiment la dimension numérique responsable.

Mais l'enjeu, c'est aussi l'acculturation à la data de tous. J'évoquais tout à l'heure le volet inclusion numérique : on a une charte territoriale des acteurs de la médiation qui a été construite avec eux. J'ai posé un acte dedans qui est de dire que les acteurs qui signent cette charte, il y en a 17 qui l'ont signée, doivent s'engager à faire de la *data literacy*, de la culture de la donnée, accompagner le grand public à la compréhension de la donnée : données personnelles, données publiques, à quoi ça sert, pourquoi certains sites sont gratuits ? Parce que derrière on capte nos données, etc., les modèles économiques de la donnée. Ces éléments essentiels à la compréhension du monde dans lequel on vit sont abordés dans nos communs numériques. On se donne comme responsabilité de former le grand public à ces enjeux-là.

Avec Ekitia on met en place des ateliers citoyens de la donnée. Dans le cadre de notre semaine du numérique responsable qui aura lieu du 13 au 18 mars, 40 évènements sur tout le territoire, on a un atelier spécifiquement dédié à la donnée. Un atelier citoyen parce qu'on se dit qu'il faut vraiment acculturer tout le monde. Par exemple, on a fait une opération citoyenne de *tracking GPS*. On a proposé à des volontaires d'être traqués via une application pour partager leurs données avec une régie de données publiques. On les a complètement sécurisés sur le *RGPD*, la manière dont on allait traiter leurs données etc. On a ainsi embarqué une centaine de volontaires qui pendant un mois ont tracé leurs trajets et validé leur parcours (trottinette électrique, voiture, vélo, etc). et le motif de ce déplacement. On a ainsi récupéré 55 000 km de traces qui ont été faites par ces volontaires. Pour arriver à ce type de projet et obtenir la contribution des personnes qui acceptent de partager leurs données, il faut qu'il y ait une culture de la donnée.

S'il n'y a pas un engagement de notre part à être souverain et éthique sur cette donnée partagée en confiance, on ne peut pas arriver à faire ce type d'opération. Si on a réussi à le faire, c'est que l'on a sécurisé le volet *RGPD*. On a tenu au courant les volontaires au fil de l'eau, on a restitué les résultats de cette opération à ceux qui le souhaitaient lors de la semaine de la mobilité parce qu'il nous semblait important de faire un retour pour maintenir la cadre de confiance

On s'est aussi engagés à la confidentialité des données, je ne connais pas ceux qui ont contribué à cette opération. L'ensemble des traces, ont été versées dans une régie de données publiques sécurisée à la direction départementale des territoires et seront détruites au bout d'un an.

On n'a pas eu accès aux données, on ne les aura que sous une forme complètement agrégée. Ils les ont anonymisées et nous ne partagerons que des grandes masses. Toutes les traces ont été coupées sur les 200 premiers mètres afin d'éviter la reconnaissance du lieu de départ, en milieu rural il est facile d'identifier les personnes.

Je vais rester sur le côté « enjeux » et après peut être plus insister sur le côté définitionnel. Aux niveaux des enjeux, tu as parlé d'interopérabilité, d'accessibilité, de transparence j'imagine aussi sur pas mal de choses. Et là j'aimerais une confirmation : au niveau des types de données, celles que tu estimes être au centre de tous ces enjeux-là, au final c'est la donnée personnelle et la donnée publique ?

Les deux !

Les deux étant dissociées. Est-ce qu'il y en a d'autres ?

Alors il y a la donnée publique, la donnée personnelle, mais aussi la donnée des opérateurs type *Enedis* par exemple. Ou la donnée de *Suez* puisque c'est *Suez* qui est notre délégataire de services publics sur les déchets. Donc on a aussi cette donnée-là et on a mis des clauses *open data* dans les marchés pour pouvoir aussi récupérer cette donnée-là.

Et par rapport à ça, puisque tu l'as évoqué tout à l'heure, c'est par rapport au cadre législatif. Tu as cité le RGPD, on a eu la loi Lemaire, on va avoir le data governance act et le data act en 2024 donc il y a pas mal de choses qui se sont faites autour de la donnée. La question que je voudrais te poser c'est

quel est ton avis par rapport à ce cadre législatif ? Est-ce un premier pas ? Est-ce limité ? Est-ce que ça a eu un impact sur tes pratiques ? Comment vois-tu cela ?

Je trouve que c'est extrêmement précieux ce que fait l'Europe sur tous ces sujets-là. On a ainsi signé la déclaration *Join Boost Sustain* signée par 169 collectivités dont 9 en France qui nous engage sur un numérique éthique, souverain. L'enjeu ? Faire l'équilibre entre ce qui est fait du côté américain, du côté chinois et les valeurs portées par l'Europe.

Dans le cadre de la signature de la déclaration on est allé à Bruxelles et on a rencontré « Monsieur RGPD » qui est donc la personne qui a réfléchi le *RGPD*, évidemment pas tout seul, et qui nous a raconté plein de choses intéressantes sur sa création, sa mise en œuvre, ses limites ...En revenant on a pris conscience que « c'est dans le cadre européen que se créent toutes les politiques du numérique en fait ».

D'où l'intérêt d'être au plus près pour mieux comprendre les ressorts de ces politiques qui impactent notre vie quotidienne. Parce qu'aujourd'hui même si les gens ne savent pas ce qu'est précisément le RGPD, on voit bien les cookies, etc., sur les sites web. D'où l'importance de donner du sens aussi « pourquoi vous voyez les cookies dès lors que vous allez sur les sites internet ? ». Et quelles sont les conséquences à donner systématiquement, son consentement pour accéder au site.

Donc je trouve que c'est extrêmement important toutes ces politiques, le *DMA*, le *DSA*, tout ce qui est en train d'être produit au niveau de l'Europe. Lors de la mise en place du *RGPD* on a pas mal râlé parce que ce cadre contraignant mais cette réglementation donne aussi une manière d'envisager le numérique en France ou en Europe qui ne porte pas les mêmes valeurs qu'en Chine ou aux États-Unis. Cette dimension politique du numérique est importante à expliquer à nos habitants ou aux collègues.

Je trouve que ces cadres européens qu'on décline après au niveau français sont indispensables.

Ils permettent aussi à nous collectivité de mobiliser en sécurité nos habitants. Il y avait une très bonne enquête qui avait été faite par *Ekitia*, à l'époque *Occitanie Data* et *Ipsos* sur l'acceptabilité de la donnée : « est-ce que vous seriez prêts à partager vos données ? ». Je ne sais pas si tu y as eu accès mais elle était vraiment très intéressante. En fait je pense que plus on sécurisera le cadre de confiance dans lequel on veut agir avec les citoyens dans le partage de la donnée, plus on aura des citoyens qui seront prêts à partager leurs données. En fait c'est cette donnée personnelle qui est riche. Mes collègues quand ils ont fait une étude mobilité, ils ont payé cher des données de *Google*. Moi je dis qu'on a quand même une responsabilité pour créer des cadres souverains de partage de la donnée qui ne soient pas ceux des américains. Il y a toute cette dimension aussi quand on parle de l'acculturation à la donnée celle du modèle économique, quel numérique veut-on ? Subi ? Choisi ?

C'est la question que j'allais poser par rapport à l'acculturation : qu'est-ce que les gens qui sont acculturés, qui ont suivi cette acculturation à la donnée, ils doivent avoir développé principalement. J'ai compris qu'il avait une connaissance globale du cadre économique, sur ce qu'est la donnée j'imagine. Est-ce qu'il y a aussi les compétences techniques ? De traitement de données, ce genre de choses.

D'accès aux sites d'*open data*. Ou est-ce que tout ça c'est secondaire et on reste dans une culture générale avec de gros guillemets ?

Alors, y a les deux. Il y a globalement pour l'ensemble, et là je parle des agents, une culture à la donnée avec ses ateliers qui sont un peu thématiques mais qui n'ont pas vocation à aller dans le détail. Et après, il y a un accompagnement. Par exemple, notre collègue diététicienne qui publie les données de la restauration collective, ce n'est pas son métier de base . Mais en donnant du sens, en expliquant que c'est pour livrer aux parents des informations de qualité. Après une formation à la publication, elle est devenue « appétente » sur la data. Comme quoi, quand on donne du sens, la motivation est là et cela donne aussi l'envie d'aller plus loin.

Ce collègue accompagne chaque agent pour aller plus loin dans la prise en main technique des outils. Et s'ils ont envie d'aller plus loin, il va un cran plus loin avec l'agent. On a aussi le collègue du *SIG* qui déploie une vision plus technique avec la voirie, l'assainissement, etc. On a ainsi énormément de jeux qui sont moissonnés sur notre portail *open data*.

Dans le cadre des travaux qu'il mène avec ses collègues, il les forme aussi sur les notions de standard de données, la mise à jour, la qualité etc. Donc il y a la double dimension : une dimension acculturation générale pour ceux qui n'ont pas forcément vocation à aller beaucoup plus loin et une vision plus technique pour les plus appétents qui veulent approfondir. Par exemple, j'ai une collègue qui publie tous les marchés publics pour faciliter le prédictif et anticiper sur 5 ans les marchés qui vont sortir. Cette collègue de la commande publique s'est formée à la data et a développé de nouveaux services.

Celle qui gère les subventions n'avait pas d'intérêt particulier à la donnée. Aujourd'hui elle a compris que ce qu'elle faisait eue un vrai sens politique parce que permettre de visualiser les bénéficiaires de nos subventions permet de déconstruire les présupposés. On avait l'impression que c'était l'économie qui était le plus fort bénéficiaire et on s'est rendu compte que c'était les acteurs du culturel.

Elle s'est ainsi dit « ça a du sens ce que je fais ». Du coup elle s'y est intéressée un peu plus. Voilà, on est dans la double dimension. Culture générale, vernis, et approche plus approfondie par rapport à des actes plus techniques.

Justement par rapport à cela aussi, parce que là j'entends qu'il y a une acculturation, qu'il y a pour certains publics une appétence, mais à l'inverse, est-ce qu'il y a beaucoup de freins par rapport à cette acculturation-là ? Il y a des personnes qui peuvent ne pas être intéressées, qui visualisent la donnée comme quelque chose de trop technique, trop compliquée, est-ce qu'il y a des freins qui ont été identifiés par rapport à ça ?

En fait, je pense qu'il y a un vrai enjeu à donner du sens. Et le collègue qui accompagne à la culture de la donnée a beaucoup de charisme... Il donne du sens opérationnel, pratique, à quoi sert la donnée ? Pourquoi je dois la qualifier. Et ils le vivent non pas comme quelque chose qui vient les perturber ou chronophage mais comme quelque chose qui a du sens pour le collectif : ce n'est peut-être pas nécessaire d'avoir trois bases de données d'associations, plutôt partageons la même ! Aussi c'est un gain de temps puisqu' il y en a plus qu'une qui est

mise à jour. Dès lors qu'on arrive à montrer qu'il y a du gain de temps, du gain de qualité et que ce que je fais peut servir à d'autres, je trouve que les collègues petit à petit, parce que cela fait des années qu'on fait ce travail sur la donnée, adhèrent. Quand on travaillait sur l'*open data*, il y en avait qui avaient peur, ils pensaient y passer trop de temps. On a objectivé et on leur a dit « bon, combien tu as passé de temps objectivement sur l'année à rentrer tes subventions ? Deux fois une demi-journée ? Bon, sur l'année c'est peut-être gérable". En fait, oui, il y a des craintes et finalement on s'aperçoit qu'elles ne sont pas toujours fondées.

D'ailleurs ce que j'entends, c'est souvent des craintes sur les usages autour de la donnée que la donnée en tant que concept ?

En effet, ils ne le vivent pas forcément comme un concept la donnée, ils les voient plutôt comme un élément technique qui permet d'aboutir à un certain nombre de services. On essaye vraiment de mettre à portée cette dimension technique qui doit être complètement gommée pour la transformer en service. Soit pour de l'aide à de la décision, parce que la donnée c'est aussi de l'aide à la décision, soit pour la production de services, soit pour de la transparence publique : on essaye de donner du sens pour que chacun puisse percevoir la donnée qu'il gère, être conscient de détenir de la donnée parce que certains n'ont même pas conscience qu'ils ont de la donnée. Déjà d'avoir cette conscience là et de se dire maintenant est-ce que ça peut servir à quelqu'un d'autre, est-ce que ça peut servir à un projet collectif, ou est-ce que c'est intéressant d'avoir des normes et des standards pour pouvoir mieux opérer et interopérer les uns avec les autres c'est un des impacts importants de l'acculturation à la donnée.

Donc dans ce que j'entends et je crois que ça a été dit tout à l'heure, c'est que ça doit concerner tout le monde en fait ? Mais à des niveaux différents selon les publics j'imagine.

Tout à fait. En fait, cette acculturation on la propose à tous les agents. Le collègue a même créé un « panini » pour valider les étapes et prouver la participation aux ateliers. Il crée tout un *wording* autour d'un jeu et devient Maître Yodata . Le côté ludique est important pour remporter l'adhésion. Et une fois que cette adhésion est faite, on peut aller sur des aspects plus techniques.

Justement par rapport à ça, j'ai une question à la fois basique et à la fois extrêmement compliquée : comment est-ce qu'on définit la donnée ? Est-ce qu'il y a une définition qui est donnée ?

Alors oui, *Open Data France* avait fait des supports sur ce point-là pour distinguer données - information – connaissance. On utilise ces supports-là aussi. Pour moi la donnée, c'est ce qui est zéro-un en fait, c'est numérique, mais bon on n'aborde pas les choses sous cet angle-là parce qu'on veut éviter d'aborder les choses sous l'angle technique, informatique. On veut donner du sens en faisant quelque chose de très concret. C'est pour ça que j'en parlais à des étudiants de la data justement il y a deux jours, je leur montrais notre petite application *chatbot*, je leur montrais « voilà, vous voyez la donnée elle n'est pas à jour puisqu'on a des évènements de septembre ». Qu'est-ce que cela signifie ? Cela veut dire que la base de données initiale est mal paramétrée pour se connecter au *chatbot*, il faut aller chercher les raisons techniques pour lesquelles cela ne fonctionne pas. Et je leur disais « voilà, si la base de données n'est pas propre, si le connecteur n'est pas correct, derrière le service il n'est pas qualitatif ». On a aussi travaillé avec les élus et un jeu d'*Ekitia* qui s'appelle *Belle Colline*

produit initialement par le département de la Haute Garonne. Ils devaient créer une mini application à partir de données. Ils ont compris que s'ils voulaient avoir des applications de qualité et riches ils devaient disposer de la donnée.

A un moment un élu a dit « tiens on pourrait créer un tel type de service ». Quelles sont les données dont j'ai besoin? Est-ce que c'est possible de le faire ou pas ? Cette démarche servicielle permet de prendre les choses à l'envers. Quel est le résultat que je veux obtenir et comment je remonte jusqu'à la source, quelles sont les données réellement disponibles ? Nous on essaye toujours de parler d'usage avant de parler technique. Et si possible on ne parle pas de technique !

J'ai deux questions qui n'ont rien à voir l'une avec l'autre. D'abord pour l'acculturation en général, c'est un projet qui est donc mené à travers différents ateliers, différentes manières de le faire, est-ce que c'est un rouage de ce qu'on entend par la gouvernance de la donnée ? Ou est-ce que ça peut faire office de gouvernance des données ?

En effet, la question de la gouvernance de la donnée, on y travaille depuis longtemps, c'est complexe. Même plus globalement, la question de la transformation numérique. On a un petit groupe de travail entre la directrice de la prospective de l'évaluation, ma *DGA*, la *DSI*, et notre élu dédié au numérique, pour essayer de mettre en place cette gouvernance de la donnée. On a partagé des vidéos qu'avait fait *Métapolis* sur la gouvernance de la donnée, très pédagogiques et très méthodologiques. On essaye d'avoir une lecture commune autour de cette question-là : la gouvernance de la donnée du point de vue politique et du point de vue technique. Et pour moi la culture de la donnée c'est l'une des briques de la gouvernance de la donnée. Pour moi la gouvernance de la donnée c'est comment à l'échelle d'un groupe, d'une collectivité, d'une entreprise, etc., on s'organise collectivement pour produire de la donnée de qualité et partager ensemble cette donnée. Et elle doit aussi bien intégrer le niveau politique que le niveau technique. Ce n'est pas quelque chose qui est dissocié, c'est ma lecture de la gouvernance.

Pour moi la gouvernance c'est vraiment ce qui permet d'organiser cette gestion de la donnée, le cycle de vie de la donnée, de sa production jusqu'à son archivage, et donc la gouvernance elle est là pour gérer tout ce cycle de vie de la donnée et le niveau également politique.

Vraiment chez nous, je pense que nos élus ont une vision très politique de la donnée, de la transformation numérique, comme un outil de de gestion publique du territoire. La donnée est le carburant de cette transformation numérique, et elle-même une dimension stratégique et politique très portée chez nous. Le président est un ancien *data scientist*. Il est vice -président de l'association des *Interconnectés*. Il est également vice-président au numérique de Intercommunalités de France. Il porte ces questions du numérique très haut et en permanence mais avec une vision politique du numérique. Et mon élu lui promeut le *big data* patrimonial. Donc on est vraiment sur une dimension éthique, souveraine, réappropriation de nos données et partage. C'est pour ces raisons de portage que tout ce qu'on décline est lié à cette vision politique de la transformation numérique.

Par rapport à ça vu que c'est revenu fréquemment, le terme d'éthique : est-ce qu'il y a derrière une conceptualisation ou est-ce que c'est juste « on va faire des pratiques éthiques, respecter une certaine transparence, une certaine intégrité » ou est-ce que c'est vraiment très conceptualisé ?

Alors pour l'éthique on se rapproche beaucoup d'*Ekitia* sur ce champs-là avec la Déclaration de Montréal. En fait, la question de l'éthique pour nous c'est la seule condition qui fera qui fera que nos habitants, nos citoyens, auront confiance. Ce n'est qu'à cette condition que l'on pourra devenir tiers de confiance. On l'a vu sur la petite opération que je décrivais tout à l'heure. L'éthique c'est le cadre qui permet de gérer le partage de la donnée. Si on ne l'a pas et si on ne la garantit pas, on ne fera rien. C'est ce qui nous différencie des modèles américains ou chinois, c'est la valeur éthique. Et c'est pour ça qu'on s'est rapproché d'Ekitia parce qu'on savait que dans ce cadre-là, se construit une éthique et une confiance autour de la donnée.

Par exemple, on commence à travailler sur les données de santé puisqu'on a un contrat local de santé, et dans le cas d'*Ekitia* on a rencontré Grégoire Mercier qui propose de faire du prédictif, et du diagnostic sur des données de santé. Il est évident que dans ce domaine très sensible de la donnée de santé, il faut être extrêmement vigilant et transparent sur l'usage de ces données. Donc pour nous, construire ce cadre de confiance, et parler d'éthique et de sécurité, etc., et de cyber sécurité c'est le prérequis pour produire des services. Et si on ne produit pas dans le cadre d'un crowdsourcing, si on n'a pas ce partage de la donnée personnelle on perdra de la donnée fine de proximité. Aujourd'hui, les plans de mobilité se font sur des enquêtes ménage déplacement d'il y a dix ans, qui coûtent plusieurs millions. On produit les infrastructures de demain donc à l'horizon 10 ans, 20 ans, 30 ans, avec des données qui ont plus de 10 ans alors que le télétravail était peu répandu et qu'il n'y avait pas eu de pandémie, qu'il n ' y avait pas ou peu de trottinette ou de vélo électrique et on construit nos services publics sur des données obsolètes. C'est pour ça qu'on a cherché d'autres types de données beaucoup plus fines, beaucoup plus subtiles qui rendent mieux compte de ce que sont les changements de comportement. Mais ces informations, on ne les a qu'à partir de la donnée personnelle. Donc si on ne capte pas cette donnée personnelle dans un cadre de confiance on n'ira pas très loin. Donc si on veut produire des services adaptés il faut garantir cette éthique et cette souveraineté de la donnée.

Ce que vous venez de dire, c'est qu'au final, la réglementation n'est pas assez coercitive ? En tout cas qu'elle n'aide pas en tant que telle à développer de nouveaux services, à développer de la confiance ? Et il faut être, je vais reprendre ce qui a été dit dans la charte de Nantes, « plus volontariste que la loi », en proposant, à travers une matérialisation concrète qui serait la charte éthique. Et l'éthique serait vraiment le catalyseur de ce que j'entends là, de justement les services, la confiance, parce que le terme de confiance revient tout le temps. Ce qui est logique, ce qui montre l'importance de tout cela. Au final l'éthique ce n'est pas juste un concept, il y a une praticité assez importante que j'avais pu déceler mais qui semble là vraiment confirmée, pour mettre en place des pratiques qui n'arrivaient pas à se mettre en place jusqu'à présent ?

Tout à fait

Notamment au terme de l'altruisme des données, c'est le terme qui est employé par la Commission Européenne, pour parler de cette volonté de donner librement ses données, ce qui n'était pas fait jusque-là, de ce que je comprends ?

Tout à fait. Alors ça peut être une vision un peu théorique mais c'est pour cette raison que j'ai donné un cadre d'application pratique pour dire qu'en effet si on veut dépasser la théorie et rentrer sur le terrain avec ces valeurs-là, c'est possible. On en a fait la preuve. Et on en a tellement fait la preuve qu'aujourd'hui à l'issue de ce *POC*, la *DDT* vient d'être lauréat d'un projet *Propulse* qui vise à créer un commun numérique qui permettrait aux collectivités de s'approprier l'application dont on parlait tout à l'heure mais enrichie de nouvelles fonctionnalités pour construire des politiques publiques. Donc on s'est dit « ça a marché chez nous, il n'y a pas de raisons que ça ne marche pas ailleurs », on peut embarquer les habitants dès lors qu'on donne un cadre « secure » pour le partage de la donnée. Ce n'est pas impossible mais ça nécessite de labourer un peu le terrain avant, Et nous on laboure beaucoup le terrain sur cette question de la donnée, y compris les plus jeunes avec la sensibilisation à la protection de la donnée personnelle. On a fait construire un jeu sur des bornes Arcade *agent data 001* pour des CM1-CM2. Un an après les élèves avait à 90% toujours les notions en tête, donc par le jeu on arrive quand même à des choses assez intéressantes. Donc on acculture vraiment des tout petits jusqu'aux plus âgés, sur toutes les dimensions de la donnée, la protection de ces données, le partage de sa donnée, pour qui, pour quoi, la dataviz, les carto partie. On espère pouvoir reprendre parce que le Covid nous a empêché d'en faire. Ce sont des supports pour pouvoir parler de la donnée, de manière concrète, pragmatique.

Ça fait partie des projets pertinents pour l'acculturation aux données ?

Oui

Le jeu et...

Oui ça c'est évident ! C'est évident que le côté ludique, le côté gamifié, le côté *challenge*, alors là c'est énorme et je pense au digital cleanup. Cela me fait penser au nettoyage de données aussi parce qu'il y a la question aussi de la sobriété, la donnée rentre aussi dans ces enjeux-là et *Open data France* a produit un guide sur la *green data*. Mais voilà, quand on parle de sobriété, on a une feuille de route sobriété numérique, on parle aussi de la sobriété de la donnée cela permet de reparler de la donnée, et du pourquoi je stocke, et pourquoi j'archive, pourquoi je m'en sépare, la donnée, la donnée utile, et pas tous types de données. Pour moi, la Loi Lemaire aujourd'hui est dépassée. Alors c'était très beau dans son concept de tout ouvrir, mais après pour des questions de chronophage, de priorité, de sobriété, on ne peut pas tout ouvrir tout le temps et plutôt ouvrons quand on sait que cela peut être utile et que derrière il y a du service qui sera proposé ou de la transparence publique, cela se réfléchit, il faut une stratégie, on ne part pas bille en tête en sortant tous les jeux de données possibles.

Il y a la réglementation et il y a l'aspect pratique, il n'y a pas forcément de mise en place, pas de moyens qui sont alloués, et c'est vrai qu'il y a actuellement encore beaucoup de collectivités, et je prends le terme de manière très généraliste, qui ont des difficultés à appliquer ne serait-ce que la loi Lemaire pour l'ouverture de leurs propres donnés publiques, ce qui est une grosse difficulté.

Bien sûr ! Mais si on ne donne pas de sens, on a encore moins envie de l'appliquer. Si on sort des jeux sans les mettre à jour et sans donner du sens à quoi cela sert. Dès qu'on a mis en place *open data*, je me suis dit « à quoi ça sert ? Comment on va motiver pour pouvoir mettre en place cet *open data* ? » et les démonstrateurs ont été une manière de prouver l'intérêt de ces données. Si on sort juste des jeux sans datavisualisations, si on ne sait pas combien de téléchargements sont faits, on ne peut rien communiquer, on perd la motivation et ils n'ont plus envie de contribuer.

Ne serait-ce que d'aller sur une site d'*open data*, le plus classique comme data.gouv, encore que maintenant il y a de la documentation, ce qui n'était pas le cas avant. Donc des gens ont commencé à mieux voir. Il y a aussi des visualisations des jeux de données. Donc il y a des choses qui sont faites, mais pour autant, ne serait-ce qu'une chose de base qui, parmi les citoyens aujourd'hui « lambdas » entre guillemets, sait qu'il existe des jeux de données ouvertes, et en plus savoir comment les utiliser.

Bien sûr !

Cela a bien été précisé tout à l'heure quand on présente aux étudiants : « vous avez un jeu de données ? ok ! Est-ce que vous savez comment le nettoyer ? Comment le faire fonctionner sur un logiciel de datavisualisation ? ». Parce que, par exemple, ne serait-ce qu'un jeu de données sur data.gouv, c'est pas dit qu'il soit bien articulé. Si vous avez des titres, des choses qui sont en-dehors du tableau, ça va faire bugguer le logiciel. Alors c'est de la technique mais on reste sur une acculturation.

Alors ce qui est bien c'est que le professeur qui travaille à Toulouse a dit à ses étudiants de faire une recherche, sans créer une seule donnée ! Ils font un master socio et data. Il leur a dit vous allez chercher sur les open data que vous voulez mais vous ne créez aucune donnée, justement pour qu'ils aillent sur ces sites d'open data, data.gouv et les autres, pour regarder, pour exploiter, pour traiter, pour datavisualiser, pour les habituer à chercher de la donnée là où elle existe et pas forcément en créer de nouvelles.

C'est toujours la même chose : l'exercice que je faisais avec mes étudiants c'est « regardez ce qui existe, donc un peu d'open data, regardez les sites qui existent, et surtout vous n'allez pas créer des données, vous allez prendre des données qui existent dessus ».

C'est cela !

Je vous montre comment on les ouvre, à quoi ça ressemble, comment on les nettoie et puis après comment on peut les représenter. Et la seule chose que je leur disais c'est « si vous voulez créer des données : vous allez prendre des rapports de l'INSEE, là où il n'y a pas un jeu de données constitué, c'est vous-même qui allez le créer ». Et alors je leur dis bon courage, ça va être long mais je vous l'autorise. Alors il n'y en a pas beaucoup qui le font. Et le but c'est de pouvoir acculturer des personnes qui n'y connaissent soit strictement rien, soit très peu de choses parce que le but c'est toujours une découverte, vraiment découverte à la dataviz, et c'est vrai qu'ils voient l'intérêt après.

Cela me parait vraiment assez évident mais c'est vrai que même des étudiants en management de projets numériques, je leur parle de data, d'*open data*, ils ouvrent de grands yeux. Je me dis « on est mal partis quand même ». Donc je pense qu'il y a vraiment besoin à tous niveaux...d'acculturer Même nos élus, la donnée si on ne donne pas de sens après avec des services, etc., ils ne visualisent pas. Au collège ils utilisent des légos, ce genre de supports pour parler de données ainsi c'est très visuel et très parlant. On comprend mieux comment agréger. Il y a un vrai vocabulaire, des techniques et des méthodes qu'il serait intéressant d'avoir parce qu'*Infolab* c'est un peu tombé en désuétude. Mais d'enrichir encore avec tous ces acteurs la culture de la donnée, créer des communs sur la culture de la donnée pour permettre à chacun de s'emparer de ces concepts et peut être du coup pousser encore plus loin dans les collectivités et pourquoi pas dans les entreprises.

[Note : la discussion quitte le cadre formel de l'entretien. L'entretien reprend suite à une discussion sur le fait que les logiques en open data influence les discours sur la culture des données.]

J'ai fait un grand état de l'art sur la culture des données, qu'est-ce qu'on entend par-là. Et notamment un passage où je dis que quand les définitions mêmes, on n'est même pas dans le sous-détail des pratiques, dans les définitions il y a « la data literacy, c'est maîtriser, gérer, etc. », des données et à l'intérieur il y a deux choses : gérer les données de manière critique, donc déjà c'est quoi l'esprit critique ? Et deuxièmement avec l'éthique. Et ça, quand on analyse les discours, on se dit que ça dépasse le simple cadre d'une culture générale pour rejoindre des impératifs un peu philosophiques.

Complètement ! Des valeurs, mais pour moi le numérique ce n'est pas technique, c'est politique ! Et ce sont les valeurs. Parce que derrière chaque outil qu'on utilise, c'est Jacques François Marchandise qui disait cela mais on embarque quelque chose, on embarque des valeurs, on embarque un modèle économique, on embarque plein de choses, et là, tout ce qu'on va mettre en place avec le département, c'est cela : la culture du numérique. Quand on utilise un outil que ce soit *Gmail*, ou un outil souverain, on ne fait pas la même chose. On ne porte pas les mêmes valeurs. Donc au moins qu'on le fasse en conscience. Moi je n'ai rien contre *Gmail* mais qu'on le fasse en conscience. Qu'on sache ce que ça veut dire. Et en fait tous les modules qu'on va créer sont là pour aider à cette prise de conscience, de ce qu'est le numérique dans ses fondements politiques, économiques, etc. C'est ça qu'on veut porter à travers ces modules qu'on va mettre en open source.

Juste parmi les dernières questions, est-ce qu'il y a des recommandations à faire sur comment un petit peu améliorer l'acculturation, toute cette culture, est-ce qu'il y a des choses à faire qui ne sont pas faites actuellement ?

Je dirais vraiment cette question de se détacher de la donnée. Parler vraiment des services, des usages, à quoi ça sert et gamifier. Je pense que c'est la porte d'entrée parce que ça fait 4/5 ans qu'on est là-dessus, et embarquer au maximum tout le monde : pas les techniciens mais tous, tous, tous. Donc je crois beaucoup à cette approche et je ne suis pas la seule/le seul parce qu'à l'époque de l'*open data* on a gamifié. Je me souviens de Loïc Hay qui en effet sur *Infolab* avait produit plein de contenus, on était vraiment sur cette idée de la gamification. Voilà je pense que c'est sans doute là et je ne sais pas qui avait créé un jeu de cartes sur l'*open data*. Un jeu de cartes qui a eu son heure de gloire, il y a peut-être 5 ans maintenant.

Peut-être [nom de personne] ?

C'est ça ! Je trouve que toute cette approche ludique est un bon moyen de rentrer dans la data.

Annexe 14 : les entretiens avec le public « universitaire »

Universitaire 1 (28/10/21)

Qu'est-ce que pour vous la *data literacy* et employez-vous ce terme ?

Oui j'emploie ce terme parce que j'enseigne à des futurs professeurs documentalistes, je travaille énormément sur la culture de l'information et sur la culture numérique. Donc la *data literacy* je l'enseigne un petit peu quand je parle de l'historique de la notion. Alors nous on utilise plutôt la notion de la culture de l'information au sens large, en expliquant pourquoi la notion de littératie est intéressante. Elle est extrêmement intéressante, elle n'est pas du tout à rejeter. Mais simplement elle est un peu compliquée à comprendre pour les étudiants ici car c'est un mot que les francophones utilisent peu. Enfin... les Canadiens l'utilisent mais pas vraiment les Français. Donc oui, on l'utilise et on utilise pas mal « culture des données » pour se faire mieux comprendre. Mais pour moi elle est tout à fait intéressante oui.

Vous avez parlé de la littératie : est-ce que pour vous c'est un concept toujours central dans la culture des données ou est-ce que c'est un terme dont le sens a été galvaudé ?

Non au contraire, je trouve ça intéressant la littératie, parce que, le concept de culture renvoie plus à des façons de penser, à des pratiques, à des systèmes de valeurs, à des imaginaires, etc. Pour moi dans la notion de culture de l'information des données, il y a la culture telle qu'elle existe en circulation chez les jeunes, chez les enseignants puisque nous on travaille, je forme les enseignants. Donc il y a l'existant, les pratiques, l'imaginaire, puis ce que l'on cultive justement donc la culture au sens élévation de l'esprit, apprentissage, etc. Donc il y a ces deux notions, pour ça c'est intéressant. Mais la littératie c'est un concept aussi intéressant parce qu'il y a cette idée de savoir lire, écrire, se débrouiller dans le monde. Il y a une dimension pragmatique que je trouve aussi intéressante, donc j'aime bien parler des deux. Et d'ailleurs on a lancé un cours, on nous a sollicité pour faire un cours dans le premier degré, pour former les enseignants du premier degré dans une option internationale et on nous a proposé de faire de la didactique comparée. Moi j'ai dit non à la didactique comparée car cela ne m'intéresse pas du tout, la didactique de l'information cela ne m'intéresse pas, mais la littératie comparée ça m'intéresse. On travaille plus justement autour de la compréhension de la notion de littératie dans ce cours.

Faites-vous une différence entre culture et littératie ?

Oui mais des nuances, je n'insiste pas trop dessus avec les étudiants pour les obliger à comprendre. Mais pour moi, parce que je vous le dis cette notion de culture elle est pour moi d'abord une notion anthropo-

logique, philosophique, etc, et ensuite une notion que l'on utilise dans les apprentissages, donc qui désigne aussi des connaissances, des compétences, etc. Tandis que la littératie est plus liée à des dimensions cognitives, d'apprentissage, de savoir-faire dans la vie.

Merci. C'est important pour comprendre la traduction française. Parfois l'on parle de littératie des données, vous préférez parler de culture des données. D'où ce point pour comprendre son usage dans la langue française.

Oui mais vraiment je trouve les deux intéressants. Je trouve que peut être dans une perspective d'enseignement et d'éducation qui est la mienne, celle de littératie est un peu plus adaptée en fait. La notion de culture est plus une notion qui va être plus désigner l'épistémologie, la recherche, mais bon nous, je sais pas, vous avez peut être vu, on a fait un carnet hypothèses qui s'appelle *data culture* parce que justement ça mélange un peu le terme data qui est plus utilisé par les anglo-saxons et culture qui est plus français. Mais c'est vraiment de la nuance, donc non je trouve les deux concepts intéressants.

Pour revenir sur la culture des données : il existe un écosystème des données. On parle de gouvernance des données, de *data science*... est-ce que vous arrivez à situer la culture des données dans cet écosystème ?

Alors là aussi il faut distinguer le public auquel on s'adresse je pense, il ne faut pas en demander trop à nos publics. Nous les enseignants, par exemple un enseignant du premier degré, c'est intéressant d'aller jusqu'à la question de la gouvernance de la donnée, des écosystèmes. Mais je pense que souvent ils ont l'impression que ça va trop loin, qu'ils n'ont pas besoin de ces connaissances-là. En revanche, pour les étudiants en documentation dont les études sont tournées autour des Sciences de l'information et de la communication, on essaye vraiment d'aborder les données sous tous les aspects. Depuis la façon dont on produit les données et justement la gouvernance, les dimensions politiques, juridiques, éthiques, économiques, jusqu'à la manipulation et les dimensions très techniques : le bidouillage des jeux de données, pas la connaissance car les nôtres sont un peu limités à part certains qui sont très très bons parce qu'ils ont fait des études en informatique, etc. Les autres cela ne va pas quand même aller très très loin dans la capacité de bidouiller dans des jeux de données. Mais quand même on leur fait produire des visualisations, des cartes, des trucs comme ça. Mais pour moi, pour ce public-là, il faut vraiment qu'ils aient conscience de l'intégralité de l'écosystème et des notions juridiques. Alors, en général, ils n'aiment pas trop, ça les ennuie. Mais pourtant pour moi c'est fondamental.

Puisque vous précisez justement que c'est fondamental, la culture des données et les données en général ont pris un nouvel essor. En quoi est-ce important aujourd'hui ?

C'est une nécessité, je vais dire des banalités mais on est dans un monde complètement gouverné par les données sauf ceux qui le refusent, mais c'est difficile d'y échapper. Là aussi dans le milieu de l'éducation, de l'enseignement, ils sont totalement fascinés par les idées de l'intelligence artificielle. Et puis ici souvent si on parle données, et même avec des experts, quand on parle données ils pensent données informatiques, algorithmes... ils pensent beaucoup *RGPD*, données personnelles... Et justement dans nos projets de la culture des données, on essaye de leur dire que c'est très important mais c'est pas tout à fait là-dessus que l'on va

travailler parce que nous c'est plutôt s'emparer des données pour faire, bon toujours banalités, un moyen d'*empowerment*, d'autonomisation, ou au moins de compréhension dans le monde dans lequel on vit, même si c'est un peu utopique de penser que l'on va faire des élèves qui vont être capable de faire des applis et qui vont être capables d'échapper au système qui les surveille, etc. Mais on se dit que, quand même, s'ils sont un peu conscients, si les élèves, ne serait-ce que si on les sensibilisait, pas forcément enseigner, mais sensibilisait à ces enjeux-là, peut être qu'un jour ils auront envie d'aller un peu plus loin. Je crois que surtout nous dans notre, sauf pour les profs de techno, sinon c'est plutôt de la sensibilisation. Et quelques petites boites à outils. Mais on ne leur demande pas d'être des spécialistes des données, mais l'idée c'est qu'ils comprennent les enjeux. Et surtout les enjeux éthiques, politiques, sociaux.

Par rapport à cela, vous avez déjà commencé à l'aborder mais la question centrale c'est : qu'est-ce qu'une donnée pour vous ?

Rapidement, c'est l'élément de base de l'information. C'est ce qui fait que cela devient une information quand on y ajoute du sens. Mais c'est aussi, au sens informatique, une donnée c'est un petit peu l'inverse, c'est une information mise sous un format qui est utilisable par une machine. Donc, voilà, y a les deux sens. Nous en Sciences de l'information c'est plutôt le premier qu'on utilise, c'est-à-dire qu'on voit le circuit de l'information à la connaissance, pardon, de la donnée à la connaissance qui passe par l'information, qui passe par le document.

En restant sur les données, y a-t-il pour vous des types de données au cœur des enjeux ?

Eh bien, il y a cette histoire de données personnelles qui est très sensible dans le milieu de l'éducation parce que les profs ont réalisé assez récemment qu'ils sont entrain de mobiliser, voire de mettre en circulation les données personnelles, les élèves sont sensibilisés à leurs propres données personnelles, avec des outils comme Pronote par exemple. Je suis presque certain(e) que maintenant il y a très peu d'enseignants qui réalisent que *Pronote* ça a longtemps était un outil totalement privé à travers lequel on héberge des données personnelles d'élèves pas forcément en France, etc. Et c'est finalement assez récemment qu'on s'est dit qu'il fallait peut-être retrouver la maîtrise française, la souveraineté sur ces données personnelles des élèves. Cette dimension-là pour les enseignants elle est extrêmement importante donc aujourd'hui je pense que c'est une des premières. Maintenant, ils ont entendu parler des données personnelles, des réseaux sociaux, de la perte de maîtrise de ces données personnelles, des choses comme ça. Et puis, l'enjeu pour moi c'est d'en faire un outil là aussi dans un milieu éducatif, d'en faire un outil d'apprentissage qui est très riche aussi, à la fois pour comprendre le monde dans lequel on vit mais aussi pour construire des connaissances. Parce que, par exemple, les profs de sciences économiques. Curieusement c'est plus les profs de sciences humaines, sciences éco, histoire-géo, qui s'intéressent, ils peuvent en faire beaucoup de choses. Cela leur donne une espèce de réservoir immense. Alors moi je trouve ça totalement passionnant : qu'ils puissent remonter jusqu'à la base de la donnée, comment elle est produite, comment on peut aller la chercher, comment on peut construire de la connaissance, comment on peut manipuler la connaissance à travers la visualisation, voir toute la chaîne pour moi c'est quelque chose de passionnant mais c'est un peu difficile d'embarquer les profs là-dedans car cela demande quand même du temps et ils en ont de moins en moins. Et les profs qui pourraient être les plus intéressés ce sont les profs de SNI, SNT, des disciplines nouvelles du lycée, et cela les intéresse très peu. Par exemple, les données ouvertes

c'est à peine mentionné dans un des programmes et ils ont une attitude très descendante avec : ils font des cours d'informatique, ils font des cours de programmation au lieu de faire des projets car je trouve que les données c'est très propice à la pédagogie de projet. Même on peut produire ces données ou on peut aller les chercher, produire une information qui correspond à ce qu'on a besoin.

Est-ce que vous avez ressenti un changement dans vos pratiques avec l'arrivée des données ? Est-ce quelque chose supplémentaire ?

Alors c'est supplémentaire si on pense éducation, si on pense à la façon dont le ministère nous traite, alors bien entendu les données sont devenues quelque chose d'essentiel. Mais si on pense éducation thématique éducative, je pense que cela soit encore, je pense pas qu'on soit allé très loin là-dedans. Et tout ce qui est éducation numérique, culture numérique, là aussi je vous parle vraiment de ma position de formatrice autour des données, je suis aussi chargé(e) de mission sur le numérique à l'*INSPE*, je vous parle du point de ce à quoi on forme les enseignants. Je pense qu'on est encore très loin de leur donner les moyens d'utiliser les données comme un élément de base de l'enseignement de l'éducation. Et après il y a un troisième élément qui est l'utopie qui fait peur sur l'intelligence artificielle. Et le fait que l'on va pouvoir récolter plein de données sur les élèves. Alors l'intelligence artificielle, plus psychologie cognitive, etc, ça donne des perspectives et des visions, qui, pour moi, font vraiment peur mais qui plaisent bien à plein de profs qui se disent « on va faire un jeu sérieux » , « on va avoir une application qui va s'adapter aux difficultés de chaque élève, ça va être super, l'élève on va manipuler et le programme va s'adapter à ses difficultés, il va apprendre à son rythme et ça va être génial ». Il y a un peu cette utopie là et donc qui repose aussi sur les données qui concernent les apprentissages des élèves.

Par rapport à l'apprentissage justement, selon vous, que pourrait-on faire pour mettre en place de façon efficace une culture des données et est-ce qu'elle devrait concerner tout le monde ?

Alors c'est une question compliquée parce que j'ai tendance à penser que l'on est pas obligé non plus... c'est comme l'histoire du mécanicien qui sait réparer un moteur : c'est important que l'on sache tous plus ou moins pourquoi notre voiture tombe en panne, mais on ne devrait pas tous être capables de réparer notre voiture. J'ai quand même le sentiment que c'est une question politique. C'est important quand même que tout le monde, pas forcément les connaissances, pas forcément les capacités à justement créer de la donnée, mais quand même la compréhension des enjeux. Et ça va jusqu'aux gens qui travaillent dans les collectivités territoriales par exemple, qui travaillent dans les services et récoltent des données dans, par exemple, la perspective de l'ouverture des données. J'ai l'impression qu'il y a eu au début du mouvement de l'ouverture des données tout un mouvement de formation des services qui ne comprenaient pas pourquoi il fallait faire autrement, pourquoi il fallait changer de façon de travailler parce que du coup ça change jusqu'aux façons de travailler, jusqu'aux outils qu'on utilise, jusqu'à l'organisation... voilà, ça a des conséquences énormes dans l'organisation d'une organisation et dire aux gens qu'il faut faire comme ça sans expliquer pourquoi il faut faire comme ça, à mon avis c'est pas terrible. Il vaut mieux que tout le monde comprenne pourquoi on fait comme ça et du coup le fasse avec intelligence. Ça, et vraiment des questions politiques qui engagent notre avenir comme la critique des utopies comme la smart city, les phénomènes de possibles dominations...

Il y a donc une part politique à prendre en compte

Oui mais cela c'est assez personnel, lié à une formation plutôt juriste, politique. Cela me paraît important. Mais c'est aussi une question de technique,

Si je vous suis bien, tout le monde devrait être acculturé à la donnée mais à un niveau technique approfondi. Il faut des gradations ?

Il faut. Et puis surtout il faut respecter les, je n'aime pas trop ce mot, les appétences de chacun. Ce qui plaît, ce qui ne plaît pas. Ce qui donne envie ou pas. Après il y a des situations où tout le monde va savoir manipuler mais je ne pense pas que ça s'impose. Par exemple, on travaille sur la robotique éducative et les algorithmes. On comprend très bien que les petits sont sensibilisés à ça et comprennent très bien la logique, mais, ce que je vous disais tout à l'heure, quand en terminale le programme est complètement tourné vers l'apprentissage de la programmation et de façon pas très intelligente en plus, je trouve que l'on laisse tout un pan de côté qui est celui plus créatif, de ce que l'on pourrait faire avec : des enjeux économiques, des enjeux politiques, y compris des imaginaires. Je pense que la donnée peut être entendue de plein de façon différentes. La sensibilisation, la compréhension, au travers d'un minimum de connaissances techniques est importante mais pas forcément la maitrise totale pour tout le monde. Je ne vois pas pourquoi… on ne demande pas à tous les élèves qui sortent du lycée d'être des grammairiens. Et donc, c'est la même chose.

Il y a justement un terme qui revient souvent, c'est celui de « critique », « d'esprit critique ». La question que je veux vous poser est : qu'est-ce que l'on peut entendre par être critique vis-à-vis des données ?

C'est être conscient de la place des données dans la construction de l'information qui vient jusqu'à nous à travers… alors si on pense esprit critique, c'est à travers les médias, notamment les médias socio-numériques. Donc c'est être capable de reconnaître quelles données sont utilisées pour construire une information, comment elles sont utilisées pour la construire ou pas, si elles sont détournées, fallacieuses, etc.

En termes de niveau général d'acculturation aux données, comment le percevez-vous ?

J'allais oublier : dans l'esprit critique, il y a la capacité à avoir du recul par rapport aux outils de visualisation. Donc la littératie visuelle est importante. Sinon, les enseignants sont très loin d'avoir une compréhension claire des enjeux. Ils ont des points de vue qui dépendent des disciplines et des didactiques. Et puis, sinon, c'est quand même des questions complexes, pour les jeunes et les très jeunes. Après, ceux qui ont plus de maturité, qui ont un peu plus d'assurance par rapport à leur métier, ça devient… ils sont plus ouverts. Mais les très jeunes qui arrivent chez nous, ils sont obsédés par des questions de didactique des maths, de didactique du français, de gérer l'hétérogénéité des élèves, etc. Ils ont un peu tendance à considérer que ça se sont des questions qui sont, soit trop complexes, soit trop éloignées de leurs pratiques quotidiennes même si, de plus en plus, on a ouvert des séminaires sur la culture numérique, sur l'esprit critique, ou sur l'innovation qui mobilisent les compétences sur les données et ils attirent de plus en plus d'étudiants. Je parle du premier degré. Après ça dépend complètement des disciplines. Je sais que les disciplines littéraires, on a du mal à les mobiliser là-dessus.

En termes justement de difficultés d'acculturation, lesquelles peut-on recenser ?

Je pense qu'il y a un manque de culture générale de la donnée. C'est encore extrêmement flou quand on en parle aux enseignants. Après, quand on rentre dans les détails, ils peuvent comprendre vite et vite être engagés. Par exemple, à travers la cartographie, je trouve qu'il y en a qui sont extraordinairement créatifs. Il y a même des profs des écoles qui font de la cartographie par *Open Street Map* et qui vont chercher des données, faire chercher des données par leurs élèves, même très jeunes. Mais globalement, souvent, ils manquent de culture de base. Et ils manquent de temps aussi. Et c'est encore assez peu présent dans les programmes. Et c'est présent dans des trucs comme *PIX*, c'est encore un outil, c'est une entrée par la norme, par la contrainte plus que par la créativité et l'ouverture d'esprit. Je pense qu'il y a un manque de culture. C'est dessus qu'on essaye de travailler. Il y a plusieurs entrées possibles. J'y pense parce qu'hier on en a parlé : on fait des cryptoparties, enfin, on a un collègue qui fait des cryptoparties. Il y a plein d'entrées possibles.

Et parmi ces entrées, lesquelles recommanderiez-vous ?

Ce n'est pas mon rôle de recommander, mon rôle c'est de donner l'ouverture la plus large possible à tous : que ceux qui sont très technocentrés ou intéressés par les mathématiques, par la logique, puissent avoir à faire. Et que ceux qui sont intéressés par les dimensions sociales, économiques aussi, donc vous avez vu que l'on a travaillé depuis un petit moment à travers ce projet data culture, on a aussi un projet spé sur la culture des données et là aussi on essaye vraiment que cela soit transversal à toutes les disciplines. Et puis on vient d'être tout juste retenus sur un *GTNUM* par la *DNE*, là aussi sur la culture des données. Mais là l'idée c'est vraiment une entrée la plus large possible, notamment je dirais par les projets. Les projets comme les *data sprint*, comme tout projet qui puisse, dont toutes les disciplines puissent s'emparer. Et la difficulté c'est d'imaginer des situations, puis des ressources et de l'information pour que même un professeur de lettres qui n'est pas forcément attiré par le numérique soit quand même capable de faire quelque chose avec ses élèves.

Vous estimez donc que c'est possible de faire de la donnée dans n'importe quelle discipline ?

Oui. Y compris dans le premier degré. Non seulement possible mais ça se fait, on l'observe.

Vous avez cité des projets, notamment ceux de l'*INSPE* : pourriez-vous m'en dire plus sur ces projets de culture des données et les approches qu'ils embarquent ?

On était parti en 2016 sur les données ouvertes, cela nous paraissait un sujet intéressant. Il était ancré dans des dimensions politiques, de gouvernance, de grands choix de société et il pouvait aller jusqu'à la manipulation par des étudiants, par des collégiens, par des lycées, par des élèves. Ce que l'on essaye de faire ici c'est de proposer...On est parti d'une idée qui s'est révélée être une impasse : de proposer des outils de médiation des données ouvertes. On est parti à l'envers : de cartographie toutes les données qui existent et on va donner envie aux enseignants de les utiliser, notamment à travers de la médiation documentaire. C'était en fait un peu impossible car la quantité de données ouvertes, et même de données, est gigantesque. Et finalement ce travail de médiation, il est fait de plus en plus, et notamment par les collectivités territoriales qui proposent des jeux de données qui sont bien faits. En 2016, ils ont fait beaucoup de progrès, ce qui veut dire qu'il y a eu aussi la formation des gens qui travaillent. Maintenant, l'idée c'est de dire qu'il y a tout cet écosystème, cet environnement à disposition : il faut imaginer maintenant ce que l'on va faire avec, en croisant plusieurs points de vue. Par

conscience économique ça les intéresse, de voir comment on peut créer une visualisation. La première entrée ça a été avec le *CLEMI*, de créer des datavisualisations : de voir comment l'on crée une visualisation à partir d'un jeu de données. Ce qui est intéressant pour l'esprit critique car si on va à l'envers, si on remonte, c'est se demander quelles données sont derrière une visualisation, comment elle a été mise en forme, quels ont été les choix esthétiques mais aussi politiques qui ont présidé à ces choix. Et puis, je pense que ce que l'on essaye aussi beaucoup de promouvoir c'est la pédagogique de projet derrière tout ça. C'est un gros potentiel de projets très ouverts donc on essaye de leur dire : il y a un minimum de connaissances à avoir sur ce que sont et comment on trouve les jeux de données, comment les utiliser et puis quels sont les enjeux par rapport aux questions d'ouverture des données, de science ouverte, etc. Puis, débrouillez-vous et laisser les élèves se débrouiller. C'est plus ouvrir sur la pédagogie de projet.

Par rapport à cette pédagogie de projet, estimez-vous qu'il y a une homogénéisation des pratiques autour des données ou est-ce assez disparate ?

Je ne pense pas qu'une organisation générale et centralisée soit souhaitable. Par rapport à 2016, par rapport à ce qui est mis à disposition c'est de mieux en mieux. Il y a plein de portails qui sont bien faits, il y a plein de collectivités qui sont militantes. On travaille avec Rennes. Dans le sud-ouest c'est pas terrible. Il y a plein de projets assez éclatés, comme *CANOPE* et le *CLEMI*. Souvent c'était quelqu'un de passionné, un documentaliste de la réunion qui fait plein de trucs. Le *GTNUM* qui vient d'être créé vise non pas à centraliser mais à fédérer pour faire en sorte que les gens sachent ce que font les autres, savoir comment l'on peut travailler avec les autres… on se forme entre nous. Mais pas forcément que l'on propose le même *MOOC* pour tout le monde, ni même le même parcours magistère pour tout le monde. L'objectif est de cartographier ce qu'il se fait et de faire en sorte que les *DANE*, les *CANOPE*, etc, soient au courant de ce que font les uns les autres et aussi sensibiliser à des enjeux qui parfois sont éloignés. Il y a une certaine hétérogénéité, par exemple dans le paysage des *DANE*, ceux qui s'occupent du numérique dans les rectorats. Il y a ceux qui sont très intéressés et d'autres absolument pas. Ils trouvent que ce n'est vraiment pas une occupation majeure. Il y a une grande hétérogénéité mais de plus en plus une grande sensibilité qui se généralise.

C'est donc plus une histoire de médiation que d'homogénéisation et de centralisation.

Oui, complètement, ça c'est sûr. Alors, avec la loi pour une République Numérique, avec le gouvernement d'ouverture des données, les normes *FAIR*, je pense qu'il y a suffisamment de normalisation au niveau central, européen et international, pour qu'ensuite ça soit plutôt ça soit la médiation qui importe. La médiation entre ce qui est à disposition et ce qu'on peut faire avec des élèves, qu'est-ce qu'il manque, c'est là-dessus que l'on travaille.

Vous avez cité la loi Lemaire, on peut parler aussi du *RGPD* : trouvez-vous que c'est une avancée pour l'acculturation aux données ou au contraire trouvez-vous que le cadre normatif est trop lourd ? Comment voyez-vous ce cadre réglementaire ?

Il y a un peu les deux. Le *RGPD* a sensibilisé tout le monde car on s'est mis à parler des données personnelles de manière plus précise, juridiquement plus précis car on en parlait de manière floue. Mais en même temps il

y a une espèce d'obsession et de confusion pour les normes. Il y a quand même des enseignants qui trouvent que l'on ne peut plus rien faire, que tout est interdit et que donc il y a un peu des deux : sensibilisation d'un côté mais excès de normativité de l'autre. Après je trouve que c'est bien la loi Lemaire, l'obligation pour les collectivités territoriales d'ouvrir les données, cela les a obligés à mettre en place des outils de gouvernance et de médiation.

En conclusion, auriez-vous des recommandations pour une meilleure appropriation de la culture des données en France ?

C'est ce que l'on essaye de faire : je n'ai pas une réponse simple. L'acculturation, la médiation, d'avoir une entrée plurielle y compris philosophique et pas seulement technique.

Le technique a-t-elle pris le dessus sur le reste selon vous ?

Dans l'esprit des gens, oui c'est certain.

Universitaire 2 (06/01/22)

Comment définissez-vous la *data literacy* et employez-vous ce terme ?

Alors, j'emploie plutôt le terme de littératie des données en français. Mais j'utilise aussi le terme de culture des données. Alors, effectivement, parfois on emploie l'un pour l'autre. On ne sait pas le quel choisir. Pour moi, la littératie des données, c'est la capacité de lire, écrire, se débrouille dans la vie de tous les jours avec des données. Faire différentes activités du quotidien, peu importe, en partant des données et en étant capable de les exploiter, de les interpréter et puis de s'en servir pour agir, pour faire des choses. Et je pense que la culture des données, je le vois comme quelque chose de plus large qui ferait partie d'une culture numérique au sens très large dans laquelle il y aurait à la fois des connaissances, des compétences, des savoirs mais aussi des pratiques qui sont à la fois personnelles mais aussi partagées avec d'autres individus. Et donc, la culture des données, cela serait tout ce qu'on est capable d'aller chercher, d'aller faire avec des données. C'est pour moi plus large, c'est une culture en termes de pratiques au quotidien. Voilà j'ai peut-être encore un peu de mal à définir.

Vous différenciez bien la littératie des données et la culture des données, ce qui m'interroge sur la « nature » de la *data literacy*. Que mettez-vous derrière le terme de culture ?

C'est à la fois une manière d'être, une manière de vivre, un ensemble de croyances. Des opinons qu'on défend. Une culture c'est ce qui nous rassemble avec d'autres, ce qui fait qu'on est en lien avec d'autres personnes, qu'on partage des intérêts communs, qu'on fait des choses en commun... c'est un peu tout ça la culture pour moi.

On parle beaucoup de la culture des données en ce moment et il existe tout un écosystème de données à l'intérieur duquel se trouvent la data gouvernance, la data science, etc. Où situerez-vous la *data literacy* dans cet écosystème ?

Je ne sais pas trop. Les *data sciences*, c'est très spécifique : ce sont des compétences déjà expertes, spécifiques. Alors que la culture des données, la littératie des données, ce sont des compétences, des savoir-faire qui pourraient correspondre à tous types de citoyens. Pas forcément des experts de la donnée au sens de *data scientists*, etc. Ni au sens de gouvernance des données. Peut-être qu'avoir une culture des données ce n'est pas forcément maîtriser, connaître tous les enjeux de la gouvernance des données. Cela implique vraiment de s'intéresser à ce sujet-là, de voir tout ce que cela implique, mais quand même il y a l'idée de connaître, de maîtriser et de s'intéresser au circuit de la donnée. C'est-à-dire d'où vient cette donnée, ce que l'on va faire avec, comment on fait avec ces données, comment on les traite, comment on les analyse, comment on les représente pour lieu les comprendre justement. C'est un peu tout ça. Cela ne se superpose pas mais en tout cas ce sont des choses différentes avec ces aspects de gouvernance, science et culture.

Cette nécessité globale d'acculturation aux données, quelle est son importance aujourd'hui ?

Quand on a commencé à travailler sur ce thème-là avec [nom de personne] en particulier, c'était un peu une recherche engagée au sens, elle me rejoint là-dessus, on pense toutes les deux que la culture des données a du sens pour les futurs citoyens, les élèves qui vont devenir les citoyens de demain, qui le sont déjà d'ailleurs, qui sont confrontés à des masses de données, à beaucoup de représentation issues de données : les visualisations que l'on voit dans les médias, les infographies, tout ça. Ils sont déjà énormément confrontés à ça. Ils sont aussi producteurs de données de par leurs activités quotidiennes avec des outils numériques, etc. Donc je pense que toute cette culture des données-là n'est pas conscientisée, je ne sais pas comment l'exprimer : peut-être que les jeunes aujourd'hui n'ont pas vraiment conscience que la donnée est au centre de toutes leurs activités, en termes de lecture, de compréhension du monde, etc, et du coup je pense que c'est un réel de les rendre plus conscients de ça pour qu'ils soient en capacité d'agir avec, de les maîtriser, de pouvoir les critiquer quand on voit des datavisualisations dans les médias ils sont capables de les interpréter et peut être même de refaire la visualisation pour voir si l'on trouve la même représentation des choses parce que c'est très complexe la représentation des données car avec des outils l'on voit que l'on peut représenter les données de manière faussée. Même sans s'en rendre compte on peut produire des informations un peu erronées, pas totalement justes, du coup il y a toute cette compréhension de la production des données, des producteurs de données qu'ils sont, qu'ils soient à même de comprendre qu'ils sont des producteurs de données, de voir éventuellement ce qu'ils peuvent en faire : il y a de plus en plus d'initiatives où l'on cherche à rendre les producteurs de données, à leur donner accès à leurs propres données pour pouvoir les analyser, voir sa consommation, voir ce que l'on produit avec ça pour pouvoir s'en servir. Il y a cet aspect-là et il y a aussi cet aspect d'être critique par rapport aux exploitations des données que l'on va retrouver dans les différents médias, donc plus le côté éducation aux médias et à l'information.

Et puis il y a aussi tous les enjeux autour de l'*Open data*. On en parle beaucoup, la France est assez exemplaire de ce point de vue-là, elle s'est vraiment faite une place au niveau européen en tant que producteur et diffuseur de données ouvertes. Bon il y a encore beaucoup de choses à faire car la qualité des données n'est pas encore

totalement parfaite : il y a encore beaucoup de jeux de données qui sont imparfaits. Cela pose encore beaucoup de problèmes en termes de réutilisations que l'on peut en faire. Mais toutes ces données-là qui sont mises en accès libre, les futurs citoyens doivent pouvoir s'en servir. Qu'est-ce que l'on fait de ces données-là, comment on est capable d'aller les chercher, de les interroger, de se poser des questions avec. Et puis, peut-être, de répondre à des questions de société avec ça, des questions de recherche peut-être, et puis produire des choses à l'initiative de nouveaux services comme l'application pour chercher un centre de vaccination, vitemadose, donc lui il s'appuie sur des données ouvertes et fraîches pour recevoir un service gratuit mis à la disposition de tout le monde, c'est assez emblématique : quand on en parle à des collégiens et des lycéens, ils se disent « chouette, je peux être moi producteur d'un service comme ça innovant et partager ça à tous les citoyens, à tous les français ». Ça peut leur donner des ambitions, des intérêts pour plus tard.

Si on devait dater ce moment où l'acculturation des données est devenu une véritable nécessité, à quelle période la situerez-vous ?

C'est pas évident en fait. Il y a peut-être eu un phénomène de mode aussi. La culture des données on peut se dire qu'elle était déjà centrale, peut-être pas avant le web 2.0 mais les profs en sciences ou en sciences économiques, pour eux la donnée était déjà au cœur de leurs activités. Sensibiliser les élèves aux données, c'est les sensibiliser à des démarches d'investigation, une démarche un peu critique et de savoir rendre compte de résultats à partir de preuves, de faits concrets, des choses irréfutables… je pense qu'il y a eu quelque chose qui s'est développé de façon exponentielle autour des données, c'est… quand on a commencé à parler du *Big Data*, quand on est confronté à des masses de données. Donc c'est là que le phénomène s'est vraiment amplifié et qu'on s'est dit qu'il faut parler d'une culture des données, d'une vraie littératie des données pour que les jeunes d'aujourd'hui et puis les citoyens soient capables d'avoir un regard là-dessus, d'avoir des réflexions sur l'éthique autour de ça. C'est un enjeu fort, l'éthique. Et puis d'être capable d'être critique vis-à-vis des algorithmes, des plateformes, des *GAFAM*, etc. Donc on peut l'associer au *Big data*, aux plateformes web 2.0, au phénomène *Facebook*, etc.

Quelle est pour vous la définition de la donnée ?

C'est tellement large et en même temps… la donnée c'est la première unité, l'unité de base, ça représente, je pense dans l'imaginaire des gens quand on parle donnée c'est surtout chiffre. Mais effectivement, on voit qu'il y a beaucoup de confusion entre des données et les statistiques, ce n'est pas la même chose. La donnée brute et les statistiques, on fait beaucoup l'amalgame. Et puis derrière donnée, cela peut être des mots, des images, cela peut être plein de choses en fait. C'est assez compliqué et c'est une des grandes questions que l'on s'est posée : comment interroger les élèves autour de ça ? Parce que parler de donnée recouvre tellement de choses et c'est tellement flou, même pour les adultes, que c'est difficile de parler de quelque chose de précis et que c'est difficile de s'entendre sur le terme que l'on emploie.

Vous avez évoqué l'*open data*, le *big data*… il y a-t-il un ou des types de donnée qui est plus important que d'autre, qui plus au cœur des enjeux que d'autres ?

J'aurais envie de dire les données ouvertes justement car je pense que c'est quelque chose auxquelles les personnes ne s'intéressent pas forcément, ne connaissent pas forcément, c'est peut-être pas forcément assez dé-

mocratisé, cette idée de l'open data et des données ouvertes. Et pourtant, ça recouvre un enjeu de citoyenneté, un enjeu démocratique, un enjeu de transparence aussi de l'information et pour moi c'est vraiment un type de données qui mériterait d'être davantage connu des élèves, des jeunes notamment, puis du grand public. Il y a encore un besoin de démocratisation autour de l'open data : ce n'est pas bien compris et même par les diffuseurs et même les producteurs de données.

Est-ce que l'arrivée des données a modifié, impacté, vos pratiques professionnelles ?

Du point de vue professionnel et pas personnel du coup ?

Les deux sont possibles.

Pas tellement justement du côté personnel. Mais c'est vrai que du côté professionnel, cela m'a amené à faire de nouveaux cours même si c'était déjà abordé dans certains des cours mais de façon différente autour des traces par exemple, autour de l'identité numérique, j'avais déjà des cours là-dessus. Mais là, c'est vrai que cette année j'ai fait un cours sur justement l'exploitation en représentant ce qu'était l'*open data* aux étudiants, les étudiants sont de futurs professeurs documentalistes en M2, et partant de là je leur expliquais les enjeux de l'*open data*, ce que c'était, les types de jeux de données que l'on pouvait trouver, savoir comment les exploiter, etc. Et partant de là je leur ai demandé d'imaginer, en sélectionnant un jeu de données, un scénario pédagogique à leurs élèves, de faire travailler leurs élèves avec justement un jeu de données ouvert. Donc en partant d'un jeu de données brutes mis à disposition sur la plateforme *datagouv* et d'autres plateformes de ce type, et cela a vraiment donné des résultats intéressants. Ils ont travaillé par groupes et j'ai eu un groupe, par exemple, qui a travaillé sur, qui a proposé une activité autour des jeux de données sur les rues dans une commune, d'aller chercher justement dans des communes des noms de personnes célèbres et de voir la représentation genrée des noms des rues. Ça a un nom, l'odonymie je crois, le fait de regarder le nom des rues, et donc là elles, c'est un groupe de filles, ont proposé de travailler là-dessus, sur cette représentation genrée, en lien avec le territoire proche, avec les communes des élèves, et se questionner sur justement pourquoi il ya autant de noms d'hommes et pas de nom de femmes dans des places, des rues, etc. Et puis de produire une visualisation pour se rendre compte de ça. Voilà, c'était un exemple, je n'ai plus les autres en tête mais que ça avait plutôt bien marché. L'autre, je crois que c'était sur de la cartographie pour représenter des lieux de culture ou quelque chose comme ça. Ils ont plutôt bien compris les enjeux je trouve. Et cela leur a vraiment paru être un enjeu à développer, à travailler avec les élèves. Au second semestre, comme ce sont des stagiaires et qu'ils ont des classes en responsabilité, ils ont tenté de déployer ces scenarii là dans leurs classes, enfin, avec leurs élèves.

Vous avez évoqué tout à l'heure le fait que la France avait engagé beaucoup d'actions notamment en ce qui concerne les données ouvertes. La question que je souhaiterai vous poser est : quelle est l'influence du cadre réglementaire comme la loi Lemaire ou le *RGPD* ? Est-ce un facilitateur de pratiques ou, au contraire, est-il plutôt restrictif ?

Alors c'est totalement subjectif ma vision des choses mais je me dis que le *RGPD* contribue à sensibiliser les gens sur l'importance des données, qu'elles soient personnelles, sensibles, ou publiques justement. Pour moi le *RPGD* c'est assez favorable justement au développement d'une culture des données dans le sens que cela

permet d'encadrer mais aussi de faire de comprendre aux personnes qu'ils livrent beaucoup de données personnelles et qu'il faut pouvoir les protéger. Donc je pense que c'est assez bénéfique en fait cet aspect législatif d'encadrement autour des données.

Pour vous, en pensant moyens, personnels, etc, comment peut-on mettre en place une culture des données ? Et est-ce que cela doit concerner tout le monde ?

Alors c'est très large. Comme organisation déjà, je vais me centrer sur l'école. La question serait comment on met en place une culture des données auprès des élèves à l'école et comment on peut développer ça. Je pense, il y a l'école et les partenaires extérieurs, parce que les médiateurs, les associations, l'éducation populaire sont très actifs là-dessus. C'est aussi par les militants du web, de l'open source, des militants autour de toutes ces valeurs-là autour du web que cela peut aussi passer donc je pense que c'est déjà par une communauté d'acteurs assez engagés. Que ce soit les enseignants, les cadres de l'éducation qui participent aussi justement à favoriser la mise en place de ce type d'action là, d'éducation autour des données. Et puis je pense aussi au monde associatif, à tous les militants qui mettent en place des journées d'action, de médiation de la donnée avec des jeunes, donc je pense que c'est en multipliant les actions de formation, de médiation, qu'on peut contribuer à ça : diffuser une culture des données. Je pense que ça passe par-là.

Vous l'avez également évoqué tout à l'heure : au niveau des compétences que les personnes doivent développer à l'issue de cette apprentissage, de cette acculturation de la donnée, vous avez évoqué l'esprit critique mais également la connaissance du circuit de la donnée. Concernant les compétences techniques vous avez évoqué les représentations. Peut-on ajouter d'autres compétences ?

Je pense que ça regroupe beaucoup de compétences, des compétences très techniques : savoir manipuler des fichiers *Excel*, savoir se repérer dans un fichier *Excel* où il y a des masses de colonnes contenant beaucoup de données ce n'est pas évident. C'est aussi comprendre, connaître, tout ce qui est métadonnées, formats de fichiers, et puis, donc ça c'est pour les compétences très techniques, je pense qu'il y a des compétences en termes de design de l'information : comment on représente, comment on arrive à choisir un type de visualisation de données, et à faire en sorte que cela réponde à une question que l'on se pose.

Et sans dénaturer l'information et les données, sans faire dire aux données ce qu'elles ne veulent pas dire. C'est quelque chose de pas évident : j'avais fait un test, une petite expérience, une fois, avec des étudiants CPE. Je leur avais fait prendre le logiciel *Excel* et le logiciel *Piktochart* pour représenter les mêmes données avec un histogramme graphique identique. Et en travaillant les données sur Excel, ils ont produit un histogramme et en travaillant sur Piktochart, qui est un outil qui simplifie, mâche un peu le travail et qui donne tout de suite les visualisations de données, on s'est aperçu qu'on n'avait pas la même représentation de données. Pourquoi ? Parce que sur *Piktochart*, les données ne partaient, je ne sais pas comment l'exprimer, elles ne partaient pas de la base. Les données commençaient à zéro en abscisses et donc la représentation était totalement faussée. On avait l'impression que l'écart entre les données était plus important ou ramassé, je ne sais plus, l'un ou l'autre, mais du coup il y avait une distorsion dans la représentation des données. Et cela, c'était lié aux logiciels qu'on utilisait. Si on n'est pas capable de vérifier ça, de se dire « mais là, est-ce que ça représente les données réelles que j'ai utilisées ? », eh bien cela nécessite des compétences au niveau d'une littératie visuelle. Parce que les

données que j'ai insérées dans ce graphique-là correspondent bien à la réalité. Donc ça, et puis qu'est-ce qu'il y aurait d'autre comme compétences… des compétences techniques, des compétences visuelles et puis après c'est plutôt de l'ordre des connaissances autour. Alors ça sort peut-être de la question.

Non pas du tout, il s'agit vraiment de savoir ce que l'on met à l'intérieur de la culture des données. J'ai parlé de compétences mais cela peut être plus large.

Après c'est un ensemble de connaissances propres au numérique, à la technique et puis aussi au cycle de vie de la donnée, ça je pense que c'est important. Et puis aux enjeux éthiques, économiques autour des données. Aux enjeux politiques aussi. Je ne sais pas si ça répond à la question.

Si totalement, l'objectif c'était de savoir si on intégrait que des compétences techniques ou si on y ajoutait l'esprit critique, des connaissances culturelles…

Oui, ça comprend tout ça. Des compétences techniques mais aussi des connaissances culturelles des enjeux, etc.

Par rapport à ces formations, arrivez-vous à voir le niveau général des personnes par rapport à la culture des données : sommes-nous sur un optimum ou faut-il renforcer les formations actuelles pour atteindre cette maîtrise des données ?

Oui, quand on regarde les programmes, c'est vrai que la donnée est parfois en filigrane de certains programmes scolaires mais ça reste un objet qui n'est pas abordé en tant qu'objet culturel. C'est souvent un moyen, cela peut être une ressource pédagogique parfois. En sciences économiques on va utiliser des données de l'*INSEE* parfois. Mais finalement les données, le *big data*, c'est très peu abordé : il n'y a que dans des spécialités au lycée que l'on traite de ça. L'open data c'est pareil, il n'y a que dans les spécialités *SNI*, sciences numériques, que cela c'est abordé. Donc il y a une nécessité de former les enseignants là-dessus surtout sur ce qui est enjeu autour des données qui est encore peu maîtrisé. L'éthique *by design*, tout ce dont on parle en ce moment, ça c'est quand même des connaissances qui sont propres aux professeurs documentalistes qui sont formés à ça, à l'information. Peut-être des profs de technologie parfois, ou des profs qui ont une culture numérique assez développée et qui s'intéressent à ça aussi dans leurs pratiques personnelles mais je pense qu'il y a vraiment nécessité de renforcer la formation des enseignants autour de tous les enjeux autour des données. Et effectivement des compétences qu'elles requièrent.

La culture des données serait donc pour tous les enseignants mais est-ce que tous les enseignants devraient enseigner la donnée ou est-ce qu'il s'agirait de disciplines spécifiques ? Vous avez cité l'exemple des professeurs documentalistes.

Non, je pense que c'est quelque chose de transversal comme la culture du numérique. Chaque discipline a un regard intéressant qui peut être porté sur les données et je pense que la formation de tous les enseignants aux données cela pourrait passer par leur faire connaitre justement toute l'offre en *open data* car c'est quand même des masses de ressources pédagogiques potentielles qu'on peut avoir sur tous sujets. Le professeur de

SVT peut être intéressé par un jeu de données en lien avec la science de la nature, le prof de physique pourrait être intéressé d'un autre point de vue, etc, en histoire géo, en lettres... je pense que tous les enseignants des disciplines auraient intérêt à connaître les données, à mieux maîtriser les données, à mieux comprendre les enjeux autour des données, et puis à connaître l'offre en open data pour voir ce qu'ils peuvent en faire du point de vue de leur discipline. Donc pour moi c'est quelque chose de transversal : c'est pas propre à une discipline ou une autre. Mais je pense que chaque enseignant d'une discipline pourrait apporter un regard particulier là-dessus qui permettrait d'enseigner cette culture numérique globale aux élèves.

Par rapport à cette nécessité de formation, quelles sont les difficultés de sa mise en place ?

Je pense qu'il y a plusieurs difficultés. Déjà une volonté ministérielle pas encore vraiment affichée, il n'y a pas de pulsion ministérielle sur cette éducation à la donnée donc c'est un premier point de blocage. En tout cas, un point qui ne va pas favoriser tout ça. Ensuite, il y aurait, alors ça c'était une difficulté par le haut, une difficulté par le bas aussi des enseignants qui se n'intéressent pas à ça en fait, qui trouve cet objet-là trop technique, qui ont peut-être une représentation des données comme quelque chose de technique qui correspond pas à leur domaine disciplinaire. Et on n'en parle peut-être pas, on en parle peut-être par méconnaissance mais aussi par manque d'intérêt pour ça. Après la mise en œuvre, la formation pour maîtriser vraiment l'exploitation de jeux de données ouverts par exemple, ce n'est pas rien donc il faudrait vraiment que les enseignants soient formés à ça. Et ça peut nécessiter du temps de formation important. C'est pas en une séance qu'on arrive à maîtriser ça. Il y a beaucoup de difficultés techniques quand on manipule des jeux de données, des difficultés cognitives que l'on peut rencontrer. Il y a des compétences que je n'ai pas évoquées qui sont des compétences statistiques qui sont essentielles et que tous les enseignants n'ont pas. Donc ça, ça peut être un frein important.

Je rebondis sur ce que vous venez de dire : vous parlez de compétences statistiques, malgré le fait que les données ne sont pas simplement des suites chiffrées, on reste encore sur une vision mathématique, statistique, qu'il faut toujours maîtriser pour comprendre les données ?

Oui parce qu'autant les profs de lettres, d'art pourraient être super intéressés pour travailler sur des jeux de données en art, en littérature dans ce que l'on appelle les humanités numériques, cela me parait vraiment une opportunité intéressante. Mais en même temps je pense qu'il y a ce blocage-là de se dire « oui mais moi, je ne maîtrise pas la statistique », il va falloir manipuler des masses de données en passant par des outils très techniques qui peut vraiment rebuter. Cela peut être un frein à ça.

Avez-vous en tête un projet de culture des données et si oui pourquoi ?

Je vais parler du projet de recherche : [passage retiré à la demande de la personne interrogée]. L'un des enjeux de ce travail-là qui va durer trois ans, ça va être de fédérer une communauté d'acteurs qui travaillent là-dessus et qui sont actifs pour diffuser la culture des données auprès des élèves. Cela va être aussi de faire des revues de littérature, des notes de veille sur, justement, ce que sont les données, ce qu'est l'*open data*, sur ce qu'est la donnée ou la littératie des données. Donc faire mieux connaître en fait toutes ces compétences et connaissances-là autour des données aux enseignants. Et puis, il y a un axe dans ce projet-là qui concerne plus les actions de médiation et comment justement on fait de la médiation des données auprès des élèves : quels scenarii pédago-

giques, avec quelles ressources, est-ce que l'on part d'un jeu de données brut que l'on va trouver directement sur datagouv ou bien est-ce qu'on toilette le jeu de données pour le rendre plus digeste et plus facilement exploitable par des enseignants et par des élèves. Donc il y a tout cet aspect de recenser toutes les actions qui sont menées en termes d'éducation et de médiation autour des données et de voir ce qui fonctionne bien, ce qui est opérant avec des élèves. Et puis de voir quels sont les objectifs pédagogiques que les acteurs mettent derrière ces différents scénarii parce que, voilà, on a beaucoup parlé des enjeux autour. Il y en a peut-être qui ne vont pas parler de ça mais qui vont plus passer par des compétences techniques opérationnelles pour rendre les élèves compétents avec des jeux de données mais pas forcément leur parler de culture.

Auriez-vous des recommandations ou des remarques pour améliorer l'acculturation aux données en France ?

Pas facile comme question. Ce n'est pas une recommandation mais je crois que c'est une question qui nous avait été posée au colloque et qui me parût assez pertinente. Au colloque auquel vous participiez, c'était d'essayer de regarder dans quelques années et de se dire « est-ce qu'il ne va pas y avoir des nouvelles technologies, des nouveaux moyens, qui vont nous permettre de rendre plus ergonomique, plus visible tout ce traitement de la donnée ? ». Alors, est-ce que cela ne serait pas une piste à explorer comme recommandation. Mais cela serait une remarque. L'objectif de l'*open data* notamment c'est de toucher le citoyen lambda, celui qui n'a pas eu de forcément spécifique en traitement de données, en statistiques, etc. Du coup, ça ne fonctionne pas aujourd'hui car ça reste très difficile d'accès si on a pas des compétences, des connaissances de base, pour traiter ce type de fichiers-là, de les comprendre et de les représenter. Donc, est-ce qu'il n' y a pas d'autres acteurs qui vont venir s'imbriquer dans la chaine pour faire en sorte qu'il y ait des outils qui nous permettent un traitement ou plus de facilité en tout cas, ou des représentations des fichiers de données plus faciles à exploiter. C'est une question très ouverte.

Universitaire 3 (01/04/22)

Comment définissez-vous la *data literacy* et employez-vous ce terme ?

Alors je n'emploie pas le terme de *data literacy*. Enfin, je l'emploie à certains moments mais ce n'est pas le terme que j'utilise de façon majoritaire. Quand j'ai commencé à travailler sur ces sujets-là, sur le sujet des données, j'ai été contactée par la *Fondation Internet Nouvelle Génération* par Amandine Brugière qui montait un projet. A ce moment-là, ils appelaient ça une campagne sur les données et, elle, à ce moment-là parle de culture de la donnée. Et en fait, lors de notre premier échange, justement, je lui pose la question : pourquoi culture ? Pourquoi pas littératie ? Et en fait pour elle c'était vraiment l'idée de prendre le terme de culture pour travailler, on va dire, à un partage de savoirs plus largement. Il y avait vraiment un ancrage politique dans sa manière d'aborder les données. Et donc d'où le fait qu'elle ait fait le choix, alors avec d'autres personnes, mais d'appeler ça la culture à la donnée. Et donc on commence à travailler ensemble, on monte un atelier. Alors je ne me souviens plus du titre exact, c'était au 104 pendant un an au *Futur en Seine* 2012 je pense, et on appelle ça

« faut-il une culture de la donnée ? ». On invite Henri Verdier, d'une part, et quelqu'un de l'*INRIA* dont je vais retrouver le nom qui a fait le livre data terae de la cité des sciences.

Et en fait ça ce moment-là, à l'issue, durant l'atelier il y a cette idée de dire « culture de la donnée cela ne va pas : on ne dit pas la donnée. Il y a tellement de types de données différentes, cela essentialise... ». Et à la fin de cela, on passe à culture des données. Avec cette idée de pouvoir prendre en considération plus largement ce champ qui est à l'époque pas nouveau mais qui en tout cas prend une importance nouvelle à la fois dans les politiques publiques et aussi dans les interrogations en lien avec la place de l'informatique à l'école ; l'apprentissage du code notamment. Très rapidement on va parler de culture des données et ça, ça va rester. Pendant le programme *Infolab*, on gardera toujours cette idée de culture des données et en fait au moment où on va vouloir justement discuter à l'international de ce que nous on avait fait au travers des différentes campagnes, on va parler de *data literacy*. Et c'est parce qu'on passe à l'international et donc qu'on va créer un évènement qui s'appelle la *Data literacy conference*.

Et là on va faire ce glissement-là parce que, à l'international, c'est le terme qui est prédominant. Et eux ne parlent pas du tout « d'éducation à » ou de « culture de » : ce n'est pas dans leur vocabulaire. Et après je me suis beaucoup posé(e) la question de la différence de culture et littératie. J'ai beaucoup lu Jack Goody et la question de la littératie. Dans la définition que peut en donner Jack Goody qui serait : les moyens, les ressources qu'on a, qu'on se donne dans une société pour accéder à une culture, à des savoirs, etc. Des moyens pour aller à la rencontre de ce qui se produit dans une société. Eh bien il y a une dimension un peu à la fois didactique. Les formes d'apprentissage que nous on n'avait pas en tête quand on disait culture des données. Vous voyez, on n'était pas sur un modèle éducatif. Et derrière le terme de littératie, je pense aussi qu'il y a un modèle : pas un modèle mais une dimension politique qui serait le fait que la littératie repose sur les pratiques d'écriture et de lecture, sur des formes instituées d'accès aux savoirs et donc une organisation, une structuration de la société autour, avec les scribes au début. Je connaissais bien Goody, je connaissais bien ces éléments-là de la définition de Goody, mais aussi du fait que j'étais dans une position de chercheur(se), peut être, sur ce projet, et que j'étais, comment dire, à la fois dans un dialogue mais que je ne voulais pas introduire les mots qui m'intéressaient seulement moi. Le mot que j'ai apporté si vous voulez c'est le terme de médiation. Quand on fait de la culture des données, l'enjeu c'est vraiment de mettre en œuvre des actions de médiation. Et c'est là où, en fait, j'ai vraiment travaillé avec eux. Voilà : pas littératie de données, culture des données au départ avec le petit *switch* que je vous ai raconté et puis après on a parlé de *data literacy* mais avec cette ambition que notre travail soit identifié à l'international et qu'on puisse dialoguer avec ceux qui faisaient l'actualité on va dire de ce champ de recherche et ce champ de militantisme.

Il existe tout un écosystème de données : on parle de data science, de gouvernance des données. Où peut-on situer la *data literacy* dans cet écosystème ?

Alors je pense très clairement que la culture des données ça doit se décorréler de la *data science*. Alors ça pour le coup, j'en suis convaincu et à plusieurs titres. D'abord parce qu'en fait ça serait une prédominance de la pensée informatique sur la culture des données. Or, je suis convaincu et ça c'est aussi ma discipline qui fait ça, que si on ne pense pas la place des données en fonction des savoirs qu'on a constitués en Sciences de l'information et de la communication, on risque une nouvelle fois d'avoir une prédominance de la pensée logistique, de la

dimension aussi « résolution de problèmes » qui est sous-jacente à cette pensée informatique la plupart du temps. Et donc évidemment, disons que la *data science* c'est un aboutissement disciplinaire et la culture des données c'est quelque chose qui devrait être partagé par tout le monde. Mais ça ne veut pas dire que c'est séparé, dans le sens où il n'y a pas de liens : il y a des liens. Mais si on pense la culture des données en fonction de la *data science*, on va faire des erreurs en termes des processus d'apprentissage ou des processus de médiation que l'on pourrait mettre en œuvre. Après sur les autres aspects on va dire, pour moi comme je disais, dans l'ancrage de l'info-com, pour moi il y a un enjeu fort. On pense l'information, on sait ce qu'est l'information (on a quand même beaucoup travaillé là-dessus), c'est à l'origine de notre discipline, en tout cas en France. Escarpit, c'est un des fondateurs de la discipline. Il a posé des bases sur cette notion de l'information. Et il y a eu énormément de travaux qui ont été fait là-dessus, donc à nous de rappeler ces éléments-là. Parce que si on pense « donnée », il faut rappeler son lien à la notion de l'information.

Et puis les médias : aujourd'hui, on ne peut pas avoir une culture des données qui ne s'appuierait pas sur une éducation aux médias et à l'information. Je placerais la culture des données plutôt comme étant un ensemble de ressources à identifier au sein de l'EMI plutôt qu'au sein de la *data science*. Quand bien même, évidemment, à un certain stade, il y a des questions : je pense que c'est super important de comprendre l'algorithme, je pense que c'est important de comprendre le mode de production des données, le plan de traitement des données, tout ça on est bien d'accord là-dessus. Ca ne néglige pas ça mais ça veut dire qu'on place dans un ensemble un tout petit peu plus complexe et surtout qu'on ne laisse pas la place à l'informatique ou à la stat. Toute la place dans cette culture des données.

Si on devait décrire la « nature » de la *data literacy*, ses caractéristiques : parlerait-on d'une culture, d'un état d'esprit, d'un ensemble de pratiques et de techniques ? Tout ça à la fois ?

Alors nous on avait un slogan : on a beaucoup travaillé sur le fait qu'on était tous des monsieur et madame Jourdain de la donnée. C'est un peu basique mais en vrai, avec cette idée que toute personne est aujourd'hui, même si les gens ne parlent pas de donnée forcément, tout le monde a bien conscience qu'on laisse des traces, que toute activité en ligne est suivie et conservée, est l'objet de traitements derrière… c'est diffus ! Mais par contre c'est quelque chose qui est présent, Et puis la culture des données, ce n'est pas simplement des tableaux *Excel*, ce ne sont pas des *CSV*. C'est plein d'éléments qu'on manipule nous-mêmes. On produit de la donnée concrètement pour soi. C'est aussi un exemple qu'on utilisait souvent : tous ceux qui font du *quantified self*, ils ont un rapport. Et là on sait, on a des gens autour de nous qui disent « moi je fais mes 12000 pas par jour », j'ai ma petite montre connectée qui me dit ça, ça veut dire qu'on a un rapport à la donnée qui est fréquent, qui est mon quotidien, pas forcément conscient mais en tout cas c'est là.

Donc nous en fait, l'objectif c'est vraiment de se faire on peut s'appuyer sur ça, le formaliser, faire prendre conscience de ça pour ensuite le transformer et que ça devienne, si je le dis un peu d'une manière qu'il y avait une dimension émancipatrice en fait dans ce qu'on a cherché à faire, surtout en plus, au départ, avec Amandine Brugière, de faire un Internet citoyen. Il y avait quand même derrière son action, une action tournée vers la place du citoyen dans la cité. Et donc, de l'ouverture des données, des mouvements militants. Après évidemment, il y a eu un glissement quand Amandine est partie et que c'est Charles qui a repris et que là on a vraiment beaucoup plus travaillé avec des organisations privées et que l'objectif était aussi de faire du salarié quelqu'un

qui va être plus à même de comprendre les politiques ou la stratégie data de son entreprise et d'y participer. Il y avait aussi un changement dans ces deux logiques. Je ne sais pas si je réponds à votre question mais en tout cas ce ne sont pas que des pratiques techniques, ce n'est pas forcément un état d'esprit mais c'est un état de fait. Nous produisons, nous sommes entourés de données. Mais après, nous on travaillait à une prise de conscience de cette culture des données : c'était vraiment cet enjeu de faire prendre conscience à des sujets sociaux que les données c'était un élément essentiel aujourd'hui dans leurs activités, qu'elles soient personnelles, politiques. Aux salariés je ne l'ai pas bien dit mais en tout cas dans leurs pratiques professionnelles.

Si on devait comparer aux autres nécessitées apparues au fil du temps comme la technicité informatique, où se situe la *data literacy* : est-ce central ou juste une compétence parmi tant d'autres ?

Alors j'aurais du mal à dire que c'est central quand même. Néanmoins, cela ne peut pas être négligé. C'est ça que j'aurais tendance à dire, que ça fait partie des enjeux contemporains de la formation dès le plus jeune âge. Mais sans oublier qu'il faut former tous ceux qui sont en train de vivre dans notre société. Donc oui, ce n'est pas central mais c'est nécessaire. Je ne dirais pas que ça doit être un passage, ce n'est pas le passage premier, mais ça doit être un des passages. Je pense que pour moi, il faut d'abord apprendre à écrire. Et dans cette pratique d'écriture, on va introduire des questions qui sont liées aux données. Parce que les données ça s'écrit, ça se produit d'une certaine manière. Et donc c'est pour ça que je le mettrais à l'intérieur d'un ensemble. Après ça fait partie des enjeux contemporains parce que, comme vous le disiez, aujourd'hui on voit bien que la place de la technique dans notre rapport au savoir est déterminante. Parce que notre rapport au savoir est médié par la technique. Elle est médiée, on peut dire, par les dispositifs informatisés. En tout cas, il est médié par ça. Donc ça serait un premier point. Ensuite, les manières d'apprendre, les manières de comprendre, s'appuient aussi sur tous ces dispositifs qui permettent de les recueillir, de les transformer et de les traiter, et de les analyser.

Et enfin, j'ai envie de dire que le troisième enjeu, c'est aussi toutes les transformations des métiers qui vont, sont en train d'advenir et que de ce fait là, alors même que tous les dispositifs qu'on manipule dès le plus jeune âge effacent leur technicité, eh bien dans les métiers on va avoir une montée en puissance de la technicité. Et ça c'est un écart qui est troublant et qui d'une certaine manière posent problème clairement sur est-ce que les jeunes manipulent. Vous savez bien qu'en fait qu'il y a à la fois une conscience et une sorte de difficulté à se représenter exactement le type de dispositif auquel ils ont affaire. J'ai fait des enquêtes récemment sur les pratiques de visionnage sur *Netflix*. On avait pris parti de ne pas parler de la question des algorithmes, etc. Et c'est très intéressant de voir comme beaucoup des personnes qu'on a interrogées, c'étaient des jeunes de 18 à 25 ans. Comment on fait ? à la fois ils sont conscients de ces algorithmes qui sont plutôt pris dans le fait de vouloir donner ces informations, de les configurer d'une certaine manière. Mais sans comprendre la manière dont ils interviennent, ni comment ils fonctionnent en fait. Donc c'est quelque chose qu'il faut, en tout cas en sciences de l'info-com, qu'on travaille sans être forcément dans une dimension critique. Les datas, c'est sale, etc. Voilà ce qui pose problème.

Arrivez-vous à situer temporellement à quel moment l'acculturation aux données est devenue une nécessité ?

En fait, je ne pense pas que j'aurais commencé à travailler avec Amandine si je ne pensais pas que c'était essentiel à ce moment-là. Donc en 2012 j'avais travaillé sur la société de l'information, des politiques publiques, et en fait il y avait un projet de loi. Parmi tous les textes qui ont été produits sur la société de l'information il y a eu un projet de loi en 2001 et le premier article concernait les données publiques. Et je ne comprenais pas à l'époque. D'ailleurs, en fait, c'est un texte que je n'ai pas analysé dans ma thèse parce que je ne le comprenais pas. Si cette loi n'est pas passée c'est que le gouvernement de l'époque avait considéré que c'était une loi trop technique et qu'en fait les gens ne comprenaient pas. Ça parlait de données publiques, ça ne parlait pas d'ouverture, ce n'était pas le terme qui était utilisé, mais de mise à disposition. Donc qui s'inscrivait en continuité avec la question de l'accès à l'information, à la transparence de l'État. C'était vraiment un texte, je veux dire, si on avait vraiment fait ça, Obama et ses mémorandums il était complètement *has been* sept ans plus tard. Ce texte était très certainement très en avance sur son temps. Mais vous voyez ce que je veux dire : c'est que à l'époque ce texte je ne le comprenais pas. Il y avait plein de passages dans ce texte de loi que je ne comprenais très bien mais cet article-là m'a laissé un peu démuni(e). Mais par contre, le moment où ce sujet commence à se développer, l'arrivée au pouvoir d'Obama qu'on a beaucoup suivi et en France il y a eu toute une sorte d'enthousiasme lié à sa personnalité, son charisme et aussi les promesses qu'il avait faites. Et en fait, à ce moment-là, je commence à me souvenir de ce texte.

Oui évidemment le point de bascule c'est pour moi ce moment où effectivement il y a des politiques publiques liées à l'ouverture des données et que l'enjeu c'est que cela ne soit pas, comment dire, que cette ouverture ne soit pas faite uniquement dans un objectif économique mais qu'elle soit bien dans un objectif de pouvoir d'agir citoyen. On jouait à ça...enfin, ce n'est pas un jeu, mais d'une certaine manière, à chaque fois, on a fait énormément d'interventions grand public, dans des entreprises, dans des associations, avec la campagne *Infolab*. Souvent, on nous demandait « pourquoi vous êtes là ? », et en fait il y avait deux groupes : ceux qui disaient « je suis là parce que c'est un enjeu politique » et ceux « je suis là parce que la donnée ça me fait tripper ». Et en fait, je sais pas si ça répond, ces deux objets, ces deux manières de considérer les données dans la société qui fait qu'on a eu un investissement en temps et aussi en production extrêmement important. Le point de bascule c'est vraiment eux, ils prennent en charge le sujet sociétal, c'est un peu grandiloquent, je pense que c'est ce moment où il y a des politiques publiques et qu'il faut absolument que ça devienne un sujet citoyen.

On arrive sur ce point central qui est la donnée : quelle définition donnez-vous de la donnée ?

Alors ça dans mon HDR justement au moment où j'ai soutenu ça a été une question qui m'a été posée puisqu'en fait vous ne trouverez pas de définition de la donnée dans mon HDR alors que je travaille sur la politique des données, donc ce n'est pas correct peut être. Pourquoi ? Pour moi justement les données c'était quelque chose qui était travaillé par les sujets sociaux que j'observais, avec qui j'étais impliqué, et cetera. Et non, pour moi, la donnée c'était vraiment un ensemble de valeurs. Cela recouvrait des logiques d'actions, s'incarner dans des dispositifs, des stratégies. Je suivais la façon dont justement, quand on parlait de données, eh bien ça faisait agir les gens, c'est que Yves Jeanneret appelle un « être culturel ».

Donc je n'ai pas de définition à proprement parler, je n'en ai en tout cas pas posé dans mes travaux. Après pour le coup, deux éléments : j'ai quand même beaucoup bossé avec des gens qui ne font que ça, de travailler avec les données. Ils ne sont pas d'accords sur la définition à donner de la donnée. Nous ça a été un sujet à

Infolab, c'était un sujet récurrent : « et quand tu dis donnée, tu parles de quoi ? Ok on est repartis ». Et on faisait des débats extrêmement longs sur « voilà comment ça se définit ». Et ailleurs, chez *Etalab*, dans les collectivités territoriales, c'était très difficile en fait de savoir si une donnée c'est des zéros et des uns, qui est une manière dont la donnée est souvent représentée. Est-ce que c'est un actif immatériel ? Ce qui m'intéressait, c'était de voir comment ces conceptions elles vivaient, elles allaient exister les unes avec les autres. Je pense qu'en fait après il y a donnée... enfin après, si on essaye de définir ce que c'est qu'une donnée, je pense que c'est une inscription qui possède un potentiel de retraitement. C'est quelque chose que l'on va inscrire, que l'on va décider de prélever et on va se donner les moyens de pouvoir la retraiter, de la retravailler. Donc je pense que je sais que ça s'entend, si je donne cette définition-là, c'est en continuité avec les travaux de Bruno Latour qui indique qu'en fait quand on parle de données ce sont des obtenues et c'est certain. Je pense que c'est la base de comprendre que toute donnée, ce n'est pas quelque chose qui est donné mais qu'on a produit. C'est quelque chose qui est produit en fonction de l'outil, du dispositif avec lequel on va le produire. Et ça pour le coup, dans les sciences de l'ingénieur, ils le savent très bien. Ils savent très bien qu'en fonction de leurs instruments pardon, la manière dont la donnée va être produite est extrêmement variée. Donc je donnerais cette définition : la donnée c'est quelque chose qui est inscrit et qui surtout va pouvoir être transformé.

Vous avez parlé des données au pluriel : est-ce qu'il y a un ou des types des données qui sont au cœur des enjeux actuels ?

Alors quand vous dites ça je vais vous demander de préciser : quand vous dites des enjeux actuels, parce que parler des industries culturelles et créatives, on peut parler des logiques de production et de recommandation autour des données. Est-ce que c'est le fichage constant en ligne...ou est-ce que c'est tout ce qu'on va pouvoir... les données c'est pas un sujet récent, les enquêtes... quand vous dites enjeux vous dites quoi ?

C'était surtout d'une part les logiques du *big data*, de l'*open data*... et de l'autre les logiques de production de réutilisation... j'aurais dû reformuler en vous demandant s'il y avait un enjeu particulièrement important aujourd'hui.

Je pense là pour le coup je reviens Bernadette Rouvroy ou tous ceux qui ont une approche critique des données. Alors ils ne se recouvrent pas tous mais en tout cas sur la gouvernementalité algorithmique l'enjeu aujourd'hui il est là : comment on fait ? Comment on se situe par rapport à cette gouvernementalité algorithmique ? J'ai l'impression typiquement que c'est extrêmement difficile d'avoir un discours de préservation de nos données. De préservation de soi dans un monde où on est calculé. Et en même temps c'est ce que je disais tout à l'heure : on est dans un monde où il y a en même temps une conscience et une inconscience volontaire. Je suis effaré(e) qu'on utilise encore *Google* comme on l'utilise ou *Facebook*. Et ça me frappe. Et j'utilise *Twitter* mais c'est pareil. Quand je regarde ces espaces dans lesquels je me dis « non mais je ne peux pas être en dehors complètement ». Je parlerai de ces enjeux-là. S'il y a un enjeu c'est celui-ci. Mais je ne sais pas : ça fait dix ans que j'ai produit un certain nombre de choses, pas nécessairement scientifiques je l'accorde mais j'ai rencontré beaucoup de publics différents et on n'a pas non plus une prise énorme sur ce qui se passe et sur ce qui advient. Vous êtes en thèse avec quelqu'un qui écrit beaucoup de bouquins et qui alarme beaucoup : qu'est-ce qu'on fait en fait ? On est tous là en disant « oui mais attention » et en fait on est quand même pris dans ces logiques de ce que d'autres appellent la gouvernementalité algorithmique. Et pour moi c'est un sujet de politique centrale pour le coup.

Si nous restons dans le niveau politique et législatif plus précisément, on a eu un certain nombre de lois et d'applications qui ont émergé. La loi CADA est très ancienne mais plus récemment on a eu la loi Lemaire, le RGPD. Avez-vous un avis sur cette évolution législative ? Et est-ce que cela va dans le bon sens ?

Alors est-ce que ça va dans le bon sens je ne saurais pas répondre à cette question-là. Par contre, ce qui m'a frappé après la loi République Numérique puis le *RGPD*, c'est que le *RGPD* a vraiment fait monter le sujet des données et beaucoup plus que la loi Lemaire. Alors que dans la loi Lemaire il y a l'*open data* par défaut, il y a vraiment des articles qui sont prometteurs. Mais c'est le *RGPD* qui a fait se poser 12 000 questions à tout un ensemble d'acteurs. Et vraiment, quand je dis tout un ensemble d'acteurs, du *freelance* qui tient un média en ligne sur le vélo à une collectivité qui s'interroge subitement comment elle collecte ses données de cantine à une coiffeuse qui a une base de données client. Les hôtels qui enregistraient les pièces d'identité *ad vitam* : « je vais scanner votre passeport » « vous n'avez pas le droit non. Mais après, vous en faites quoi ? » « je sais pas » « bah oui mais c'est juste mes données, tout va bien… ». Je rigole avec ça mais en tout cas le *RGPD* a fait vraiment mais pas vraiment dans un bon sens, c'est à dire que ça a été l'affolement général : « ah mais comment on fait ? Du coup, les gens il faut qu'ils signent des trucs, des cases… ». Je suis très critique du *RPGD*. Et figurez-vous que je fais partie d'un centre Internet et société avec Julien Rossi et Christine Chevret-Castellani avec qui j'ai beaucoup travaillé sur l'intelligence artificielle, à la régulation de l'intelligence artificielle. Eux disent que le *RGPD* est une bonne chose alors que moi je ne trouve pas. Finalement je réponds à votre question indirectement. Et c'est prévu qu'on ait un débat ensemble… alors je ne sais pas quels sont leurs arguments exacts, mis à part que ça a certainement limité certains trucs, certaines pratiques mais pour moi là les preuves du consentement c'est un enfer sur le web. Donc et surtout qu'en fait les gens cliquent juste sur oui. On a créé un consentement généralisé ultraviolent. Pour moi on a inventé la liberté d'importuner du numérique. Je le dis en rattachant un peu des sujets un peu lointains mais très clairement c'est hallucinant. Et sous couvert d'avoir protégé les gens c'est une duperie phénoménale.

Et après sur la loi Numérique, ce qui était dit au moment où elle est passée, au moment où on a fêté deux ans après parce qu'il y a eu prolongation. Il n'y avait pas eu tous les décrets mais elle a été promulguée au bout de deux ans. Ce qui était critique pour les acteurs militants, c'était son absence de contrainte sur tous les sujets : que ça soit la publication des algorithmes, l'*open data* par défaut. Alors le droit à l'oubli aussi, il n'y a pas beaucoup de contraintes qui sont mises en œuvre. La portabilité des données aussi : c'est possible, mais quelle portabilité ? C'est possible de faire de l'*open data* par défaut mais faut-il encore que les collectivités aient les moyens de pouvoir mettre à un moment donné des données, non pas telles qu'elles sont produites justement mais parce que ça ne marche pas. Et pour le coup, le travail de Samuel Goëta et de Jérôme Denis le montre bien : il y a toujours un retravail nécessaire. Je ne sais plus comment ils appellent ça, le terme m'échappe. Oui, ce qu'ils appellent la brutification des données, un travail de renettoyage. En tout cas, l'absence sur la République Numérique, c'était ce qui était souvent soulevé. L'absence de contraintes qui a conduit finalement à une certaine passivité. Alors que le *RGPD* il y en avait tellement que ça a conduit à une certaine frénésie et surtout avec des avocats qui n'étaient pas formés et qui disaient « attention ! ». J'ai vu des interventions, je passais derrière : comment parler d'*open data* alors qu'avant on avait dit « il faut sécuriser, il faut absolument tout fermer » parce que vous n'avez pas de données anonymes. C'était vraiment une période assez compliquée. Clairement. Et donc je pense que le système législatif, l'ensemble des textes législatifs autour des données en France

et un peu en Europe puisque la France *lead* pas mal ces sujets-là, ne sont pas comment dire, ne conduisent pas en tout cas à des pratiques concrètes dans les administrations du fait du manque de moyens et du manque d'infrastructures. Et je dis juste ça, typiquement les collectivités publiaient des données. On va leur répondre souvent que il y a data.gouv : il suffit de créer un compte pour publier ses donnée sur data.gouv. Mais pour une collectivité, ce travail-là, c'est invisibiliser les choses.

Comment peut-on s'y prendre pour mettre en place une culture de la donnée dans une organisation et est-ce que ça doit concerner tout le monde ?

Je pense deux choses : je pense effectivement qu'il faut former les profs avant de changer les programmes. Formons les profs, ça prend du temps c'est sûr mais il faut former les formateurs. Je ne sais pas pourquoi on dit qu'on veut toujours former les enfants et en fait on se retrouve après avec des gens qui ne savent pas faire. Ce n'est pas facile. Et pour avoir fait beaucoup d'ateliers et avoir vu beaucoup de gens faire des ateliers, ce n'est pas facile de former à ces questions-là. Il y a un gros travail à faire et là on y travaille. Vous le savez d'ailleurs : le *GTNum* sur la littératie des données, ça va permettre de renforcer les ressources à disposition des profs et ça va permettre de mieux penser leur rôle. Le rôle des données dans l'enseignement. La classe, pas le rôle, des données dans l'enseignement. Après on verra aussi ce que ça donnera. On sait qu'il y a des études qui sont faites à des moments et qui disent « voilà ça c'est super » et c'est pas maintenu forcément. Mais en tout cas, ça je pense que c'est un premier axe. Vous disiez que c'est pas très structuré, ça se travaille ? Pas vraiment, je pense qu'il y a un petit mouvement de rapprochement entre nous tous qui avons travaillé sur ces sujets-là. Et donc ça peut aussi favoriser comment dire une meilleure adéquation des productions et des réflexions scientifiques avec des productions à destination des personnels de l'éducation. Donc ça c'est le premier point.

Le deuxième point, je vais rebondir sur ce que je disais tout à l'heure : c'est de penser les données dans les espaces dans lesquels elles sont produites et donc de ne pas strictement réduire les données à des questions d'informatique. Je pense que c'est le deuxième aspect central. C'est d'être dans une réflexion qui permet d'inscrire les données en lien avec l'éducation aux médias et à l'information d'une part, et d'autre part de le raccrocher à tout un ensemble d'aspects techniques, la dimension, la place des technologies, leur rôle et leurs conditions de manipulation. En tout cas c'est ce mouvement entre *EMI* et la technicité, les gestes techniques qui sont liés à l'usage de ces deux dispositifs d'instruments liés aux données, au traitement de données.

Pour rebondir sur ce que vous avez précisé sur le fait de s'émanciper de la dimension informatique : à l'issue d'une formation à la culture des données, que doivent avoir développé les personnes qui y ont participé ?

J'ai envie de dire qu'il y a un enjeu autour de la fameuse définition de la donnée dont vous parliez tout à l'heure. C'est-à-dire qu'ils aient une vision claire du fait qu'une donnée, ce n'est pas quelque chose qui existe : les *raw data* ça n'existe pas. C'est le titre d'un bouquin : *Raw data is an oxymoron*. Donc c'est l'enjeu de départ, de bien comprendre qu'une donnée c'est produit et que comprendre ses conditions de production c'est super important pour pouvoir ensuite s'en servir et être en capacité d'action par rapport à ces données. Ce serait le premier point. Après là on va travailler à la dimension de physicalisation des données dans le petit projet à l'intérieur du gros projet *GTNUM*. On va travailler un peu sur la physicalisation. Et je suis assez persuadé(e) que c'est un

dispositif de représentation des données qui peut être intéressant. Je comprends bien l'enjeu de technicité très souvent, il faut pouvoir ouvrir des logiciels, manipuler des logiciels, etc. Mais je reste aussi persuadé(e) tout comme en fait : est-ce que ça sert à quelque chose d'être formé à *Photoshop* si on n'a aucune culture graphique ? je n'en suis pas sûr(e). Je pense que les gens produisent des images parce qu'ils ont une culture de l'image de dingue. Là c'est pareil. Du coup, il faut savoir se représenter en fait. Ce n'est pas parce qu'on va leur dire qu'il faut savoir programmer que ça sera suffisant ou que ça nous apportera quelque chose d'un point de vue politique.

Politique aussi dans le sens manque de professionnels, de praticiens, et tout ça. Je pense qu'un travail sur l'idée qu'on a c'est aussi que grâce à la physicalisation : on va essayer de faire comprendre les conditions de production des données, c'est-à-dire que c'est pas juste « je représente, c'est beau », ça permet de dire : c'est quoi les variables ? Qu'est-ce que j'ai comme éléments ? Est-ce que cette donnée elle est claire ? Comment elle a été structurée ? Est-ce que je peux en faire quelque chose ? Est-ce que je peux la questionner ? C'est ça notre enjeu. C'est pas juste de mettre des balles dans des tubes pour faire des comparaisons. Mais ça va demander de réfléchir un petit peu à ces dispositifs de médiation que l'on va mettre en place.

Un terme revient souvent, celui d'esprit critique vis-à-vis de la donnée : est-ce que ça vous parle et comment le définiriez-vous ?

Alors, on travaille justement là dessus : vous avez vu le rapport Bronner qui ne parle pas beaucoup de données, surtout d'esprit critique, et puis les lumières à l'ère du numérique qui parlent un peu de données, en tout cas qui parlent beaucoup des algorithmes et la façon dont ça modifie notre manière de comprendre et de percevoir l'information. Alors je fais ce rapprochement là parce que je pense effectivement que la question de l'esprit critique aujourd'hui est centrale d'abord parce qu'il est préempté par des gens qui ont une vision de l'esprit critique qui est assez problématique de mon point de vue. Parce que la notion de l'esprit critique est souvent réduite à des enjeux d'objectivité. Or, pour moi l'objectivité n'existe pas. Quant aux enjeux de vérité : qu'est-ce qui est vrai aujourd'hui ? Cela dépend les points de vue, où l'on se situe, ça dépend les moyens de que l'on a à chercher l'information. Et puis les enjeux d'autorité. Cela serait pas autorité d'ailleurs, cela serait plutôt la question de la fiabilité. Je ne sais pas si vous me suivez mais en tout cas je pense qu'il y a un enjeu autour de cette définition de l'esprit critique, surtout si on la rapporte à la question des données. Ça revient à la question de la donnée brute, etc. Mais ça revient aussi à ça, c'est-à-dire : est-ce qu'à un moment donné, la donnée n'est pas un élément sur lequel on est, on va dire que c'est quelque chose de stable mais on sait bien les chiffres du chômage, de la délinquance…il y a tout un pan de la sociologie qui a montré combien notre rapport à des sujets qui sont centraux est travaillé par notre capacité à comprendre la manière dont ces données sont produites et dont elles circulent dans les médias, dans les conversations, etc. Donc je pense que ça en fait partie : quand on travaille à la culture des données, la dimension esprit critique doit en faire partie mais qu'elle est la conception de l'esprit critique que l'on va mobiliser ? la question de la vraie fiabilité ? Est-ce que c'est la question de l'autorité qu'on accorde à une source avec tous les problèmes que ça pose à nouveau sur la dimension médiatique ? Si on regarde la page leMonde.fr aujourd'hui, il y a des articles du *Monde* qui ne sont pas du *Monde*. Mais on est quand même sur le Monde.fr. Comment on apprend à reconnaitre ce travail éditorial alors même que on va dire que « le Monde c'est un journal sérieux, etc ».

Dans quelle mesure aujourd'hui tous ces processus de qualification, d'enjeux de reconnaissance, de la scientificité, ou de la bonne construction d'une information : c'est plus compliqué que de tracer des lignes entre ce qui appartiendrait à une source d'autorité et ce qui ne l'est pas. Tous les débats aujourd'hui sur la désinformation tiennent bien de ça et puis tous les débats qu'il y a eu sur l'islamo-gauchisme tiennent encore de ça. On est à un moment très charnière d'un point de vue du rapport aux sciences liés et notamment liés à ces conceptions circulantes de l'esprit critique. Ramener ça aux données, cela complexifie la chose, bien évidemment. Je me répète mais je pense que quand on dit culture des données, il y a forcément à un moment donné une réflexion à apporter en termes d'esprit critique et donc, si je définissais l'esprit critique, c'est être en mesure de qualifier, d'avoir des capacités pour qualifier les informations et donc les données, que je rencontre dans différentes espaces médiatiques, pédagogiques, et que je puisse ensuite me positionner par rapport à cette information ou cette donnée. Aux États-Unis aujourd'hui, d'abord à certains endroits on n'enseigne pas le darwinisme et on enseigne à côté les théories créationnistes. On parle des États-Unis, le pays qui a souvent une aura particulière. Il y a du travail quand même, de la marge.

Si on reste dans ce point de vue, voyez-vous le niveau d'acculturation de votre public aux données ?

Qu'est-ce que j'observe ? Alors en tout cas dans les étudiants, je fais beaucoup de cours sur les données, cela ne s'appelle pas cours sur les données, l'année prochaine on en fait un, je vous raconterai ce que ça donne. Ce que j'ai observé en tout cas, ce sont des étudiants en communication, j'enseigne peu dans d'autres matières pour avoir un vrai retour. Mais en tout cas, chez les étudiants en communication « on veut faire de la com, les données c'est pas notre champ, on comprend pas ». Et ça je l'ai observé assez tôt, ce refus ou de mise à distance parce qu'ils font de la com'. C'est oublié uniquement quand on veut être *community manager*, comprendre les chiffres d'audience peut être intéressant, c'est de la donnée ! Et en fait pour eux *community manager* c'est d'abord être dans l'interaction, dans la production de slogans, de petits phrases : ils délaissent l'aspect stratégique qui va être lié évidemment à la capacité de comprendre l'audience, de sortir des prémâchés que peuvent fournir les *analytics* en termes d'engagement, etc. Et ça c'est un peu compliqué. Après, ceux qui vont être confrontés, et ça c'est ce que j'observe quand ils partent en stage, ils reviennent et ils disent « ok, là ça y'est on a vu donc on va plus se concentrer et on va voir l'intérêt ». Et ça c'est vraiment dans les trois dernières années. C'est vraiment monté en puissance.

Avant j'avais vraiment l'impression de brasser du vide. Il y a vraiment un enjeu et il y avait un enjeu énorme pour eux en tout cas. Sur le tas ils récupèrent et font le pont avec ce qu'ils ont vu à la fac. Mais disons qu'il y avait comme une décorrélation aujourd'hui, en tout cas les étudiants m'écrivent, ceux qui m'écrivent me disent « on a compris ». J'ai un étudiant, il m'a écrit un mail en m'expliquant que finalement, le cours que je lui avais donné lui servait quasiment quotidiennement. Il exagère certainement mais parce qu'il travaille chez *Sony* avec des artistes mais c'est lui qui négocie leur présence sur *Spotify*, *Deezer* et autres plateformes. Et en fait il n'arrête pas de travailler avec la question des données, comment l'artiste va être recommandé, à quel taux... c'est intéressant qu'il soit poussé un peu. Et là je me suis dit qu'il y a tellement de choses que je ne sais pas qui se passent dans les négociations entre les labels et ces boites de musique et il me dit « je parle data sans arrêt ». Et je pense que mon cours est très en-dessous de ce qu'il fait maintenant mais au moins il avait la possibilité de comprendre deux trois trucs. Donc c'est un exemple, je n'en ai qu'un mais pour le coup c'était extrêmement

patent. Il est dans la culture, industrie culturelle, il a fait la transposition, le transfert du monde académique vers le monde professionnel. Ça c'est une observation plutôt récente.

Vous l'avez évoqué à plusieurs reprises mais j'aimerai vous interroger sur les difficultés d'acculturation aux données : quelles sont les difficultés actuelles ?

Le fait d'avoir du mal à se représenter et c'est un peu lié à ce que vous disiez quand on parle donnée : entre les données à caractère personnel, les données de fréquentation, les données budgétaires. On peut ajouter la trace du clic, du temps passé sur une page. Et je pense qu'il y a une difficulté de représentation et une difficulté aussi à pouvoir en faire quelque chose sur lequel eux ils vont avoir prise. Je pense que c'est vraiment ça le souci : c'est-à-dire comment à partir de là je vais pouvoir le transformer pour faire de la com et de ne pas faire du traitement statistique ? Il y a ce passage-là qui est extrêmement complexe pour eux à se représenter et que certainement nous, on n'a pas suffisamment de cas si je fais mon autocritique. Ce que je vous disais tout à l'heure, ces indépendants, ces médias qui sont indépendants avec des *freelances* qui en fait ont été en négociation à un moment donné sur le *RGPD*, est-ce que je fais un *paywall* ? J'ai besoin de la pub pour vivre, c'est mon gagne-pain, je dépends d'une régie publicitaire, quelle est ma marge de manœuvre ? Il faudrait que j'aille rechercher des histoires comme ça pour que ça soit plus concret pour eux, pour qu'ils puissent le lien entre comment ces données, qui sont liées à une activité communicationnelle, comment on s'en sert, comment on va chercher aussi à se les approprier pour en faire derrière des éléments de positionnement, des éléments de négociation, etc.

Est-ce que vous avez en tête un projet en *data literacy* et en dire deux trois mots dessus ?

Alors ça sera trois mots. Je me souviens de Catherine d'Ignazio sur le féminisme, les données genrées. Je regrette de pas y avoir pensé. C'est quelque chose que j'aurai voulu faire et Amandine, je pense je ne lui en ai jamais parlé, mais Amandine aussi. Ce sont des sujets qui nous intéressent l'une et l'autre. Ce sont plus des activités personnelles que des activités liées et qu'on n'avait pas mis dans nos activités de praticiennes des données. Et ça c'est comment la donnée peut devenir un moyen de se saisir d'un sujet et de devenir acteur de son féminisme. Elle ne le dit pas comme ça, c'est maladroit. Mais s'il y a quelque chose qui m'a et inspiré et que j'aurais voulu faire, c'est ça.

Pour terminer, avez-vous des recommandations pour une meilleure acculturation aux données ?

J'y ai un peu répondu mais déjà avoir des cas à étudier, en tout cas pour les étudiants en com. Ce que je disais tout à l'heure : partir de cas concrets où les données jouent un rôle déterminant dans des métiers liés à la communication et qui ne soient pas que des mesures d'audience. Mais qui soient des choses, des objets de négociation parce que je pense que ce sont des enjeux. D'une part.

Après le petit projet sur la physicialisation des données, l'enjeu c'est justement de partir, de s'appuyer sur les données pour que les collégiens et lycéens aient une capacité d'action renforcée sur leur orientation et ça c'est pas le *data feminism* mais ça pourra donner quelque chose d'intéressant. Je parle de manière très concrète mais ça ne peut passer que par des actions de médiation concrètes. Ce partage de la culture des données. Mon

enjeu, et je finirais sur ça, c'est que ça ne soit pas le pouvoir de quelques-uns, le risque en tout cas en France c'est vraiment que ça soit une élite qui maîtrise les données et que ça soit soit une élite, soit des industriels. Permettre par des sujets qui concernent les gens. C'est important que ça les concerne. Sinon le risque c'est comme ceux qui disent que les maths ça sert à rien. « Pythagore ça sert à rien ». Si ça sert en fait. D'abord à la construction de ton esprit et c'est extrêmement formateur pour toi : si tu comprends Pythagore, tu comprendras plein d'autres choses qui n'ont rien à voir avec Pythagore et même rien à voir avec des maths. Parce que c'est un processus cognitif en fait que tu as appris aussi. Et ça je pense que sur les données, il y a un enjeu de cet ordre-là, c'est-à-dire que ça ne soit pas capté par des élites ou, et ou uniquement par le capitalisme médiatique mais que ça soit partagé et donc il faut trouver des modalités de partage, de prise de conscience des enjeux liés à ça et de faire donc de la donnée, les données, comme modalité de réappropriation de son parcours scolaire. En tout cas ça sera notre objet. De la sa réappropriation de sa place dans la société en tant que femme. Cela sera mon ouverture.

Universitaire 4 (20/04/22)

Comment définissez-vous la *data literacy* et employez-vous ce terme vu que vous avez parlé de culture des données ?

Déjà pour situer un peu mes recherches sur les données ouvertes. La régularisation des données ouvertes c'est plutôt sous l'angle des pratiques professionnelles, des acteurs professionnels qui réutilisent les données, que ce soit des *data scientists*, *data analysts*, des développeurs donc bon on a aussi des administrations qui ouvrent des données. Et cette entrée-là effectivement ne suppose pas le travail direct avec la question de la *literacy* et de la culture des données puisque ce sont des personnes qui ont déjà non seulement cette culture des données mais une culture avancée des données.

Mais c'est vrai que cette question de la *data literacy* peut se poser dans le cadre de la réflexion de la réutilisation des données, il peut se poser quand on s'interroge sur les réutilisateurs finaux, c'est-à-dire les citoyens. Et c'est là-dedans que se pose la question de la culture des données. Je parlais plutôt de la culture des données dans mes travaux mais les deux termes me conviennent : je n'ai rien contre *data literacy*. Pour moi, *data literacy*, ça implique une référence plus importante à la question de la pédagogie. Avec la culture des données, on n'est pas obligé de passer par l'éducation en fait. Alors que la *data literacy*, ça suppose une démarche éducative et prescriptive quelque part. Pour moi, il y a moins de liberté dans la *data literacy* que dans la culture des données.

Vous faites vraiment le distinguo entre *data literacy* et culture des données qui serait plus large.

C'est ça. Pour moi, la culture des données c'est plus large. C'est aussi moins prescriptif, voire moins restrictif, c'est-à-dire que c'est un peu la culture qui peut se constituer de différentes manières par le faire, par les discussions… *data literacy*, je l'associe plus à des démarches éducatives ou pédagogiques qu'on peut développer dans le cadre plutôt institutionnel.

Et donc cette *data literacy*, vous venez de le dire, elle s'inscrit dans un ensemble de concepts autour de la question des données avec la gouvernance des données, la *data science*, et à l'intérieur, où est-ce que, selon vous, on pourrait la situer ? Est-ce une chose supplémentaire ou quelque chose similaire à la *data science* avec un volet pédagogique ? Est-ce que vous arrivez à le situer de votre côté ?

Pour moi c'est un pôle à part : avoir une culture des données, développer une *data literacy* ça nécessite de parler ou d'éduquer à la question de la gouvernance aussi, de la prédiction des données, de la réutilisation des données donc ça nécessite une vision de l'écosystème en quelque sorte. Donc c'est pour ma part un peu surplombant : on est pas aux mêmes échelles en fait que *data science* ou gouvernance des données. C'est plus quelque chose de surplombant qui nécessiterait en fait d'aborder les différents aspects que vous évoquez.

La *data science* est poussée d'un point de vue technique et la *data literacy* à côté vous avez cité que c'était plutôt quelque chose d'éducatif donc est-ce, si on devait donner une « nature » de la *data literacy*, qu'elle serait-elle ? Un ensemble de pratiques, un état d'esprit, une éducation ?

C'est plutôt un ensemble de pratiques par lesquelles on éduque en fait. Je reste sur cet aspect d'intervention et d'éducation en fait qui pour moi est associée à la notion de *literacy*.

On note un certain engouement pour les données. Est-ce qu'aujourd'hui la maîtrise des données est plus importante ou équivalente avec d'autres nécessités comme l'information, l'informatique.

On a plus d'attentes par rapport à la *data literacy* et qui viennent de tous horizons. On a beaucoup de sollicitations là-dessus parce que c'est le terme même de donnée/data qui est au cœur de l'économie d'aujourd'hui, au cœur des réflexions sur le numérique aujourd'hui donc effectivement on est passé de la culture numérique à la culture des données. C'est un effet de la mode en quelque sorte mais aussi une évolution de nos sociétés puisqu'en même temps on constate qu'on exploite davantage des données qu'avant et on a des moyens beaucoup plus poussés aussi pour les exploiter. On a des outils qui, avant, étaient réservés qu'à une poignée de personnes. Aujourd'hui, c'est un ensemble de phénomènes dans nos sociétés qui fait que la question de la culture des données devient quelque chose de très important.

Et est-ce que vous pouvez, de votre point de vue, situer temporellement depuis quand ça a vraiment pris de l'ampleur ?

En France en tout cas, parce qu'il faut toujours se situer par rapport à des horizons géographiques. En Europe du Nord je donnerais d'autres chiffres mais pour la France ça serait à partir des années 2014-2015 parce que l'ouverture des donnés ça commence en 2010 si on prend l'exemple de l'*open data* mais ça n'est réservé qu'à des administrations qui les ouvrent, à une poignée d'acteurs dans ces administrations qui sont porteurs de ces mouvements-là. A quelques acteurs d'*Etalab*. Mais quand on regarde : on avait travaillé sur des portails de données ouvertes à l'époque et c'est vrai que c'étaient des portails qui comportaient des jeux de données et personne ne savait comment les exploiter dans des formats csv par exemple que personne ne connaissait, on ne savait pas les manipuler. Et donc la deuxième génération des portails open data, puisque ce sont des portails qui évoluent, justement arrivent dans les années 2014-2015, notamment à Lyon, et intègrent petit à petit de

plus en plus d'outils qui permettent de faire des choses avec ces données, de faire des visualisations ou de se situer sur une carte, des choses simples, basiques, mais qui du coup contribuent à sa manière à cette culture des données des citoyens lambda qui peuvent faire quelque chose de ça. Je situerais ça dans les années 2014.

Vous avez cité les données à plusieurs reprises et j'aimerais aborder avec vous ce point épineux qu'est la donnée en débutant par une question de base mais qui est centrale : qu'est-ce que c'est la donnée pour vous ?

C'est une question très compliquée. La réponse la plus simple et sociologiquement la plus intéressante c'est de dire que les données se sont les obtenues, comme le dit Bruno Latour, de déconstruire toute la mythologie sur la construction des données en disant que les données sont brutes, alors que quand on travaille avec les acteurs on sait très bien que ce n'est pas le cas, mais bon on peut aussi partir de définitions qu'on donne dans des cours donnés aux étudiants en disant que les données ce sont des informations obtenues à partir de différents systèmes de mesure et donc des signaux qui peuvent se traduire dans des données et qui sont enregistrés [...] et sont extraites pour leur donner du sens.

Vous l'avez cité au début, vos travaux s'ancrent principalement dans les données ouvertes, et j'ai une question par rapport aux données en général, c'est les enjeux actuels selon vous liés aux données. Est-ce que vous arrivez à en définir certains qui sont plus prégnants que d'autres ?

Les enjeux restent les mêmes qu'il y a encore cinq ans ou dix ans, parce qu'on aurait l'impression que certains enjeux peuvent disparaître : ce sont les enjeux économiques et les enjeux de la gouvernance. La question de la gouvernance reste criante parce que l'écosystème des données c'est un écosystème très complexe qui fait converger les acteurs publics et privés et la question de la gouvernance est la question la plus épineuse à laquelle on ne sait pas très bien répondre, ni en matière de données ouvertes, ni en matière de *big data*. On peut aussi aborder la question juridique : les données personnelles cela serait plutôt des enjeux juridiques, mais si on les regarde c'est aussi l'action de la gouvernance. Le terme de la gouvernance des données en fait c'est le terme prégnant en fait et qui va rester central encore quelques années parce que ce sont des intérêts tellement divergents des acteurs qui sont concernés.

Donc, les problématiques actuelles ça dépasse même les types de données et quand je dis type c'est vraiment très large : *open data*, données personnelles, etc. Pour vous la problématique actuelle c'est vraiment une problématique de gouvernance.

Oui parce que la question de la gouvernance touche beaucoup, concerne tous les acteurs : aussi bien les producteurs qui se posent la question de qu'est-ce que leurs données vont devenir ? Est-ce que ça vaut le coup de les prédire ? Est-ce que ça vaut le coup de les diffuser dans un format ou dans un autre ? Et le coût de production de ces données de stockage, comment peut-il être réparti ? Est-ce que le modèle économique va fonctionner ? Donc voilà, cette question-là va se poser pour les producteurs et ça se posera aussi pour les réutilisateurs : est-ce que j'ai le droit de réutiliser la donnée ? sous quelles conditions ? comment ? dans quels formats ? est-ce qu'il y aura une redevance à payer ? est-ce que la donnée va être pérenne dans le temps ? Et tout ça, ça dépend justement de la gouvernance.

Par rapport à ces questions globales d'éthique, il y a une question qui me vient sur le cadre législatif. On a en France des réglementations pour encadrer la pratique autour des données : la loi Lemaire, le *RGPD* et tout ce qui est en cours de discussion au Parlement européen, à savoir le *data act* et le *data governance act*. Est-ce que vous constatez une différence dans les pratiques du fait de cet éventail législatif ? Et considérez-vous que c'est un bon début ou qu'il y a encore des choses à améliorer ?

C'est un bon début, en tout cas en termes de protection des données personnelles d'une manière indéniable en France et en Europe, on a bien avancé sur ce sujet-là. Effectivement, ça nécessite le changement de pratiques d'acteurs en termes de collectes de données, notamment dans des sociétés qui utilisent des *big data* et qui avaient l'habitude de revendre leurs fichiers. Non, on ne fait pas de la même manière. En tout cas, on prend des précautions. On trouve aussi des moyens de détourner le système pour continuer à travailler. Mais ça pose des questions et ça redéfinit les pratiques professionnelles des acteurs dans le domaine du marketing par exemple c'est clair. Après, est-ce que c'est suffisant ? Je pense que la législation en termes d'*open data* est bien posée en France. C'est plus une injonction qu'un impératif. On a l'*open data* par défaut. C'est surtout la réutilisation qui pose le plus de soucis. Parce que c'est là où on voit les données ouvertes, les *big data*... En tout cas il faut partir des pratiques des acteurs et non pas des cadres idéals en termes [coupure de son].

Si on entre sur un angle plus pratique, est-ce que la *data literacy* / culture des données doit concerner tout le monde ou être réservé qu'à certains publics ?

Aujourd'hui, je pense que c'est devenu incontournable pour tous parce que pour les jeunes, les collégiens, les lycéens, c'est nécessaire. Parce que dès qu'on devient usager des objets numériques, connectés, la question des données se pose. C'est comme la question des usages des autres objets. Dans les rêves de sécurité, on va avoir des connaissances, qu'est-ce qu'on fait ? Qu'est-ce qu'on peut faire ? Je pense que ça concerne tous les publics du moment où on utilise ces objets connectés. Ça concerne les jeunes, les vieux, ça concerne toutes les catégories de la population. Après peut être au niveau du volume de connaissances des domaines, on peut avoir des nuances. Au niveau des acteurs et des agents publics, au sein des administrations je pense à tous les acteurs que j'avais rencontré dans les différents services de la métropole qui produisaient des données sans savoir qu'ils les produisent. Au sein des administrations, il y a des agents publics qui produisent des données sans savoir qu'ils produisent des données, et donc que peut-être il est nécessaire d'avoir des formations plus spécifiques, professionnalisantes, sur les données, sur l'acculturation aux données, va dépendre si on est un citoyen lambda ou si on est concerné par la problématique des données d'une manière plus ou moins directe.

Et vous, est-ce que vous avez une vision globale de cette mise en place de cette acculturation ? Est-ce qu'il faudrait avoir des moyens spécifiques, des personnes spécifiques ? Ou est-ce que ça va être spécifique à l'organisme dans lequel on va se situer ?

Je crois beaucoup à l'Education Nationale. Si on veut sensibiliser les jeunes à différentes choses, il faut commencer là-dessus et travailler ces sujets-là à l'école. Et là tout le monde sera concerné, bien évidemment, parce que tout le monde sort de l'école et passe par l'école. Après, il y a déjà des initiatives qui sont déjà portées, il y a des choses qui se font au sein des tiers-lieux. A Lyon, on a l'organisme qui est le *Tubà* qui est un tiers-lieu d'innovation et donc de culture des données aussi et il y a des animations qui sont organisées : il y a des jeux de

cartes, des *escape games*, des trucs comme ça qui sont sur des cookies, sur autre chose qui sont organisés pour sensibiliser les différentes populations à ces enjeux-là. Il y a aussi le *CLEMI* qui faisait de l'éducation, qui continue à faire de l'éducation aux données. Il y a des organismes plus spécifiques, plus traditionnels, d'éducation au numérique qui aujourd'hui se penchent sur l'éducation aux données. Après, au sein même des organisations publiques, ce ne sont pas des administrations par exemple, c'est des acteurs qui s'y intéressent plus et qui sont aussi moteurs pour lancer des initiatives. A la préfecture du Rhône, il y a une cellule interne, ils ont créé un *datalab* en fait interne à la préfecture pour justement acculturer leurs agents à des outils de *design thinking*, très modernes d'ailleurs et innovants. Donc voilà ça peut venir de l'intérieur et souvent en fait ça vient des personnes, des individus qui sont sensibles à ces points-là et qui amènent des initiatives qui se traduisent ensuite par des décisions un peu organisationnelles. Je pense que ça peut venir de tous les niveaux, à toutes échelles.

On parle depuis le début de cette acculturation aux données. Et moi la question que je me pose, c'est que doivent développer les personnes à l'issue de cette acculturation ? Oui on doit connaître la donnée mais qu'est-ce qu'on doit connaître exactement ?

Il faut avoir un état d'esprit certainement, connaître un peu les grands enjeux en fait, connaître un peu ses droits en fait. Et avoir un peu la connaissance de la boite noire : je ne dis pas tout, on n'a pas besoin de savoir comment ça fonctionne mon ordinateur ou mon smartphone mais en revanche, quels types de données peuvent être récupérés puisque, en acceptant des conditions d'utilisations, on ne lit pas en fait, on lit mais pas tout. Expliquer en général, le fonctionnement général, le cadre général, quitte ensuite que chacun se débrouille par rapport à des outils qu'il utilise lui-même, mais connaître le cadre général je pense que c'est ça la base. Après en termes de manipulation des données, je ne suis pas convaincu que tout le monde sache coder, je pense que c'est une utopie aussi. Déjà il faut que tout le monde sache écrire, c'est déjà suffisant, donc voilà il n'y a pas besoin de ça. En revanche, connaître les grands enjeux, je pense que c'est ça la culture des données.

Et justement il y a un terme qui revient tout le temps, c'est cette notion d'esprit critique, mais là vis-à-vis des données. Vous, qu'est-ce que vous comprenez par l'esprit critique par rapport aux données ?

On vit dans une société de *fake news*, etc., et donc effectivement, à l'époque des *fake news*, l'esprit critique c'est le mot que l'on nous répète tout le temps. Et on exige de nos étudiants d'avoir un esprit critique, à partir des données je trouve que c'est compliqué. C'est très compliqué en fait. L'esprit critique par rapport à la donnée, il faut avoir la capacité à évaluer la donnée en tant que telle et c'est des données qui viennent de différentes branches, de différents domaines, c'est très compliqué. La donnée, qu'est-ce que la donnée quand vous l'avez sur votre ordinateur, votre donnée c'est juste un chiffre. C'est tout, c'est juste un nombre. Avoir un esprit critique par rapport à ce nombre, eh bien c'est compliqué. Après je vois à quoi ça renvoie. Je dirais c'est plutôt avoir un esprit critique par rapport à des dispositifs qui produisent et diffusent les données que par rapport à la donnée. Avoir un esprit critique par rapport à la donnée, je le garderai pour les *data analysts* et les *data journalists* que j'avais beaucoup interviewés et effectivement j'ai vu un esprit critique par rapport aux chiffres et aux nombres parce qu'ils savent interroger les chiffres et les nombres avec des outils, croiser les données, croiser ces chiffres et ces nombres, vérifier, etc. Mais pour un citoyen lambda, c'est un esprit critique par rapport aux dispositifs qui utilisent des données.

En tant qu'enseignant-chercheur, voyez-vous une évolution des compétences générales, en tout cas sur le niveau général d'acculturation aux données chez vos étudiants ?

Je dirais quand même que les étudiants aujourd'hui sont beaucoup plus conscients de la question des données. Ils sont conscients, ils en retiennent l'idée qu'il faut être attentif, vigilant, il y a donc de la méfiance qui est là. L'idée de l'esprit critique, avoir un esprit critique, ça ils vont le sortir. Mais après en termes de pratiques, c'est un peu plus compliqué : je me rends compte que, quand on creuse un peu, j'ai le cours de la culture numérique où on aborde la question des données, quand on creuse, ils ne savent pas grand-chose en fait. Ni sur les dispositifs législatifs. Ils savent qu'il y a une boite noire, mais savoir comment ça marche dans la vie réelle, ils ne savent pas. C'est juste l'esprit de méfiance en fait qu'ils ont et qui peut-être contraste avec l'esprit d'insouciance des étudiants d'il y a encore quelques années. Peut-être est-ce déjà le début, je n'en sais rien. Donc les étudiants d'il y a quelques années ne se méfiaient pas autant en fait, étaient beaucoup insouciants là-dessus. Donc là maintenant, on a l'insouciance mais il manque toujours des éléments concrets de quoi il faut se méfier, qu'est-ce que ça implique, quels sont les dangers réels ? Et par rapport à des collègues, ça change pas : ceux qui étaient des habitués à manipuler dans le domaine, oui ils sont bien conscients. Après, je côtoie aussi des collègues qui ne sont pas dans l'info-com et j'ai pas l'impression qu'ils ont évolué là-dessus.

Est-ce que vous sentez une plus grande imprégnation de ces logiques-là ?

Je ne sais pas si c'est perceptible. Peut-être à l'époque du Covid, si on l'a quand même vécu un peu. Et on a eu l'occasion d'avoir des discussions avec des collègues, notamment par rapport au choix de la plateforme qu'on utiliserait pour communiquer entre nous ou avec les étudiants et c'est là-dessus où on était obligés en fait de donner des arguments un peu plus concrets : pourquoi utiliser telle ou telle plateforme ? Quels sont les dangers d'utiliser telle ou telle ressource ? Donc là il fallait argumenter, documenter, renvoyer à des articles, donc c'était aussi… je pense que le Covid nous a amené beaucoup d'interrogations sur ce sujet et était un point d'interrogation sur les datas. Les datas en général, les datas numériques, les choses sensibles que l'on peut laisser en termes de traces numériques. Et puis, il y a des collègues qui peuvent demander, poser des questions parce que chacun a ses compétences donc on peut avoir aussi des discussions sur ces sujets.

A contrario, quelles sont les difficultés que vous percevez à cette acculturation ? Vous avez cité, pour les étudiants, certaines difficultés qu'ils avaient en termes de connaissances, de connaissances législatives, mais de manière plus générale, où situerez-vous ces difficultés ? Sur ces connaissances ? Ou sur une vision conceptuelle de la donnée ?

Il y a les deux volets. Déjà, définir la donnée c'est compliqué, à distinguer qu'est-ce qui est donnée, qu'est-ce qui est information. Même les étudiants en parcours info-num, ils ont du mal avec tout ça. Donc je pense qu'il y a des questions conceptuelles sur la donnée mais ce sont des questions qui viennent a posteriori. D'abord on part des usages en fait, ou des cas concrets, ou des cas où on s'est retrouvés dans des difficultés par rapport à des données qui étaient dévoilées par exemple ; par rapport à des données qui n'étaient pas forcément protégées. Et ça, ça peut devenir un point de départ sur qu'est-ce que la donnée, est-ce de la donnée ou pas ? Donc on partira plutôt des cas d'usage pour aborder la vision conceptuelle. Mais les manques de connaissances concernent les deux aspects.

Cela me fait rebondir sur une de mes dernières questions qui concerne les projets autour des données : en avez-vous un en tête qui a bien fonctionné et pourquoi ?

A quelle échelle ? Des étudiants ?

Les étudiants mais cela peut être aussi à un niveau plus global

Actuellement, je travaille sur un projet qui s'appelle *Adoprivacy* qui porte les données privées des jeunes, des adolescents. C'est un projet financé par le défenseur des droits qui a démarré il y a un an à peu près. On va avoir des journées de restitution le 4 et le 5 juillet de cette année à Paris et il se trouve que c'est un projet qui est intéressant parce que c'est une recherche action en fait qui nous a amené à aller dans les lycées, auprès des jeunes, pour faire des *focus group*, interroger les jeunes sur leurs pratiques, qu'est-ce que c'est les données privées pour eux, qu'est-ce qu'ils considèrent données privées, données publiques. Qu'est-ce qu'ils font avec ces données. Comment ils se protègent. Donc il y avait des *focus group*. Puis on a construit des saynètes de mises en situation en quelque sorte qu'on a proposé à leurs enseignants. La deuxième phase c'était de jouer ces saynètes en classe : on leur proposait des situations. Par exemple « Léa a créé un faux profil, elle a utilisé les données de quelqu'un d'autre, elle a usurpé l'identité de quelqu'un d'autres, il y a des données personnelles » donc qu'est-ce qu'elle risque, qu'est-ce que ça implique, qu'est-ce qu'on lui conseillerait. Des saynètes, il y en avait sur différents sujets. L'idée de les faire jouer en classe, de les mettre les élèves en situation de réfléchir sur un cas qui ne les concerne pas je pense que ça a plutôt fonctionné pour un dispositif pédagogique parce que du coup voilà elle n'était plus...c'était assez concret, assez simple. Mais en même temps, il faut voir ce qu'ils en retiraient mais j'ai l'impression que ce type d'échanges ne se cantonnait pas dans la prescription, dans les bonnes pratiques « ça c'est bien, ça c'est mal », ça, ça ne fonctionnera jamais dans un modèle comme dans l'autre, je pense plutôt que ce sont des discussions participatives, qui laissent la parole, où on ne prescrit pas, on ne juge pas mais on voit comment on fait, on apporte des éléments. Après ces saynètes on a apporté des fiches. Qu'est-ce qu'on pourrait imaginer. Comment le cadre légal protégerait Léa ou ses camarades. C'est une manière moins intrusive.

Après qu'est-ce qui fonctionne d'autre... le *Tubà* aussi en fait. Il y a un jeu qu'ils ont construit qui s'appelle « gare aux cookies ». C'est un jeu de cartes... jeu de plateau en fait. On peut jouer peu importe l'âge. Dans une séance d'essai, il y a deux petites vieilles qui sont venues qui habitaient dans le coin, qui sont venues avec leur panier légumes, il y a eu des jeunes, donc je pense que c'était intergénérationnel. Et je crois que c'était aussi une manière un peu ludique aussi de matcher les différentes situations. Je pense que ce qui fonctionne bien, c'est la mise en situation, en tout cas en termes de données personnelles, privées, ça fonctionne très bien.

Ma dernière question qui n'est pas véritablement une question, mais plus une ouverture : auriez-vous des recommandations pour améliorer cette acculturation aux données aujourd'hui ?

A différentes échelles, ce qu'il manque c'est les moyens en fait. Les moyens financiers, ça manque beaucoup : au *Tubà*, ou à d'autres structures mêmes qui souhaitent se lancer là-dedans. Ce ne sont pas les compétences humaines, car elles existent, il faut les rémunérer. Ce qui manque c'est les moyens financiers. Ma recommandation, c'est une prise en charge de ces dispositifs par l'État. Par exemple, à l'école, ça peut être introduit, cela

peut être soutenu. Tous les enseignants nous disaient « on n'a pas le temps dans les cours d'aborder ces sujets, ce n'est pas dans le programme, ce n'est pas prioritaire, donc on n'en parle pas ». Et pourtant on sait que les étudiants utilisent leur smartphone depuis l'âge de dix ans ou même avant. Donc ça créé des biais en fait entre des usagers et des objets techniques, des objets connectés, et par les dispositifs qui collectent les données. L'État pourrait introduire dans le programme scolaire des éléments là-dessus.

Universitaire 6 (01/07/22)

Comment définissez-vous la *data literacy* et est-ce que vous employez ce terme ?

Je l'utilise peu. Cela m'est arrivé de l'utiliser essentiellement avec certaines équipes de recherche ponctuellement lors de communications. Et pour moi, la *data literacy* finalement, c'est un des pans de la littératie numérique. Et pour moi, la *data literacy* c'est la transposition de ce que l'on trouve dans les littératies en général appliquées aux données, aux jeux de données, en sachant que ce que j'y vois toujours derrière *data literacy*, c'est ce qui est intangible avec les littératies en général et ce qui est bien entendu spécifique aux datas. Et pour moi dans la littératie c'est essentiellement des habiletés à écrire et lire les données, dans le cas des données. On est à la fois sur des modes scripturaux et à la fois sur des modes de lecture et d'appropriation à travers la lecture.

Et du coup, si vous n'employez pas le terme de *data literacy*, est-ce que vous employez un terme particulier ou pas du tout ?

Moi je revendique plutôt, en tout cas par rapport à mon parcours, la *data culture*, c'est-à-dire la culture des données. Mais on n'est pas sur le même registre : c'est-à-dire qu'on est sur quelque chose de plus générique qui est quelque chose justement de l'ordre de l'anthropologie politique ou les modalités de communication autour notamment de la *data literacy*. Mais c'est pas par synonyme : on n'est pas sur le même positionnement, on n'est pas sur les mêmes attentes.

Il existe tout un ensemble de concepts liés aux données : la gouvernance de la donnée, la *data science*... est-ce que vous arrivez à situer la *data literacy* dans cet écosystème ?

Pour moi, je dis pour moi car c'est très personnel, la *data literacy* est davantage sur l'approche compétences donc on n'est pas sur un niveau qui serait un niveau générique. Ce que je disais, la *data culture*, on est sur quelque chose de plus spécifique qui est quand même très voué à réfléchir aux incidences et aux formes de compétences. On est quand même sur un niveau opératoire parce que derrière il peut y avoir du référentiel, il peut y avoir des dispositifs d'accompagnement, de médiation ou pour la pratique des datas. On est sur un volet opératoire et très pragmatique finalement. Beaucoup plus pragmatique que la gouvernance des données et la culture de données, etc. Je suis dans le courant classique des compétences informationnelles, des pratiques informationnelles. Puis après, à travers ça, on en est arrivé aux littératies où on a, vous le savez mieux que moi, importé ce concept nord-américain. On s'est amusé à le traduire ou pas. Cela n'ouvre pas les mêmes champs : pour moi c'est très opératoire. Et après on a travaillé la génération des translittératies où on a essayé de voir

un peu la circulation des littératies. Et après est arrivé ultérieurement la *data literacy* encore une fois qui est autour du champ des compétences et de mise en œuvre autour des données, des jeux de données, des services liés aux données, etc.

Suite à ce que vous venez de dire, si on devait définir la nature de la *data literacy* : est-ce que ça serait une culture, un ensemble de pratiques, un état d'esprit, tout ça à la fois ?

A mon avis c'est plutôt un ensemble de pratiques, clairement. Selon moi, je suis vraiment dans le périmètre. Quand on commence à observer, bon dans notre cas nous c'est plutôt autour des question d'éducation : quand on entre dans des classes, des pédagogies de projet que peuvent avoir des lycéens, étudiants ou autres, voilà on va être en immersion dans ce qui se passe dans la classe, les consignes qui sont données, les méthodologies qui sont proposées, comment les acteurs que l'on observe en supposant que les élèves s'approprient ça, qu'est-ce qui échappe à ce qu'on avait prévu et pourtant se fait. On est vraiment dans les théories de l'activité et pratique. Encore une fois, je ne dis pas que c'est la bonne définition, surtout pas, mais c'est comme ça qu'on l'appréhende.

Comment vous définissez l'importance de la *data literacy* comparé à d'autres nécessités (*information literacy*, technicité informatique) ? Est-ce une nouvelle compétence qui doit s'ajouter au reste ou est-ce plus important que ce qu'il y a pu avoir auparavant ?

A mon avis, elle se rajoute au reste comme vous disiez, sachant que pour moi une des grandes questions c'est : qu'est-ce qui est commun, finalement, à travers des datas ? Qu'est-ce qui est commun du rapport que l'on a au document et à l'information en général et qu'est-ce qui en fait une spécificité, sur quel registre, quelle nature de contenu, etc.

Après la deuxième remarque qui me semble importante c'est : je pense que la *data literacy* et les incidences que l'on peut observer ont un impact très fort sur les métiers de l'information. C'est-à-dire que je pense qu'à partir du moment où on est entré autour de ces questions de jeux de données, avec lesquels on travaille, on manipule, etc. Je pense que ça repositionne toute une partie de l'activité des professionnels de l'information que sont les documentalistes, les bibliothécaires, les bibliothécaires spécialisés, etc.

Après, la troisième remarque c'est qu'à partir du moment où il y a quand même ce grand domaine de la libération des données, des données publiques, etc., ça a une incidence très forte pour le coup de citoyenneté numérique, c'est-à-dire que : bon certes, on met à disposition les jeux de données, mais après qu'est-ce qu'on en fait globalement quoi. Et comment on acquiert un peu un regard critique, comparatif, critique, un rapport distancié, réfléchit, etc., au regard des données que l'on libère et que l'on met à disposition. Parce que finalement, si on n'accompagne pas un petit peu en termes d'outils de lecture critique, etc., et alors ? On prend les données et on leur fait dire même le contraire de ce qu'elles affichent réellement. On est sur une dimension, pour le coup, de formation que l'on retrouve dans les littératies selon moi et qui est indispensable. Ce que je ne sais pas à ce jour, c'est l'incidence que ça a très réellement dans le champ social des acteurs, que ce soit le citoyen, etc.

J'en parlais avec des collègues par exemple de Lille. Avec certains collègues on se disait : on voit bien que des choses bougent et sont en train de changer. Est-ce qu'on les a clairement identifiées ? Pour le moment c'est un peu fluctuant. J'ai l'impression qu'il y a un discours un peu idéal de ce qu'on est supposé faire, etc., je ne suis pas sûr qu'on ait encore des éléments suffisamment fins et précis de voir l'incidence réelle sur, notamment comme je dis, le monde professionnel ou les acteurs, les citoyens, etc. Et après, derrière remarque, je trouve que l'on met en œuvre… finalement c'est marrant quand on voit les approches, le cycle de la data et toutes les schématisations, vous savez mieux que moi, vous vous amusez à transposer ça avec des cycles plus anciens du traitement de l'information, etc., et globalement on a remplacé « information » par « data » mais on n'a rien de nouveau globalement. J'exagère mais le nombre d'étapes, les vocables utilisés, la notion de cycle, etc., on est en train de transposer un peu. Voilà, ça c'est la grande question : est-ce une simple transposition des littératies ? Est-ce que c'est plus compliqué ? Est-ce qu'il y a l'émergence de nouvelles formes de littératies, de compétences derrière, etc. ? Moi j'ai le sentiment que oui mais faut-il arriver précisément à les repérer.

Est-ce vous arrivez à situer dans le temps depuis quand c'est devenu une nécessité ? Et si oui, est-ce qu'il y a des raisons spécifiques ?

Ce que j'ai constaté très modestement à partir de mes activités, c'est que finalement l'entrée je ne l'ai pas eu des collègues en sciences info-com et je l'ai pas eu par des injonctions dans le cadre des programmes d'enseignement par exemple de formation à l'université et tout ça. Je l'ai eu par une chaîne d'acteurs très particuliers que sont les auto entrepreneurs, etc., qui se sont petits à petits investis, qui se sont rapprochés des réseaux académiques que je fréquentais : réseaux académiques, rectorat, les directions du numérique, certaines entreprises de la région Aquitaine, etc. J'ai vu arriver des acteurs que je connaissais un peu et qui ont vraiment pris à bras le corps la question des datas. Et ils sont souvent arrivés avec des profils, c'était très particulier, de super experts qui étaient pas du tout académiques, c'est-à-dire dans le champ de l'éducation, ou de la chambre de commerce et d'industrie, etc. Et qui sont venus se positionner comme experts externes en proposant justement de l'accompagnement des formations, etc. Et là j'ai bien vu que finalement il y avait une nouvelle économie qui se mettaient œuvre, qui était une sorte d'économie intermédiaire entre des formations assez classiques que l'on avait dans des organismes accrédités, habilités, etc., et ces gens qui arrivaient et qui se positionnaient. Et ils se positionnaient avec un discours alternatif. Un discours un peu de rupture, de vraie nouveauté sociale, etc. Et ce que j'ai vu avec le temps, progressivement, même si ça leur fait pas plaisir, je leur ai souvent dit que finalement, à travers ce discours de l'auto entrepreneur, de la *new generation*, etc., finalement, au bout du compte, on reproduisait à la fois ce que je vous disais à la fois des cycles assez classiques de la formation autour de l'info et puis une nouvelle économie qui, quand même, était une économie qui restait quand même assez capitalistique parce qu'on avait beau être entrepreneur, il faut bien vivre et dégager un peu de bénéfice.

Et puis, petit à petit, je les ai vu commencer à proposer du clé en main, ils ont créé de vrais espaces de pouvoir à l'échelle par exemple des métropoles, etc. Et ils devenaient des incontournables, avec beaucoup de travaux, etc. Et ils ont compris très vite de la nécessité de s'appuyer sur les chercheurs aussi, sur les réseaux de chercheurs clairement parce que les chercheurs allaient nourrir un peu ce qu'ils faisaient. Cela leur donnait une sorte de crédibilité assise, etc., et ils ont compris aussi que dans les appels à projets et tout ça, il y avait de l'argent à faire. Alors le dire comme ça, ça peut donner l'impression que je suis très critique avec eux : pas du tout. Pas du tout car je trouve qu'ils ont joué un rôle de grande nécessité sociale et notamment d'articulation entre le

monde social, le monde de l'entreprise, et celui de l'université. Et en plus, ce sont souvent des personnes hyper volontaires, super motivées, c'est vraiment agréable de travailler avec eux. Mais si on met à distance la relation professionnelle, affective, etc., ce phénomène là m'a intéressé. Et on voit après que tout le monde s'est engouffré dedans : toutes les collectivités territoriales, etc. Et il y a une forme d'injonction de la data qui fait qu'après tout le monde a pris le train en marche parce qu'on jouait de la visibilité, on jouait aussi de la visibilité politique, etc.

Du coup je reviens juste sur le niveau temporel : c'est assez récent pour vous ?

Cette question, je pourrais vous renvoyer à des textes majeurs sur la libération des données : pour ça oui indirectement. Dans mon propre parcours et cheminement, c'était vraiment au niveau des auto entrepreneurs, c'était une petite dizaine d'années je dirais. Depuis 2012, quelque chose comme ça. Mais les premiers travaux je dirais une dizaine d'année, à peu près 2012. Et puis il y a eu un autre évènement mais ça vous pourrez repérer, avec la création de la cité du numérique à Bègles. Il y avait tout un volet numérique : à l'intérieur, il y avait toute cette question de travailler avec les données et tout ça, de faire travailler les auto entrepreneurs en lien avec des projets municipaux. Donc une dizaine d'année je dirais.

Sur l'objet central qu'est la donnée, je vais vous poser une question classique : qu'est-ce que la donnée ?

Alors moi il se trouve que dans mon parcours, j'ai souvent travaillé régulièrement avec les informaticiens, bien que je le dis clairement, je n'ai aucune compétence informatique. Je suis un super usager mais pas plus, clairement. Je suis en sciences info-com, avec une dimension anthropo-sociologie de l'usage très forte. Jusqu'à ma thèse, j'étais en pure sociologie anthropologie, j'ai fait un double cursus là-dedans. Donc comment je la définis : pour moi, il n'y a pas de synonymie entre le schéma classique donnée-information-connaissance-savoir, je distingue clairement les quatre niveaux. A la différence de ce que font beaucoup de mes collègues, pour moi la donnée ce n'est vraiment pas un synonyme d'information. C'est-à-dire que je reviens à un pan de l'informatique d'origine : pour moi, la donnée seule, elle est hyper sèche. Et finalement, c'est par l'agrégation de données qu'on commence à avoir ces fameux jeux de données, et on n'est pas encore dans l'information puisque pour moi l'information, il y a un mode d'éditorialisation derrière : la manière d'agréger des jeux de données, la manière éventuellement d'expliciter, la manière de scénariser, etc, fait qu'on en vient à parler d'information. Voilà comment je me situe clairement. Après j'ai des collègues en informatique quand ils me parlent de données, pour eux ce sont déjà des éléments structurés et affichés, des modes d'affichage, parfois même *designés*, etc. Pour moi, pas du tout.

La donnée c'est la source de l'information et on passe par la phase d'agrégation de données pour arriver à créer une information. Après la chaîne classique : c'est-à-dire qu'une information, même en circulation sociale, ne constitue pas une connaissance. Pour qu'il y ait connaissance, il faut qu'il y ait pour le coup le facteur humain : c'est-à-dire qu'il faut qu'il y ait des modes d'appropriation, de circulation des savoirs, etc. Et ensuite ça devient vraiment un savoir à partir du moment où il y a ce vocable qu'on retrouve beaucoup maintenant depuis quelques temps. Et moi ça m'a fait rire car j'utilisais l'idée depuis très longtemps de ce qu'on appelle l'institutionnalisation : pour moi un savoir il rentre bien par le processus d'institutionnalisation, c'est-à-dire

la mise en place de programmes, d'experts, de modes de validation, de référentiels, etc, qui fait finalement qu'une connaissance peut être non institutionnalisée et socialement échappée, ou alors elle devient un savoir dans le sens où il y a une forme d'académisme, d'autorité et donc automatiquement de transfert aux autres, de formation, etc.

Quels sont pour vous les enjeux actuels autour des données ? Et comment la data literacy peut y répondre ?

Pour moi, j'allais trois mais non, il y a deux dimensions. Il y a d'abord une grosse dimension technique derrière, c'est-à-dire d'abord être capable de retrouver ces jeux de données, être capable de les travailler, de les comparer… oui, de les travailler à travers des outils classiques, etc. Et puis effectivement, travailler aussi les modes d'affichage, etc. Donc vraiment il y a de la compétence technique, je vais dire outils, logiciels, pour travailler ça. Ça c'est indéniable, pour moi c'est dans la *literacy*.

Et le deuxième volet après, c'est, après avoir fait tout ça, après avoir travaillé ces données, être capable justement de les comparer et d'être capable de repérer les systèmes d'intention. La genèse des données mises à disposition et finalement comment elles ont été agrégées, mise à disposition. C'est là que pour moi il y a une dimension vraiment analyse critique distanciée des jeux de données. Je vais prendre un exemple tout bête : les municipalités qui veulent mettre à disposition des jeux de données sur leur consommation d'eau par exemple pour montrer qu'ils sont extrêmement vertueux en termes d'environnement, etc., donc voilà : interroger tous ces schémas-là. Qui les met à disposition, sur quelles bases ? Qu'est-ce qu'on a mis dans la consommation d'eau ? La relativité du jeu de données que je vais mettre, c'est-à-dire de pondérer au prorata de la population, par exemple de la commune, etc. Voilà, il y a tout ce travail de « retrouvage » des données, de le travailler, de le mettre en perspective, de le comparer et de chercher les systèmes cachés. Pour moi, ça fait partie quand même de la littératie de données, effectivement c'est ce que je vais dire : une formation analytique et critique des données qui sont mises à disposition que je récupère, que je travaille. Et donc on est sur ce double niveau : compétences et analyse critique.

Et est-ce qu'il y a, à travers les enjeux que vous venez de citer, des types de données qui en sont au cœur ?

Je les ai citées : plutôt les données publiques. Mais ça c'est en fonction de ce que j'ai pu faire et ce sur quoi je travaille. Que ce soit en éducation, en santé. Là je vois on est en train de construire des réseaux par exemple en francophonie. Ce qui m'intéresse aussi, c'est de voir justement comment, par exemple, des jeux de données vont être reçus dans la sphère francophone différemment en fonction du filtre culturel des usagers. C'est essentiellement ça. Je ne dis pas que c'est que ça encore une fois mais vous m'interrogez sur ma propre expérience.

Comment ressentez-vous sur vos pratiques le cadre législatif ?

Toujours pareil, ça dépend la sphère qu'on observe. Dans celui de la formation et de l'éducation, je trouve que plus il y a une législation et plus on est dans un rapport justement qui brille de l'usage, c'est-à-dire on met en avant la sécurité personnelle, la sécurité des systèmes, la sécurité de l'environnement professionnel dans lequel

est la personne qu'on forme, etc. Alors autant je suis pour ne pas laisser un monde sans règles et sans lois et qui devient extrêmement concurrentiel, voire sauvage, etc., le problème c'est qu'on entre par la contrainte.

Par contre, je trouve qu'il y a un effort européen qui mérite d'être souligné. C'est-à-dire à la fois de légiférer pour mettre de l'ordre dans tout ça et en comptant beaucoup sur, justement, la formation, l'accompagnement, ce que les acteurs vont en faire progressivement, etc. Après la question, voilà, je pense qu'il faut bien distinguer les niveaux d'expertise des scientifiques, ceux des médiateurs dont je parlais : les documentalistes, etc. Et puis le citoyen ordinaire. Il y a du boulot si j'ose dire. Je ne suis pas sûr que ce je dis soit partagé de tous. Mais globalement, il y a du bon sens dans les textes. Et au niveau européen, ce que fait l'état français aussi, c'est plutôt assez courageux, un peu inconscient quand on voit l'ampleur de la tâche, mais il n'empêche que les lignes bougent, et notamment grâce à l'Europe et la France.

Comment peut-on mettre en place une culture des données dans une organisation et est-ce que ça doit concerner tout le monde ?

Je crois qu'il y a trois choses qui sont indispensables finalement, qu'on soit un jeune étudiant, un élève, un professionnel, etc. Je pense que déjà c'est montrer le champ des possibles, c'est-à-dire vraiment dresser la cartographie de l'offre telle qu'elle est, qu'on soit dans un secteur d'activité ou un autre. Je vais prendre un exemple : j'ai une étudiante qui est en responsabilité, elle est responsable d'un institut de formation des infirmiers. On est dans le secteur de la formation et de la santé, je pense que quand on forme nos futurs infirmiers, il faut les sensibiliser aux jeux lmassifs de données qui circulent autour de la santé publique. Finalement ça concerne un enseignant dans son domaine, une infirmière, un futur technicien ou ingénieur en aéronautique ou autre : c'est d'être capable, dans le secteur où l'on nous forme, de montrer ce fameux champ des possibles : qu'est-ce qui existe, quels sont les gisements, comment ils sont affichés, interrogeables, téléchargeables, etc. Cela me semble indispensable parce qu'il y a beaucoup de fantasmes sur ce qu'on est censé trouver et pas trouver, où des formes de médiatisation de la donnée qui seraient affichées avec ces cartographies, ces affichages désignés, etc., et des fois c'est hyper sec, difficile à trouver, très hétérogène, etc.

Donc déjà il y a être capable dans un cycle de formation par exemple, etc, de montrer ce qui existe à l'instant T et comment on peut s'y prendre pour veiller sur ce qui sort et va sortir à l'avenir. Ça c'est le premier volet. Après, je pense que c'est indispensable effectivement, je reviens à ce que je disais tout à l'heure, de former un minimum à quelques outils qui permettent de travailler la donnée, de l'extraire, etc. Et puis après j'en reviens aussi à ce que je vous disais c'est de progressivement instiller cette capacité à chercher, à comparer, à confronter. Je ne vais pas raconter ma vie mais par exemple depuis quelques mois une expérience de vie personnelle fait que je suis en train de m'intéresser à une question médicale que je ne connaissais pas il y a quelques mois. Je ne savais pas que ça existait tout simplement parce que je suis en train de vivre ceci à titre personnel. Mais qu'est-ce que je fais actuellement ? je suis en train d'essayer, sans y passer trop de temps, d'aller chercher des données médicales les plus fines possibles, sur des études randomisées dans le monde anglo-saxon ou le monde francophone, et d'essayer de voir finalement si vous voyez justement des données mises à disposition par un laboratoire sont confirmées par un autre. Avec cette idée que ça va me permettre moi-même de mieux comprendre ce qui m'arrive mais aussi de mieux comprendre comment il faut que je compare les résultats qui sont rendus publics par ces organismes, sérieux, habilités, pour m'aider moi-même à faire un choix finalement dans

des protocoles. Et cela, à travers une expérience de vie comme ça, on va chercher des données. Bon, il y a de l'affichage, il y a de l'histogramme, il y a du pourcentage, il y a des littératures plus ou moins accessibles, des fois des visualisations, etc. Mais c'est extrêmement compliqué parce on ne peut pas comparer les carottes et les pommes de terre. Donc ça oblige à avoir les mêmes bases. Mais faire ça, cela veut dire des niveaux de capacité à chercher l'info, à sélectionner, la fiabilité. Et les données, pour y revenir spécifiquement, cela me semble particulièrement compliqué parce qu'elles ne sont pas toujours explicitées, elles ne sont pas analysées et donc on prend des données et on peut leur faire dire à peu près tout et leur contraire. Et cette posture-là, à mon avis, on est au cœur de la *data literacy*. Et on est au cœur de la problématique de la formation qui est effectivement nécessaire, quel que soit le secteur d'activité. Parce que sinon on choppe à peu près tout ce qu'on veut, tout et son contraire, et cela peut être même très préjudiciable en fonction des domaines d'activité. Pour moi on y viendra. Et là, vous voyez ce que j'essaye de vous dire, on est bien sur, encore une fois, les données et les jeux de données. Si on est sur du document, parce que des fois j'ai vu des communications où franchement on n'était pas dans la data, on était dans le rapport au document et à l'information. Quand on commence déjà à être sur des extractions de données qui sont citées, sourcées, mais où il y a un écrit, un article, une analyse, on est déjà sur autre chose, on est sur la documentarisation des jeux de données et sur de la comparaison de documents, de la recherche documentaire. On est vraiment en amont pour moi.

Vous avez un peu répondu à la question suivante qui est : que doivent développer les personnes qui ont été acculturées à la donnée, et si je résume ce que vous avez dit c'est d'abord une posture intellectuelle vis-à-vis de la donnée, augmentée après de pratiques...

Oui et non, en filigrane il y a cette question de la formation analytique et critique mais qui ne peut pas faire l'économie de compétences techniques. Parce que finalement, par la non compétence technique, on passe à côté et donc c'est indissociable. Et je ne veux pas faire trop long mais le problème que l'on a c'est que, volontairement je tire le trait pour bien me faire comprendre, on a trois postures : on a une première posture de gens, de formateurs, d'acteurs, de chefs de service, qui ne mesurent pas les enjeux. Donc effectivement par l'ignorance on n'en parle pas et on continue de reconduire ce qu'on fait habituellement. Ou alors comme je vous disais on prend tout et on diffuse sans réfléchir.

Les deux autres postures, c'est celle d'avoir cette dimension critique mais être tellement déficitaire sur la compétence technique qu'au bout du compte on construit un discours qui est très éloigné des possibilités. Ou alors l'autre entrée, vous me voyez venir, elle est hyper technique, c'est de la maîtrise d'outils, on joue avec les jeux de données, on éditorialise, mais on ne réfléchit pas à ce qu'on fait non plus forcément. Parce que finalement on en vient à jouer avec des jeux de données mais encore une fois sans avoir cette dimension analytique. Et donc c'est indissociable les deux.

Et c'est pour ça qu'à mon avis, une anecdote, on vous l'a peut-être déjà donné, nous on suivait dans un petit projet qu'on avait financé du *CNRS*, on avait regardé un programme d'enseignement qui était les sciences économiques et sociales. Dans les sciences économiques et sociales, donc en lycée, est arrivé dans le programme de *SES*, un référentiel programme : du jour au lendemain, on a imposé aux professeurs de *SES*, ça remonte à 7/8 ans peut être, où on a imposé l'injonction de traiter les données, les jeux de données, en économie bien entendu. Il fallait former les lycéens aux jeux de données. Et les professeurs en sciences économiques et sociales ont été

en détresse absolue. C'est-à-dire que quand ils ont vu ça arriver, eux-mêmes ils ne savaient pas ce qu'étaient les jeux de données. Ou alors ils en utilisaient, mais des jeux de données qui étaient un peu anciens parce qu'édités dans des grands répertoires de l'*INSEE* par exemple. Voilà le type de problèmes. Donc globalement à l'époque quand on a regardé ça, il y avait soit des professeurs qui avaient complètement écarté le truc : on fait comme-ci on ne voit pas. Justement, on transpose ce qu'on faisait de manière traditionnelle à travers le document, l'article, etc. Et ceux qui ont essayé de prendre le truc à bras le corps, eh bien clairement ils ont accompagné les classes en apprenant eux-mêmes. Et c'était aussi très intéressant de le voir. Alors petit à petit, après bien entendu les enseignants, l'institution s'était organisée pour commencer un peu à former. Et les collègues qui ont pris ça à bras le corps ont été obligés de monter en compétences techniques. Mais ce qui est intéressant c'est que vu que ce sont des spécialistes en économie, cette dimension analyse critique, ils l'avaient bien entendu, de manière traditionnelle. C'est hyper intéressant parce qu'on voit que des fois, les injonctions, les textes liés à une modernité, du moment, si j'ose dire, ne réfléchissent pas à l'incidence que ça a sur celui qui doit s'initier lui-même, se former parce qu'il travaille dans un collectif.

Il y a un terme qui revient souvent, c'est celui d'esprit critique, être critique vis-à-vis de la donnée, est-ce que vous vous mettez quelque chose derrière ça ?

C'est vrai que maintenant, je partage le diagnostic, il y a un retour de l'esprit critique partout. Je travaille avec des gens qui sont plutôt spécialisés dans l'esprit critique, parce qu'ils viennent de la philosophie, de l'info-philosophie, de l'éducation, ou du courant des humanités. Après, plus que d'esprit critique, moi je parlerais d'une dimension critique. C'est-à-dire, ce que j'aime bien rappeler dans la dimension critique, c'est la posture du doute dans ce que l'on fait. Et de pas être séduit par un principe de facilité parce que je le trouve facilement, parce que ça réconforte mon système de croyances, de valeurs que j'ai. Donc je vais prendre plutôt ce qui ne fait que renforcer ce que je sais déjà et qui me donne plus d'assises dans mes certitudes. Pour moi, plus que l'esprit critique finalement, c'est le doute et c'est le système de confiance que je vais attribuer aux données que je vais utiliser, manipuler et les gisements que je vais m'approprier. C'est-à-dire jusqu'où je fais confiance, comment je bâtis mon système de confiance par rapport à ce que je vais trouver et bien entendu c'est là, si j'ai le doute, que je relativise mon système de confiance : c'est là qu'à mon avis je suis dans la dimension critique.

Après je ne suis pas spécialiste de ce qu'est l'esprit critique mais je dirais plutôt : confiance, défiance et posture du doute. Mais dans le sens positif, ce n'est pas de se mettre en rupture mais d'aller comparer, chercher. Là dans l'exemple que je vous donnais sur les protocoles médicaux, j'ai trouvé deux études, une quarantaine de cobayes, etc. Et ce qu'il y a d'hyper intéressant c'est que les deux ce sont des études internationales référencées, tout ce qu'il y a de plus sérieux. Mais au final, elles arrivent à des conclusions totalement différentes. Donc ça veut dire quand même quand on a ce doute et qu'on va comparer en essayant de voir derrière qu'il y ait pas trop d'enjeux de vendre tel produit, on voit que les études les plus sérieuses sur des éléments un peu inexplorés, au bout du compte, peuvent arriver à des conclusions qui sont presque contraires. Et ce n'est pas que l'une est plus sérieuse que l'autre, puisque référencée partout, ce sont des équipes robustes. Mais c'est tout simplement qu'on est sur des terrains instables. Et donc le but n'est pas d'instiller la certitude, c'est justement bien ce que je vous disais de garder le doute et la comparaison et de soupeser. On est dans la modération.

Arrivez-vous à voir l'évolution en termes de connaissance et de compétences de votre public par rapport à la culture des données ?

Je vois pas trop non. Pas pour l'instant. Et je ne vois pas le retour d'investissement pour tout ce que l'on a fait bien avant, notamment sur justement les pratiques informationnelles puisque je vous rappelle que la plupart sont passés par les collèges, les lycéens, ils ont été initiés. Et franchement, je vois pas de niveaux d'expertise forcément plus distanciés. Je suis un peu pessimiste là-dessus. Par contre ce que je vois quand j'ai des étudiants de master, voire qui commencent en écoles doctorales, là oui. Là je trouve que quand même il y a un niveau de compétences un peu plus élevé. Mais seulement, on est sur des micro-populations de jeunes qui ont fait des vrais parcours du combattant. Et qui en plus, sont dans le domaine des sciences de l'info-com et sont assez, on peut l'espérer, assez robustes et ils ont au moins fait le choix de cette discipline. Donc dans leur esprit, quand même, ça mouline un petit peu et ils sont assez sensibles, vraisemblablement, aux comparaisons, confrontations.

Mais globalement après sur la masse de publics plus larges, non. Et je vois souvent des travers majeurs. J'ai par exemple un enseignement, de niveau master M1, sur la communication scientifique et alors moi je m'amuse à les tester là-dessus, quels sont leurs critères de recherche de la scientificité, les grands organismes qui publient de la donnée, etc. Je travaille et sur le document, et sur la donnée, parce que l'objet ce n'est pas que la donnée, c'est donnée et document. Et pour tous les grands pièges, j'ai les trois quarts des groupes qui tombent à pieds joints dedans. Donc ça montre que quand même, ce n'est pas du tout maîtrisé par des étudiants qui sont en M1. Alors après on les forme pendant six mois et je vois, il faut six mois pour qu'ils soient dans ce doute, dans cette confrontation, etde ne pas se laisser embarquer par un joli *design* ou un joli accompagnement qui en plus est assorti d'un grand spécialiste. Et de voir la traçabilité des données qui sont produites. Et on y arrive. Mais il faut du temps. Et je ne suis pas sûr qu'à l'issue de ces six mois, ils passent à autre chose : ils ne me revoient plus et la question que je me suis posée c'est « est-ce qu'ils réinvestissent ceci dans d'autres contextes ? » ça je ne peux pas le dire, je n'en sais rien.

Donc comme je leur dis à ces étudiants, on va apprendre pendant six mois à être des sortes d'enquêteurs. On rentre par des thématiques, on rentre par des profils de scientifiques. On rentre aussi par des objets qui les intéressent particulièrement puisqu'ils vont construire un mémoire et je rentre par ces trois niveaux-là. Et j'essaie de déconstruire progressivement avec eux mais il y a des grands travers. Je vois beaucoup mes collègues qui travaillent le *design* de la donnée, et je trouve ça super. Mais il ne faut pas oublier que le *design* dans sa forme générique induit un maximum de biais cognitifs et de biais d'appropriation. Donc c'est intéressant le *design*, l'approche *design*, c'est nécessaire, y compris en termes d'utilisabilité. Mais ça créé un biais de séduction et d'attirance qui est hyper piégeant parce qu'on peut refourguer de la donnée hyper superficielle, voire manipulée, et avec un bel affichage *designé* ça passe quoi.

Vous l'avez souligné à plusieurs reprises : sur les difficultés de la mise en place de cette *literacy*, voyez-vous en particulier ?

Encore une fois, au départ je suis du domaine de l'éducation. Le principal obstacle, c'est que l'institution ne comprend pas. Elle ne comprend pas un double niveau, elle ne comprend pas comment s'organise le monde

professionnel et elle ne comprend pas les enjeux que ça a notamment en termes d'employabilité sur les années à venir, sur le prospectif finalement. Pour moi, c'est le principal filtre. Par exemple j'ai des responsabilités dans un institut national de formation des enseignants, l'*INSPE* : pour l'instant on affiche culture numérique, que du numérique, du projet numérique. Je ne dis pas que la donnée n'entre pas dans la formation mais elle entre en fonction du groupe auquel on appartient et du formateur qu'on va avoir.

Alors maintenant, je n'attaque personne à travers ce que je veux dire mais on voit maintenant que certains ont bien choppé la chose : le data journalisme par exemple. Bien entendu c'est attirant mais j'ose espérer qu'à la sortie on aura des journalistes qui vont davantage mobiliser les jeux de données pour mettre en discussion la qualité du traitement informationnel. Si c'est pour faire de l'affichage infographique de type BFM, bonjour les dégâts. Ce que je vois là-dessus quand vous mettez quand même la dimension économique, quand on voit quand même comment les agences de presse originelles ont disparu dans le temps, de moins en moins de reporters de terrain, on a vite fait de faire du journalisme sur des jeux de données qui sont disponibles, qu'on va mettre en confrontation. Mais si vous avez des data journalistes qui ne sont pas en mesure d'avoir du temps pour fouiller, pour comparer, et qu'il faut travailler dans l'urgence, trouver des données qui vont travailler un discours, c'est dévastateur. Et ça pour le coup c'est un vrai enjeu de société. Et pour l'instant il n'y a pas de data professeur ou de data professeur des écoles, ce n'est pas dans les tuyaux. Mais pour le coup l'éducation, elle est totalement absente là-dessus. Et quand elle le décrète c'est dans les programmes mais après il n'y a pas cette étape de la traduction comme je le disais tout à l'heure pour les *SES*. Voilà pour moi les principaux obstacles.

Avez-vous en tête un projet en *data literacy* qui a bien fonctionné et pourquoi ?

Le projet que je trouve intéressant, c'est la suite de ce qu'on avait fait sur les *data cultures*. Ces *data cultures* où on regarde, c'est dans le cadre, pour que vous puissiez repérer, c'était un *GTNUM* qui est sur les *data cultures* qui associe Rennes, Bordeaux, etc., et on travaille justement sur l'observation qui se fait en classe,. Ça vraiment je trouve que c'est une belle recherche et ça ferraille beaucoup entre nous parce que justement c'est ce que je disais, en fonction des profils de chercheurs, on a quelques chercheurs finalement qui ne connaissent pas la classe et ce qu'on y fait, et qui transposent ce qu'ils ont appris en documentation à la donnée. Et ça c'est les deux biais majeurs.

Une fois qu'on a passé cette étape-là on voit bien qu'il y a des dynamiques qui se mettent en œuvre, on voit bien qu'il y a des réseaux de soutien derrière, et ces réseaux de soutien de type *Canopé*, ils ont besoin aussi des données pour le coup scientifiques qu'on recueille pour pouvoir éventuellement après envisager de la production de ressources, de méthodes,je trouve que ça marche assez bien. Ce qu'on fait là, je le dis d'autant plus que je ne suis pas le porteur de ça, ce que je trouve d'intéressant dans cette recherche aussi qu'on ne voit pas souvent ou pas du tout, c'est vraiment un objet transactionnel et commun, c'est-à-dire que l'air de rien ça fédère un maximum d'acteurs avec des profils totalement différents : il y a du monde de l'entreprise, il y a du monde de la production, d'éditeurs, il y a des acteurs de l'école, il y a des universitaires, et l'air de rien, ça boucle avec ce que je vous disais au début. On sent bien qu'il y a des choses qui se font, on sent bien qu'il y a des choses qui nous échappent, on sent bien que ça a des incidences sur notre manière de faire et d'accompagner : ce qu'il y a d'intéressant c'est ça. On choppe en quelque sorte l'activité à chaud.

Pour conclure, avez-vous des recommandations pour améliorer l'acculturation aux données en France ?

Alors moi je me suis souvent refusé à fonctionner sur des préconisations. Je n'ai pas de préconisations malheureusement pour vous « il faut faire ça, ça, ça » parce que derrière je n'ai pas de dogmatisme et ça reste très fragile de toute façon en plus parce que les études qu'on fait en plus on nous ouvre des terrains. Ce sont des terrains où on fait des choses, où il y a des acteurs mobilisés, où il y a vraiment des pépites de collègues hyper actifs qui ont envie d'innover. Mais ça n'est pas le quotidien des élèves. Je suis entouré de jeunes, de lycéens ou de jeunes étudiants quand je discute avec eux, là dans ma sphère personnelle. Ils ne me racontent pas qu'ils ont été initiés avec une progression pédagogique sur plusieurs mois, sur plusieurs semaines sur les datas, clairement. Et par contre quand ils vont chercher de la data, c'est en catastrophe, quand ils ont un dossier à rendre pour la semaine prochaine et qu'ils se disent que les datas c'est pas mal, ça habille et ça donnera de la scientificité, vous me voyez venir.

Donc non, je n'ai pas de préconisation. Et puis en plus ce n'est pas trop dans mon registre. Par contre, je crois que véritablement, il faut par tous les moyens arriver à sensibiliser aux enjeux, à montrer un peu le potentiel qui se dessine, et tous les biais qui sont en jeu. Et ça c'est vrai que c'est là-dessus que j'ai mon système d'alerte, c'est-à-dire, ce n'est pas en lien avec les données je m'en excuse mais c'est pour bien me faire comprendre : hier j'étais dans un comité avec des gens importants qui décident de l'avenir d'un organisme universitaire sur les filières *SNI* et il y a quelqu'un qui rappelait, j'étais tout à fait d'accord avec ça : quand on ouvre une filière de Service Numérique Informatique, bien entendu que c'est du numérique et de l'informatique, mais ce n'est pas que de l'opératoire, c'est-à-dire que derrière on a rappelé qu'il y avait cette dimension critique, culturelle, méthodologique, qui est bien plus que la seule technicité. C'était pour une discipline, la mise en place de *SNI*, mais ceci reste valable, selon moi, autour de la *data literacy.*

La *data literacy* bien expliquée serait de montrer, j'aime bien cette idée des biais qui sont connectifs, des biais de lecture, de compréhension, etc. Et après qu'on sensibilise : c'est vraiment se demander comment on l'instille, comment on le met en œuvre, quand on forme des infirmières, des étudiants, des futurs journalistes, etc. Et j'en reviens à ça, si le data journalisme est axé là-dessus, il n' y a pas de soucis. Si c'est accaparé parce que ça draine du public et que ça va faire gagner du temps aux chaînes et aux médias, là on a tout faux et ça sera même dévastateur. Donc voilà mes préconisations si j'ose dire.

Universitaire 7 (11/07/22)

Comment définis-tu la *data literacy* et emploies-tu ce terme ?

Data literacy est-ce que je l'utilise ? Oui, je pense que je peux l'utiliser. On parle plus de compétences pour la donnée et je parle peut-être plus aussi d'intégration de la donnée dans nos modes de raisonnement et nos prises de décision. C'est peut-être ça qui me motive dans tous les cas : penser qu'il faut monter en compétences sur

la donnée. Donc comment je la définis cette *data literacy :* ce sont les compétences, à la fois pour intégrer les données dans nos prises de décision, pour nos activités, et puis en même temps, pour être acculturé à lire.

Lire d'abord d'un point de vue plus « grand public », à intégrer la lecture des livrables sous forme de données, ce qui n'est pas toujours encore évident. Et du côté des professionnels, cela serait plus être capable d'intégrer la donnée comme le document dans tout ce qui relève de fonctions transverses pour accompagner les autres processus, que ce soit la décision, l'éclairage, le risque, l'anticipation. Je crois qu'on parle autant de la donnée aujourd'hui parce qu'on n'a jamais eu autant de données dues à la numérisation de la société et de toutes ses activités. Donc je pense que la définition est variable en fonction du public cible. Si c'est le public du « grand public », cela sera plus une forme de culture, d'intégration du raisonnement , de lecture des livrables de la data et de recours à la data pour s'orienter dans nos activités. Et du côté des professionnels conformes à notre organisation, cela serait plutôt une montée en puissance des compétences concernant le traitement de la donnée mais cela on y reviendra certainement après.

On a une panoplie de termes qu'on pourrait intégrer dans un écosystème de données comme la *data science*, la gouvernance des données. Est-ce que tu arrives à situer la *data literacy* dans cet écosystème ?

Alors elle n'est pas au même niveau que data gouvernance où on est déjà dans les métiers qui ont affaire à la donnée. Quand tu parles de *data literacy* c'est comme l'alphabétisation des sociétés. C'est plus un niveau, un objectif d'éducation finalement : un objectif de maîtrise du socle de compétences de base et cela relève plus de ce niveau-là pour moi la *literacy* quand on parle du grand public. Après encore une fois, dans la sphère professionnelle, c'est un peu différent selon moi : la *data literacy* c'est être capable de maîtriser les outils et de maîtriser les types de traitements intellectuels qui sont statistiques, qui sont aussi qualitatifs de la donnée.

La *data literacy* en fonction du contexte, c'est de toute façon à chaque fois des compétences particulières : pour l'éducation, cela sera plus lié au fait de mobiliser, d'aller lire de façon aussi critique. Et pour les professionnels c'est monter en puissance. Ce qui est différent essentiellement c'est aussi par rapport au document. Il est clair que les statistiques, l'informatique ainsi que les dimensions éthiques et de droit sont quand même très importantes dans les entreprises. Notamment, ces montées en puissance, ces types de compétences par rapport au document.

Je vais rebondir sur ça en te demandant, et je prends des guillemets quand je dis ça : si on devait définir une « nature » de la *data literacy*, quelle serait-elle ? Des compétences, un état d'esprit, une nature ?

Je crois que c'est très flou car ça dépend des contextes. Pour moi c'est vraiment très contextuel. Mais dans tous les cas je pense que c'est peut-être globalement, à un niveau macro, le retour à une forme d'objectivation des sociétés contemporaines du fait qu'il y a de plus en plus de données. On a vécu dans l'ère des discours et on continue à y vivre bien sûr, mais je pense que c'est une forme de retour à l'objectivation et à des données statistiques et factuelles qui quelque part contrebalancent des prises de positions parfois très idéologiques. Je pense qu'on a un balancier qui revient au centre et qui tient de l'idée qu'on peut de plus en plus avoir recours

à la *data literacy* pour cadrer, pour gouverner, parce qu'on a de plus en plus de données. Et je pense qu'il faut les mettre en balance avec des raisonnements plus qualitatifs et politiques.

Donc je pense que c'est ce mouvement global et massif du retour des données statistiques et factuelles qui marque la *data literacy* dans son ensemble. Quel que soit le métier d'ailleurs : que ce soient les journalistes, même si ce n'est pas toujours vérifié, que ce soient les dirigeants qui embauchent de plus en plus, qui créent des entreprises, créent des *fablabs*, des *fabdatas* plutôt. Des *fablabs data*, il y en a de plus en plus de façon transversale : on le voit avec les étudiants qu'on a dans les entreprises. Après en fonction des contextes cela sera plus une visée de performance : l'entreprise cherche toujours à être plus performante. Il y a des enjeux qui sont différents mais là aussi tu vas peut-être m'interroger là-dessus.

En termes d'importance, est-ce que tu estimes que c'est devenu central de s'acculturer aux données ou bien est-ce que c'est une compétence comme une autre ?

Pour moi, la *data literacy* est intégrée globalement dans la *digital literacy*: si on a des données, c'est parce qu'on a du numérique. C'est le côté un peu science dure des littératies contemporaines. Après, tout est un petit peu connecté : dans le cas de la littératie informationnelle, l'information se diversifie de plus en plus et les data font partie de l'information aujourd'hui. C'est une proximité même si quand on parlait de l'*information literacy* on faisait plus référence aux textes, aux documents écrits, alors qu'avec les data c'est plus factuel. Mais je crois qu'il y a une convergence maintenant : on le voit bien dans les processus qu'on avait l'habitude de traiter en info doc sur la veille ou l'archivage. Tout le monde se pose la question de l'information textuelle et de la data aujourd'hui de façon simultanée. Il y a une forme de convergence qui est en marche, même s'il y a des niveaux de granularité de la data qui sont quand même variables. Redis-moi le centrage de ta question que je me perde pas.

C'était au niveau de l'importance de l'acculturation aux données aujourd'hui. Si c'est plus important que le reste ou au même niveau.

Cela reste quand même moins important pour moi que l'*information literacy*. Nous sommes toujours dans des sociétés qui sont, et heureusement, pilotées par l'argumentation, la discussion, le texte et même la data est convoquée dans ces domaines-là. La *digital literacy*, elle est fondamentale aujourd'hui dans nos sociétés parce qu'il y a de plus en plus de services numériques. Donc la fracture numérique est mal vécue par beaucoup de gens : cela les isole considérablement. Je dirais qu'aujourd'hui, s'il fallait mettre une graduation entre tout ça, il faudrait quand même mettre l'*information* et la *digital literacy,* qui sont beaucoup globales, au-dessus de la *data literacy* qui me semble quelque part incluse dans les deux autres. Elle est en-dessous dans tous les cas.

Arrives-tu, à un niveau temporel, à définir depuis quand c'est devenu une nécessité ?

Je pense que ce n'est pas si vieux que ça même si il y a des choses que l'on redécouvre. Je pense que la politique publique sur tout ce qui est *open data*, tout ce qui relève des données de la recherche, tout ce qui est transparence, ce sont les moments forts qui ont fait qu'on peut dire aujourd'hui que c'est peut-être à partir de l'émergence de ces mouvements-là que la data est redevenue très importante. Ainsi que le problème

de la souveraineté de la data. Je pense que ça date d'une petite dizaine d'années à peu près. On redécouvre l'importance des données internes avec la *BI* qui elle est beaucoup plus ancienne dans les entreprises : valoriser les données internes de l'entreprise, on n'a pas su beaucoup le faire jusqu'à présent mais c'était un sujet la *BI*. L'informatique décisionnelle existe depuis les années 80. Mais en même temps, je pense que la technologie fait que les sujets deviennent de plus en plus importants. La technologie est de plus en plus mûre pour traiter ces données, les stocker. De nouvelles architectures sont là. On a quand même les *data warehouses* qui commencent à s'implémenter un petit peu partout dans les entreprises donc je ne pense pas qu'on peut dissocier l'importance du sujet sans sa corrélation avec la maturité des technologies et l'accompagnement du discours politique. On est toujours la résultante de quelque chose.

Je pense que les forces qui ont consolidé la *data literacy* viennent du politique, viennent de l'économie, de la techno. Dans les sociétés numériques, le « capital matériel » entre guillemets est quand même maintenant aussi important pour avoir l'avantage concurrentiel. Les entreprises du numérique et leur importance aujourd'hui ont accompagné aussi ce poids de la data. Selon moi, cela ne naît pas de rien : cela naît de forces conjoncturelles fortes qu'il faut trouver au croisement de l'économie, du politique et de la maturité des technos.

Si on se centre sur l'objet central de la *data literacy* qui est la donnée, la première question que je vais te poser c'est comment tu définis la donnée ?

Ça, c'est une bonne question. C'est vraiment un fourre-tout. C'est aujourd'hui devenu tout. On est bien d'accord : il y a là aussi un mouvement de fond qui fait que la donnée est à la fois structurée, textuelle, non structurée, non textuelle, etc. C'est du numérique aujourd'hui, même si on peut le regretter parce que tout est mis au même niveau dans une forme de [son inaudible] un peu général mais c'est le mouvement de force de l'informatique aujourd'hui, même dans les organisations.

Il y a des départements aujourd'hui autour du *data management.* Il y a aussi les archives. La sauvegarde des données. Cela se manifeste différemment selon les entreprises : les grands groupes, notamment anglophones, sont en train de tout rassembler sous les mêmes départements. Soit je vais avoir une définition puriste et universitaire et je te dirais alors : « le document c'est le document, la data elle est non interprétée et elle est généralement brute ». On va réitérer la différence entre document, data, connaissance et information. Tout cela se justifie au sein d'une vision universitaire avec le niveau de traitement intellectuel qui est fait sur le matériau de base. D'un point de vue opérationnel, pour les entreprises, c'est du matériau transverse en appui aux activités principales avec des niveaux de traitement différents et quelque part c'est un peu mis au même niveau tout ça.

Est-ce que tu as en tête les enjeux actuels les plus prégnants autour des données et comment la *data literacy* peut y remédier, y répondre en tout cas ?

Oui, c'est sûr que, selon les contextes, la data porte des enjeux différents. Alors si on considère la data au sens de ses réceptions, il y a toujours un enjeu aujourd'hui. Un enjeu fort qui a été celui de la transparence, de la confiance, de l'*open data*. Ça c'est clair et c'est fort. C'est le contexte des politiques publiques : la data porte

quand même de gros enjeux de transparence, de confiance et de lutte contre tout ce qui s'apparente à la fraude. Pour l'entreprise, la valeur c'est la performance. Il ne faut pas se leurrer, l'entreprise est toujours dans des logiques de gain de productivité et l'enjeu principal c'est la performance dans les entreprises, que ce soit la *BI*, que ce soit le *reporting* quotidien, pour développer de nouveaux marchés, de nouveaux segments, etc. Pour les scientifiques, pour les gens qui sont plus dans le savoir, c'est l'intégrité, c'est la qualité de la preuve, c'est la vie scientifique.

On a donc trois registres un peu différents mais forcément encore une fois qui se greffent par rapport à ce qui anime profondément et prioritairement ces communautés-là, ces objectifs-là en tous cas. Donc après, du point de vue médiatique, je pense que pour nos sociétés, l'enjeu de la data c'est la performativité du message que tu transmets. Elle a quelque chose de fort dans la persuasion, dans la performativité du message, dans les arènes publiques, la sphère publique. En info-com c'est important ça, d'un côté média. On pourrait trouver d'autres choses. Personnellement, ce sont les univers que j'ai traversés. Après on pourrait se dire que dans le numérique, dans l'univers des start-up, c'est la possibilité de développer des services, d'améliorer leurs services : la data géographique, la data de géolocalisation, tout ceci est fondamental aujourd'hui dans le domaine applicatif.

Tu viens un peu d'y répondre mais vois-tu des données qui sont au cœur de ces enjeux ?

C'est important ça ! Tout d'abord, on voit que les données les plus réutilisées sont celles qui font office de qualité. Il ne suffit pas de libérer les données, il faut qu'elles puissent être jugées comme étant légitimes, bonnes, irréprochables, en tous les cas fiables. On se rend compte que les grands jeux de données les plus réutilisés sont souvent ceux produits par des organismes dont ce sont les métiers de générer de la donnée. Que ce soit les instituts de statistiques, que ce soit tout ce qui est cartographie, que ce soit pour la population avec le *CNED*, *l'INSEE*, que ce soit des données géolocalisation, etc. Après ce sont les données aussi qui sont stratégiques avec par exemple les données de vente immobilière : tout ce qui peut permettre à une entreprise de se positionner sur un marché et de donner de nouveaux indicateurs pour ses clients. Toutes les données qui ont été libérées et qui permettent de qualifier un marché, un secteur, c'est important.

Après quand on est dans la science, toutes les données qui permettent de progresser : que ce soit les données génomiques, sur l'environnement, les données de sismographie, des données qui ont du vrai sens et qui sont capitalisables entre scientifiques pour aller plus vite dans la recherche scientifique, ça c'est évident. Tout ce qui est bien corrélé à des enjeux de société, pour résoudre des problèmes de société, relève du bien commun finalement ; qu'il s'agisse de la découverte des nouvelles maladie, du partage de données cliniques sur les grandes cohortes de patients, c'est fondamental. Après, maintenant, on est aussi dans des enjeux économiques de mise au point de médicaments : tout cela n'est pas simple parce qu'on n'est pas dans un monde de bisounours.

Il y a pas mal d'activités économiques et les données scientifiques sont en prise avec des logiques économiques. Donc il faudrait un peu réguler tout ça sans tuer l'innovation. Et c'est un peu le problème. Je ne suis pas de ceux qui pensent que tout doit être ouvert, ce n'est pas facile et on peut comprendre qu'il y a de la recherche privée et que parfois il faut capitaliser sur ses propres résultats pour pouvoir continuer d'avancer. Toutes ces données me semblent importantes. Après, on peut pendre en compte la vie citoyenne : les bus, les rues, tout ce qui permet de faire de la géolocalisation.

Tout ne peut pas être complètement ouvert en revanche. La frontière pour moi c'est dans la zone où si on ouvre, on détruit une activité économique qui portait une valeur et qui finalement était supérieure à l'ouverture des données parce qu'il faut aussi laisser l'innovation se développer. Après on peut partager des données dans des cercles restreints entre acteurs. Ils se partagent entre eux des données pour créer de la valeur. La finalité ce n'est pas l'ouverture, c'est la création de valeur et le progrès.

Je vais rebondir sur ce que tu viens de dire il y a quelques secondes : j'aimerais t'interroger sur la réglementation, sur le cadre législatif sur les données en général. Comment vois-tu ce cadre législatif et est-ce que tu ressens des impacts sur tes pratiques ?

La régulation est très importante parce qu'elle prend soin des équilibres. J'avais beaucoup suivi le RGPD, le règlement sur l'intelligence artificielle qui est en cours de consultation et le *data act* j'en ai entendu parler. Le *data act* qui considère un peu tous les cercles : les données d'entreprises et pas juste les données publiques. Je pense qu'effectivement à partir du moment où les données sont des actifs fondamentaux pour le développement des économies, elles le sont aussi pour le développement des sociétés contemporaines, autant du côté de l'éthique, du côté de la préservation de la vie privée qui finalement est à considérer par rapport à l'innovation. On voit que les données de santé ont toujours été des données extrêmement protégées en Europe et beaucoup moins aux États-Unis. Avec notamment un tollé des start-up de la santé : « ils vont plus vite que nous parce qu'ils ont accès à plus de données grâce au *machine learning*, nous on peut pas en faire autant ». Il y a des enjeux, des équilibres à trouver dans cette société numérique.

Penser que la régulation est importante dépend encore une fois des valeurs fondamentales que l'on met en premier dans une société. Après, le *Data act*, je ne suis pas au fait des dernières lectures du *Data act* mais moi dans ma vie d'enseignant(e)-chercheur(se) je n'ai pas été du tout gêné(e) par le fait que des données soient fermées, il faudrait plutôt voir du côté des gens de la santé ça c'est sûr. Les start-up de la santé.

Donc que dire, je pense que c'est important qu'il y ait une régulation et forcément cette régulation elle est corrélée aux valeurs prioritaires qu'une société se fixe pour se développer. Je pense que l'Europe fait un boulot très bien, très intéressant parce qu'en plus il y a beaucoup de consultations, toutes les parties prenantes sont consultées. Alors peut-être que les décisions sont déjà prises et que ce n'est qu'une mascarade mais je n'y crois pas trop. Et que je pense qu'on doit tenir compte de l'ensemble du point de vue des acteurs. Je ne suis pas un(e) grosse utilisateur(rice) de données : je forme beaucoup aux données mais je ne suis pas un(e) grosse utilisateur(rice).

As-tu une idée sur la façon dont il faut mettre en place une stratégie de *data literacy* dans un organisme et est-ce que ça doit concerner tout le monde ?

Oui, je suis un(e) militant(e) du *no code*. Je pense qu'il faut qu'on ait de plus en plus de services numériques tout d'abord. Plus on aura des applications, plus on aura des plateformes qui permettront de rendre limpide finalement la visualisation de données. Alors effectivement, ça renvoie à une notion de boite noire et pour le grand public on ne peut pas former tout le monde aux algorithmes statistiques qu'il y a derrière pour faire du *data mining*, du *clustering*, et ça n'intéresse pas tout le monde. C'est une culture qui est difficile à faire pas-

ser contrairement à la culture littéraire et même là parfois c'est difficile. C'est compliqué parce que c'est plus technique. Qu'on est peut-être dans une société un peu moins technicienne et un peu moins en prise avec l'informatique statistique que d'autres sociétés plus anglophones qui ont un rapport au chiffre plus important. Donc globalement pour la développer il faut avoir des plateformes, des services, une fluidité et ça permet d'approcher les métiers.

Les métiers de toute façon, le grand public, tout ça n'a jamais autant marché que quand on a tout un écosystème de services qui rendent transparent la technicité nécessaire pour visualiser des données. Donc ça je pense que c'est important. Après, il ne faut pas non plus aboutir à des sociétés qui soient envoûtées par les chiffres et la dataviz sans qu'il y ait une vision plus critique de la façon dont tout ça est construit et ça j'allais dire qu'en *back office* il faut vraiment qu'il y ait des professionnels de la data et des régulations de la data pour qu'en bout de chaîne on ait une nation en qualité de services. Dans les organisations, il faut rendre visible les départements, les départements transverses à la data. Je ne pense pas que les métiers ça va se faire en quelques années. L'intégration complète de la data dans tous les métiers par exemple, on en est encore loin. Je pense qu'il y a la place pour un certain nombre d'années avant de rendre visible une vision transverse avec des savoir-faire transverses sur la data qui vont travailler eux à rendre de plus en plus accessibles et mettre en place cet écosystème un peu *no code* dont je parle pour que les métiers s'en servent pour aller effectivement replonger dans de la technicité. Je pense que la culture se développe qu'à partir du moment où c'est facile d'accès et que ça parle à notre raisonnement et notre capacité de décision. Et s'il faut traiter les données avant, il faut qu'on arrive à avoir du *no code* pour qu'il y ait une vraie culture de la data qui se développe. On n'y arrivera pas sinon, ça restera un truc de spécialiste.

Les personnes à l'issue d'un apprentissage en *data literacy*, qui sont acculturées à la donnée, que doivent-elles développer ? C'est quoi ce qu'ils doivent savoir : une culture générale sur la donnée ? Avoir des compétences techniques / pratiques ? Avoir un état d'esprit ?

Je pense qu'il y a plusieurs niveaux. D'abord il faut qu'ils sachent d'où viennent ces données, qu'il y ait une représentation de ces données. D'une façon générale il faut bien sûr avoir une posture sur comment ces données ont été construites, quels sont les biais possibles de ces données, donc s'interroger. Mais bon on le faisait déjà sur les textes aussi : donc c'est important. Savoir aussi que quand on veut traiter des données, forcément c'est moins facile. Alors je parle des données non textuelles, je pense les données au sens renouveau. Si on parle des documents comme des données, on revient à la case départ. Mais moi je parle des données qui sont réellement des données nouvelles, des données non structurées, marchandes, des data plutôt traitées par des outils statistiques, par des outils de traitement de textes, résumés, etc.

Donc je pense qu'il faut un minimum. Et ça peut s'acquérir dans l'éducation jusqu'au lycée, voire les premières années aussi. Il faut que les techniques traditionnelles, je sais pas, d'agrégation de données, de clusterisation, il n'y a pas trente-six méthodes quand on parle de régression linéaire. D'ailleurs un traitement de régression linéaire je n'ai pas besoin de le rentrer, je n'ai pas besoin de le savoir, mais je sais à quoi ça correspond derrière. Je pense qu'il y a une culture minimum de traitement des données, en statistiques sur les données qu'il faut enseigner au même titre qu'on a enseigné la condensation de textes, le résumé, le commentaire de texte. Depuis

la seconde on ne cesse de faire du traitement de texte. Je pense que le retour des mathématiques dans le tronc central dans l'information des lycéens ça serait bien.

On ne sera pas tous des statisticiens mais on doit être tous capables de comprendre que quand on fait une classification ascendante par exemple à quoi ça correspond. Un minimum de représentation des traitements statistiques : même si on ne les fait pas soi-même et qu'on ne va pas les optimiser, on sait à quoi ça correspond derrière. Et je crois que c'est ça aussi qu'il faut arriver à mettre en place pour les données pour ça puisse être beaucoup plus mobilisé finalement que cela ne le sera sinon cela sera vécu comme quelque chose venant de l'extérieur et pas comme faisant partie de notre propre décision interne. Donc la montée en puissance de ces compétences, de statistiques mais à un niveau macro et quand on a une plateforme qui nous propose trois types d'algorithmes, c'est quand on appuie sur un bouton à quoi ça correspond quoi.

On sait aussi que si on nous dit de jouer avec des classes, qu'on en met cinq ou six, on sait à quoi ça va correspondre derrière, les regroupements. On sait jouer les potentiomètres et on sait derrière ce qui va être fait par l'algorithme qui a été mis en place, auquel on va déléguer une confiance, on ne pourra pas faire autrement de toute façon. Après c'est pareil avec les visualisations : savoir choisir, dans la sémiotique ce qui correspond le plus pour traduire le sens qu'on a envie de voir dans les données. Ce sont des outils intermédiaires qu'il va falloir un peu investir. Les simplifier aussi parce que je ne crois pas du tout qu'on va former les gens à l'informatique, au code ou aux statistiques, mais il faut que ça devienne un tronc commun de représentation macro : on sait à quoi ça correspond, à les mobiliser. Je pense qu'il faut qu'on travaille à ça. Pour le faire, il faut des outils de plus en plus ergonomiques, pratiques, fluides. Je pense que ça contribuera en tous les cas à développer cette mobilisation de la data dans nos activités. Sinon ça restera un univers de spécialistes.

Tu as parlé d'être critique : est-ce que le terme d'esprit critique te parle et si oui, que mets-tu derrière ?

Oui, l'esprit critique c'est comme le document : d'où viennent ces données ? quelles sont les sources ? Toujours dire, toujours penser qu'il peut y avoir des biais donc ne pas aller trop vite dans son raisonnement. C'est quand même une démarche en Sciences de l'information : c'est pour ça que j'ai en horreur *Twitter*, tu ne peux pas imaginer. Il y a toujours des stratégies d'influence, toujours des stratégies d'occuper un espace médiatique qui n'ont rien à voir avec la confrontation de données, une vision plus équilibrée et objectivée des choses. Donc l'esprit critique c'est ça : c'est d'avoir conscience de ça. Avoir conscience aussi des traitements qui sont derrière. Ça renvoie à ce niveau macro de traitement qu'on doit maîtriser pour se dire « ah oui d'accord, derrière ils ont fait ça, je vois ». Donc c'est savoir ouvrir la boite noire sans devenir un spécialiste. Mais en tout cas pouvoir engager un dialogue avec les spécialistes, ça c'est sûr. Donc c'est essentiellement ça : c'est la véracité des sources, le questionnement des biais, la conscience de la boite noire et savoir l'ouvrir, voir ce qu'il y a dedans et avoir une connaissance *a minima* pour engager un dialogue avec des spécialistes de cette boite noire.

Comment vois-tu le niveau général en culture des données des publics que tu observes : les étudiants déjà, j'imagine. Quel est ton avis là-dessus ?

C'est compliqué, je n'ai pas un point de vue très objectif. Alors pour les étudiants, je pense qu'il y a un goût de plus en plus fort. Dans mes classes il y a un goût plus prononcé. Alors est-ce que c'est un mouvement général de la société, parce qu'on en a de plus en plus partout et qu'on communique de plus en plus avec de la dataviz. Dans tous cas, on voit que ceux qui étaient vraiment sur les documents s'intéressent de plus en plus à la donnée. Je ne parlerai pas des étudiants de master parce qu'ils sont là pour se perfectionner.

Après, dans la société en général, je pense que la donnée d'un point de vue médiatique fait une percée considérable en termes d'accompagnement des messages. Faire passer un message, c'est quand même essentiel. Après ça reste effectivement à un niveau plus compliqué : ça reste technique pour beaucoup de gens. C'est plus facile de lire et de recevoir un message qui vient d'un traitement de données que de vouloir en produire un ; ce n'est pas encore facile de se dire « mince, je vais essayer de produire un message avec des données disponibles, là il faut être formé ». Alors que produire un texte, le mettre sur le *web*, tout ça c'est à la portée de tout le monde presque. La data ça reste une difficulté un peu plus grande : savoir chercher les données, les récupérer, les agréger, tout ça ce n'est pas encore très facile et cela ne le sera jamais parce qu'il y a des contextes très compliqués et très différents pour récupérer des données.

Les difficultés actuelles à l'acculturation aux données, à la mise en place de la *data literacy* en général, lesquelles sont-elles pour toi ?

Oui, je pense que ça a des racines profondes d'abord sur l'éducation dès le départ : ça part de très loin la façon dont notre cerveau se modèle. Ça part presque de l'école maternelle. Donc il y a une difficulté peut-être éducative avec une intégration qu'il faudrait peut-être plus travailler très jeune sur la donnée parce que c'est là que ça prend source. Après, répète-moi la question…

Les difficultés actuelles sur la mise en place d'une culture des données.

Après ça vient peut-être aussi des enseignants. Quelque part on a une génération d'enseignants qui n'est pas vraiment acculturée à la donnée. Donc comment rompre ce système infernal ? Parce que il faut peut-être l'intégrer aussi dans la formation des enseignants, ça passe aussi par le poids donné à ces données, au traitement statistique dans les diplômes généraux de base, ça c'est clair. C'est une grosse erreur d'enlever les maths, les stats, je pense que c'était antagoniste au fait de vouloir développer une *data literacy*.

Après les difficultés, pour moi il faut qu'on arrive à avoir du *no code*. Je suis vraiment un(e) obsédé(e) de ça. Si la dataviz a autant de place aujourd'hui, c'est parce qu'il y a des tas de logiciels qui permettent à une majorité de gens un minimum d'investissement et ils n'ont pas une ligne de code à faire. Pas forcément des bibliothèques R studio à mobiliser parce que ça c'est bien mais c'est un niveau de professionnalisme un peu au-dessus. La culture de la donnée au sens où moi je l'entends il faut qu'elle soit aussi accessible à un public élargi, qu'elle prenne vraiment racine. Il faut développer des appli simples, développer des plateformes et des logiciels simples et puis certainement en *open source*. Il faut penser les connexions aussi : on a aussi beaucoup de logiciels qui fonctionnent. Je pense à l'entreprise par exemple : les connecteurs entre les banques de données diverses et variées et puis des logiciels qui font de la visualisation. Ça aussi que ça soit de la tuyauterie un peu plus facile : ce n'est pas toujours simple. Plus il y aura une fluidité de ce côté-là, plus on aura une culture de données qui se dé-

veloppera. Il ne faut pas que ça reste une histoire de spécialistes, ça c'est sûr. Mais peut-être que dans quelques années tous les métiers de données : *data steward*, *data manager*, ça aura disparu. A partir du moment où il y aura une culture de la donnée qui se sera propagée, on pourra dire que finalement c'est devenue une composante dans bien des métiers et une compétence intégrée dans tous les métiers. Et c'est peut être vers ça qu'il faut tendre.

Est-ce que tu as un projet de *data literacy* en France qui a bien fonctionné et si oui lequel ? Ou en tout cas basé sur une idée de *data literacy* ?

C'est compliqué car il y a des publics différents. Pour l'université je dirais notre master parce que je pense qu'on a contribué dans tous les cas à cela parce que c'était l'objectif du master. Après dans la sphère grand public, je pense que la covid-19 a été un bon catalyseur d'acculturation à la donnée : les gens se sont renseignés. Le data journalisme en général. Comme ça touche beaucoup de monde et que c'est très performatif, dans des conditions où les gens lisent de moins en moins, le message doit être visuel et court. Donc je pense que dans cette communication organisationnelle, institutionnelle, je pense que ça a été une acculturation à la *data literacy* assez forte : les gens sont allés voir les sources, ils sont allés voir les choses en ligne, je pense que c'est une forme de la *data literacy* importante en termes de lecture. Ce n'est pas pour autant que les gens sont devenus des spécialistes. Mais à ce niveau-là en tout cas ça a monté d'un cran, le poids des données pour pouvoir faire passer un message.

Après, qu'est-ce que je pourrais dire d'autre ? Est-ce que l'*open data* était un phénomène important pour développer la *data literacy* ? Peut-être aussi. La récupération de données, peut-être de faire travailler les jeunes. Je pense qu'il y a des choses bien en lycée-collège, surtout en collège, sur la réappropriation de données de proximité, de la ville, de l'histoire, dans l'expérience de *data literacy* avec les collégiens, il y a des choses intéressantes. Je pense que c'est aussi des enjeux importants, c'est-à-dire que la data doit prendre aussi du sens par rapport à un contexte très proche, de les faire travailler avec l'*open data* de la ville, ou en histoire, les morts de la première guerre mondiale dans sa ville par exemple, je pense qu'il y a eu des choses intéressantes, je ne sais plus qui les a menés, peut-être *Canopé*. Il y a en tout cas des choses intéressantes. Donc voilà, pour mener une *data literacy*, il faut vraiment que ça se greffe sur un contexte motivant. Dans l'entreprise c'est effectivement ses données internes et ce qu'elle peut en faire pour être meilleure. Pour les collégiens, c'est effectivement cette idée de valoriser avec les données de proximité et tout prend sens parce qu'on valorise des données qui nous appartiennent presque. Et puis pour le grand public en général avec les crises sanitaires, ça les a familiarisé même s'il y a eu des tas de critiques. A la limite encore une fois même si les données sont quand même discutées, de se rendre compte pourquoi les données étaient erronées par exemple : les données de mortalité, et quand on s'est rendus compte que toutes les données ne remontaient pas, que quand on comparait les pays entre eux c'était pas comparable, ça c'est de la vraie mise en débat de la *data literacy*. Ça permet à la sphère publique de se poser de vraies questions sur la data en tous les cas.

Est-ce que tu as des remarques pour aider à l'acculturation en France sur les données ?

Il faut que ça parte de très loin. Il faut remettre les stats, les maths, les données, dans les programmes, lycée, collège, les faire parler et pas que d'un point de vue technique. Montrer en quoi ça peut être intéressant, convo-

quer les sources de données maintenant qu'il y en a beaucoup pour raisonner. Moi je pense que tout part de là. Après nous à la fac, je pense que ce qu'on fait nous aujourd'hui dans le master ça peut disparaître si finalement. Cette culture de la data elle part de beaucoup plus près en termes de programmes éducatifs. Donc j'aurais envie de dire ça : je pense qu'il faut qu'on soit effectivement une mobilisation un plus forte de ces données. Alors on peut les critiquer, de se dire qu'on va être de plus en plus pilotés par rapport aux données mais est-ce que ce n'est pas mieux que d'être piloté par des visions idéologiques, par des visions orientées, stratégiques… ça contrebalance un petit peu beaucoup de choses. Même si elles peuvent être incomplètes.

Universitaire 8 (13/07/22)

Comment définis-tu la *data literacy* et est-ce que tu emploies ce terme ?

Je n'emploie pas ce terme parce que c'est un terme anglo-saxon et qu'il y a des très bons termes en français pour dire à peu près la même chose. Après quand il existe un terme qui fait consensus comme le terme français quand on dit management, c'est vrai que ce n'est pas très gênant. Je pourrais l'utiliser mais pour l'instant c'est vrai que je n'ai pas encore vu un consensus évident. Mais je pense que ça va avancer. Comment je le définirais ? C'est l'aptitude à pouvoir manipuler la donnée. C'est un point de vue plus personnel plutôt qu'une définition générale mais c'est être capable de la collecter, de la mettre en forme, de la faire parler. Pour moi, c'est le spectre de la *data literacy*.

On a tout un ensemble de concepts autour des données : la gouvernance des données, la *data science* par exemple. Est-ce que tu arrives à situer la *data literacy* dans cet écosystème-là ?

La *data literacy* c'est quelque chose qui est utilisé en *data science* effectivement. En data analyse aussi. Tous les métiers qui sont liés plus ou moins l'utilisent mais n'ont pas les mêmes *skills*, les mêmes aptitudes. Ce ne sont pas les mêmes termes mais ça couvre un spectre qui est utilisé par à peu près tous les métiers en lien avec la donnée. Le *data scientist* il va être à l'aise sur la transformation, la récupération. Le *data analyst* va être capable de prendre de la donnée qui a plus ou moins été transformée, d'en faire des analyses avec une réflexion, avec un point de vue derrière. Le *data scientist*, lui, va avoir cette possibilité de faire parler les données mais en utilisant un petit peu des méthodes qui seront des méthodes statistiques, algorithmiques, qu'il va un peu *customiser*.

Et tous les métiers vont être comme ça : ils vont avoir leur propre *data literacy*. C'est par rapport aux aptitudes de chacun et ce qu'ils veulent faire avec la donnée. Il y en a qui vont vouloir plutôt en faire une représentation graphique afin que d'autres puissent mieux la comprendre et prennent des décisions typiquement dans le type de l'*intelligence*, de l'*intelligence business*, de la *BI*. On va récupérer avec un outil qui s'appelle *Power BI*, quelque chose qui va être un classeur Excel et on va faire en sorte d'avoir des tableaux, de pouvoir visualiser des informations et puis de comprendre comment on orienter son *business* par rapport aux informations qu'on en tire. Donc, la *data literacy* ce n'est pas un bloc que tout le monde a : chacun développe la sienne.

Si on reprend du point de vue de la gouvernance de la donnée et de la *data science*, ça reste assez séparé de la *data literacy* ?

Alors la gouvernance de la donnée, c'est soit en entreprise, soit dans l'administration, soit plus globalement à l'échelle d'un État. C'est la façon dont on va décider de collecter, valoriser, partager la donnée et la traiter : là effectivement, on est plus sur de la gouvernance. La *data literacy*, ça va être l'aptitude de chaque personne à traiter sa part de ce qu'il a à traiter avec ses propres méthodes. Méthodes qui peuvent évoluer avec le temps, avec des systèmes d'information ou juste se former sur le tas.

Tu as commencé un peu à l'évoquer mais on va aller plus loin : si on devait donner une « nature », des caractéristiques de la *data literacy* , est-ce que tu dirais que c'est une culture, un ensemble de pratiques et de techniques, un état d'esprit ? Comment tu verrais ça ?

C'est les deux mon bon monsieur. C'est à la fois une culture mais également des prérequis mais qui pourraient être épistémiques : si on se met sur un plan scientifique, il y a des sciences différentes qui font appel à la donnée et qui la manipulent : il y a donc de la *data literacy* dedans. Et il y a un peu des *motos*, des proverbes locaux à chaque discipline. Par exemple, il y en a un dans la gestion des données de laboratoire qui disait « garbage in, garbage out ». C'est l'idée de dire, en termes de *data literacy*, que si on ne prépare pas des données de qualité, les résultats qu'on aura en recherches ne seront pas meilleurs du fait de ce qu'on fait rentrer dans les algos, dans les outils, et donc là c'est plus une partie « culture de la qualité de la donnée ». Après il y a plus ceux qui sont sur de la transformation technique. Donc les *data analysts,* on va leur donner de la data : ils ont la connaissance de certains logiciels, éventuellement de deux / trois algorithmes et ils peuvent corréler les données entre elles. Et donc ils vont utiliser un peu de connaissance technique mais ils ne vont pas entrer finement dans la manipulation. Donc là, il va y avoir à la fois un savoir-faire mais qui va être plutôt de l'ordre d'une sociologie culturelle de la data avec un vernis technique. Il va y avoir le *data scientist* qui va avoir sa propre façon de travailler la donnée : cela va être une matière première sur laquelle il va, en tant qu'artisan, manipuler en créant ses propres outils. Et là on sera dans la manipulation technique avec une orientation qui ne va pas être tellement analytique mais purement statistique en disant « quand on est dans tel cadre, on utilise tel type d'algorithme ». Il ira donc après sur l'optimisation de l'algorithme.

Alors qu'en gouvernance des données dans l'entreprise, on va plus s'interroger : « on a ça qu'on a posé là, qu'est-ce qu'on fait ? Est-ce qu'on peut le valoriser ? Quel va être le périmètre des personnes qui vont avoir accès à ça ? Surtout éviter que ça sorte de l'entreprise. Que ça n'aille pas vers les mauvaises personnes dans l'entreprise parce que parfois il y a de la compétition et il faut éviter que ça aille dans un endroit plutôt qu'un autre ». Ça c'est intéressant. Certains vont dire : « est-ce qu'on va plutôt prendre cette donnée et la poser dans un *data lake ?* » : on la pose et les gens vont en faire ce qu'ils veulent. Après ils les récupèrent, ils en prennent des morceaux, ils les analysent, ils les agrègent.

Ou alors on va voir des *data frames*, des trucs qui vont être un peu plus carrés. Il y a des étagères. Dans les étagères il y a des petits tiroirs. Puis on les met proprement, on sait les retrouver, c'est étiqueté, c'est déjà classifié. C'est un peu ce que mon grand-père faisait quand j'étais petit : il mettait ses différents outils et il en dessinait le

contour sur le mur avec un stylo comme ça il savait où c'était rangé et il avait une vision globale. Et ça demande beaucoup de travail : c'est une façon un peu différente de voir les choses.

En termes d'importance aujourd'hui de s'acculturer à la donnée, est-ce que tu dirais que c'est devenu une nécessité comme une autre aujourd'hui comme la technicité informatique, l'*information literacy*, ces compétences de base qu'il faut obtenir ou est-ce qu'on est vraiment dans un autre niveau ? C'est-à-dire que ça surplombe un peu tout ça ? En termes d'importance, comment tu vois les choses ?

Je vais te donner deux points de vue : un point de vue qui serait plus le mien et un point de vue que j'ai lu dans un mémoire de master qui m'a fait un petit peu sourire. Sourire mais qui en même temps m'a fait réfléchir. C'était quelqu'un qui travaillait dans une entreprise de comptabilité et qui gérait la data d'un point de vue comptable. Cette personne-là avait l'air de dire que la vision dans son entreprise c'était que le moindre comptable maintenant devait être aussi un *data analyst* en analyse financière : que les données, les classeurs *Excel* dans lesquels il y a des données comptables, c'était aussi un boulot de *data analyst* et qu'il fallait à la limite être capable aussi de faire des régressions. C'était écrit tel quel dans son mémoire. Et moi, sans aller jusque-là, je me suis dit « oui, effectivement, maintenant c'est devenu quelque chose que tout le monde devrait être capable de faire ».

La preuve en est c'est qu'avant pour faire une corrélation entre des personnes, entre des outils, il fallait vraiment des *data scientists :* on implémentait ça dans des logiciels qu'on programmait en C. On récupérait des données, c'était compliqué de récupérer des données, c'était des objets binaires. Maintenant, n'importe qui, n'importe quel étudiant en licence, même à la limite au lycée, prend un classeur *Excel*, compare deux colonnes, sélectionne les deux colonnes, prends une cellule, et puis fait « égal » puis « somme », colonne 1, colonne 2 et ça donne le coefficient. C'est-à-dire que maintenant il est plus ou moins indispensable d'avoir une vision de ce qu'on fait mais ce n'est pas la peine de forcément maîtriser le détail des algorithmes qui font le job. A la limite, t'as pas forcément besoin de savoir ce que c'est : si tu sais ce qu'est une moyenne, une racine carrée, des choses comme ça, pour implémenter l'algorithme, tu n'as pas tellement besoin de le connaître réellement : tu peux le faire en quelques secondes avec des outils *dataclick*. Ce qui est un avantage à la fois parce que tu n'as pas besoin d'être pointu pour arriver à le faire mais si tu ne le comprends pas tu peux arriver à des aberrations.

Et en termes de culture de la donnée : on le répète, on le rabâche mais quand on fait par exemple des analyses, on a vite fait de mélanger ce qui est une corrélation avec une causalité. Et on en voit tout le temps : on a beau le dire, on a beau sourire en le disant, ça arrive tout le temps parce que les gens ont une culture de la donnée et ont eu une culture de l'usage des outils mais n'ont pas forcément derrière une connaissance fine parce qu'ils n'en ont pas toujours besoin. Mais ils ont quand même besoin d'un petit vernis. Donc, est-ce qu'on en a besoin ? Oui et non, jusqu'à un certain niveau. On peut utiliser avec les connaissances, on va dire une petite culture plutôt qu'une connaissance pointue, on peut utiliser les outils qui vont très vite mais on n'en a pas autant besoin qu'avant. Par exemple, j'aurais dit avant qu'un *data analyst* devait aussi avoir les compétences d'un *data scientist* : il devait avoir les deux. En fait, les deux pouvaient devaient être la même personne parce qu'un *data analyst* sinon il devait aller voir un mathématicien ou un statisticien et il devait lui demander de faire les régressions pour faire des évaluations. Alors que maintenant, il a des outils qui font ça très bien : il a du

python, il a du R. A la limite, il a besoin d'une petite formation. Cela peut être fait avec *Excel* à la limite, s'il n'y a pas trop de données, s'il y a un volume raisonnable, s'il y a quelques dizaines de milliers de lignes.

Et toujours de ce point de vue là, toujours dans ce même registre, est-ce que tu as senti une bascule en termes temporels de cette nécessité-là, de cette acculturation à la donnée ? Est-ce que tu arrives à dater depuis quand ça a pris de l'ampleur ?

Je dirais aux alentours des années 2010 : il y avait un paradigme qui existait déjà, c'était le « not only SQL ». C'était de dire qu'on ne manipulait plus les mêmes masses de données . Avant ce qu'on faisait dans l'informatique à la papa c'était de mettre des données dans des structures de qualité et on pouvait travailler dessus très simplement. Et dans les trois V du *big data*, ou les 5 V selon les versions qu'on utilise, il y a la vélocité et le volume. Donc maintenant on travaille en mode *stream*, en mode flux : c'est-à-dire qu'on travaille avec tellement de données et que ça va tellement vite qu'on n'a plus le temps de se poser pour faire tous les calculs en utilisant des requêtes SQL pointues dans des bases de données relationnelles. Arrivé à un moment, on n'avait plus besoin d'être spécialistes de ces langages-là, de cette précision-là, pour arriver à tout faire : il y a eu des nouveaux outils qui sont arrivés. Déjà il y a eu le *cloud et* le *software as a service* pour dire « tu poses toute ta data ici et moi je te mets un logiciel qui va être capable de faire tous les calculs dont tu as besoin, tout est déjà préprogrammé : tu sélectionnes tes colonnes, tu les mets face à face et tu dis ce que tu veux faire avec et j'ai déjà des outils tout fait ». On est arrivé à ces moments là, autour de cette bascule-là, pour moi, pas réellement dans les métiers, mais on commençait à le voir au début des années 2010. Puis aux alentours 2014/2015 c'est devenu un gros sujet ! Et il y a eu plein de formations qui ont commencé à fleurir sur le *big data* en cessant de réfléchir à ce que j'ai appris à l'école fin des années 90 début des années 2000, c'est-à-dire de faire de la donnée de qualité, bien structurée. On l'apprend toujours, mais on apprend aussi à se défaire de ce point de vue pour aller vers quelque chose du type : « ok, cela ne sera pas parfait, mais cela sera posé à un endroit et on va pouvoir travailler sur la qualité : on éliminera les données sur lesquelles on ne pourra pas bosser et il nous en restera suffisamment pour faire quelque chose ». Pour répondre à ta question en deux étapes : 2010, un paradigme qui est là et en 2015 un paradigme qui commence à être ancré dans les pratiques métiers. Et ensuite ça s'est vraiment intensifié. Après je n'ai pas le recul pour analyser l'impact du covid : que tout le monde ait travaillé chez soi, qu'on n'ait pas le matos pour travailler sur les jeux de données sur son petit ordinateur de bureau, est-ce que le fait qu'il y ait le cloud, est-ce que tout ça a changé les mentalités ? Je n'ai pas l'idée générale. En tout cas, pour moi deux marches : 2010 et 2015. Après dans la recherche, il y avait certainement des choses avant.

Justement, parce que tu en parles, si on se focalise sur l'objet central de tout ça qui est la donnée : comment tu définis la donnée ?

Il y a pas mal de définitions et il y a aussi le fait de savoir si les documents sont considérés comme de la donnée ? Est-ce que les métadonnées sont de la donnée aussi ? Ce sont des choses qui sont un peu complexes. A la limite, ça provoque de la fâcherie entre les différentes personnes qui en parlent. Et même dans les sciences humaines on commence aussi à utiliser de la donnée. On utilise aussi des langages liés à la donnée sans forcément le poser, le définir. Pour moi la donnée c'est quelque chose qui se mesure, qui se capte, et qui s'enregistre. Alors c'est vrai que j'ai une grosse culture document initialement parce que il y a Pédauque, et à l'intérieur il y a cette

idée du document, des inscriptions, de la date, du support, etc. Mais pour moi, on est un peu dans le même principe avec la donnée : on mesure des choses, on les enregistre et après on est capable de les quantifier et de travailler dessus. Donc pour moi c'est pas complètement délirant de dire que le document serait de la donnée aussi. Pour moi, c'est quelque chose qui va s'enregistrer, se mesurer et être posé sur un support. C'est un peu simple comme définition mais initialement avant d'être dans l'info-com, l'info-doc, j'ai été informaticien donc je fais des trucs simples en fait : ce qui s'enregistre sur un support, principalement de manière électronique avec des 0 et des 1, et qui peut être manipulé ensuite toujours avec des portes logiques.

Toujours selon ton avis, quels sont les enjeux actuels autour de la donnée, en tout cas des données, et comment la *data literacy* peut un peu y répondre ?

Les enjeux avec la donnée, c'est que les personnes dans la construction de sens dans leurs métiers respectifs, et là je parle pas seulement de chercheurs, je parle aussi de métiers du quotidien, comment arriver à déjà, à la récupérer, la traiter, la traiter c'est-à-dire comment on peut la manipuler.

[pause de l'entretien]

Ce que je disais dans les enjeux, c'est qu'on devait trouver des sens à ce qui est enregistré, à ce qui est capté : cela peut être des données météo, des données de pollution, des données sur un ordinateur connecté à un site web. Ou alors des données plutôt qualitatives qui ont été transformées dans une étude sociologique de quelqu'un qui fait des entretiens qualitatifs et qui ensuite transforme le texte, l'enregistre et le classe en le mettant éventuellement dans des variables : ça, il faut arriver à le collecter, le qualifier pour pouvoir l'enregistrer et le récupérer avec autre chose pour qu'il y ait un traitement automatisé et ensuite faire du sens avec ça. C'est un enjeu de sens.

Et l'autre enjeu, c'est un enjeu encapacitatif, et pas tout le monde est capable de le faire. Il y a ce terme aussi anglais « empowerment », «l'enpouvoirement » : de dire que chacun dans un absolu merveilleux devrait être capable de produire du sens par lui-même, c'est-à-dire d'avoir au moins une notion de ce que fait l'algorithme : quel algorithme on utilise dans tel cas ? Comment est-ce qu'on qualifier les variables ? Etc. De pouvoir, par lui-même, prendre le pouvoir sur les données et être capable de répondre aux questions qu'il se pose Qu'il se pose dans son métier, à la limite d'un point de vue personnel sur les aspects de la vie qui l'intéresse. Que la personne soit chercheuse en sociologie, en sciences de l'information et de la communication, ou analyste financier pour voir si quelqu'un est crédible financièrement pour pouvoir lui attribuer ou pas un crédit. Ça et être capable de pouvoir le faire.

Et à l'intérieur de ces enjeux, est-ce que tu arrives à déceler des types de données en particulier qui sont vraiment au cœur de ces enjeux-là ?

Tu veux parler de typologies de données c'est ça ?

Alors c'est un peu la question « piège » : qu'est-ce que toi tu entends par types de données et ce que tu estimes important comme types de données. Cela peut être des types spécifiques en termes de formats, soit de l'open data... tout dépend ce que tu mets derrière.

La question est très ouverte. Quand on est dans une entreprise, il y a des données produites par l'entreprise et au sein des données produites par l'entreprise, il y a des données qui peuvent être des données d'entretien, des données qualitatives. Par exemple, quelqu'un qui fait un compte rendu, une réunion, un tête à tête RH. Tout ça, ça va possiblement être codé dans des variables et être intégré à un progiciel de gestion. J'ai eu le plaisir de suivre des mémoires dans de très grandes entreprises, du luxe par exemple, qui avait une question d'attrition à régler, c'est-à-dire que les cadres allaient partir et ils ne savaient pas comment la gérer dans ce cadre-là. Et donc ils ont mis au point des questionnaires et des suivis de carrière : là il était question de la donnée d'entretien. De logs aussi. Les logs sont des traces, des enregistrements automatiques : donc là on n'est pas dans de la donnée qualitative, de la donnée déclarative, on est dans « il a fait ça tel jour », point. On peut essayer de l'expliquer si on est *data analyst*, on peut essayer de trouver les explications. On va essayer d'en faire du sens. Ça c'est de la donnée avec une interprétation. On a des données journalisées. C'est-à-dire des logs : ça pour l'interpréter, il faut avoir des connaissances du système dans lequel l'information s'est classée. Mais aussi soit c'est du social et ça traite des êtres humains, soit ce sont des trucs extrêmement métiers, si on est dans du météorologique par exemple. Si on a des capteurs météo, il faut connaître la météo pour pouvoir le traiter. Mais dans tous les cas, ça reste de l'analyse avec une connaissance métier. Donc ça, c'est du log.

Après on a des données qui sont déclaratives. Et là, on se pose des questions. Tant sur les données déclaratives. Des données d'impôt par exemple, c'est déclaratif : la réalité qui se cache derrière la donnée c'est la réalité qu'il y a derrière la volonté de la personne. C'est quelque chose qui a une fiabilité. Donc, dans le cas des logs, on a de la métrologie, c'est-à-dire « est-ce que c'est calibré ? ». ça peut être très grossièrement calibré : par exemple, quand on a une montre, une montre connectée selon qu'on ait une Dochoutsou, un truc qui existe pas, qui vient de Chine et qui a été calibré avec les pieds ou qu'on ait, je sais pas, un modèle *Samsung*, *Apple*, etc., ça reste du calibrage et on a plus ou moins des marges et ça correspond à une certaine réalité. Des données déclaratives, ça peut soit être un mensonge, c'est-à-dire que quelqu'un va, pour une raison ou pour une autre, pour avoir un intérêt ou un autre, une honte ou ne pas avoir honte dans un entretien, va dire la vérité ou pas. Et dans ce cas-là, si on se fie à des données déclaratives, c'est assez compliqué.

Qu'on soit dans une entreprise ou dans une étude, les recherches c'est très compliqué. Ça, pour le valoriser, on a un autre type de données. Donc toutes ces données internes à l'entreprise ou les données du chercheur aussi, on peut les croiser avec des données comme celles qui existent dans l'*open data*. Donc soit de l'*open data*d e la recherche, soit de l'*open data* gouvernementale : tout ce qui produit de la donnée ouverte qui va permettre d'aller vérifier, d'aller affiner, ou d'aller poser d'autres questions. Dans ce cas de figure, on est sur un autre type de données. Souvent, dans l'entreprise, on a nos données internes et on a besoin, par exemple si on fait de la *BI*, d'aller récupérer d'autres données pour valoriser nos propres données ou pour aller investir d'autres marchés. Si on a des informations issues de l'*open data*, ou encore d'autres informations qui là ne sont pas du tout *open* mais qui sont stratégiques, on est toujours sur de la donnée : mais c'est de la donnée qui a été vendue, vendue au poids, c'est-à-dire que là on va être sur des données qui viennent d'opérateurs de téléphonie, de n'importe quoi qu'on utilise sur un site web. C'est le *data broker* qui va acheter deux kilos de données de personnes. Et

ça fonctionne aussi pour valoriser la donnée de l'entreprise. On a de la donnée externe qui est *open* ou qui est achetée et dans ce cas là on peut aussi avoir des problèmes très forts. On peut être soi-même très bons en data analyse, en *data science*, on va avoir nos propres jeux de données et on va se retrouver à acheter un autre jeu de données. Dans la revue *The Lancet* des chercheurs avaient acheté des données médicales qui étaient peu, voire pas fiables du tout, voire fabriquées, donc ces données en fait, croisées avec d'autres données, qui n'étaient pas fiables, ça donne un résultat « garbage in, garbage out » et un résultat : une rétractation d'article. Donc on est sur du* skills* qui n'a pas tellement été mis en cause.

Par contre, il y a l'éthique : est-ce qu'on fait une étude en achetant des jeux de données par exemple ? . En termes de *skills* aussi, des chercheurs américains avaient prédit la dette, le problème lié à la dette grecque, avec une erreur, dans une cellule, dans un classeur *Excel*, et ils se sont fait rabrouer par un doctorant. Doctorant qui leur a montré leur erreur : là c'était juste l'open data, le *dataset* qui était mis à disposition a été récupéré par le doctorant qui a dit : « en fait votre article-là est intéressant, ce n'est pas le souci, c'est juste que votre formule dans Excel est fausse ». Donc les prémisses étaient justes, l'analyse bonne, mais les conclusions complètement erronées parce qu'ils n'ont pas utilisé le bon algo, ils ont mal implémenté dans un classeur Excel. Et donc ça, on en arrive à : *data literacy*, oui, mais deux chercheurs ne peuvent pas faire ça. Moi je n'ai eu aucun problème de demander de l'aide à un collègue quand j'écris un article parce qu'initialement je ne suis pas du tout statisticien. Quand j'ai eu besoin pour des articles d'utiliser des méthodos un peu hors de mon champ, je suis quand même allé voir ceux qui savaient faire pour vérifier que je ne disais pas de bêtises. En particulier pour des valeurs de corrélation où on peut se poser la question : « si c'est à – 1, est-ce que ce n'est pas corrélé ou est-ce que c'est fortement corrélé mais négativement », ça ce sont des questions qu'on peut se poser. Et en disant une erreur comme « c'est – 1 donc ce n'est pas corrélé », c'est faux. C'est très corrélé mais négativement. Donc ça c'est des *skills* qu'il faut avoir, donc ça fait partie de la *data literacy*. A minima, si on ne connaît pas les choses, on manipule, on a des résultats et on se retrouve avec des interprétations aberrantes, du moins fausses, erronées parce qu'on ne maîtrise pas tout. Quand on ne maîtrise pas tout, les outils qui font directement, qui disent à tout le monde les résultats, que [nom de personne] appelle les « clique boutons ». Et quand on utilise les cliques boutons, c'est bien, ça donne des résultats mais après il faut être capable de les lire et de sélectionner le bon algorithme, le « bon » bouton à cliquer. C'est sélectionner les bonnes data et les bons boutons. Donc pour le périmètre des données, je t'en ai donné trois différents : les données internes, les données externes ouvertes et les données externes achetées avec derrière cette capacité à les traiter ou ne pas les traiter.

Toujours dans le cadre des données, on a un cadre législatif. A défaut d'un cadre législatif, un cadre réglementaire. La loi Lemaire, le RGPD, la loi CADA avant cela. Est-ce que toi au quotidien tu ressens ce cadre-là et qu'est-ce que tu en penses de ce cadre-là ?

Alors, je le ressens depuis les années 90 : quand on m'a appris la gestion de projets, les bases de données. On m'a dit : tu pré-remplis une déclaration CNIL pour le projet, tu le télécharges, tu le remplis et tu le rends avec ton devoir. Après, quand j'ai travaillé effectivement sur ce type de projet mais pas en tant que chercheur, parce qu'à l'époque j'étais ingénieur, cela faisait partie du boulot et cela en fait toujours partie. C'était un cadre qui était moyennement contraignant parce que la CNIL n'avait pas la force qu'elle a maintenant. Maintenant qu'elle a été renforcée par le RGPD depuis 2018, elle a la possibilité de taper fort : il n'y a plus seulement des standards d'éthique et de bonnes pratiques. Il y a aussi un cadre législatif et des décrets d'application. Et ça fait

toute la différence. On en arrive à des amendes qui sont épaisses : c'est là où ça existe réellement et je pense qu'il fallait en passer par-là pour que ce soit dissuasif en fait. C'est de la dissuasion nucléaire: si tu ne joues pas le jeu, comme ont pu le faire de très grosses entreprises des GAFAM, tu es puni. Alors ils ne payent pas directement, ils négocient derrière, mais : 1, c'est une mauvaise publicité, ce n'est jamais bon pour ces grosses boites, et 2, tous les *Bill Gates, Bezos*, tout ceux qui maintenant ne voient ça plus que de loin parce qu'ils n'ont plus le nez dedans, ils ne font plus la gestion par eux-mêmes mais leur boite en tout cas a dû faire face à des problèmes d'image et des problèmes financiers. Mais ce n'est pas ça qui est le plus violent. C'est le fait d'être condamné par l'Europe ou par un pays par le biais des règles de l'Europe, auxquelles s'ajoutent le gendarme interne qu'est la CNIL et qui vient sonner le coup de semonce avant de lancer devant les tribunaux. Pour moi, cela a une réalité.

Donc on a toujours ces questionnements. Les entreprises, leur intérêt c'est toujours de récupérer les données personnelles, les *logs* pour pouvoir faire des prescriptions, de suivre les gens, de traquer, de faire comme avant mais que cela ne se voit pas ou que cela ne soit pas assimilé à des infractions. C'est un peu comme le *green-washing* : à un moment donné tu payais jusqu'à un certain point pour pouvoir polluer. Là ils se demandent si cela leur coûte moins cher de payer une petite amende et de pouvoir continuer à utiliser les données des téléphones, des sites web, etc. La transition par exemple maintenant pour les webmestres, cela a fini par arriver mais il s'agit de pratiques qui avancent vraiment à tous petits pas. De se dire que maintenant, par défaut, il faut lire les 12 000 pages. Pour une application par exemple c'est moins contraignant que ça l'a été. Maintenant tu lis 12 000 pages sinon tu te tais. A une époque, dans les conditions d'utilisation de *Gmail* que je me suis amusé à lire pendant ma thèse : tu acceptes qu'on te fasse de la pub personnalisée en lisant tes mails ! Ils ne disent pas qui lit tes mails. Est-ce que c'est un algorithme ? Probablement un algorithme, mais en tout cas on lit tes mails, on crée de la pub personnalisée en lien avec les partenaires. C'est normal, ils ont un modèle économique : *Gmail* est gratuit, sauf quand tu utilises *Gmail* en entreprise où tu payes dans ce cas-là. Les données sont plus lues, sont plus disponibles, c'est quelque chose qui avance petit à petit mais qui commence à être un peu plus contraignant : le fait sur les sites web, le *privacy by design* avec le *privacy shield,* qui commence à être réel : c'est que maintenant tu n'as plus à cliquer 12000 fois, tu veux ou tu ne veux pas, avec des petits trucs en bas, des petits astérisques, t'es un peu moins embêté. Aussi, j'ai entendu dire que maintenant, ils en sont arrivés à un point où ils n'essayent plus contourner la loi. C'est comme *Allociné* : soit tu acceptes nos conditions d'utilisation et t'acceptes qu'on fasse du *business* avec tes données, soit tu payes ! Et on en revient à ce que j'avais écrit dans mon premier article : les gens étaient prêts à payer pour avoir des recommandations si elles étaient de qualité. On en arrive à cela maintenant. Soit tu donnes tes données, soit tu les vends. Ce sont des modèles économiques qui se construisent à partir des données. Donc ta question c'était « comment est-ce qu'on voit ça ? Comment ça évolue ? » : ce sont des choses auxquelles on a pas mal réfléchies avec [membres d'une organisation]. Dans les métiers de la communication, quand on fait des sites web par exemple, on arrive à ce type de réflexion. Parce que tout le monde ne fait pas des sites web pour l'*open data* culturelle : il y a des modèles économiques souvent, et si on ne travaille pas pour l'amour de l'art ou si on ne travaille pas pour le gouvernement, il faut trouver un moyen de vivre et c'est normal, c'est le jeu. *Allociné* qui fournit un service, qui fournit des prestations, des recommandations personnalisées, il faut bien que ça se nourrisse d'une manière ou d'une autre.

On arrive sur la dernière partie de l'entretien qui est plutôt le côté applicatif. Et la première question que je voudrais te poser là-dessus c'est : comment selon toi on peut mettre en place une *data literacy*, une culture des données, dans un organisme et est-ce que ça doit concerner tout le monde ?

C'est un peu piégeux comme question parce qu'on aurait envie de dire oui. Par exemple, c'est très clairement le cas avec [nom d'organisation], quand on travaille avec [nom de personne] qui est en fin de doctorat aussi. Ce qu'il faudrait pour que les données de l'entreprise soient valorisées dans, par exemple, le cadre de l'industrie 4.0 où on a vraiment tout le process industriel qui est transformé numériquement, qui est accompagné numériquement, qui est enregistré, qui est loggué, qui a des jumeaux numériques, etc., qui a une sorte de GED centralisée qui regroupe tout, en fait il faudrait que tout le monde soit acculturé. La personne qui va casser les cailloux sur la voie à la limite ça n'existe plus comme métier mais à la limite oui, il faudrait que ces personnes-là soient capables de rentrer les données dans le système, de les lire, de les interpréter. Il y aurait une visée humaniste qui serait celle de [nom de personne] qui est de dire que tout le monde est capable de s'intégrer dans un système sociologique qui fait que chacun apporte sa pierre, chacun partage. Après j'ai une vision : celle de [nom de personne] est vraiment très belle mais je la trouve un peu utopiste, de dire que tout le monde est concerné et que tout le monde participe. Tout simplement parce que tout le monde ne veut pas participer. Et est-ce que tout le monde est en capacité aussi de s'approprier un tel système, je suis pas confiant à 100% là-dessus parce que cette *literacy* que chaque agent devrait avoir, là je vais donner l'exemple de [nom d'organisation], de produire des données ou des documents qui sont de mon point de vue aussi un peu des données dans un process, dans une entreprise comme une administration, ça pourrait être [nom d'organisation] avec des personnes qui font de l'ingénierie pédagogique, partagent des documents. Et on en arrive à des moments où les données sont enregistrées, elles sont fausses, elles ne sont pas mises à jour et ça complexifie le process. Il faudrait, alors je ne dis pas que cela doit être une élite mais en tout cas, ceux qui travaillent sur des données qui vont être cruciales, qui vont impacter ce qu'on va appeler le cœur de métier, qui va impacter techniquement la réussite de l'entreprise, qu'il y ait une formation en tout cas. Que cette *literacy* ne soit pas quelque chose qui soit apprise sur le terrain mais quelque chose qui procède d'une vraie formation avec, plus ou moins, un certificat à la fin. C'est dire « voilà, je suis capable sur le *Excel* en ligne de rentrer les bonnes lignes, de les valider et d'avant de les envoyer il faudrait qu'il y ait des personnes qui soient habilitées à le faire réellement ».

En lien avec ça, tu avais déjà un peu répondu à cette question en début d'entretien mais je la pose ici plus en détails : que doivent développer les personnes à l'issue d'un apprentissage en *data literacy*, à l'acculturation de la donnée selon toi ?

Alors, je pense que de manière générale, il faut une culture générale de la donnée pour avoir une compréhension de ce que c'est que la donnée dans le système d'information et dans la société, le comprendre pour pouvoir l'utiliser ensuite. Cette culture générale n'est pas moins importante, de mon point de vue, que d'aller utiliser les choses un peu pointues, un peu techniques derrière qui arrivent en deuxième partie dans le sens elle vient après la première ! Une fois qu'on a cette base théorique qui est de dire :« voilà ce que c'est de la donnée dans l'entreprise ou de manière générale dans la société, quelle est son importance ? Et voilà maintenant quelle est ma place et qu'est-ce que je vais devoir faire sur cette donnée dans la chaîne complète de la donnée, voilà de quelle manière j'interviens ». Et là on apprend que les personnes qui sont dans la chaîne, si on prend les sciences de gestion, qu'on a une *supply chain*, une chaîne d'approvisionnement de la data. Pour faire une comparaison : on

a des *inputs*, on a de la donnée. Il faut comprendre que cette donnée, il faut la regarder d'une manière critique, pas en avoir une foi absolue, c'est ce que je disais tout à l'heure, c'est de dire « garbage in, garbage out » et de dire qu'on a des données qui vont peut-être être des données déclaratives : aux impôts, ils ont très bien compris que les données étaient déclaratives. Ils ont des services d'analystes derrière pour voir si ces données déclaratives n'ont pas d'erreurs intentionnelles ou non dedans. Ils ont des services d'analystes qui vont même voir sur les réseaux sociaux, qui vont aller comparer d'autres flux de données : des données visuelles, des données autres. Et ça ce sont des choses qui existent réellement, ce sont de vrais services. On va avoir la personne qui possède cette culture générale, qui va penser la donnée, qui va penser réellement, de manière correcte d'où vient la donnée. Et une fois qu'elle aura apporté sa pierre à l'édifice, où ça va aller ? Est-ce que ça correspond bien à ce qui est demandé ? On a donc ces deux aspects-là : cette formation initiale à la donnée globale et cette formation après d'apporter, de qualifier. Pour moi ce sont ces deux choses qui sont séparées mais complémentaires. C'est un non-sens total d'apprendre à quelqu'un de faire une régression sur *Excel* et de lui dire ensuite à quoi ça va servir. Comprendre la donnée en société, comprendre la donnée dans l'entreprise, et comprendre où est sa place, et comment on travaille dessus : tout ça c'est l'aptitude, la prise de pouvoir sur les données. Prise de pouvoir par la connaissance puis par la compétence. La connaissance globale de l'écosystème de la donnée et la compétence, c'est-à-dire sa capacité personnelle à transformer la donnée, en tout cas la récupérer, se l'approprier et d'en faire quelque chose, soit pour soi-même si on est en bout de chaîne, soit pour quelqu'un d'autre derrière.

Il y a un terme qui revient souvent, on parle d'esprit critique ou du moins d'être critique vis-à-vis de la donnée, est-ce que ce terme pour toi il est pertinent et qu'est-ce que tu y mets derrière ?

Alors pour moi il est extrêmement pertinent. L'esprit critique, c'est comme en terminale : on fait une formation à la philosophie et on voit différents philosophes. On en voit qui sont bienveillants comme Rousseau, d'autres qui sont un peu plus critiques sur l'esprit humain, d'autres critiques comme Max Weber sur la sociologie, sur l'ordre établit, etc. Il faut être armé de la même manière, il faut avoir une philosophie de la donnée et je pense que même la philosophie initiale n'est pas délirante pour appréhender la donnée. C'est un peu ce que faisait [nom de personne] avant qu'il n'intègre [nom d'organisation] : il est philosophe de formation ce gars-là à la base et il avait un esprit critique sur la donnée. Donc pour moi, c'est très important d'avoir ce regard critique. Critique, ce n'est pas critiquer pour critiquer : c'est avoir un regard sur les conditions de production de la donnée, sur ce que peut ou ne peut pas faire un algorithme, sur ce qu'il est pertinent ou pas de faire avec des données. Ce regard critique est juste salvateur pour l'esprit pour ne pas avoir ces grosses œillères de, on va dire de geek pour qui la vérité c'est la data initiale.

C'est ensuite le traitement algorithmique avec éventuellement une petite [perte de son] dans l'algorithme et une optimisation sur la complexité de l'algorithme, d'aller faire quelque chose qui va se traiter plutôt de manière itérative sur un seul niveau que de descendre, etc. Et effectivement, ça c'est vraiment important de savoir techniquement faire les choses. Mais se poser la question de savoir si c'est souhaitable de le faire de cette manière-là ou pas, est-ce ça va avoir un intérêt, ou est-ce que ça va nous orienter sur une mauvaise piste, alors ça c'est encore autre chose. Et ce sont des bonnes questions que se posent réellement les collègues que je vois dans les humanités numériques : au début, ils ne sont pas forcément toujours enthousiasmes et quand on leur dit « en fait, l'esprit critique c'est vous qui l'avez, c'est vous qui faites l'algorithme, c'est vous êtes l'algorithme. L'algorithme ce n'est pas quelque chose d'informatique. L'algorithme c'est la mise en œuvre du traitement

d'une donnée ou d'une information. Donc ça c'est vous l'algorithme. Derrière, il y a un traitement qui va être un traitement de la donnée accéléré par le traitement automatisé. Si vous validez le traitement, si vous validez la chaîne de traitements, c'est bon ». Mais prendre un logiciel, pousser de la data dedans, regarder ce qui sort sans recul, c'est là où on arrive vers des erreurs monumentales et ce n'est pas ça qu'on souhaite. Ce qu'on souhaite, c'est que la personne soit plus *analyst*, donc *analyst* par rapport à *scientist* qui lui va aller tripoter des algorithmes, les implémenter différemment, les optimiser. Mais l'analyste, il va réfléchir, il a une vision, en utilisant le terme que [nom de personne] aime bien, « sociétale », c'est-à-dire que ça vient de la société. Il y a aussi les questions sur la métrologie : est-ce qu'elles sont mesurées proprement, est-ce qu'elles sont réelles ? Des données qui sont récupérées par exemple dans les entretiens, la manière de les calibrer, de les mettre dans des classeurs : il y a toute une réflexion derrière et il faut avoir l'esprit critique. Quand on récupère un jeu de l'*open data*, c'est d'abord de savoir si cela correspond à ce qu'on veut. Et ça, c'est fondamental. C'est salvateur. C'est salvateur dans l'entreprise mais encore plus dans la science. Le fait de récupérer des données, des données expérimentales dans certaines disciplines ou des données d'entretien dans d'autres, d'avoir l'esprit critique sur la donnée en elle-même, c'est fondamental. Et le recul sur le traitement c'est fondamental aussi. Et cela on le voit dans les disciplines des sciences dures, de se poser la question : « ok il a tablé deux algos pour faire ça », par exemple Pierson ou Pierman, laquelle des deux tu as appris, pourquoi tu as appris celle-là ? Et dans le résultat, il y a une petite divergence entre les deux et peux-tu l'expliquer ? Cela, c'est l'esprit critique sur 1 - la donnée, 2 – sur la méthode de traitement de la donnée et la qualité du résultat. Pour répondre à ta question : oui, c'est fondamental.

Comment vois-tu la situation en termes d'acculturation générale à la donnée dans les publics que tu fréquentes ? J'imagine principalement les étudiants, et peut être pas qu'eux. Toi est-ce que tu trouves que c'est un niveau qui augmente ou est-ce qu'on est arrivé à un plateau ? Comment tu vois les choses ?

Pour moi, dans la *data literacy*, il y a ce regard critique par rapport aux informations et les données qu'on récupère par exemple sur les réseaux des médias sociaux. Cette question liée à une information qui arrive : est-ce que c'est une vraie information ? Est-ce que c'est une fausse information ? Et est-ce qu'on a la capacité de critiquer cette information par rapport à des sources de données ? Est-ce qu'ils vont penser directement à aller faire des recherches d'images inversées ? Des recherches de vidéos ? Regarder les métadonnées dans les photos, dans les vidéos qu'ils récupèrent. Est-ce que ça ils le font ? Est-ce qu'ils ont cette culture de la donnée ? Il y a une collègue qui a écrit un bouquin sur une étude longitudinale assez importante qui est « grandir connectés » et clairement, ils ne l'ont pas. Pour eux, il y a un média, comme la télé en fait, et il y a assez peu de recul là-dessus. Le fait de dire que les données pourraient éventuellement corroborer ou pas des informations, je ne suis pas certain que ça arrive avant le CELSA, Science-Po, ou quelques grandes écoles de journalisme ou d'études politiques par exemple. Cela est compliqué, ce n'est pas évident. Donc dans ces publics-là, on n'y est pas trop.

Après sur les collègues. Du côté des informaticiens, on est plus sur de la réflexion des optimisations des algorithmes, sur les emplois derrière des données. Sur la donnée elle-même il n'y a pas forcément de réflexion parce ce n'est pas leur sujet d'études, ce n'est pas ça qu'ils étudient, donc à part quelques-uns qui sont un peu entre les deux, qui sont sur les sciences humaines et sociales et les sciences techniques et de médecine, on en a pas mal

des personnes comme ça qui se posent des questions et on a également des chercheurs en sciences humaines et en sciences sociales. En sciences sociales, ils ont quand même une bonne culture de la donnée et de la manière de travailler sur la donnée. Ils n'ont pas trop de problèmes, c'est assez correct. En sciences humaines, ils ont un peu peur parfois. Alors quand je dis sciences humaines, là je prends littérature, histoire, lettres, langues et puis les collègues d'histoire. Dans leur cas, ce n'est pas dans leur cursus initial théoriquement. Après, il y a ce fameux groupe des humanités numériques et moi je n'ai pas trop de recul là-dessus parce j'ai le nez dedans tout le temps. Je suis avec eux, je travaille avec eux, je leur propose des solutions, et donc c'est compliqué pour moi d'avoir un avis là-dessus. Et c'est quelque chose que ma collègue me disait : ceux que je fréquente, ils ne sont pas majoritaires. Donc je ne vois qu'eux, je ne vois que ces personnes-là qui sont des spécialistes de traitement de la donnée : des gens qui sont au Luxembourg, des gens qui sont sur la côte Est des États-Unis qui ont créé les logiciels comme *Zotero*, qui ont créé les logiciels comme *Omeka*, qui donc sont à la fois des spécialistes de l'ingénierie mais également des spécialistes de la documentation avec un un cœur de métier comme on utilise ce terme en informatique. En ingénierie informatique, on dit « le métier », c'est-à-dire le spécialiste du traitement. Et ce traitement peut être l'histoire, cela peut être l'histoire de l'art, cela peut être la littérature et ces gens-là sont très au fait. Ils font des revues spécifiques : cela peut être des *data papers* mais cela peut carrément être aussi des J*upiter notebooks* en histoire. Je sors cet exemple parce qu'il est aberrant. C'est très positif mais c'est ce qu'on appelle une donnée aberrante dans un ensemble de statistiques, c'est-à-dire qu'il est à la marge. Il est à la marge mais c'est quand même hyper rare de voir ça, y compris en sciences dures, de quelqu'un qui écrit un article et intègre ces figures non pas sous la forme d'une capture d'écran ou d'enregistrements mais directement sous la forme d'un algorithme qui produit des documents, qui produit une image, qui produit des graphes, qui produit tout ça. Qui produit la carte en direct devant toi et tu peux vérifier et critiquer les résultats au fur et à mesure que tu lis l'article. Ce sont des historiens donc je les vois ces gens-là, je sais qu'ils existent. Mais de ce que me disent mes autres collègues, y compris dans le labo, il ne s'agit pas d'une majorité. Et il y en a beaucoup qui ont peur car ils ne comprennent pas. Je pense que c'est une interprétation personnelle qu'ils ne comprennent pas, qu'ils ne savent pas comment on fait et ça les saoulent de se lancer dedans parce que quand t'as commencé ta formation peut-être dans les années 80/90, que t'es pointu dans ce que tu fais, que tu changes de travail en fait, que ton travail est complètement transformé par de nouvelles méthodes. Quand on te dit « ben maintenant, les budgets vont être coupés, c'est ce que dit Pierre Mounier en 2018 quand il a écrit les humanités numériques et là-dessus je partage complètement son analyse, pour se crédibiliser et garder de l'argent pour pouvoir conserver les formations ouvertes, pour pouvoir continuer à faire de la recherche aussi, on « s'est mis », entre guillemets, à singer les techniques des STM ».

Donc *data papers*, utilisation technique de l'analyse des données, c'est-à-dire qu'on va faire aussi du quanti : c'est-à-dire qu'on va récupérer des données, on va les travailler, on va essayer de donner une explication à ces résultats alors que, historiquement, dans ces matières-là, dans ces disciplines-là, c'est pas ça qu'on fait. C'est qu'on fait c'est de l'interprétation, on s'appuie sur ce qui a déjà été écrit. Et donc du coup ça transforme le métier. Et cette transformation du métier, ça c'est le discours officiel depuis 2010. Dans les humanités numériques, il y a Johanna Drucker qui parle de ça, il y a ses petits camarades qui disent « voilà, le métier, y compris pour les littéraires, y compris pour les historiens de l'art, le métier est transformé ». Et donc pour ceux qui l'acceptent, qui vont l'avancer, c'est super, ça devient un petit peu des stars dans leur domaine. Et puis ils ont des relations avec d'autres disciplines : on lance l'interdisciplinarité comme un mantra aussi au niveau gouvernemental, donc ça il y a cette transformation, et il y a ceux qui refusent. Alors je ne dis pas qu'il y a une bonne

et une mauvaise méthode : juste que le métier tel que le connaissaient ces chercheurs dans d'autres disciplines (lettres, langues, histoire, histoire de l'art, et puis aussi beaucoup en archéo), ce métier, il est en fait scindé en deux. Il y a des « archéologues numériques », entre guillemets, qui sont capables d'utiliser des nouveaux outils. Il y avait cet exemple du nouvel outil qui est capable de redessiner un ancien, qui est capable de regarder la topographie d'un lieu. Tu peux voir là où il y avait des habitations avant et ça te le transforme, ça te le modélise, ça te le met et tu vas écrire des articles scientifiques par rapport à ça. Tu n'as même plus besoin de creuser pour voir ce qu'il y avait avant. Tu sais où creuser mais tu vas le faire effectivement. Là tu vas avoir un ancien archéo qui va faire le job mais à un moment donné, les données sont modélisées numériquement, elles sont mises à l'écran et tu peux commencer à la limite à écrire ton article avant d'avoir posé ton premier coup de pelle en tant qu'archéologue. Ce sont des choses qui n'existaient pas avant.

Donc cette transformation des métiers, elle existe potentiellement, comme dans tous les autres métiers d'ailleurs. Elle existe aussi en sciences. Est-ce que c'est la majorité ? Est-ce que tout le monde s'y fait ? Est-ce que cette *data literacy* tout le monde est ok avec ? Je dirais clairement non. Il y a plusieurs écoles. Et ce n'est pas forcément les jeunes les mieux lotis sur le sujet parce que sur les collègues en histoire de l'art que je peux côtoyer ou en histoire tout court, il y a des personnes qui sont très *old school*, qui restent uniquement en bibliothèque et qui utilisent uniquement un crayon à papier, et des personnes qui ont mon âge pour des collègues que je côtoie dans les milieux plutôt germaniques, qui sont ultra pointus du côté numérique, c'est-à-dire que ce sont des *digital humanists*, en plus de leur spécificité, mais ils ont cette double casquette, et ils militent à la limite, comme pourrait le faire Fatiha Idmhand, pour les humanités numériques comme discipline. C'est-à-dire que la *data literacy* cela ne serait pas un aspect de formation potentiel dans un métier de recherche. Cela deviendrait un objet d'études en soi et il y a des personnes qui militent pour ça en disant qu'il y a une telle transformation, qu'on n'est plus sur le même métier. Alors il y en a qui disent exactement le contraire avec une très grande fermeté et avec des arguments tout aussi audibles. Mais cette *data literacy* dans les milieux que je fréquente, donc tu m'avais également posé la question avec les jeunes, avec les étudiants, il y avait ce que je t'avais répondu sur « grandir connectés ». Et puis également pour ces chercheurs qui sont plutôt côté sciences humaines et sociales avec cette transformation qui se fait avec le numérique, par le numérique. Et donc cette *data literacy* qui fait maintenant théoriquement partie des nouvelles formations.

Alors c'est bien parce que tu as quasiment à la dernière question que je voulais te poser avant la conclusion, qui était : est-ce que tu arrives à cibler les difficultés actuelles à une acculturation aux données, à une acculturation à la *data literacy* mais tu as quasiment tout dit, sauf si tu veux ajouter autre chose.

Alors j'ai une collègue dans le labo que j'adore qui est très pointue en termes d'informatique, qui est capable de modéliser des bases de données, de les réaliser, de les utiliser, qui a travaillé sur des gros systèmes d'information. En recherche elle utilise du web sémantique, elle est capable de modéliser ses propres objets de recherche en RDF par exemple. Cependant, elle dit que l'informatique et que la gestion de la donnée, ça l'intéresse un peu mais ce n'est pas son métier. Et qu'on n'a pas de méthode générale pour transformer un métier qui préexiste depuis longtemps et qu'effectivement certaines personnes peuvent utiliser ces méthodes-là mais que ce n'est pas obligatoire et qu'il faut arrêter d'être obnubilé par ça. Par exemple « transformation des métiers de l'écriture de la recherche », « nouveaux objets d'écriture » comme les *data papers* ou les *Jupiter notebooks*, utiliser des outils

comme Stylo, etc., ce sont des objets de recherche, des objets d'étude. Éventuellement c'est un outil mais cela ne doit pas intégrer le cœur de la recherche. Donc là il y a un gros freinage de la part de certains chercheurs. D'autres par contre sont totalement ouverts à ces nouveaux modèles et se disent « ouais ben tant mieux. Si jamais je fais une étude, je crée des jeux de données, forcément j'ai un corpus, il y a de la donnée, je le décris et ça fait un *data paper*, ça me fait un article, ça me fait des points. En sociologie, en anthropologie de la recherche, ça marche : à la prochaine évaluation HECRES, ça fait des bons points. Et tout le monde est content. On a ces deux profils.

Pour la conclusion, j'ai deux choses à te demander. Est-ce que tu as en tête un projet en _data literacy _en France qui a bien fonctionné et si oui lequel et pourquoi ?

De *data literacy*...

Ou assimilé en tout cas.

Je serai tenté de dire dans l'accompagnement à la recherche un projet de *data literacy*, il y a Huma-Num où les chercheurs sont accompagnés par d'autres, alors je ne vais pas dire des chercheurs forcément mais des ingénieurs de recherche qui ont une forte connaissance du milieu de la recherche quand ils sont pas chercheurs eux-mêmes, qui proposent des services pour utiliser les méthodes de *data literacy.* Ils proposent la *data literacy* « vous avez besoin de la data. Vous ne savez ni la manipuler, ni comment la stocker. A la limite vous savez la récolter mais après vous ne savez pas dans quel format la mettre ». Et il y a un accompagnement là-dessus, c'est Humanum. Cela marche en tout cas pour les humanités numériques et pour les sciences humaines et sociales de manière générale. Ce projet, je saurais pas dire s'il était réellement un succès, si ça a vraiment marché, c'est une bonne question. Je pense que si on se met dans le CNRS, oui ça a évidemment bien marché mais je ne suis pas certain que ça rencontre à ce point-là son public. En tout cas ce sont des retours que j'en ai d'autres chercheurs qui ont des points de vue mitigés sur le sujet.

Il y a un autre de projet qui est en démarrage et qui est pas mal, c'est DORANUM, donc toujours CNRS avec l'INIST et puis avec les différents centres de formations à la formation scientifique et technique, qui est de comprendre tout l'écosystème de la data. DORANUM, c'est droit de la documentation sur le numérique et savoir comment est-ce qu'on la stocke, comment est-ce qu'on la manipule, quelles sont les obligations légales tout ça : c'est du concret, c'est vraiment de la *literacy*. Je ne peux pas te dire si ça marche. Il y a déjà des formations qui sont organisées, il y a des tutos qui sont organisés mais on n'a pas tellement de recul parce que là je suis en train de faire un tuto avec un.e ingénieur.e de [nom d'organisation] et je sais pas si cela correspond réellement à un besoin. Est-ce que ça trouvera un écho ? est-ce que ça sera utilisé ? ça je ne peux pas le dire. J'espère que ça développera la *literacy :* je pense pour des générations de doctorantes et de doctorants, peut être pour des jeunes chercheurs aussi, peut être que ça trouvera son public mais ça je ne peux pas vraiment, c'est de la prospective. En tout cas, c'est prévu pour. Pour reparler d'Humanum, je trouve que c'est bien parce qu'il y a un ensemble de collègues qui n'auraient pas pu faire des entrepôts de données, qui n'auraient pas pu fouiller des jeux de données s'il n'y avait pas eu les outils proposés par Humanum en fait, par le CNRS.

Et toute dernière chose du coup : est-ce que tu as recommandations ou ne serait-ce que des remarques pour améliorer la mise en place de la data literacy en France ?

En France de manière générale, pas dans le milieu de la recherche ?

C'est comme tu veux, selon toi.

Je pense que ça devrait être intégré à la culture générale, la culture de la *literacy*, de la donnée. Donc ça devrait être un élément de culture générale. Pour moi, on apprend à faire un certain nombre de choses à l'école, on fait un peu de stats, on fait un peu de maths, on en fait jusqu'au bac à la limite, et ça devrait en faire partie pour pouvoir être un citoyen, pour prendre ces décisions en connaissance de cause, il faut être capable d'analyser un discours et pour analyser un discours t'apprends à l'école les rudiments de la rhétorique, de comprendre ce qu'il se dit, la typologie des arguments, un peu de philo pour prendre un peu de hauteur et de recul sur ce qui se dit et sur ce qui se fait, la manière de le faire, sur l'homme dans la société également. Donc ça c'est une vision philosophique de la sociologie et de l'anthropologie. Il faudrait la même chose pour la data, pour prendre des décisions éclairées dans à peu près tous les aspects de sa vie, il faudrait être capable de travailler la data. Savoir de quelle manière tu vas remplir ta déclaration d'impôts, ça peut passer par un classeur Excel, ça peut passer par savoir à la limite, est-ce qu'il est plus valable d'utiliser telle ou telle méthode.

C'est pas forcément des maths ! C'est juste un peu de *literacy* sur les données dont on dispose pour prendre des décisions éclairées. Donc ça, ça devrait un peu être appris à l'école. C'est en cours. Après, c'est un peu décrié de dire « ok, t'as du Python au lycée » et les gens regardent ça en disant « mais pourquoi ? ». Pourquoi est-ce qu'ils voient *scratch* au collège et en primaire. Pour moi, ça va dans un bon sens. Après il faudrait l'enseigner conjointement avec d'autres disciplines pour avoir du recul et ne pas faire des cas pratiques. Pas forcément au collège, en primaire peut être mais au lycée peut être faire une interdisciplinarité entre des cas d'éducation citoyenne. Tu prends des jeux de l'open data, t'as une réflexion dessus et ton Python tu le fais sur des jeux de l'open data et tu demandes en termes de gouvernance qu'est-ce qui serait mieux, est-ce pertinent de faire-ci, est-ce que ce n'est pas pertinent de faire ça. Donc pour moi cela serait en tout cas une préconisation, puisque tu parlais de préconisation. C'est pour le citoyen responsable de sa vie en société et de ses prises de décision et également dans le cadre d'un parcours doctoral par exemple, tu as un certain nombre de crédits et tu les valides plus ou moins à la carte. T'as des cours d'épistémo qui sont obligatoires. Moi quand j'ai passé mon doctorat j'avais des cours d'épistémo qui étaient obligatoires. Pour moi ça en fait partie en fait. A la limite il faudrait une épistémo de la data et qu'elle ait un peu de pratique aussi parce que dans la *literacy* il y a de la pratique. Et puis surtout à [nom d'organisation] on est capable de faire avec la main, de manipuler. Cela ne serait pas mal que ça soit intégré, que cette *data literacy* fasse partie du parcours doctoral. Pas avec un gros coefficient mais au moins qu'il y ait cette capacité à traduire des objectifs de recherche de manière potentiellement chiffré, ou avec pourquoi pas des corpus textuels de manière chiffrée et essayer de faire un peu de traitement du langage dessus, de dire « ça existe ! ». C'est un peu l'affrontement qu'il y a eu, c'est dans le livre de Mounier, où il y a le télescope de l'esprit : de dire que tu as le télescope de l'esprit pour voir ton objet de recherche comme-ci tu voyais une galaxie. Tu le vois de loin et tu le vois en entier. Donc en littérature il y avait ce télescope de l'esprit de ce côté-là et il y avait [nom de personne] qui était un chercheur *old school* et qui disait « non tu gardes la main ». A aucun moment tu n'as la finesse des raisonnements, la beauté de l'écriture qui peut être détectée

dans des séquences par des algorithmes et ramenée. Il y avait donc cette dualité qui était mise en regard par Pierre Mounier à travers les différentes lectures qu'il avait synthétisées et ça c'était très intéressant. Donc de dire qu'effectivement cette *data literacy* va permettre de faire émerger les choses et que tu gardes ton libre arbitre en tant que chercheur pour évaluer si c'est pertinent ou pas. Donc voilà pour ma préconisation.

Universitaire 9 (25/11/22)

Comment définis-tu la *data literacy* et emploies-tu ce terme ?

Alors je vais inverser la question. Oui j'emploie ce terme parce que c'est un terme anglais. Et effectivement, la définition anglophone se rapproche de ce qu'on veut dire. Je trouve que c'est le terme que les gens ont l'habitude d'utiliser ou en tout cas qui fait écho à ce qu'ils ont envie d'entendre. Alors je dis volontairement « qui fait écho à ce qu'ils ont envie d'entendre » parce que bon il y a l'anglicisme prégnant au quotidien. Donc oui c'est un terme que j'utilise beaucoup mais que moi j'essaye de véhiculer et que j'associe à travers ce dont on a déjà discuté, c'est-à-dire que c'est un terme que volontairement je rapproche très fortement et de manière similaire à ce qu'on appelle « acculturation à la donnée » ou « culture de la donnée » même si en vrai par définition je pense que ce sont des choses différentes ; à quel point et quel détail près est-ce que c'est différent, je sais pas si je serai vraiment capable de définir et de mettre le curseur.

Par contre, ce que je suis capable de te dire, c'est que je pense que *data literacy* fait partie et est englobée dans ce que j'appelle l'acculturation de la donnée qui est englobée dans ce que j'appelle la culture de la donnée. Je ferai un peu trois cercles un peu comme ça. Mais c'est un terme que j'utilise parce qu'il fait écho facilement aux gens. Peut-être moins que si on parle justement de culture de la donnée, même si l'analogie avec la culture d'entreprise peut se faire assez facilement. Mais ça parle moins en fait. Alors est-ce que c'est parce que c'est un terme qui est très fortement véhiculé dans la littérature ou dans des posts ou sur les réseaux type *Twitter*, *LinkedIn* et compagnie ? Très certainement. Est-ce que c'est la vocation due au fait que beaucoup utilisent cet anglicisme volontaire ? Très certainement. Je ne sais pas. Mais ça serait intéressant de voir pourquoi est-ce qu'il y a une prégnance assez rattaché à ce terme-là plus qu'à d'autres.

Ok, et du coup, comment tu le définis ce terme ?

Si on essaye de trouver une définition simple, ça s'apparente à à la fois une et plusieurs compétences qu'il y a derrière mais surtout à une capacité à comprendre un environnement dans lequel on a basculé, pour essayer de résumer en une phrase. Comprendre un environnement dans lequel on a basculé qui est beaucoup plus orienté technologique et qui du coup génère davantage de données. Au final, on pourrait se dire que les données existent depuis toujours puisqu'on a toujours collecté des données.

Je donne comme exemple à mes étudiants : dans une boîte vous avez toujours eu des données, le cahier de comptes de boutique de votre arrière-grand-mère c'est une base de données au final : c'est juste qu'elle est pas de version numérique, digitalisée, qu'on peut consulter à travers un ordinateur, etc. Mais au final elle est

répertoriée, elle est catégorisée, elle est tout ce que vous voulez au final. Donc au final, c'est cette capacité à pouvoir comprendre le monde dans lequel on a basculé : données, des grosses masses de données, mais aussi typologie / variété, donc tout ce qui est *big data*, et comment est-ce qu'elles sont collectées, comment est-ce qu'elles sont utilisées, quels sont les enjeux qu'il y a derrière ?

Après c'est une définition simpliste mais à travers laquelle je vois plusieurs niveaux. C'est-à-dire qu'il y a le premier niveau de la prise de conscience et de la compréhension. Pour moi, il y a le deuxième niveau de l'appropriation. C'est-à-dire qu'on est capable finalement d'avoir cette première base de compréhension de ce monde-là et de comment c'est collecté et à quoi ça sert. Et l'appropriation, c'est-à-dire qu'on est capable finalement de développer des compétences de manipulation, d'articulation, de jouer avec, etc. Là je fais un peu l'analogie avec un *data analyst*, c'est-à-dire quelqu'un qui est capable finalement d'utiliser un jeu de données, d'utiliser des outils qui sont plutôt interfacés, plutôt simples d'utilisation, qu'on utilise beaucoup aujourd'hui dans des entreprises type on peut citer *Power BI*, *Tableau*, *Google Data Studio*, *Dataiku* et tout ça. Donc, qui permet de manipuler, d'ouvrir *Excel* tout simplement, de manipuler des données et de pouvoir comprendre finalement ce qu'il y a dedans, comment est-ce qu'elles sont articulées, comment est-ce qu'elles sont structurées, de se poser la question à quoi elles correspondent. Cette nécessité de renvoyer à quoi elles correspondent, à quoi correspond son contenu et en termes de métier à quoi ça correspond, quelles données on a collecté, à quoi elles servent. Et puis de mener une première analyse, en tout cas une première exploration.

Et puis pour moi, la troisième étape, c'est cette compétence et cette capacité d'aller plus loin. C'est-à-dire d'être dans une démarche où on peut utiliser des outils plus haut niveau, de manipulation, d'articulation, de jouer avec les données et là on passe rapidement sur de la programmation, du code et des solutions finalement moins propriétaires, plus *open source*, même si sur le deuxième niveau on peut utiliser des solutions *open source* mais il y en a un peu moins qui se déploient.

Et du coup cette compétence de pouvoir là jouer, articuler, faire un peu ce qu'on veut et de contrôler. Finalement ce qu'on fait parce que l'enjeu c'est que derrière ces solutions type interface, on sait à quoi ça correspond, on peut savoir à quoi ça correspond mais on n'est pas toujours capables de maîtriser exactement comment ça fonctionne, pourquoi ça fonctionne, même si aujourd'hui sur des outils type IA, on n'est pas toujours en capacité, quand on déploie des modèles, de tout contrôler et de savoir tout ce qu'il se passe. Néanmoins, on a un peu plus quand même le contrôle.

Voilà donc tu vois : multi-niveaux. Fait un peu simple parce qu'on pourrait déployer dans chacun des trois niveaux des sous-niveaux, mais c'est comme ça que je le définis. Et pour ça je m'appuie pas mal aussi sur les travaux d'*Open data institute*, avec les niveaux et les compétences, c'est ce qu'on utilise pas mal aussi au sein de *mydata*, du coup structure associative internationale dans laquelle [nom de personne] fait partie mais moi aussi et pour laquelle on a monté justement le groupe de travail *data literacy* mais l'angle dans *mydata* qui s'intéresse à plutôt qu'est-ce que la *data literacy* pour les données personnelles, ce qui est encore un champ spécifique pour lequel on pourrait appliquer tout ce qu'on vient de se dire là.

Si on devait définir les caractéristiques de la *data literacy*, est-ce qu'on parlerait d'un ensemble de compétences, de connaissance particulières, d'un état d'esprit, tout à la fois ?

C'est tout à la fois en fait pour moi : ça fait appel à des compétences effectivement. Mais encore une fois développer ou utiliser des compétences en-dehors d'un cadre et en-dehors d'une compréhension, en-dehors quelque part d'une culture finalement, c'est comme, je vais essayer de trouver une analogie, c'est comme utiliser un couteau mais ne pas savoir cuisiner. C'est un peu ça : on a la compétence de savoir un peu découper des trucs mais si on sait pas qu'en fait finalement couper des trucs, on peut couper du bois, on peut couper des légumes, couper des choses différemment, etc., ça n'aura pas le même effet et ça ne sera pas tout à fait pareil, le geste sera pas tout à fait pareil donc il faudra l'adapter.

Donc c'est un peu ça finalement : ce sont des compétences, mais des compétences qu'il faut développer vis-à-vis soit d'une culture, soit d'un sujet spécifique et finalement ces compétences qui vont devoir s'adapter soit à ce sujet spécifique, soit à cet environnement spécifique. Et donc c'est un ensemble finalement. Ce sont des compétences, une capacité à comprendre un environnement, une capacité de s'adapter, une capacité de capter des choses et quelque part une capacité à interagir et à connecter parce que c'est comprendre des ramifications parce que tous ces outils data sont toujours au service de quelque chose. Il y a des chercheurs qui font de l'IA et de la donnée pour faire de l'IA et de la donnée mais ils sont quand même un nombre très restreint. Et on sait que c'est toujours au service d'optimiser une action, une tâche spécifique de faire de la prédiction sur des solutions spécifiques, etc. Mais voilà, c'est toujours au service de. Donc pour moi, ce sont des compétences mais qui s'encadrent finalement dans une culture, dans une nécessité d'adaptation, dans un contexte qu'il faut remettre.

D'un point de vue conceptuel, il existe un certain nombre de concepts dans l'écosystème des données : la gouvernance des données, la *data science*. Arrives-tu à faire un distinguo entre ces concepts et à savoir où s'y imbrique la *data literacy*?

Alors, il y en a beaucoup de concepts et d'éléments. Pour moi je dirais : la *data science*, ou la science des données, ça se définit par elle-même, c'est une discipline. C'est cette discipline d'être en capacité d'analyser, de manipuler des données, pas seulement d'un point de vue technique ! Pour moi la science des données c'est aussi tous les enjeux, des enjeux géopolitiques, des enjeux juridiques, des problématiques éthiques de morale, de tout ce qu'on veut mettre là-dedans, de déploiement, d'implication, de nécessité de gérer les ressources financières, humaines aussi, dans cette intégration. Voilà, science des données, ça implique, ça inclue l'IA, ça inclue tout ça. Et pour moi c'est une vraie discipline qui va venir se connecter avec autre chose.

Mais quelque part la *data literacy*, si on en revient à cette segmentation par niveau et par échelon, si on considère ce premier socle de compréhension, je le sortirai en fait de la discipline et j'en ferai plutôt comme un liant et une connexion avec les autres champs d'activité, l'environnement, les autres disciplines. Si on prend juste le socle de compréhension et de compréhension de l'environnement, de ce qu'il se passe et de comment ça marche et pourquoi ça fonctionne comme ça, quelque part, même si ça peut faire partie de la *data science*, je le vois plutôt comme un liant et un connecteur finalement, avec un pont nécessaire qui devient quasiment une nécessité du quotidien aujourd'hui au final. Ce qui n'empêche que ça concerne juste ce premier niveau et que dès qu'on rentre dans les niveaux supérieurs de ce qu'on a défini, ou de ce que j'ai essayé de définir tout à l'heure, finalement on bascule très rapidement dans cette bulle finalement qu'on appelle science des données.

Donc je ne mettrais pas *data literacy* exclusivement dans la bulle science des données, j'en sortirai le premier niveau justement comme quelque part un fondamental à développer aujourd'hui. Alors je dis à développer parce qu'au final on se rend compte il y a beaucoup de gens qui utilisent beaucoup de services et de solutions sans même comprendre ce qu'il y a derrière et comment ça fonctionne et on en arrive à ce qu'on observe aujourd'hui, à toutes les problématiques de pratiques mais aussi de dérives et d'abus et d'arnaques qu'on observe. Donc là c'est le lien *data science – data literacy*. Après tu évoquais d'autres termes, c'était lesquels ?

Alors j'évoquais la gouvernance de la donnée mais si tu as d'autres termes en tête et qui sont importants, n'hésite pas à les dire.

Alors moi la gouvernance de la donnée, je la mettrais dans la science des données parce que finalement c'est une des briques de la science des données, de la discipline, comme on pourrait avoir le management des données d'un point de vue tech. J'ai beaucoup de termes en anglais parce que j'ai tout ce qui est *data policy*, donc surtout des règles de gouvernance, des règles d'utilisation. Donc tout ça je le mettrais dans science des données, comme vraiment la discipline.

Et après quels autres termes ça peut faire écho ? J'en ai parlé : tout ce qui est IA, je le mets dans science des données. Je le mets dans cette bulle qui est une très grosse bulle. Soyons d'accord, c'est énorme. Néanmoins ce dont je me rends compte : ce n'est pas parce que c'est une bulle qu'il faut la voir comme une bulle isolée. Vraiment, quand on voit les disciplines, on voit bien quand on parle, je ne sais pas, de biologie, de neurosciences, ou autres, on voit bien que les bulles sont interconnectées, il y a toujours des zones de recouvrement, il y a toujours des zones où on parle des deux sujets, des deux disciplines ensemble et qu'on peut pas parler de l'une sans l'autre.

Donc on a aussi la bulle qui va se rapprocher sur science des données : on a la bulle technologique, plutôt ingénierie, c'est-à-dire ingénierie de comment fonctionne un téléphone, de comment fonctionne un ordinateur, de savoir ce qu'est le réseau internet, comment est-ce qu'il fonctionne, comment il échange des données, etc. Et qu'on oublie souvent que si on n'a pas un câble ou si on a pas un réseau ou si on n'a pas une antenne, tous ces échanges de données-là n'existent plus. Donc quelque part, de comprendre une bande passante et qu'on ne peut pas mettre dans un tube ou dans un tuyau toutes les données qu'on voudrait, etc., fait partie aussi de la compréhension nécessaire même si c'est très ingénierie en fait, même si c'est très sciences de l'ingénieur. On va avoir un *overlap* tout comme on a un *overlap* avec évidemment la gestion des entreprises, tout comme on a un *overlap* avec la perception, la sociologie, tout ce qui est *design* aussi, *UX*, et au final, tout ça pour dire qu'il ne faut pas le voir comme une bulle isolée volontairement à mettre de côté : il y a toujours des *overlap* avec d'autres aspects et des connexions à créer.

Au niveau de son importance pour une organisation, comment arrives-tu à qualifier l'importance aujourd'hui d'être acculturé aux données, d'avoir une certaine *data literacy*, comparé à d'autres choses qui existent déjà ? J'entends par-là une certaine culture informationnelle, une certaine culture informatique, technique… comment vois-tu les choses ?

Là encore je dissocierai en fait les niveaux dont on a parlé juste avant puisque toutes les organisations n'ont pas vocation à développer des modèles, à créer des outils, à analyser profondément des données, pour revenir à la définition, à la distinction par échelon précédente. Le niveau trois n'est pas nécessaire, voire même le niveau deux même si analyser une veille ou de voir ce qu'il se passe à titre informationnel par exemple, c'est toujours nécessaire. Et ça nécessite aussi un niveau d'analyse et une compréhension mais pas forcément un niveau de manipulation d'outils spécifiques. Néanmoins le premier socle, j'en reviens, de compréhension de comment on génère des données, de comment elles sont générées, comment est-ce qu'elles perturbent ou améliorent ou font avancer aussi les choses en termes organisationnels, en termes de fonctionnement puisque aujourd'hui les entreprises se réorganisent, elles s'adaptent à ce qu'il se passe au niveau technologique.

Et on le voit bien au final, cette réorganisation impacte énormément les interactions entre les gens, les modes de fonctionnement et on voit bien que le monde informatique numérisé, ou digitalisé, je n'aime pas trop le mot digital pour ça, tel qu'on le connaît aujourd'hui a énormément impacté le mode de fonctionnement des organisations. On compare juste le monde de fonctionnement des organisations il y a cinquante ans avec celui d'aujourd'hui, ça n'a rien à voir en termes d'échange d'information, en termes d'interaction, en termes de mode de fonctionnement, on voit que ça a beaucoup évolué, que ça a fait changer les choses. Donc ce premier socle, il me semble à mettre au même niveau aujourd'hui que savoir dans une entreprise ou une organisation manipuler des outils de base type un ordinateur pour ouvrir un e-mail, ou pour savoir ouvrir des fichiers de type document *Word*, document office ou fichier *Excel*, etc. Quelque part, cette base, elle est nécessaire. Mais tout comme au quotidien au final ! Là je sors des organisations.

Au quotidien, utiliser un téléphone, il y a plus beaucoup de téléphones non smartphone aujourd'hui. Ça existe toujours mais au final, nous qui avons grandi dedans, grandi avec ces projets technologiques, quelque part on a cet avantage d'avoir vu l'évolution et d'avoir suivi un petit peu pourquoi ça fonctionne comme ça et pourquoi finalement c'est intéressant d'avoir ces nouvelles solutions, ces nouveaux services. Mais c'est vrai que les nouvelles générations sont nées directement avec : avec un ordinateur, avec des smartphones, etc. Et n'ont pas toujours ce réflexe de se dire « oui mais comment ça fonctionne et pourquoi est-ce qu'on a fait ça ? Il y avait quoi avant ? ». Donc au final c'est quelque peu de l'histoire de la science des technologies et du progrès. C'est toujours intéressant de remettre en perspective la perspective passée et la perspective future à venir de comment est-ce que ça a évolué et comment est-ce que ça a fait bouger les choses.

Parce que ce qu'on observe, c'est que ce monde et cet univers dont on parle depuis tout à l'heure de la donnée, il est beaucoup plus récent que ce qu'on imagine. Il bouge beaucoup, ça bouge très vite. Mais on oublie souvent qu'au final l'utilisation massive des données, la réutilisation massive des données, les outils et les services qui existent aujourd'hui, ils ont vingt ans, à la louche. Mais Internet, c'est 80, 90, à titre chercheur en plus, donc le temps que ça se déploie avec les premières connexions grand public au début des années 90. Donc remit dans l'ensemble, c'est trente ans. Et trente ans à l'échelle d'une vie, c'est que dalle. Et à l'échelle de ce qu'on vit, le temps que ça se déploie, et que ça impacte et que ça génère tous ses changements, on les voit aujourd'hui au final. On les a vu il y a dix ans et on les voit aujourd'hui en termes d'impact, d'adaptation et de nécessité.

Tu me fais penser du coup à la question que je voulais poser juste après : est-ce que tu arrives à situer dans le temps depuis quand c'est devenu une nécessité, en tout cas depuis quand on en parle des données en tant que telles ?

J'en reviens à ma discussion au début : ça existait depuis toujours dans le monde d'ingénieur. On a des outils pour mesurer, pour collecter des données, pour faire des études scientifiques, pour suivre l'état d'une machine, ou d'un robot industriel, ou d'une voiture, la manière dont fonctionne une voiture, ça on l'a. Mais c'est le moment où tout devient un peu interconnecté. Et quelque part, finalement, je me rends compte que l'émergence de cette nécessité, elle apparaît le jour où l'utilisateur individuel a accès soit aux données, soit partage ses données sans le savoir, soit a accès à ces outils et ces services. Et donc c'est quelque part l'avènement de ce que j'appellerai la démocratisation et le déploiement à titre individuel de l'accès au web à la fin des années 90, mi-fin des années 90, avec une vraie démocratisation début des années 2000. Quand je dis, mais il faudrait regarder les chiffres, que le déploiement à hauteur de 60, 70% de foyers connectés ça doit être à peu près à cette époque-là fin des années 90 le vrai déploiement, la vraie démocratisation.

Et donc pour moi, quelque part avant c'était plutôt un enjeu et un sujet d'ingénieur ou de spécialiste tu vois parce que les financiers ont toujours eu accès à des données, tout comme ceux qui pilotaient des entreprises à la marge, à des transactions ou des choses comme ça. Donc ça a souvent été un sujet d'experts jusqu'à cette démocratisation d'accès et de déploiement où finalement on s'est rendus compte qu'on pouvait collecter énormément de données sur les individus, de la manière dont ils utilisent les outils, la manière dont ils utilisent les navigateurs, la manière dont ils utilisent des applications, des solutions, des services, et de se rendre compte « ah tiens on pouvait utiliser ces données pour soit améliorer notre propre mode de fonctionnement, améliorer notre service, ou vendre de nouvelles solutions ou mieux cibler, etc., donc là on est clairement dans le marketing. Et quelque part cette démocratisation a aussi généré cette nécessité de démocratiser le comment ça fonctionne.

Qu'est-ce qu'une donnée pour toi ? Comment définis-tu ce qu'est la donnée ?

Qu'est-ce qu'une donnée ou la donnée ?

La donnée.

Parce que une donnée, je fais la distinction, si tu as le temps on peut le faire. Une donnée, on peut la définir par plein de manières différentes : par son contenu, sa nature, etc. Mais une donnée, j'aime bien faire référence notamment à un bouquin « data revolution » de Rob Kitchin, qui dit du coup : une donnée c'est une brique, un bloc, une représentation particulière d'un bout du phénomène que vous avez observé. Et si vous connectez tous les bouts ensemble, tous les blocs de toutes les représentations ensemble, vous avez créé une information qui va être une partie de la représentation d'une réalité, d'un bout de phénomène, etc. et ça c'est vraiment une donnée.

La donnée ça fait souvent référence à, et je pense que c'est un peu comme l'IA, c'est-à-dire que quand on dit la donnée c'est une définition un peu glissante à travers le temps. C'est en fonction des progrès techniques et tech-

nologiques, tout comme avec l'IA. Ce qu'on définit comme l'IA aujourd'hui, ce qu'on définissait comme IA il y a 10 ans, il y a 15 ans, il y a 20 ans, ce n'est plus la même chose que ce qu'on définit comme IA aujourd'hui.

Du coup, il y a eu un curseur, il y a eu un glissement technique et technologique, qui fait que ce qu'on appelait IA il y a 20 ans : aujourd'hui on appelle ça outil informatique ou statistique ou analyse de données et puis finalement la définition de l'IA a glissé pour quelque chose de technologiquement ou de techniquement plus élevé avec des capacités plus hautes. Donc pour en revenir à la donnée, si tu veux je le mets dans un cadre où aujourd'hui quand on parle de la donnée, ça englobe pour moi le phénomène *big data*. En général quand les gens parlent de data, ils parlent de *big data*. Et donc ils parlent de cet enjeu de toutes les données massives que l'on collecte, qui sont générées au quotidien, qu'on peut utiliser, qu'on peut manipuler, avec lesquelles on peut jouer, etc.

La donnée, ça fait référence à tout ce que l'on trouve et tout ce qui est généré aujourd'hui au quotidien, quelles que soient les thématiques, quels que soient les outils utilisés. Et donc derrière évidemment ça fait appel aux outils, ça fait appel à toute l'analyse qu'on est capable d'en faire, mais ça fait aussi appel à toutes les données qu'on collecte et qui ne servent à rien, ou qu'on sait pas utiliser, ou qu'on ne sait pas qu'on collecte. Donc quelque part ça fait appel à tous les enjeux qu'il y a derrière. Donc la donnée aujourd'hui c'est plus un ensemble et un environnement plutôt qu'un objet spécifique. L'objet spécifique c'est une donnée et c'est la massification de cet ensemble qui a donné la donnée. Alors ça se trouve demain la donnée aura une autre définition et un autre cadrage parce que techniquement, technologiquement, on aura *shifté*. Ou pas que techniquement ou technologiquement mais en termes de cadrage on aura *shifté* aussi et on aura recadré. Là on voit tout ce qui se passe au niveau européen avec les mouvements de « qu'est-ce qu'on aura le droit de faire, quel est le cadrage à la fois en termes de politique publique ? en termes juridiques, en termes de gouvernance des données, ce que les organisations, les entreprises ont le droit de faire, sont capables de faire, sont cadrées dans ce qu'elles peuvent faire par rapport à tous ces mouvements de données finalement ». Donc la donnée quelque part c'est cette référence à cette massification de collection de données qui est générée et qui est notre quotidien aujourd'hui et qui fait appel à toutes ces ramifications d'enjeux, de nécessité, d'analyse, de nécessité de collecte, d'outils à la fois techniques, non techniques, de gens qui sont derrière pour pouvoir jouer avec, décider que c'est important de la collecter, de pas la collecter. Je pense que j'ai fait le tour.

Quels sont les grands enjeux pour toi autour des données et est-ce que la *data literacy* peut y répondre ?

Pour moi les grands enjeux, ils ne sont pas tant liés à la technicité et à la technique et aux outils parce qu'aujourd'hui si on regarde les outils techniques qui se déploient, les fonctionnalités, en vrai ça marche. Ça marche très bien. On peut faire à peu près tout ce qu'on veut en termes de manipulation et d'analyse. Mais ce qui pose problème c'est souvent l'avant. Je m'explique : pour moi un des grands enjeux, alors ce n'est pas par ordre de priorité même si on pourra, si tu veux, les ordonner, mais pour moi un des enjeux majeurs aujourd'hui c'est de sortir du paradigme dans lequel finalement on a grandi parce que ce n'est pas si vieux, qui est que la donnée est un sujet technique. L'idée de ce paradigme là c'est que pendant longtemps quand on va dans une organisation, une entreprise, un institut, une collectivité ou l'État tout simplement, quand on parle de données, on nous renvoie systématiquement vers le DSI. Finalement la problématique c'est qu'on a

toujours pensé que tous les sujets autour de la donnée, vu que ça provient d'outils technologiques de collecte, c'est un sujet tech ou c'est un sujet d'ingénieur donc on revoie systématiquement à la DSI.

Pour moi l'enjeu il est qu'il faut qu'on se sépare de ce paradigme parce que finalement je regroupe avec ce que je disais tout à l'heure, finalement tous ces outils sont au service d'un métier, d'un sujet, d'une thématique. Et donc, le paradigme dans lequel on devrait basculer c'est ce paradigme dans lequel la donnée est un sujet thématique ou métier qui doit être piloté par les métiers, en tout cas de ceux qui collectent cette donnée et en s'appuyant sur le support de la DSI, de l'IT, pour pouvoir se poser des questions : quels sont les meilleurs outils pour les collecter, quels sont les meilleurs outils pour les analyser, quels sont les meilleurs outils pour les structurer, pour les stocker, etc.

Par contre le sujet quelles données on collecte, qu'est-ce qu'on collecte, avec quelle précision, comment on la collecte, à quel moment, pourquoi, où, etc., c'est un sujet thématique métier, cela ne doit pas être un sujet tech en fait. Parce qu'au final, vous pouvez vous tourner vers un ingénieur, un *data scientist* pur ingénieur : que tu lui donnes des données covid ou des données de comptage de patates, ou des données de comptage de cochons, c'est pareil pour lui. Il va prendre les données, il va les analyser, il va les structurer. J'exagère un petit peu : le paradigme commence à bouger mais cette analyse, cette collecte, la manière dont on les collecte, à quoi ça va servir, comment, pourquoi, c'est un regard métier qu'il faut avoir. Ce n'est pas un regard tech ou statistique ou analytique.

On l'a vu par exemple avec le covid : on a vu plein de gens interpréter les données de covid et il y en avait pas beaucoup qui étaient médecins, et du coup on peut les interpréter dans un sens ou dans un autre et au final tout le monde aura raison si on regarde que les données. Mais le problème c'est qu'il ne faut jamais regarder que les données : il faut regarder le contexte, l'environnement qui va avec, qui les a collectés, où, comment, pourquoi, etc. Donc ça c'est un premier sujet et le premier enjeu c'est de sortir la data à tous les sujets de collecte, de qu'est-ce qu'on veut, comment on le veut, des besoins de données de la DSI et de l'IT. Ça c'est le premier enjeu.

Ça commence à bouger et c'est le deuxième enjeu pour moi dans les organisations : des gens qui sont capables d'avoir ce double regard. Un regard thématique métier-sujet et un regard de savoir échanger et interagir avec cette DSI et ces IT pour finalement comprendre et se connecter, créer des passerelles. Parce que le problème c'est que si on ne sait pas échanger et on ne sait pas interagir, on ne va pas se comprendre. Et si on ne se comprend pas chacun va faire un peu à sa sauce et évidemment ça ne va jamais marcher. Donc quelque part c'est créer cette passerelle et créer cette connexion du coup en termes d'échanges avec les équipes métier, thématiques, techniques, et les équipes métiers data, techniques, ingénierie, etc. Donc ça c'est vraiment quelque part le premier enjeu. Et en déployant et en accentuant la culture ou l'acculturation à la donnée, la *data literacy*, on va favoriser ce changement et ce paradigme.

Alors je disais que ça a déjà commencé puisqu'on en parle de plus en plus. Avant, on parlait data, *big data*, ça faisait plein de choses, oui mais ça restait tech, data c'est ingénieur donc ce n'est pas grave. Et puis avec l'émergence de cette capacité à aller fouiller dans ses données, de scandales qui au fur et à mesure sont ressortis « tiens, on utilise les données ! Tiens, pourquoi est-ce que tel acteur a telle donnée sur moi ? Pourquoi ?

Comment ça marche ? ». Donc on a un peu fait émerger des enjeux autour de « qui a ma donnée ? Pourquoi ? Comment ? Etc. ». Je ne dis pas que c'est tout le monde, je ne dis pas que c'est une grande vague monstrueuse. En tout cas les gens sont de plus en plus sensibles à « tiens, la donnée… mais c'est quoi ? Ok, ça contient quoi ? Mais comment ? Pourquoi ? ». Et au final, c'est par cette première étape que je trouve qu'on commence à couvrir ce premier niveau d'acculturation, de compréhension, de ce que c'est, à quoi ça sert, comment, pourquoi. Et ça, ça permettra pour moi d'être un liant et d'être une première marche pour faire en sorte que les gens vont comprendre « cette donnée j'en ai besoin pour faire ça ! Pour faire telle action ou pour mesurer telle chose, tel phénomène » et donc après je vais connecter avec les équipes tech. Donc ça participe finalement à ce déplacement de paradigme. Donc premier enjeu.

Deuxième enjeu, je suis un peu utopiste dans l'âme et optimiste mais c'est de quelque part clarifier ce qui est collecté, comment, pourquoi. Je parle de données personnelles, certes, mais au final c'est en insufflant cette acculturation à la donnée que les gens vont être capables de pouvoir dire « ok finalement je comprends ce qui est collecté, comment c'est collecté » à la fois à titre individuel mais aussi à titre d'acteur membre d'une organisation que ça soit une entreprise, une collectivité, une instance X ou Y, et de comprendre que quand j'organise un évènement et que je note des noms, il faut peut être le faire de manière plus structurée, de manière un peu standardisée, et quelque part structurer cet ensemble. Deuxième enjeu qui va derrière, qui va avec quelque part, c'est arrêter de croire (même si on l'a déjà commencé, encore une fois, ça *shifte*) la donnée, un petit peu ce qui a été vendu avec le *big data*, avec tous les rapports des grands acteurs et des grands acteurs conseil de cabinets de conseil.

Donc à l'époque : *big data*, fin des années juste avant 2010 et début 2010 avec tous les rapports qui sont sortis du type « non mais le *big data* c'est génial ! ça va nous faire gagner des milliards ! », « si vous vous mettez à la data, tout le monde va s'y retrouver et vous gagnerez plein de sous », etc. Et au final, ça a été le cas pour certains et ça n'a pas été le cas pour d'autres. Et au final de sortir de cette promesse un peu surréaliste du « mettez-vous à la data et ça va marcher » : ben non, ce n'est pas comme ça que ça marche. C'est encore une fois la donnée au service de et les outils et l'analyse au service de pour finalement faciliter la compréhension d'un phénomène observé, faciliter les prises de décisions, parce qu'au final c'est toujours ça. On collecte de la donnée pour pouvoir générer des informations et faciliter cette prise de décision. Donc la nécessité c'est de recentrer le débat autour de « pourquoi on collecte la donnée, à quoi elle va nous servir, et comment on se projette là-dedans en fait ? Comment est-ce qu'on va l'utiliser par la suite ? Et comment est-ce qu'on l'intègre dans notre stratégie ? ».

Alors je parle en termes d'organisation et en termes d'entreprise mais on peut tout à fait aussi l'avoir en d'autres termes. Il y a eu un projet que nous on a mené, on avait pas mal échangé avec les membres d'*AMTO* par exemple qui est la structure et l'organisme de l'observatoire de la qualité de l'air en Auvergne-Rhône-Alpes mais du coup il y a un *AMTO* par région. Et on avait discuté d'un projet qu'ils avaient lancé et qu'ils avaient mené notamment autour de donner un capteur d'air à tout le monde et on s'était dit « mais comment on peut déployer un projet de ce type-là ? ». Et c'est un projet notamment qui tourne aujourd'hui à la [nom d'organisation] à [nom de ville] avec chacun qui peut avoir un petit capteur dans son sac à dos et du coup on collecte des données, on mesure, etc. Et l'analogique qui avait été faite, c'était de donner un petit capteur,

notamment de mesure de réseaux qui passent à travers la maison. Donc : réseaux wifi, internet, ou réseaux téléphoniques, etc.

Et chez certaines personnes, ça avait généré une phobie des mesures donc elles continuaient à mesurer absolument tout le temps alors qu'avant elles n'avaient aucune réalité des réseaux qu'il y avait. Et chez certains ça a même été poussé loin, que ça a dérivé vers une phobie de « non il y a trop de réseaux ! Il faut que je me coupe de tout ! ». Et le fait de mesurer mettait le doigt sur l'existence de quelque chose. Le cadrage et l'enjeu qu'il y a derrière c'est : mesurer, ok. Observer ? ok. Mais pour en faire quoi ? Comment ? Et qu'est-ce qu'on va en faire après ? Il y a un vrai enjeu derrière parce que le numérique c'est à la fois une solution et un problème d'un point de vue environnemental. Par exemple, toutes les données qu'on collecte, il faut bien les stocker, il faut bien les organiser, il faut bien les maintenir, etc. Mais on sait que la plupart des données qu'on collecte, elles servent à rien parce qu'elles ne sont pas de bonne qualité, parce que les gens ont commencé à la collecter puis ils ont changé de projet et ça ne sert plus à rien, elles sont tombées à l'oubliette. Ou parce qu'on ne sait pas qu'on les collecte. Et au final, on ne sait pas trop ce que ça devient. Donc quelque part, re-réfléchir et re-reposer le cadre de « ok les données c'est bien mais à quoi ça sert, pour qui, pour quoi ? », cela permettrait de recadrer et de reposer les choses. Je t'ai donné plein d'enjeux, je t'en ai donné deux principaux je pense, et quelque part la *data literacy*, et l'acculturation à la donnée, elle est UN des leviers pour faciliter ce recadrage même si cette *data literacy* pour le changement de paradigme elle va davantage jouer sur la redéfinition de qu'est-ce qu'on doit connecter, comment, où, pourquoi, d'un point de vue stratégique parce que ça c'est d'un point de vue : qu'est-ce qui est utile et qu'est-ce qui est pas utile. Ça participe à l'acculturation à la donnée mais il n'y a pas que ça.

Après ce que tu viens de dire, est-ce que tu identifies des types de données en particulier qui sont au cœur de ces enjeux ?

Il y a un enjeu autour des données personnelles sachant qu'en termes de perception c'est assez vague et j'ai des étudiants qui sont assez intéressés, qui commencent à s'intéresser à ça dans un de leurs mémoires, à se dire « quel est l'équilibre entre je suis conscient que je partage mes données personnelles mais j'accepte quand même parce que l'outil que j'utilise me sert à quelque chose ou je suis dans un déni de conscience de partager mes données parce que j'ai quand même envie d'utiliser ce service et cette solution ». Donc on est un peu là autour des usages du numérique et sur des études autour de la psycho-sociale qui sont super intéressantes mais pour revenir à cette nécessité des typologies de données : les données personnelles ça c'est clair c'est un vrai sujet. Après néanmoins, on est sur une frontière autour des données, on est sur un questionnement qui va être nécessaire à un moment donné sur qu'est-ce qu'on entend comme données publiques, données à partager, données accessibles. Alors, quand je dis accessibles, je pense que tous les gens que tu as dû rencontrer ont dû mettre trois couches dessus : accessibles ne veut pas dire je mets juste un pdf en ligne et c'est bon tu y as accès. Non, accessible ça veut dire dans un format qui est peut être standardisé mais en tout cas qui est accessible au sens où je vais extraire chacun des éléments, je vais pouvoir manipuler, je vais pouvoir aller jouer avec les données qui sont partagées, en tout cas je vais pouvoir les restructurer, les analyser, et je vais pouvoir les comprendre.

Typiquement accessibles je donne un exemple : *Legifrance* c'est accessible. Après si on n'est pas juriste, c'est complexe comme sujet et comme facilité d'accès à la donnée et à l'info. C'est pareil, je donne un autre exemple. Le rapport du *GIEC* : ça c'est accessible, on peut le trouver facilement, il n'y a pas de problèmes. Par contre, être en capacité de se farcir les 518 pages du rapport, ce n'est pas très accessible comme contenu on va dire. Ni les rapports de l'*ONU*, ou les choses comme ça sur rendre accessible et rendre facile l'appropriation non pas des données mais du contenu qu'elles véhiculent ou le message qu'on a envie de véhiculer à travers ces données. Donc là il y a un sujet de « qu'est-ce qu'on doit mettre accessible à tous ». Ça fait appel évidemment derrière à toutes les démarches autour des communs numériques.

Et puis après ce n'est pas seulement qu'un sujet de données : c'est aussi un sujet d'outillage. Tout dépend de l'outil qu'on utilise. Ce n'est pas parce qu'on a accès à des données qu'on est capable de reproduireles résultats qu'on a derrière. Donc c'est aussi un accès à une compréhension, une transparence sur ce qui a été utilisé pour générer un résultat. Et comment ça a été utilisé. C'est marrant parce que d'un point de vue académique et de publication, à chaque fois qu'on publie un contenu basé sur des données on nous demande de démontrer la méthodologie de A à Z et finalement de savoir par quelle étape on est passé, quel type de résultat et quel type d'erreur vis-à-vis de ces résultats ou l'incertitude par rapport à ces résultats et d'animer cette discussion. Tout ce qui est fait d'un point de vue académique, on ne le retrouve pas.

Alors, on est d'accord ça brouillerait le message vis-à-vis de ceux qui produisent des outils ou des contenus. Évidemment, c'est difficile à *marketer* de se dire « on pense que notre système de recommandation vous suggère ça. Bon alors, c'est à 98% » mais au final les utilisateurs ils s'en foutent, soyons d'accord. Mais en tout cas de rendre facile et compréhensible ou appropriable cette phase entre les deux : ce n'est pas juste la donnée, ce sont tous les outils qu'il y a derrière et qui sont directement connectés. Si on avait juste la donnée, sans les outils on n'en ferait pas grand-chose. On essaierait d'analyser des tableaux à la main comme on faisait avant d'avoir les outils numériques. Et ça redeviendrait un sujet d'ultra spécialistes.

On a depuis un certain nombre d'années des réglementations sur les données, la plus récente c'est la loi Lemaire et le *RGPD*. Il y a le *data act* et le *data governance act* qui sont en cours de discussion à la commission européenne. Est-ce que tu vois une différence dans les pratiques, les tiennes et celles de tes collègues, depuis qu'il y a cette législation-là et globalement quel est ton avis sur cette législation ?

Le *RGPD* on le voit, très clairement, mais aussi parce que c'est entré dans les mœurs. Parce qu'il y a eu aussi un temps nécessaire d'appropriation, de compréhension. Aujourd'hui on se rend compte que tous les outils, tous les services, toutes les entreprises, toutes les organisations : tout le monde n'est pas *RGPD*. Aligné en tout cas. Donc ça va prendre un temps d'appropriation de ces réglementations. Un temps de compréhension de ce que ça signifie et de comment je dois l'appliquer dans mon contexte et dans mon environnement spécifique. Donc pour moi, c'est une chance qu'on ait ces réglementations.

Néanmoins on le voit très clairement sur les nouveaux textes qui vont être validés : ils sont déjà en partie validés par la Commission. Mais le temps que ça ruisselle et que ça redescende au niveau quotidien, au niveau d'une organisation, que ça soit une collectivité, une entreprise ou autre, je pense qu'on en a au moins pour dix ans, le temps que tout le monde soit au moins label data / *RGPD* / *governance act* / *data act* / *digital service act*,

parce que c'est souvent complexe, très générique et très général. Parce que ça donne de grandes directions et de grandes lignes, mais en termes d'application et de mise en application, c'est très complexe. Je fais l'analogie avec les lois qui sont votées au niveau national, à l'Assemblée. On le voit : mêmes les magistrats, juges et autres disent « on est obligés de s'adapter, on est obligés en fonction du contexte, de l'environnement, des gens qu'on a en face de nous, de s'adapter et on ajuste ces lois et ces définitions par grands critères qui permettent de poser un cadre autour duquel il faut un peu jouer à la marge parfois ». Il faut savoir interpréter, et d'ailleurs c'est pour ça aussi que je pense que les textes en partie sont interprétables. Pas comme on le souhaite mais interprétables en tout cas pour permettre d'ajuster aussi en fonction du contexte et de l'environnement et de la situation que l'on observe ou qu'on est censé juger.

Donc, c'est important mais parfois ce cadre il est un peu loin de la réalité ou en tout cas il est difficilement applicable et posable sur les diverses situations en termes d'organisation, de management, de mode de fonctionnement, d'organisation des services entre eux, là quelle que soit l'organisation, parce qu'il faut changer les modes de fonctionnement, il faut changer les outils qu'on utilise, ça apporte plein de contraintes parfois auxquelles ne pensent pas forcément ces lois et ce cadrage qui vient de tomber. Après je pense que parfois c'est un peu loin des situations et des problématiques locales. C'est bien de cadrer, c'est bien de standardiser. Mais après ça peut créer, générer des situations un peu complexes qui ne permettent pas certaines explorations. Je pense recherches, je pense expérimentations pour explorer davantage, pour voir ce que l'on peut faire, comment on peut faire évoluer les choses, etc.

Je pense qu'on en revient à ce que je disais tout à l'heure : que c'est tellement récent qu'on peut difficilement projeter les usages sans les avoir vu et du coup la réglementation court un peu après les usages. Donc du coup ça sort toujours après, mode de fonctionnement oblige. Et on ne peut pas tout réguler. Il va falloir à un moment peut-être faire appel aussi au bon sens ou à la structuration ou à l'usage en lui-même, mais ça générera toujours des dérives et ça générera toujours des enjeux supplémentaires. Mais encore une fois, c'est important. C'est bien qu'il y ait une prise de conscience aussi parce que le fait qu'il y ait des réglementations et des régulations qui se font, c'est qu'il y a une prise de conscience sur les risques que ça peut générer, les enjeux qui vont avec. Mais parfois ces réglementations sont soit un peu hors sol, soit un peu loin du terrain, ou des usages. En fait c'est banal ce que je dis parce que c'est valable pour les datas et pour tout le reste.

On va passer à la dernière partie de l'entretien avec une question très générale pour commencer sur le volet applicatif : comment peut-on faire pour mettre en place une *data literacy* dans un organisme et est-ce que ça doit concerner tout le monde ?

Encore une fois je renvoie vers mes différents niveaux. Sur le premier niveau, ça doit concerner tout le monde. Ça doit concerner tout le monde tout comme une éducation sur l'évolution des progrès, l'histoire des sciences, etc., normalement ça devrait concerner tout le monde. Donc ça doit concerner tout le monde, même les individus, même si tu ne fais pas partie d'une organisation, parce que de toute façon, si je reviens dans les organisations, la manière dont tu vas rentrer tes éléments et tes données, tes informations dans un système d'information, parce qu'aujourd'hui la plupart du temps c'est comme ça que ça fonctionne, va impacter ce que tu vas être capable de faire, va impacter la qualité de ta donnée.

Donc si les gens ne sont pas informés, ou en tout cas n'ont pas conscience de l'enjeu de ce que eux font, là je parle par exemple d'une personne qui fait partie d'une organisation où on fait des appels téléphoniques et où on collecte des infos. On collecte des données de base et en fonction de la manière dont on va intégrer ces données dans le système d'information, si on ne respecte pas un standard, si on ne respecte pas une structuration, on voit bien que par la suite ceux qui vont les analyser, les utiliser, les manipuler, soit vont considérer « on ne peut pas l'utiliser donc du coup on est obligé soit de les mettre de côté, soit les jeter ou on n'aboutira pas à un bon résultat analytique et ça va se sentir sur la prise de décision. On va pouvoir l'utiliser mais en se voilant la face : oui il y a 50% du jeu de données qui ne sert à rien mais il a mal été collecté et du coup les données sont inutilisables ». Donc au final, on oublie souvent tout cet aspect de données qui ne sont pas de bonne qualité et qu'on ne peut pas utiliser.

De la même manière, quelqu'un qui collecte sur le terrain de la donnée, il faut savoir comment bien la collecter. Après est-ce que c'est un enjeu de former forcément à comment ça va impacter par cascade derrière, d'un point de vue donnée je ne sais pas. Est-ce que ça passe par des formations de bonnes pratiques « vous allez collecter telle info, telle donnée, voilà comment on le standardise et comment on met des cadres et des standards et il faut le suivre, point ». La première solution est peut-être plus facile. Évidemment. Et donc en fonction des niveaux, cette acculturation n'est pas forcément nécessaire, elle ne concerne pas tout le monde même si par curiosité c'est toujours intéressant d'aller voir ce qu'il se passe mais je peux tout à fait comprendre que ce n'est pas nécessaire pour tout le monde. En fonction des métiers, en fonction des individus, ce n'est pas nécessaire pour tout le monde. Clairement pas. Comme pour d'autres sujets je veux dire.

Justement, les personnes qui vont avoir cette formation-là à l'acculturation à la donnée, qu'est-ce qu'elles doivent développer en priorité pour toi ? Des connaissances, des compétences, qu'est-ce que tu mettrais derrière ?

Je le mets dans les compétences. Pour moi le premier niveau, ce sont les bonnes pratiques. Quand on reçoit un jeu de données, c'est quoi la bonne pratique. Qu'est-ce qu'on en fait ? Qu'est-ce qu'on fait ? Quelles sont les premières étapes qu'on doit explorer pour déjà se faire une première idée sur les données qu'on a reçues, sur le jeu de données, qu'est-ce qu'on va en faire, pourquoi, comment ? Donc quelque part c'est une bonne pratique. Et puis, je me suis perdu dans la question, tu peux me répéter la question ?

Quelles sont pour toi les choses primordiales que les personnes qui ont cette formation à la *data literacy*, qu'est-ce qu'elles doivent développer à l'issue de la formation ?

Très clairement : développer des bonnes pratiques. Pratiquer, parce que le fait de ne pas pratiquer fait qu'on oublie. On perd finalement les réflexes, on perd les bonnes pratiques, faciliter la compréhension par les autres donc quelque part c'est une rediffusion, un repartage auprès des autres et faciliter le « oui mais je ne comprends pas pourquoi. Est-ce que on doit faire ça, on fait ça comme ça ? ». C'est cette diffusion de la culture et de la bonne pratique qui fait partie des rôles des gens qui sont formés. Car ils vont être ambassadeurs mais aussi quelque part connecteurs et intermédiaires à la fois avec les gens qui ne maîtrisent pas, ne connaissent pas et ce monde plus tech finalement avec lesquels ils vont connecter plus ou moins en fonction de la formation. Donc diffusion, bonnes pratiques, et puis continuer de mettre en pratique, de définir peut-être des standards, des

modes de faire, et qui fait partie de cette diffusion au final. En diffusant et en transmettant. C'est en essayant de cadrer un petit peu les choses, d'essayer d'adapter par rapport à leurs métiers, à leurs quotidiens, à leurs organisations, à leurs modes de fonctionnement et donc c'est d'adapter ce qu'ils ont vu qui est forcément, très certainement généraliste. Et donc d'essayer de l'adapter à leur quotidien et d'essayer de comprendre comment ça se ramifie avec ce qu'ils font et ce qu'ils ont fait et comment ils le font.

Il y a deux termes sur lesquels je voudrais t'interroger de manière globale : quand on parle d'éthique de la donnée, est-ce que toi c'est quelque chose qui te parle et qu'est-ce que tu lui mets derrière ?

Comment le dire ? Il y a une manière provocante de le dire et peut-être une manière plus diplomatique. En fait, je vais essayer la manière simple : même si je comprends complètement ce qu'on met dedans, je ne suis pas un grand fan du mot éthique. Parce que pour moi le mot éthique, il renvoie à la morale. Au final l'éthique est-ce qu'elle est générique, standardisée ? pour toutes les personnes, toutes les cultures et pour tous les modes de fonctionnement ? Je ne sais pas. Ça se trouve on a chacun notre éthique et notre mode de fonctionnement. Et on trouve que l'éthique de l'autre n'est pas la même et pas la bonne donc qu'est-ce que cela signifie ? qu'il faut standardiser et du coup la conformer pour créer un cadre. Ou qui dit qu'on n'a pas le droit d'aller-là et on fait ça. C'est un peu comme la morale, chacun a un peu la sienne. Donc oui on peut le faire avec des réglementations et des régulations. Je pense que l'éthique de la donnée est plutôt liée aux réglementations qui existent aujourd'hui et qui se déploient. Alors pas au niveau européen mais au niveau du monde en prenant en compte qu'est-ce qu'on peut se permettre de collecter ou pas et dans quelles conditions avec quelles modalités. Ceci est pour mon premier point.

Et puis surtout après, comment est-ce qu'on peut utiliser ? Quels sont les cadres qui nous limitent à ce qu'on a le droit de faire en termes d'utilisation. Un peu comme la bioéthique justement : sur les lois de bioéthique qu'est-ce qu'on peut faire, qu'est-ce qu'on peut ne pas faire et où sont les limites. Et cette éthique-là elle doit être liée à une transparence de ce qu'il y a derrière le *RGPD* avec cette demande de consentement éclairé et d'expliciter, qu'est-ce qu'on fait avec les données ?

C'est pareil : c'est de mettre à voir et de rendre accessible facilement telle donnée et ce qu'on va faire avec. Alors, rendre accessible ça veut dire aussi rendre compréhensible à tout un chacun ce qu'on va en faire et comment on le fait, ce qui n'est pas toujours le cas. Et puis, un des problèmes et des enjeux ce n'est pas tant l'interlocuteur direct avec qui tu vas échanger tes données mais c'est ce qui se passe après. Et est-ce que ce tiers de confiance il va repartager avec quelqu'un d'autre ? Ou est-ce qu'il va finalement les revendre et ça on n'y a pas accès. On ne sait pas au final jusqu'à quel point ça se diffuse et ça se partage et ça s'interagit. Donc l'éthique de la donnée elle a plusieurs niveaux :

- qu'est-ce qu'on peut collecter, comment, pourquoi ;
- qu'est-ce qu'on peut faire avec ;
- comment l'expliciter ce qu'on peut faire avec ;
- et qu'est-ce qui se passe après finalement, quel est le cycle de vie de la donnée, c'est je la garde combien de temps, pour qui, pour quoi, comment, où.

Mais de toute façon, même en faisant tout ça, je ne suis pas sûr que les gens vont le faire. Exemple basique : typiquement aujourd'hui, moi je fais très attention sur les navigateurs, les réseaux, on accepte les cookies, etc. Ma compagne ça fait dix ans, j'ai beau lui expliquer chaque fois, premier truc qu'elle fait « j'accepte, de toute façon ça prend 15 centièmes de seconde, c'est plus rapide que se dire où est-ce qu'il est le bouton de je continue sans accepter ? ». Et pourtant, je lui explique tous les jours. Mais, pourquoi est-ce que je donne cet exemple, c'est pour essayer de montrer le niveau à la fois d'ambiguïté et de paradoxe qui est derrière : c'est que on aura beau expliquer, on aura beau former, machin, il y aura quand même en termes d'usages, et en termes d'appropriation de l'usage de ces outils-là, un vecteur à générer qui va être hyper nécessaire sur « comment est-ce qu'on facilite », tu vois il y aura beau y avoir le *RGPD*, il n'y aura pas le bouton « j'accepte » en gros puis le « continuer sans accepter » en haut à droit qu'on ne voit pas parce qu'il est en police 8. Du coup, blanc sur fond blanc, j'exagère à peine. Mais voilà, il y a aussi des pratiques qui vont avec, des usages qui vont avec, et d'être plus transparents sur les pratiques et les modes de faire mais ça je dirais que c'est plutôt lié aussi à un paradigme plutôt capitalistique de collecte, de conquête, de concurrence, de capacité à vendre un service.

Je vais évoquer la posture critique, l'esprit critique vis-à-vis des données. Pareil : est-ce que ça te parle et si oui qu'est-ce que tu mets derrière ?

Posture critique vis-à-vis de la donnée, je reviens à ce que je disais tout à l'heure : pour moi ce sont des compétences qu'il faut avoir, qu'il faut développer au niveau un et qu'il faut avoir au niveau deux. C'est-à-dire cette capacité d'évaluer une donnée par rapport à ce qu'on veut en faire. D'évaluer, d'être capable de recontextualiser comment est-ce que cette donnée a été générée, collectée, par qui, comment, où, quoi, dans quel contexte. C'est-à-dire : est-ce que ça va m'être utile, je n'ai pas des biais de collecte, je n'ai pas des enjeux d'analyse quand je vais collecter cette donnée, d'échantillonnage, de représentativité, de problématique vis-à-vis de mon phénomène observé, d'échantillon de ce que je veux faire, d'échelle.

Tu vois par exemple, quand on analyse les données météo, tu collectes des données météo, tu peux faire la météo globale, France, Monde ou continent. Sauf qu'en faisant ça, tu lisses soit les micro-climats, les variations locales, etc. Alors oui, parce que tu veux donner une *big picture* de ton phénomène mais tu ne vas pas véhiculer le même message parce la donnée ne va pas te servir à la même chose que si tu veux faire une vision hyper locale au mètre près.

Donc cet enjeu de compréhension et de regard critique sur la qualité de la donnée vis-à-vis de ce que l'on a envie d'en faire en termes de représentativité, de problématique de contenus, d'échantillonnage, qu'est-ce que tu dois mettre de côté en termes de qualité. On en revient à ce qu'on disait tout à l'heure : si ta donnée elle est de mauvaise qualité et que tu ne peux pas l'utiliser et que tu n'as qu'une partie de ton jeu de données, ça va impacter ton analyse et la manière dont tu vas jouer avec. Donc développer ce regard critique, cela veut dire aussi ne pas avoir ce regard donnée, statistique, outillage, mais un regard donnée, thématique, de ce qu'on observe, de comment on l'observe, de quel est le phénomène qu'on veut analyser, et quel est l'objectif derrière en termes d'analyse. Ça au final, ça s'acquiert et ça se développe avec le métier et avec l'expérience de ce qu'on observe et de savoir ce qu'on observe et pourquoi on l'observe.

Comment vois-tu le niveau de culture des données des personnes que tu côtoies, ça peut être tes étudiants, ça peut être tes collègues... est-ce que ça s'améliore, est-ce qu'on est arrivé à un plateau, est-ce que ça n'augmente pas ?

Non, on est en progression. J'ai de plus en plus de collègues au sein de l'organisation qui viennent nous voir, nous équipe, service administratif, pour des conseils « on aimerait bien restructurer tout ça. Vous êtes spécialistes de la donnée, est-ce que tu peux nous donner des conseils ? ». On sent qu'il y a une prise de conscience qui est de se dire « je sais que si je fais mal ou que si je structure mal, ça va avoir un impact sur la manière dont je fonctionne et sur la réalité et la représentativité de ce qu'il se passe réellement ». Donc très clairement, je suis dans une démarche de croissance. Tu vas me dire « oui mais on part de loin ». Oui mais c'est tant mieux. Auprès des étudiants aussi. De plus en plus, on a une prise de réalité sur les données personnelles, sur les données en tant que telles, l'environnement, et tous les outils qu'il y a derrière. De plus en plus on a des étudiants qui ont ces sujets en bouche et qui les évoque, et qui sont de plus en plus capables de les connecter mais on a encore un manque de compréhension et de connexion d'outillage technologique du progrès technique et scientifique, de ce qu'il y a derrière. Parce qu'ils ne savent toujours pas comment un téléphone se connecte à une antenne. Ou alors c'est quoi la différence avec le *WIFI* et le *Bluetooth* et le *RFID*, j'exagère mais c'est un peu ça l'idée. Ces différentes technologies impactent aussi la donnée qu'on va collecter, la typologie de données qu'on collecte et ce qu'on peut en faire.

Quels sont les freins et les difficultés que tu as pu noter à la mise en place de cette acculturation ?

Pas dans les cours parce qu'on essaye de mettre en pratique mais on l'a évoqué en comité scientifique : il y a plein d'outils de formation, d'acculturation, où on peut retrouver des formats types *MOOC*, type vidéo, type écrit, type atelier jeu workshop, on peut faire plein de formats et il y en a plein qui existent aujourd'hui qui ont été développés depuis plusieurs décennies. Néanmoins la problématique derrière c'est comment on le rend concret pour les gens individuellement pour qu'ils s'en saisissent. Parce qu'ils peuvent comprendre mais comment ils l'appliquent derrière, comment ils le mettent en action de manière concrète dans leur quotidien, que ça soit leur quotidien professionnel ou leur quotidien personnel parce qu'on n'a pas toujours les outils pour le faire au final. Et donc derrière l'idée c'est qu'un des freins c'est la manière dont on l'applique, comment on le met en action, surtout quand on a des interlocuteurs qui n'identifient pas forcément des situations dans lesquelles ils sont confrontés à des données. Par exemple la dernière fois on avait fait un atelier à la journée *Open Data France*. On a animé une mini-conférence au début et puis deux ateliers l'après-midi. Et en fait on demandait aux participants quelles sont les situations dans lesquelles vous êtes confrontés à de la donnée. Et il y a des situations qui n'ont pas été évoquées, typiquement des acteurs, des collectivités, ou des acteurs qui sont face à « tiens je collecte des données, via un formulaire j'ai des gens qui répondent à un questionnaire. ». Pour eux ce n'était pas forcément de la collecte de données. Si, c'est de la pure collecte de données. Je collecte juste des noms, prénoms, adresses, emails. Vous collectez de la donnée. En fait, rendre concret qu'à tout moment on collecte de la donnée et préciser quelles sont les bonnes pratiques. Donc cette mise en situation et reconnaitre les situations dans lesquelles on est face à de la donnée, ce n'est pas toujours simple et c'est parfois un frein aussi. On peut être sensibilisé, on peut être acculturé, mais encore faut-il détecter les situations dans lesquelles c'est le cas. Et puis après l'appliquer et le mettre en action, ça ce n'est pas toujours simple.

As-tu en tête un projet en *data literacy* qui a bien fonctionné et si oui lequel et pourquoi ?

Ça veut dire quoi « bien fonctionné » ?

C'est la question : est-ce que c'était l'objectif qui était recherché de pouvoir par exemple sensibiliser les personnes au parcours de la donnée, est-ce que c'est les former à des logiciels particuliers et avoir compris comment la donnée transite à l'intérieur. C'est en gros un projet en *data literacy*, tout dépend de ce que l'on met derrière. Est-ce que le but recherché est atteint ou pas ?

Je regarde ce qu'on a fait avec la *FING*, avec la métropole de Lyon, les projets *self data*, tous ces projets-là. Au final, je pense que les projets fonctionnent quand ce n'est pas centré sur la donnée. La donnée ou le numérique. Alors pour donner un exemple. Je pense à *PIX*. Je trouve que *PIX* typiquement a eu un certain écho pour permettre à chacun d'évaluer du coup son niveau de compétence vis-à-vis du numérique. La vocation n'était pas forcément *data literacy* même si au final on pourrait le mettre dedans, mais en tout cas ça a eu un écho sur le fait de « je suis capable d'évaluer moi-même mes compétences avec des questions, des quiz, des machins. Et je suis capable de prendre conscience que ça veut dire tout ça derrière et que j'ai ce niveau de compétence par rapport à d'autres ou voire par rapport à un niveau ou un échelon spécifique et donc du coup ça participe à la prise de conscience ». Même si je trouve que typiquement c'est le genre d'outil qui n'est pas assez déployé, qu'on ne trouve pas assez, qui n'est pas un élément qu'on demanderait à tout le monde de passer comme une certif ou quelque chose.

Mais j'en reviens à nos exemples de *self data* avec les projets que l'on a menés. Avec mes infos, je me rends compte que quand on l'évoque il y a une prise de conscience dans le projet ou dans la démarche mais ça ne veut pas dire encore une fois qu'il y a une appropriation ou en tout cas une récupération par les participants pour l'appliquer et le mettre en œuvre derrière. Donc oui, pour répondre à la question, de manière très normande, il y a des démarches et des projets autour de la *data literacy* qui apportent des prises de conscience, mais encore une fois, je ne suis pas sûr(e) et on n'est pas allé mesurer en fait. On est pas allé mesurer six mois après, un an après, deux ans après, 5 ans après, ce qu'on en fait, ce qu'on a compris, est-ce qu'on continue de de le réutiliser ou pas, etc.

Donc au final je ne sais pas si on a des moyens de mesurer. Ça sera que des moyens finalement qualitatifs, ou intuitifs, ou subjectifs. Et du coup la validité de ces éléments-là, on peut la remettre en cause quoi. Mais je sais pas exemple si sur *RUDI* à Rennes il y a des étapes prévues pour l'évaluation de la compréhension ou de l'application. Peut être qu'on est en manque effectivement d'éléments et de données sur la mise en perspective : « est-ce que ça a fonctionné ? Est-ce qu'il y a une réappropriation ? ». Et qui dit réappropriation dit du coup repositionnement de cette compréhension, de cette nouvelle compétence et application au quotidien.

En guise d'ouverture, as-tu, à défaut de recommandations, des remarques pour améliorer cette appropriation de la *data literacy* en France ?

Déjà que ça soit quelque chose d'abordé dans le parcours scolaire. On avait une bonne idée des formats type cours de technologie mais ça pourrait peut être se transformer en ce fameux cours d'histoire des sciences, his-

toire des progrès techniques et technologiques. Et pour mieux repositionner tout ce qu'il se passe et pour peut-être aider les plus jeunes à mieux comprendre et à remettre en perspective, ce qui serait déjà une excellente base à cette compréhension et à cette appropriation du monde de la donnée.

Pour ceux qui sont là, quelque part on n'a jamais trop mieux fait que répéter, refavoriser, remettre à disposition des choses, des éléments et réexpliquer et à chaque fois. Cela fonctionne comme ça, voilà comment c'est utilisé. Même si au final on se rend compte qu'il y a quelques accélérateurs qui sont issus de contextes spécifiques comme le covid, comme les scandales comme *Facebook*. Il y en a d'autres qui sont passés plus inaperçus, je pense notamment à *Avast* avec la récupération des données de navigateurs et revente des données, alors que l'outil est censé te protéger des acteurs extérieurs. Mais oui, il y a des accélérateurs très clairement. Alors de là à dire qu'il faudrait un bon vieux scandale pour que du coup ça accélère l'appropriation, non ! De manière très objective ce sont des accélérateurs quisqu'on est face à une situation dans laquelle ça nous force à réfléchir ou ça nous force à nous reposer la question.