THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITE DE MONTPELLIER

En Informatique ED I2S

École doctorale : Information, Structures, Systèmes

Unité de recherche UMR AMAP

Vers une démarche ontologique pour la gestion des bases de données en agroforesterie

Présentée par Raphaël CONDE SALAZAR

Le 15/12/2021

Sous la direction de Isabelle MOUGENOT et Alexia STOKES

Devant le jury composé de

Nathalie Hernandez, Professeur, IRIT-UT2J, Toulouse	Rapporteur
François Pinet, Directeur de recherche, UR TSCF, Clermont-Ferrand	Rapporteur
Christian Pichot, Chargé de recherche, URFM-UR629, Avignon	Examinateur
Sylvie Ranwez, Professeur, IMT Mines Alès, Alès	Examinatrice, Présidente du Jury
Isabelle Mougenot, Maître de conférence, UR Espace Dev, Montpellier	Directrice de thèse
Alexia Stokes, Directrice de recherche, UMR AMAP, Montpellier	Directrice de thèse
Fabien Liagre, Ingénieur, SCOP AGROOF, Anduze	Invité

UNIVERSITÉ DE MONTPELLIER

RÉSUMÉ :

Vers une démarche ontologique pour la gestion des bases de données en agroforesterie.

Dans un contexte général de développement durable, les recherches menées autour de l'agroforesterie prennent tout leur sens en rendant intelligibles les interactions plante-plante dans un environnement pouvant être dégradé. Depuis plusieurs années maintenant, des données issues d'observations sur le terrain sont acquises en partenariat avec plusieurs catégories d'acteurs (en particulier des forestiers, agriculteurs et éleveurs). Il s'agit maintenant d'ouvrir ces données et de les rendre disponibles depuis le web à d'autres communautés. De plus, les études menées en agroforesterie nécessitent des approches systémiques pour par exemple mieux comprendre les mécanismes de résilience des plantes face aux changements climatiques ou encore face à la pollution des sols. Il s'agit également de rendre les données liables à d'autres jeux de données relatifs au climat et aux territoires. Le double enjeu de partage et d'interconnexion des données en agroforesterie nous rapproche de ce qui se pratique aujourd'hui au sein du web sémantique avec différents outils et méthodes à même de promouvoir la mutualisation de sources de données ouvertes et liées. Les technologies du web sémantique fournissent les mécanismes standards pour décrire et accéder aux ressources présentes sur le web. Le travail de thèse s'inscrit dans ce cadre, nous réutilisons les formats standards du web sémantique et exploitons un éventail d'ontologies à la fois cadres, de domaines et terminologiques afin de proposer un modèle de connaissances, ouvert et flexible à même de rendre compte de la complexité des données déjà collectées en agroforesterie. Ce modèle se compose d'une ontologie de domaine nommée "Agroforestry". L'importance est donnée à la publication de données parfaitement décrites en privilégiant la mise en capacité de lier ces données à des données déjà disponibles sur le web (données climatiques, profils pédologiques, données sur le fonctionnement de communautés végétales, etc.). Nous avons aussi construit une ontologie terminologique dédiée à l'agroforesterie nommée "AgroforestryVoc". A terme, cette ontologie se consacrera à la qualification linguistique des concepts clés partagés par toute la communauté et sera rendue publique sur le portail "AgroPortal" (http://agroportal.lirmm.fr/). Les derniers volets du travail de thèse se concentrent sur l'interprétation des résultats dégagés sur la base de l'interconnexion des données. L'ontologie de domaine "Agroforestry", formalisée en OWL 2, est construite de manière à pouvoir raisonner sur l'ensemble des données intégrées. L'objectif est donc de faire ressortir des tendances relatives aux associations entre arbres et plantes cultivées dans l'exploitation des terres par mise en application de mécanismes inférentiels sur l'ontologie.

ABSTRACT :

Towards an ontological approach for the management of agroforestry databases.

In a general context of sustainable development, research on agroforestry takes on its full meaning by making plant-plant interactions intelligible in a potentially degraded environment. For several years now, data from field observations have been acquired in partnership with several categories of stakeholders (in particular foresters, farmers and breeders). It is now a question of opening these data and making them available from the web to other communities. Moreover, studies conducted in agroforestry require systemic approaches in order to better understand, for example, the resilience mechanisms of plants to climate change or soil pollution. It is also necessary to make the data linkable to other data sets related to climate and territories. The double challenge of sharing and linking data in agroforestry brings us closer to what is practiced today within the semantic web with different tools and methods to promote the sharing of open and linked data sources. Semantic web technologies provide standard mechanisms to describe and access resources on the web. The work of this thesis is part of this framework, we reuse standard formats of the semantic web and exploit a range of ontologies both frameworks, domains and terminologies to propose a knowledge model, open and flexible to account for the complexity of data already collected in agroforestry. This model is composed of a domain ontology named "Agroforestry". The importance is given to the publication of perfectly described data by privileging the ability to link these data to data already available on the web (climatic data, soil profiles, data on the functioning of plant communities, etc.). We have also built a terminological ontology dedicated to agroforestry called "AgroforestryVoc". Eventually, this ontology will be dedicated to the linguistic qualification of key concepts shared by the whole community and will be made public on the "AgroPortal" (http://agroportal.lirmm.fr/). The last parts of the thesis work focus on the interpretation of the results obtained from the data interconnection. The "Agroforestry" domain ontology, formalized in OWL 2, is built in such a way as to be able to reason about the integrated data set. The objective is therefore to highlight trends in associations between trees and crops in land use by applying inferential mechanisms on the ontology.

« Les forêts précèdent les peuples, les déserts les suivent »

François-René de Chateaubriand

1768–1848

SOMMAIRE

1 Introduction **1**

1.1 Introduction générale 1

1.1.1 Contexte 1

1.1.2 Problématique 1

1.1.3 Objectifs 2

1.2 Cadre institutionnel 4

1.2.1 UMR AMAP 4

1.2.2 UMR Espace-Dev 6

1.2.3 Agroof SCOP 6

1.3 Plan du mémoire 7

2 État de l'art **9**

2.1 Les systèmes de gestion des connaissances 10

2.1.1 Les listes d'autorité 11

2.1.2 Les taxonomies 12

2.1.3 Les thésaurus 13

2.1.4 Les folksonomies 14

2.1.5 Les ontologies 15

2.1.6 Normalisation des systèmes de gestion des connaissances 18

2.2 Quelques exemples de vocabulaires contrôlés 20

2.2.1 Vocabulaires en Agroforesterie 20

2.2.2 Vocabulaires en agriculture et écologie 23
2.3 Quelques exemples de systèmes d'information et bases de connaissance 28
2.3.1 Dans les domaines "grand public" 29
2.3.2 Dans le domaine agroforestier 30
2.3.3 Dans le domaine agricole 31
2.3.4 Dans les autres domaines scientifiques 32

3 L'agroforesterie 35
3.1 Introduction 35
3.2 Définition 36
3.3 Les aménagements agroforestiers historiques en Europe 37
3.4 Problématique de l'agroforesterie en France et dans les pays de l'UE . . 38
3.5 Les services fournis par l'agroforesterie 39
3.5.1 Stabilisation des sols 39
3.5.2 Restauration des sols 40
3.5.3 Bien-être animal 41
3.5.4 Retour de la biodiversité 41
3.5.5 Stockage du carbone 42
3.5.6 Limitation de la pollution 42
3.6 Les aménagements agroforestiers 43

4 Les ontologies 47
4.1 De la philosophie à l'informatique 48
4.2 Définition de l'ontologie 49
4.3 Contenu des ontologies 50
4.3.1 Les concepts 50
4.3.2 Les relations 52
4.3.3 Les contraintes 52

4.3.4 La hiérarchie des concepts 52
4.3.5 Les différentes informations exprimées par les ontologies 53
4.4 Les différents types d'ontologies 53
4.4.1 Les ontologies terminologiques 54
4.4.2 Les ontologies cadres 54
4.5 Rôle des ontologies 55
4.5.1 Représentation et uniformisation de la connaissance d'un domaine 55
4.5.2 Réutilisation et échange de la connaissance 55
4.6 En résumé 55

5 Les formalismes logiques 57

5.1 Les réseaux sémantiques 57
5.2 Les langages à base de frames 59
5.3 Les graphes conceptuels 61
5.4 Les logiques de description 63
5.4.1 Les constructeurs. 64
5.4.2 Les différentes logiques de description. 64
5.4.3 Raisonnement avec les logiques de description. 66
5.4.4 Les domaines concrets. 67

6 Le web et les principes du web sémantique 69

6.1 La naissance du web 70
6.2 Présentation du W3C 71
6.2.1 Les principes du W3C 71
6.2.2 Des outils et standards de qualité 71
6.3 Le web sémantique 72
6.4 Ressources, IRI, URI, URL et URN 72
6.5 Les langages de description 74

6.5.1 XML 74
6.5.2 RDF 75
6.5.3 RDFS 77
6.5.4 OWL 80
6.5.5 SKOS 83
6.6 SPARQL : un langage de requête pour les données du web sémantique . 84
6.7 La pyramide du web sémantique 86

Introduction aux contributions de la thèse **89**

7 Première contribution : un référentiel terminologique dédié **91**
7.1 Introduction à la modélisation des connaissances liées à l'agroforesterie 91
7.2 Les concepts et termes de l'agroforesterie 92
7.3 Acquisition des jeux de données d'Agroof 93
7.4 Les données issues des expérimentations du domaine de Restinclières . 97
7.5 Concepts obtenus à partir d'un travail d'extraction de textes 98
7.6 Utilisation de référentiels terminologiques existants 100
7.6.1 La classification des systèmes agroforestiers selon P.K.R. Nair . . 100
7.6.2 La classification des systèmes agroforestiers selon E. Torquebiau 100
7.6.3 Le thésaurus de l'Agence Française d'AgroForesterie 102
7.6.4 Les vocabulaires du web sémantique 103
7.7 Une ontologie terminologique pour l'agroforesterie 104
7.8 Démarche retenue pour la construction d'un référentiel terminologique dédié 104
7.8.1 Description générale 104
7.8.2 Résultats 109
7.8.3 Justifications concernant la construction d'un référentiel 116
7.8.4 Extraits du référentiel en cours de construction 118
7.8.5 Discussion autour de la version actuelle 122

7.8.6 Les éléments d'importance retenus pour l'élaboration d'un cadre général de modélisation des connaissances 123

8 Deuxième contribution : construction d'une ontologie de domaine pour les aménagements agroforestiers 125

8.1 Introduction 125

8.2 Définition d'un modèle conceptuel 128

8.2.1 Définition du modèle conceptuel : Agroforestry 128

8.2.2 Une démarche mobilisant l'existant 129

8.2.3 Intégration de l'ontologie SOSA 130

8.2.4 Extension de l'ontologie GeoSPARQL et de l'ontologie OWL-Time 131

8.3 Passage au modèle formel 140

8.3.1 Utilisation des logiques de descriptions 140

8.3.2 Les fondements logiques d'Agroforestry 141

8.4 Passage au modèle opérationnel 143

8.4.1 Premiers essais d'inférence 144

9 Troisième contribution : Une application pour la validation de notre démarche 149

9.1 Validation et confrontation du modèle avec la réalité 149

9.1.1 Le domaine de Restinclières 151

9.1.2 Description de l'étude 155

9.1.3 Diagrammes de cas d'utilisation 157

9.1.4 Diagrammes d'instances du modèle Agroforestry 160

9.1.5 Implémentation du modèle Agroforestry 165

9.1.6 L'application AOBRA 173

9.1.7 Mode opératoire avec les axiomes OWL 175

9.1.8 Mode opératoire avec l'application de règles Jena 180

9.1.9 Études réalisées avec AOBRA et leurs résultats 199

9.2 En ce qui concerne la validation et la confrontation de notre modèle avec la réalité 210

10 Conclusion et perspectives **213**

10.1 Difficultés rencontrées 215

10.1.1 le paradigme du monde ouvert a un prix 215

10.1.2 Mais où sont les données de l'agroforesterie ? 216

10.2 Perspectives 218

10.2.1 Rapprochement des domaines 218

10.2.2 Unification des sources de données complexes voire contradictoires 218

10.2.3 Développement d'une application pour la gestion des données liées 219

10.2.4 Développement d'une plateforme modulaire pour la gestion des données en agroforesterie 220

Remerciements **223**

Bibliographie **225**

Liste des figures **249**

Liste des tableaux **251**

Annexe A Amménagements du domaine de Restinclières **255**

Annexe B Échantillon de règles JENA utilisées avec l'application AOBRA **259**

B.1 Création de nouvelles observations sur la base d'observations initiales . 259

B.2 Création de nouvelles observations sur la base de critères imposés 261

B.3 Création de nouvelles observations sur la base d'une agrégation en fonction de la valeur observée d'une propriété particulière 264

B.4 Autres possibilités de traitements 267

B.4.1 Utilisation de la taxonomie 267

I

Introduction

1.1 Introduction générale 1
1.1.1 Contexte 1
1.1.2 Problématique 1
1.1.3 Objectifs 2
1.2 Cadre institutionnel 4
1.2.1 UMR AMAP 4
1.2.2 UMR Espace-Dev 6
1.2.3 Agroof SCOP 6
1.3 Plan du mémoire 7

1.1 Introduction générale

1.1.1 Contexte

Le développement de l'agriculture intensive en Europe conduit, depuis une cinquantaine d'années, à un épuisement graduel des sols [CALAME, 2016]. Les effets du réchauffement climatique commencent également, à peser sur l'agriculture Européenne [O'NEILL et al., 2017]. En parallèle, l'augmentation de la population humaine et l'appauvrissement des ressources nécessaires au fonctionnement de l'agriculture (eau et énergies fossiles en particulier), se poursuivent de façon continue. Ces constats nous conduisent à repenser l'agriculture, à la fois pour préserver les ressources et renforcer l'usage de pratiques respectueuses de l'environnement, dans le but de la rendre à la fois compétitive et durable. Parmi les alternatives proposées, l'agroforesterie semble offrir des solutions efficaces pour relever les défis d'une agriculture raisonnée.

1.1.2 Problématique

Si certaines pratiques agroforestières ont pu résister en Europe, au cours des cent dernières années, à l'exemple des prés-vergers de Normandie [DUCROS et al., 2005],

la plupart sont tombées en désuétude. Face aux politiques de mécanisation et de remembrement, pour lesquelles l'arbre est considéré comme un obstacle, le savoir-faire en agroforesterie a peu à peu disparu au fil des générations. Afin de mettre en place une pratique moderne, et intégrée de l'agroforesterie, nous avons besoin aujourd'hui d'indicateurs pour gérer au mieux les ressources nécessaires à ce type de production agricole. Acquises en partenariat avec plusieurs catégories d'acteurs comme en particulier les naturalistes, les forestiers, les agriculteurs et les éleveurs, les données issues des observations de terrain dans le cadre d'expériences agroforestières s'accumulent depuis plusieurs années déjà [LABELLE, 1987]. La gestion et la réutilisation de ces données sont rendues difficiles par la multiplicité des supports et des formats utilisés, et par la diversité des acteurs et de leurs vocabulaires métiers. En outre, les études agroforestières nécessitent des approches systémiques pour comprendre, par exemple, comment mieux gérer un site, en réponse au changement climatique, aux ravageurs, ou à la pollution du sol. Ces réponses se doivent de considérer les liens étroits tissés avec d'autres domaines de connaissance tels que la climatologie, la zoologie ou encore la pédologie.

1.1.3 Objectifs

Le but est ici de démontrer que l'utilisation d'une démarche basée sur les ontologies peut apporter des solutions pérennes et novatrices pour faciliter l'exploitation et l'analyse des données actuelles et futures issues des expérimentations menées dans le cadre de l'agroforesterie. L'importance est donnée à la capitalisation des savoirs experts en agroforesterie, dans ce que nous considérons comme une reconquête d'usages anciens, qui sont revisités pour les adapter aux besoins des sociétés actuelles.

Un des premiers objectifs du travail de thèse est d'"ouvrir" les données diverses et variées de l'agroforesterie, en les rendant disponibles et partageables à différents acteurs, à la fois du domaine, et de disciplines connexes. Les expérimentations menées en agroforesterie, sont basées sur le long terme, à savoir sur la durée de vie des arbres. L'acquisition des connaissances dans ce contexte ne peut se faire qu'au travers d'un travail combiné autour de sites expérimentaux reconnus, pour étudier en profondeur les processus de cohabitation entre l'arbre et la culture ou la présence d'animaux, mais également grâce à un réseau de parcelles pilotes rendant compte de la diversité des situations pédoclimatiques et des systèmes de production. La complexité d'un tel projet demande un effort important pour pérenniser et capitaliser les données obtenues et nécessite des protocoles et des bases de données fiables et pérennes. Pour valider la qualité de l'information, la possibilité de nourrir mais éventuellement de modifier ou ajouter des variables caractérisant les projets agroforestiers devra être donnée aux expérimentateurs de terrain, agriculteurs ou techniciens. Pour ces utilisateurs, l'objectif est double : comprendre les interactions entre l'arbre et le système de production agricole, mais aussi identifier de nouveaux itinéraires techniques pour améliorer l'efficacité des projets.

Le deuxième objectif est de rendre ces données issues de l'agroforesterie liables à d'autres jeux de données relatifs à des domaines connexes comme le climat ou les territoires. Ce lien avec d'autres jeux de données devrait permettre de réaliser des approches systémiques pour par exemple mieux comprendre les mécanismes de

résilience des plantes face aux changements climatiques ou encore face à la pollution des sols.

Ce double enjeu de partage et d'interconnexion des données en agroforesterie nous rapproche de ce qui se pratique aujourd'hui au sein du web sémantique avec différents outils et méthodes à même de promouvoir la mutualisation de sources de données ouvertes et liées [HEATH et BIZER, 2011].
Le web sémantique est la vision du World Wide Web Consortium [1] d'un web global où toutes les données seraient liées entre elles et à partir desquelles des agents logiciels pourraient fournir des services à forte valeur ajoutée [BERNERS-LEE et al., 2002]. Les technologies du web sémantique fournissent les procédures standards pour décrire et accéder aux ressources présentes sur le web [LAUBLET et al., 2002].
Ce travail de thèse s'inscrit dans ce cadre, nous avons eu à charge de réutiliser les formats standards du web sémantique [GRAU et al., 2008] et d'exploiter un éventail d'ontologies terminologiques afin de proposer un modèle de connaissances ouvert et flexible à même de rendre compte de la complexité des données déjà collectées en agroforesterie.

Pour atteindre nos objectifs, avons construit un système de gestion des connaissances [ALAVI et LEIDNER, 2001] nommé "AOBRA" [2] entièrement dédié à l'agroforesterie, qui réutilise des ontologies standards empruntées au web sémantique [DAVIES et al., 2003].
La construction du système de gestion des connaissances a fait l'objet de trois grandes étapes :

La première étape a consisté en la mise place d'une ontologie terminologique [LAPORTE et al., s. d. ; GARNIER et al., 2017] de l'agroforesterie que nous avons nommé "AgroforestryVoc". Cette ontologie décrit les différents concepts utilisés en agroforesterie, avec les termes qui y sont rattachés et les relations entre ces concepts (principalement des relations hiérarchiques).

Dans une deuxième étape, sur la base de cette ontologie terminologique, nous avons créé une ontologie cadre des aménagements agroforestiers nommée "Agroforestry". Cette ontologie cadre est un modèle d'organisation des connaissances dans le domaine des aménagements agroforestiers qui nous sert de modèle pour notre système de gestion des connaissances AOBRA. Agroforestry s'appuie sur le formalisme des logiques de description ce qui nous permet de mettre en place des mécanismes d'inférence.

Dans la troisième étape, nous avons construit l'application AOBRA en nous appuyant sur le modèle "Agroforestry" et en prenant en compte la possibilité d'exploiter d'autres ontologies dans différents domaines présentes sur le web.

En exploitant les capacités d'inférence permises par le formalisme logique d'Agroforestry et les liens vers d'autres ontologies sur le web, AOBRA peut effectuer des raisonnements sur la base de faits initiaux et générer ainsi de nouvelles connaissances.

1. W3C : World Wide Web Consortium. https://www.w3.org/.
2. AOBRA : **A**pplication layer with **O**ntology-**B**ased **R**easoning for **A**groforestry.

1.2 Cadre institutionnel

Mon travail de thèse est principalement mené au sein de l'UMR AMAP [3], et a bénéficié d'un double financement provenant de la Fondation de France [4], dans le cadre de son programme de soutien pour l'environement et en particulier pour l'agroforesterie, et du projet européen ECOMED [5], qui se concentre sur l'accompagnement des sociétés d'ingénierie en génie civil dans la pratique d'aménagements écologiques.
Deux autres structures, l'UMR Espace-Dev [6] et Agroof SCOP [7], m'ont également accueilli et accompagné pour la réalisation de mes travaux.

1.2.1 UMR AMAP

L'UMR (Unité Mixte de Recherche) AMAP dépend de cinq tutelles, qui sont le CIRAD, le CNRS, l'INRAe, l'IRD et l'Université de Montpellier. Elle est rattachée aux départements et institut suivants :

- Département scientifique Systèmes Biologiques (BIOS) du CIRAD (UR 51)
- Institut Écologie et Environnement (INEE) du CNRS (UMR 5120)
- Département Ecologie et Biodiversité (ECODIV) de l'INRAE (UMR 931)
- Département Ecologie, Biodiversité et Fonctionnement des Ecosystèmes Continentaux (ECOBIO) de l'IRD (UMR 123)
- Département Biologie-Ecologie-Evolution-Environnement-Sciences de la Terre et de l'Eau (B3ESTE) de la faculté des Sciences de l'Université Montpellier

L'UMR AMAP rassemble un personnel de quatre-vingt permanents (43 chercheurs, 3 enseignants-chercheurs, 1 professeur agrégé, 33 ingénieurs et techniciens) et est basée principalement prés de la ville de Montpellier en France. D'autres sites secondaires se trouvent en Guyane et en Nouvelle Calédonie et du personnel est expatrié à La Réunion au Cameroun et au Kenia. AMAP a pour particularité d'être une unité interdisciplinaire, dont le but est d'acquérir des connaissances fondamentales sur les plantes et les végétations afin de prédire la réponse des écosystèmes à des changements environnementaux, tel que le réchauffement climatique. Les disciplines exploitées sont la botanique (systématique et structurale), l'écologie végétale, l'agronomie, la foresterie, l'informatique, et les mathématiques appliquées. Les milieux étudiés sont les végétations méditerranéennes, tempérées et tropicales. La présence sur un même lieu de travail d'informaticiens et de mathématiciens auprès d'écologues, d'agronomes et de botanistes à facilitée depuis la création de cette unité la mise en place d'approches transdisciplinaires.
Les recherches de l'UMR AMAP s'organisent autour de trois axes thématiques transversaux :

- Axe Biodiversité : systématique, biogéographie, écologie

3. Umr amap : botanique et modélisation de l'architecture des plantes et des végétations. http://amap.cirad.fr.
4. FdF : Fondation de France. https://www.fondationdefrance.org/fr.
5. ECOMED : news alliances and dynamics between mediterranean higher education institutions and ecoengineering enterprises. http://ecomedbio.eu.
6. UMR Espace-Dev : Observation spatiale, modèles & sciences impliquée . http://www.espace-dev.fr.
7. Agroof SCOP : Société Coopérative et Participative spécialisée dans l' étude et le développement des systèmes agroforestiers. www.agroof.net.

Afin de luter contre l'effondrement de la biodiversité, les objectifs de cet axe sont l'acquisition des connaissances de base sur les espèces, le développement d'outils et méthodes pour caractériser leur diversité à différents niveaux d'observation (lignées évolutives, formations végétales), et la construction de modèles prédictifs de la dynamique spatio-temporelle de cette diversité en réponse aux changements globaux.

– Axe Biomasse : dynamique et production des plantes et des peuplements
  Afin d'anticiper l'impact des changements globaux en cours et à venir, l'objectif de cet axe est l'étude des processus fondamentaux de la croissance et du développement des plantes et la prédiction de l'organisation et de la dynamique des couverts végétaux, à différentes échelles spatio-temporelles et sous différents scénarios de gestion ou de modification environnementale.

– Axe Plantes numériques : modèles, analyses et données de l'organe aux écosystèmes
  Afin de modéliser les processus complexes qui participent à la dynamique de systèmes dans les domaines de l'écologie, de l'environnement, de la biodiversité et de l'agronomie, l'objectif de cet axe est le développement d'outils et de méthodes pour l'étude et la modélisation des systèmes complexes issus du monde végétal.

A ces axes sont associés des thèmes comme diversité biologique, biologie intégrative, agriculture et foresterie durables, agroécologie et changements globaux. Les chercheurs de l'unité se positionnent généralement sur plusieurs thèmes au travers d'un ou plusieurs axes .
Mon travail de thèse a été conduit principalement au sein de l'axe "Biomasse" et du thème "EASE" (Écologie Appliquée et Services Écosystémiques) dont l'objectif est d'appréhender, de prévoir et d'évaluer les conséquences dues aux changements globaux sur les services écosystémiques (comme la conservation des sols, la séquestration du carbone ou la biodiversité) fournis par des écosystèmes naturels ou anthropisés pour en permettre une gestion efficace et durable. Pour atteindre ce but il est nécessaire d'avoir une approche systémique afin de comprendre les dynamiques de ces services et leurs interactions. Parmi les écosystèmes étudié à l'UMR AMAP les systèmes agroforestiers offrent des services écosystémiques nombreux et interdépendants et mon rôle dans cette équipe est de participer à l'élaboration de concepts et d'outils adaptés à la gestion durable des ces services en agroforesterie. Mon travail s'inscrit donc parfaitement dans un des objectifs du laboratoire qui est la création de modèle et d'outils pour permettre, aux différents acteurs (agroforestiers,ingénieurs, forestiers, agronomes, agriculteurs) de mieux gérer, reconstruire ou restaurer des environnements fragiles ou vulnérables.
J'ai également été intégré à l'équipe du thème "BIAS" (Informatique de la Biodiversité et Intelligence Artificielle en Sciences agro-écologiques) de l'axe "Plantes numériques" pour des échanges sur la question des outils informatiques dont les ontologies informatiques utilisables pour la gestion collaborative de données complexes et hétérogènes, non centralisées et en grand volume.
De par le fait que les recherches au sein du laboratoire se déclinent aujourd'hui dans un contexte d'augmentation extrêmement rapide des volumes et des flux de données très diversifiées mon travail à suscité un intérêt au sein de l'ensemble de l'unité pour l'utilisation d'outils sémantiques dans le traitement de ces données afin d'en augmenter l'interopérabilité et le partage.

1.2.2 UMR Espace-Dev

L'UMR ESPACE-DEV est placée sous les tutelles de l'IRD, l'Université de la Guyane, l'Université de la Réunion, l'Université des Antilles et l'Université de Montpellier. Elle étudie les dynamiques spatiales caractérisant les éco-sociosystèmes. Son but est de définir des indicateurs pour mieux appréhender ces dynamiques. Les dynamiques étudiées sont la bio-géophysique, l'évolution des sociétés, les risques liés aux maladies émergentes en fonction de paramètres environnementaux, et les changements et vulnérabilité des territoires aux changements globaux. Elle développe ces activités de recherche autour de l'acquisition, du traitement et de l'exploitation des données satellitaires en y intégrant des connaissances multidisciplinaires. Trois équipes composent cette unité :

- OSE (Observation Spatiale de l'Environnement)
  Réception et maîtrise des flux de données des données satellitaires et extraction des objets d'intérêt pour déterminer les dynamiques environnementales
- AIMS (Approche Intégrée des Milieux et de Sociétés)
  Recherches sur les dynamiques socio-environnementales et culturelles des territoires vulnérables aux changements globaux.
- MICADO (Modélisation, Ingénierie des Connaissances et Analyses de DOnnées spatiales)
  Modélisation, analyse et aide à la décision pour des systèmes spatio-temporels.

Parmi les axes de recherches transversales, l'axe 'Ontologie ou modélisation de la connaissance' favorise l'interdisciplinarité de cette UMR en construisant et en réutilisant des modèles génériques de données et de connaissances pour les sciences de l'environnement et de la santé.

1.2.3 Agroof SCOP

Agroof SCOP est un bureau d'étude spécialisé dans l'étude et le développement des systèmes agroforestiers. C'est une Scop [8] constituée de 9 salariés-associés aux compétences complémentaires (chargé de partenariat, chargé de projet, chargé de recherche, conseiller-formateur, gestionnaire, ressource digitale). Domiciliée dans la ville d'Anduze en France, elle est actuellement dirigée par Fabien LIAGRE (Chargé de projet R&D) qui à été à l'initiative de ce projet en l'an 2000.
Agroof exerce plusieurs activités:

- La recherche
  Dans le but de découvrir les pratiques les plus adaptées aux différents contextes pédologiques et climatiques rencontrés sur les exploitations agroforestières, Agroof collabore avec des chercheurs et des agriculteurs pour obtenir de nouvelles connaissances sur ces systèmes. Dans le cadre de cette activité de recherche, elle met en place des projets agroforestiers expérimentaux impliquant les agriculteurs dans les phases de construction, de mise en place et d'analyse des résultats obtenus.

- L'ingénierie de projet
  Agroof exerce son activité d'ingénierie auprès d'un large public composé d'agriculteurs, de collectivités, de particuliers, d'organismes techniques et d'organismes de recherche en les accompagnant dans la création de leurs projets agroforestiers

- La formation
  Agroof propose des formations professionnelles à l'intention des agriculteurs, tech-

8. Scop : Société coopérative et participative. https://www.les-scop.coop/les-scop.

niciens, collectivités, administrations et entreprises ainsi que des formations dans le cadre de l'enseignement supérieur et professionnel.

– L'édition de contenus
  A partir de son site internet, Agroof propose des ressources gratuites ou payantes sur l'agroforesterie sous formes de livres, vidéos, dvd et pdf.
  La description des projets de recherches et leurs résultats, sont accessibles sur le web via leur site.

Au sein de cette équipe, j'ai pu acquérir des connaissances en agroforesterie et en particulier en agrisylviculture. Ces connaissances ciblent l'installation de nouveaux aménagements et leur entretien ce qui m'a permis de cerner les notions essentielles à la réussite d'un bon aménagement comme l'importance de l'organisation spatiale ou du choix des espèces.
Il m'ont aussi fait part de leurs besoins et de leurs difficultés dans le cadre de leurs activités en matière de gestion de données et d'accès à l'information. Au niveau de la gestion des données, l'aspect le plus problématique était l'absence d'un modèle fiable pour la capitalisation et l'exploitation de l'ensemble des données collectées sur le terrains au cours de leurs activités d'ingénieries et de recherche. Au niveau de l'accès à l'information, il s'agissait dans le cadre de leur activité d'édition de contenus d'accéder plus aisément aux ressources du web concernant l'agroforesterie. Parmi les renseignements, qu'ils ont accumulés durant leurs vingt années d'existence, sur les sites agroforestiers, on trouve entre autres:

– La description des sites avec leur découpage en parcelles et blocs et leur localisation

– Une description administrative avec par exemple les noms des propriétaires et des contacts recherches lorsqu'il s'agit de projet expérimentaux

– Une description des aménagements réalisés avec par exemple leur type, leurs caractéristiques et leur finalité

– Une description des traitements sur ces aménagements avec par exemple l'irrigation, le paillage et les traitements phytosanitaires.

– Une description des éléments biotiques de l'aménagement comme par exemple la caractérisation des espèces végétales ou animales présentes sur les aménagements

– Une description des éléments abiotiques de l'aménagement par exemple le pH et l'humidité du sol

– Une description topologique comme par exemple la pente du terrain ou son altitude

Leur expertise en agroforesterie et l'accès à leurs données collectées m'ont permis de mieux cerner les connaissances qu'il s'agissait de capitaliser dans le domaine de l'agroforesterie.

1.3 Plan du mémoire

Dans les chapitres suivants de ce mémoire nous commencerons par présenter une série d'états de l'art pour permettre la compréhension des principes essentiels relevant de l'ingénierie des connaissance et de l'agroforesterie :

Le chapitre 2 présente un état de l'art sur différentes initiatives concernant les vocabulaires contrôlés, les systèmes d'information et les bases de connaissance en agroforesterie et dans d'autres domaines. Cet état de l'art n'a pas vocation d'être exhaustif mais présente les initiatives qui ont retenu notre attention pour la réalisation de ce travail de thèse.

Le chapitre 3 définit l'agroforesterie et différents services écosystémiques pouvant être apportés par sa pratique. Un exemple d'aménagement en agrisylviculture sera présenté.

Le chapitre 4 présente les ontologies en tant qu'outils de modélisation et de formalisation de la connaissance d'un domaine.

Le chapitre 5 introduit différents formalismes de représentation dont les logiques de description sur lesquelles s'appuient les outils du web sémantique.

Le chapitre 6 dresse un historique du web sémantique et présente les différents outils mis à disposition par le World Wide Web Consortium.

Par la suite, les trois chapitres suivants (chapitres 7, 8 et 9) présentent nos contributions par ce travail de thèse. Dans ces chapitres, nous détaillerons les différentes étapes de construction de notre ontologie sur les aménagements agroforestiers "Agroforestry". Cette ontologie sera le cœur d'une application "AOBRA" que nous utiliserons pour la réalisation d'un cas d'étude avec des données provenant du site expérimental de Restinclières afin de valider l'efficacité de notre modèle. Nous terminerons par nos initiatives pour mettre notre modèle à disposition de tous, et initier la mise en place d'une ontologie terminologique évolutive et partagée par la communauté agroforestière accessible via le web.
Pour finir le chapitre 10 résume les principaux résultats des travaux effectués durant cette thèse. Nous identifions ainsi les avantages et les inconvénients d'une démarche ontologique pour la gestion des données en agroforesterie. Pour clore ce dernier chapitre nous aborderons les pistes de futurs développements envisageables en ayant recours à l'ingénierie des connaissances dans le cadre de l'agroforesterie et dans d'autres domaines.

II

État de l'art

2.1 Les systèmes de gestion des connaissances 10
2.1.1 Les listes d'autorité 11
2.1.2 Les taxonomies 12
2.1.3 Les thésaurus 13
2.1.4 Les folksonomies 14
2.1.5 Les ontologies 15
2.1.6 Normalisation des systèmes de gestion des connaissances . . . 18
2.2 Quelques exemples de vocabulaires contrôlés 20
2.2.1 Vocabulaires en Agroforesterie 20
2.2.2 Vocabulaires en agriculture et écologie 23
2.3 Quelques exemples de systèmes d'information et bases de connaissance 28
2.3.1 Dans les domaines "grand public" 29
2.3.2 Dans le domaine agroforestier 30
2.3.3 Dans le domaine agricole 31
2.3.4 Dans les autres domaines scientifiques 32

Inspiré par la pensée d'Aristote, reconnu aujourd'hui comme un des précurseurs les plus influents de notre pensée scientifique moderne [SMITH, 2000], nous recherchons depuis tout temps à définir et classifier les choses qui nous entourent de façon universelle. Les travaux d'Aristote comme son traité intitulé « Parties des animaux »[REMACLE, 2010] dans lequel il expose ses méthodes pour la classification des animaux influencera des générations de naturalistes pendant plus de deux mille ans et posera les bases de nos systèmes contemporains de classification. La classification est un outil essentiel pour organiser nos connaissances et est donc de ce fait présente dans de nombreuses disciplines comme les sciences de la nature, les sciences sociales et économiques ou la gestion des bibliothèques. L'évolution des méthodes de classification parallèlement aux avancées informatiques a permis la création d'applications qui en utilisant ces classifications renforcent leurs capacités à cataloguer, récupérer et traiter les connaissances.
Nous commencerons ce chapitre en exposant cette problématique de la classification au travers de la gestion des bibliothèques. Les bibliothèques par leur long passé dans la

recherche d'organisation de grandes collections de documents sont une bonne représentation des organisations de gestion de l'information de notre monde actuel. Pour organiser et permettre l'accès à leurs multiples ressources documentaires, les bibliothèques utilisent des dispositifs variés regroupés sous le terme de "Systèmes de gestion des connaissances" (SOC) [HODGE, 2000]. Nous opposerons brièvement cette notion de systèmes de gestion des connaissances à celle d'ontologies que nous présenterons plus en détail au chapitre 4 et terminerons ce chapitre par quelques exemples.

2.1 Les systèmes de gestion des connaissances

Il existent différentes types de systèmes de gestion des connaissances pouvant être analysés et comparés [ZACKLAD, 2007 ; ZACKLAD, 2010]. Nous aborderons dans cette section la famille des vocabulaires contrôlés avec les listes d'autorités, les taxonomie et les thésaurus, puis celle des folksonomies et enfin des ontologies. Le nombre et la taille des collections d'objets et de documents (textes, images, enregistrements sonores, ...) gérés par les bibliothèques augmentent sans cesse. Les facilités de création et de diffusion apportée par l'utilisation des techniques du numérique participe à cette augmentation. Sous leur forme numérique, internet et le web les ont rendu plus facilement accessibles, mais il n'en reste pas mois difficile de les cibler. Une information pertinente n'étant d'aucune utilité si celle-ci ne peut être facilement retrouvée, l'enjeu principal des systèmes de gestion des connaissances est de faciliter le ciblage de cette information pertinente au sein de ces nombreux documents.
Pour cela les systèmes de gestion des connaissances ont un double objectif :

- L'indexation des documents
  Parallèlement au stockage matériel et immatériel des documents, les bibliothèques doivent permettent à leur utilisateurs de les retrouver facilement en fonction de leur contenu. Pour cela, elles constituent des catalogues rassemblant les principales caractéristiques intellectuelles et physiques de leurs documents. L'analyse d'un document, son identification et la description des notions et concepts qui se rapporte à son contenu constitue l'indexation de ce document. L'indexation est donc une approche sémantique puisqu'elle détermine le sens d'un document. Mais c'est aussi une approche sémiotique car ce sens sera symbolisé par un signe (un terme) choisi, Le descripteur. L'indexation permettra la création d'une fiche, la notice documentaire, qui décrira le document. Cette identification des notions et concepts du document par un descripteur repose sur deux principes:
  - L'unité de sens
    Au sein d'un même catalogue, les descripteurs choisis représentent toujours les mêmes notions ou concepts.
  - Le principe de spécificité
    Les descripteurs choisis ont la même extension sémantique que les notions ou concepts qu'ils représentent

- La recherche des documents
  Une fois indexé un document pourra être retrouvé dans le catalogue de la bibliothèque à partir des renseignements de sa notice documentaire qui permettent de l'identifier, de le localiser et de décrire sommairement le contenu de celui-ci. L'indexation facilite la recherche d'une information pertinente au sein d'un document puisqu'il n'est plus nécessaire de parcourir ce document pour savoir si cette information s'y trouve mais de rechercher dans sa notice si un descripteur correspond au descripteur présenté dans la recherche.

Les systèmes de gestion des connaissances sont donc avant tout des lexiques de la terminologie d'un domaine. Les termes de ces lexiques appelés généralement descripteurs servent à identifier et a signaler les notions et concepts contenus dans les divers documents. Le choix du descripteur peut être laissé au libre arbitre de la personne réalisant l'indexation, mais dans ce cas des problèmes peuvent apparaitre si l'indexeur et le chercheur ne partage pas le même vocabulaire. En effet des résultats faussement positifs peuvent survenir en cas de polysémie, ou ne pas être retrouvé en cas de synonymie, des termes employés dans l'indexation et la recherche. La solution est alors d'utiliser une même liste de descripteurs servant de référentiel tant pour la personne responsable de l'indexation que pour celle effectuant la recherche. De plus pour s'assurer du respect des principes d'unité de sens et de spécificité les lexiques sont généralement constitués de termes prédéfinis par leurs concepteurs, en opposition au langage naturel où il n'existe pas de restriction quant au choix des termes utilisés, afin d'éviter la synonymie ou la polysémie de notre langage. De telles listes sont appelées des vocabulaires contrôlés. Les systèmes de gestion des connaissances peuvent aussi exprimer des associations entre leurs différents termes. Ces associations sont plus ou moins fortes, ainsi certaines peuvent être des relations sémantiques explicites comme par exemple les relations généralise/spécialise, partie-de ou cause-de.

2.1.1 Les listes d'autorité

Les listes d'autorité font partie de la famille des vocabulaires contrôlés, les termes de ce vocabulaire sont appelés vedettes-matière. Parmi ces termes certains sont retenus pour l'indexation , on dit qu'il font autorité, et d'autre sont exclus, ce sont des synonymes. Des combinaisons de termes peuvent être faites au cour de la création du catalogue de la bibliothèque selon des règles et une syntaxe précises. La liste d'autorité RAMEAU [1] de la bibliothèque nationale de France (BnF) est un exemple de ce type de liste. La figure 2.1 montre la notice d'autorité du terme agroforesterie indiquant un terme synonyme "agrosylviculture" qui ne devra pas être utilisé pour l'indexation d'un document mais pourra être utilisé lors d'une recherche à la place du terme agroforesterie, des termes génériques "foresterie" et "agriculture", des termes associés "plantations ligneuses" et "agriculture itinérante", ainsi que des correspondances avec d'autres vedettes-matière de la librairie nationale américaine et de l'université Laval au Canada .

1. RAMEAU : BnF catalogue générale - notices d'autorité. https://catalogue.bnf.fr/recherche-autorite.do?pageRech=rat.

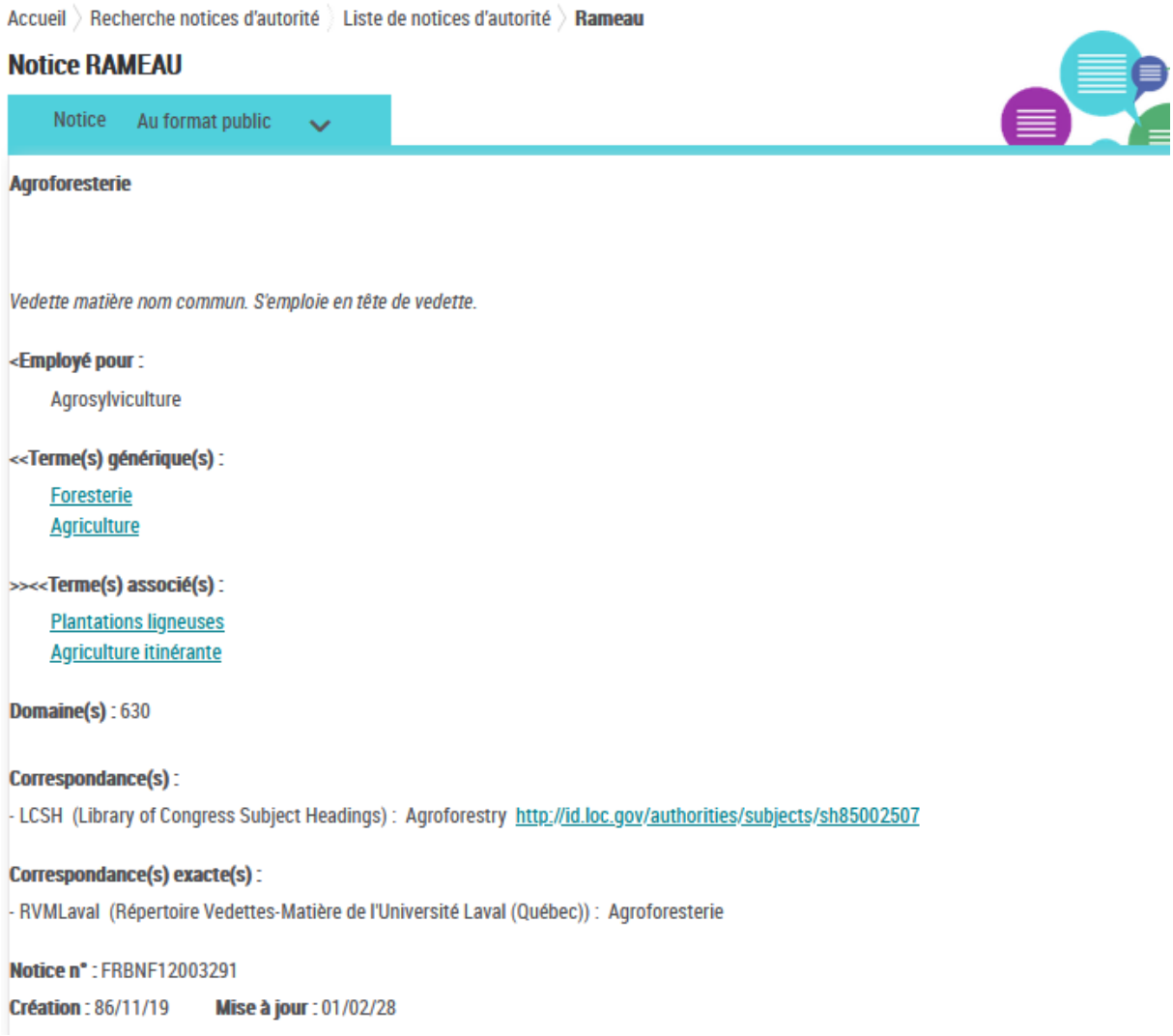


FIGURE 2.1 – *Notice d'autorité de l'agroforesteriea de la liste d'autorité RAMEAU .*

2.1.2 Les taxonomies

Contrairement aux listes d'autorité, les taxonomie ne retiennent que les termes nécessaires à l'indexation, en excluant leurs synonymes, et les organisent de façon hiérarchique. Cette hiérarchisation se base uniquement sur une relation de généralisation ou de spécialisation, ainsi il est possible d'exprimer le fait que le terme "être vivant" généralise le terme "plante" et le terme "animal" et que le terme "arbre" spécialise le terme "plante ligneuse" (voir figure 2.2). On passe donc d'un vocabulaire contrôlé à un vocabulaire contrôlé et organisé. Lors d'une recherche à partir d'un mot, les termes parents seront utilisés afin d'optimiser celle ci. Pour exemple si un thésaurus associe les mots "arbre" et "plante ligneuse" et les mots "agrisylviculture" et "agroforesterie", une recherche sur les arbres en agrisylviculture sera étendue aux documents indexés avec les descripteurs "plante ligneuse" et "agroforesterie".

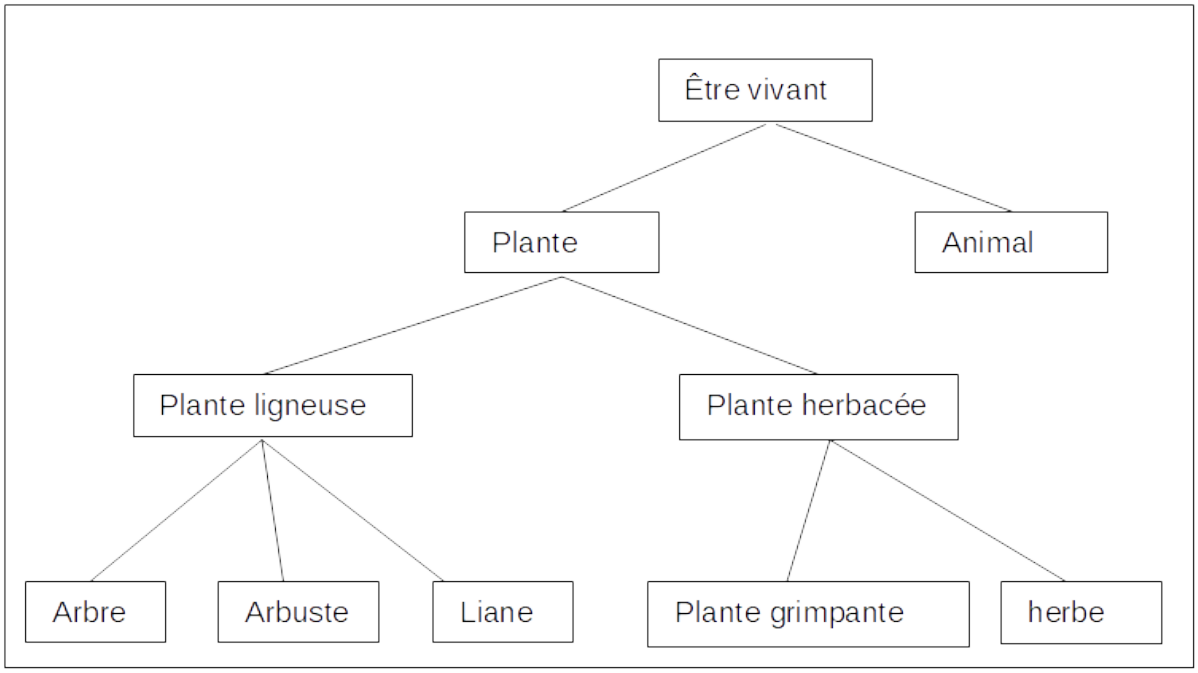


FIGURE 2.2 – *Un exemple de taxonomie.*

2.1.3 Les thésaurus

A la fin des années 50 les thésaurus sont exploités sous leur forme moderne pour la recherche d'information [ROBERTS, 1984]. Aux relations hiérarchiques présentes dans les taxonomies, les thésaurus rajoutent des relations associatives mais une autre distinction importante qui distingue ces deux dispositifs de classification est dans le fait que les thésaurus visent à décrire des concepts plutôt que les sujets d'un document. Les sujets de documents étant à priori en nombre illimité vis à vis des concepts en nombre plus restreint qu'ils représentent, cette distinction confère aux thésaurus un avantage pour la cohérence du descripteur alors considéré comme le meilleur représentant de ce concept et pour la constitution de thésaurus multilingue ou selon [MANIEZ, 1999] le concept est alors utilisé comme « l'opérateur d'équivalence » de la traduction.
La figure 2.3 nous montre le diagramme UML d'un thésaurus. Au sein de ce diagramme, on retrouve au centre la classe **Terme** qui définit l'ensemble des mots contenus dans un ou plusieurs thésaurus comme le montre sa relation d'agrégation avec la classe **Thésaurus** qui définit l'ensemble des thésaurus. Cette classe **Terme** possède deux classes filles, la classe **Descripteur** et la classe **Non descripteur** qui représentent respectivement l'ensemble des termes retenus pour l'indexation des documents et l'ensemble des termes exclus pour l'indexation. Ces deux classes sont liées par la relation *aPourSynonyme* traduisant le fait qu'un terme descripteur puisse avoir zéro ou plusieurs synonymes parmi les termes non descripteur. Deux individus distincts de la classe **Terme** peuvent être reliés l'un à l'autre d'une part par les relations *généralise* ou *spécialise* indiquant une relation hiérarchique entre les deux termes et d'autre part par le relation *estProcheDe* indiquant une relation de proximité entre les deux termes. La classe **Concept** définit l'ensemble de tous les concepts et sa relation *désigne* avec la classe **Descripteur** nous renseigne sur le fait qu'un même terme choisi comme descripteur ne désignera qu'un seul concept, et ce quel que soit le thésaurus dans lequel il se trouve.

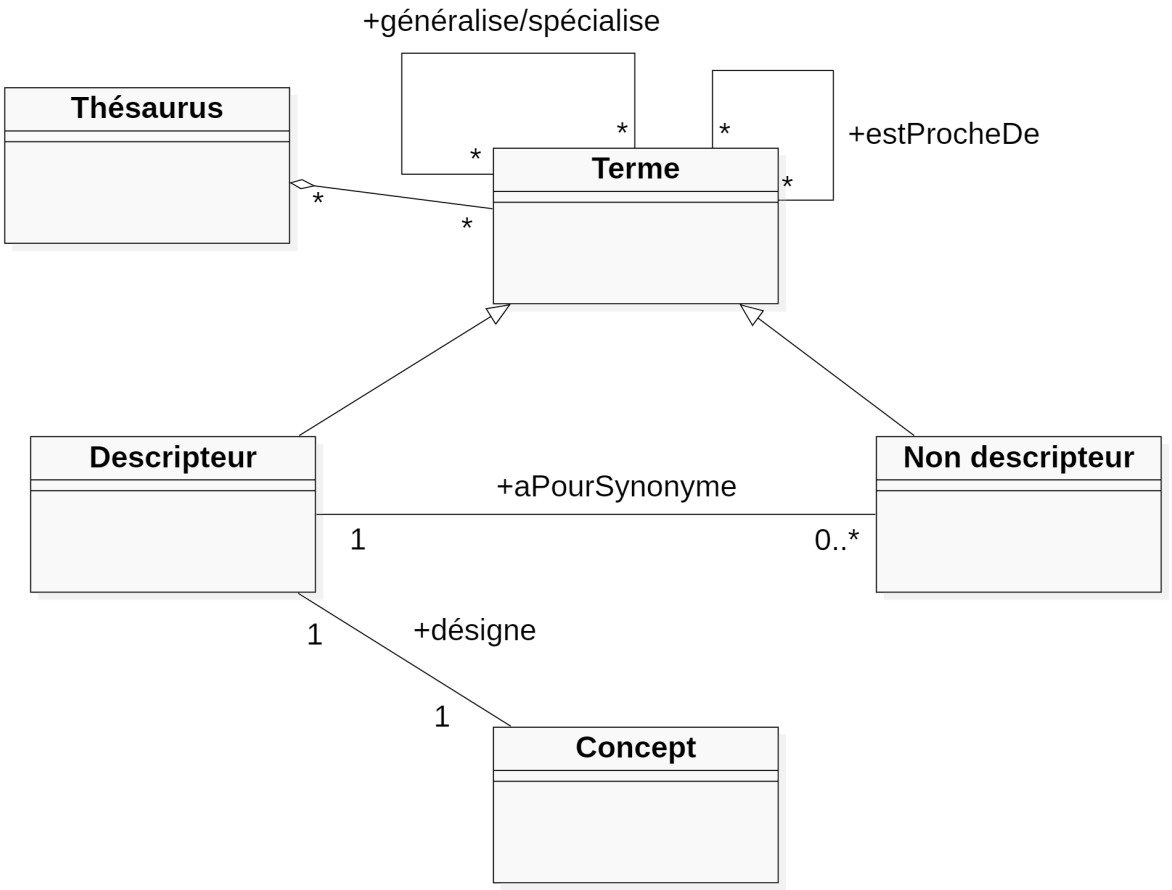


FIGURE 2.3 – *Diagramme général pour la modélisation d'un thésaurus.*

2.1.4 Les folksonomies

Plus récemment, les folksonomies sont apparues avec une évolution du web nommé web 2.0 [O'REILLY, 2009] caractérisée par une plus grande interactivité avec les internautes. Dans les usages participatifs du web 2.0, les utilisateurs lambda peuvent indexer eux même des documents afin de pouvoir facilement les rechercher [LE DEUFF, 2006]. Le vocabulaire n'est donc ici ni contrôlé, ni organisé et les termes utilisés prennent les noms de mots clés, tags ou étiquettes. On peut citer en exemple la platforme Flickr [2] où les documents indexés par les utilisateurs sont des photographies. Les étiquettes apposées aux photos sont utilisées pour classer celles ci, afin de faciliter les recherches thématiques des autres utilisateurs (voir figure 2.4). N'étant nullement prédéfinis, les termes dépendront donc des différents utilisateurs et un même document pourra être étiqueté avec différents mots clefs et différents synonymes, ce qui entraîne une faible cohérence du descripteur (synonymie, polysémie et absence de relation hiérarchique et sémantique). Cependant la multiplication des tags pour un même document augmente la possibilité qu'un de ceux-ci soit utilisé par l'utilisateur qui effectue la recherche et le faible coût de cette indexation puisqu'elle ne nécessite pas l'intervention de spécialistes de la classification font des folksonomies un dispositif encore très étudié dans le contexte actuel d'une surabondance de documents disponibles sur le web.

2. Flickr : Service de gestion et de partage de photos en ligne. http://www.flickr.com.

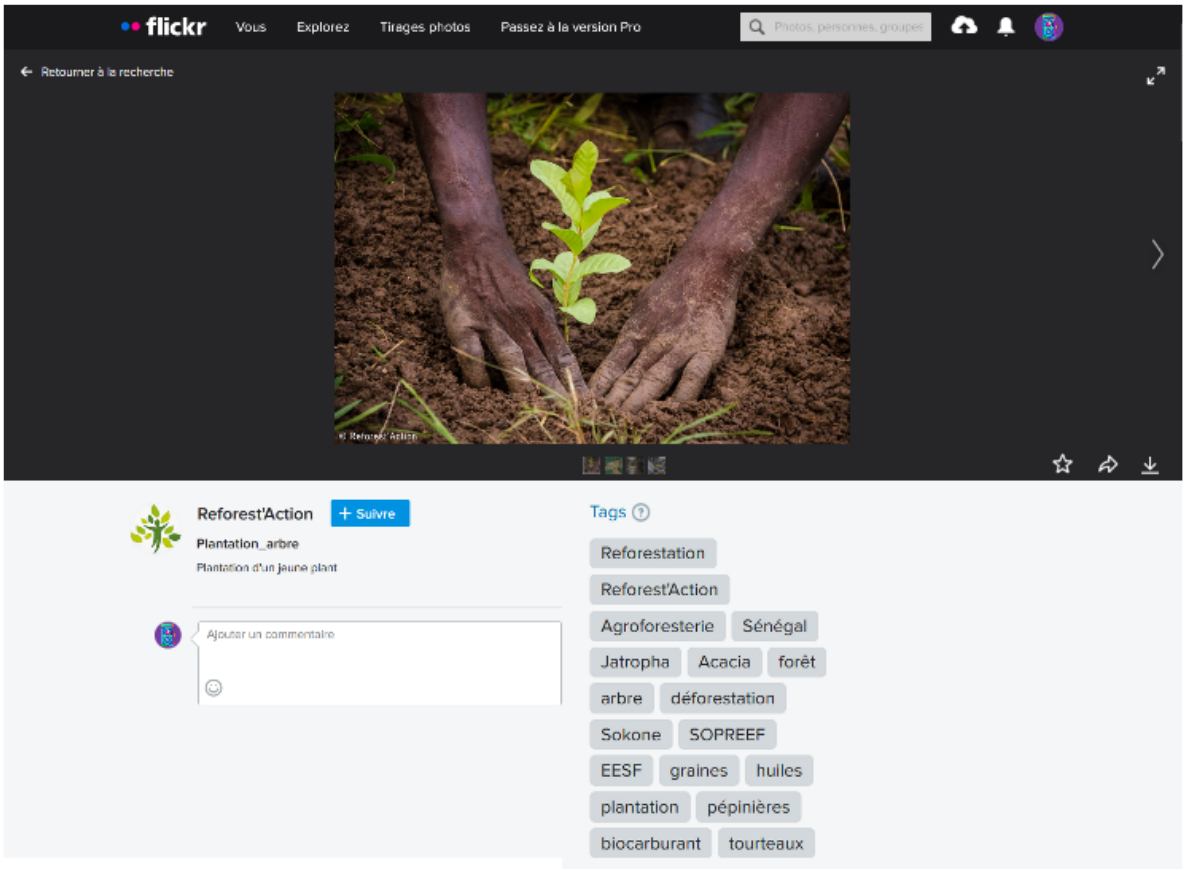


FIGURE 2.4 – *Exemple d'une photographie étiquetée avec Flickr.*

2.1.5 Les ontologies

Bien que nous les comparons ici aux listes d'autorité, taxonomies et thésaurus les ontologies n'ont pas une vocation exclusivement liée à l'indexation et à la recherche de document. Développée en ingénierie de connaissances, elles sont considérées comme des « spécifications explicites de conceptualisations » [Tom R. GRUBER, 1993]. Pour Roberto Poly, « une ontologie n'est pas un catalogue du monde , une taxonomie, une terminologie ou une liste d'objets, de choses ou autre. En fait, une ontologie est la structure générale dans laquelle les catalogues, les taxonomies, les terminologies peuvent être organisées de manière appropriée. » [POLI, 1996]. En faveur de cet argument, il n'est pas strictement nécessaire d'utiliser en ensemble de documents à indexer pour les constituer car une ontologie peut être constituée à partir de l'analyse d'une base de données, d'entretiens avec les experts d'un domaine ou d'un modèle déjà existant dans une spécialité. De plus la finalité des ontologies est qu'elles soient exploitées par des agents logiciels et non par des utilisateurs.
Mais les ontologies ont aussi des points communs avec les vocabulaires contrôlés. Si l'on compare les ontologies avec les thésaurus, on constate qu'ils partagent la notion de concepts et de termes qui désignent ces concepts et que pour ces deux dispositifs des relations pré-définies existent entre termes pour l'un et entre concepts pour l'autre. La différence notable se trouve dans les relations, là ou elles sont limitées pour les thésaurus à de simples relations hiérarchiques et associatives, les relations présentes au sein d'une ontologie sont sémantiques et ne se limitent donc qu'a la richesse de notre langage. Si l'on considérait donc que les ontologies sont des représentations de vocabulaires contrôlés au même titre que les listes d'autorité, les taxonomies et les thésaurus, la forme la plus aboutie serait représentée par les ontologies.

La figure 2.5 retrace les différents dispositifs selon un expressivité croissante. Cette expressivité repose sur la signification des concepts qui sont définis non seulement par le contenu textuel de leurs étiquettes mais aussi par les liens qu'ils partagent ensemble. L'expression "note sur application" englobe la définition textuelle d'un terme, une note historique ou un commentaire qui fournit à l'utilisateur plus de détail sur le terme ou le concept.

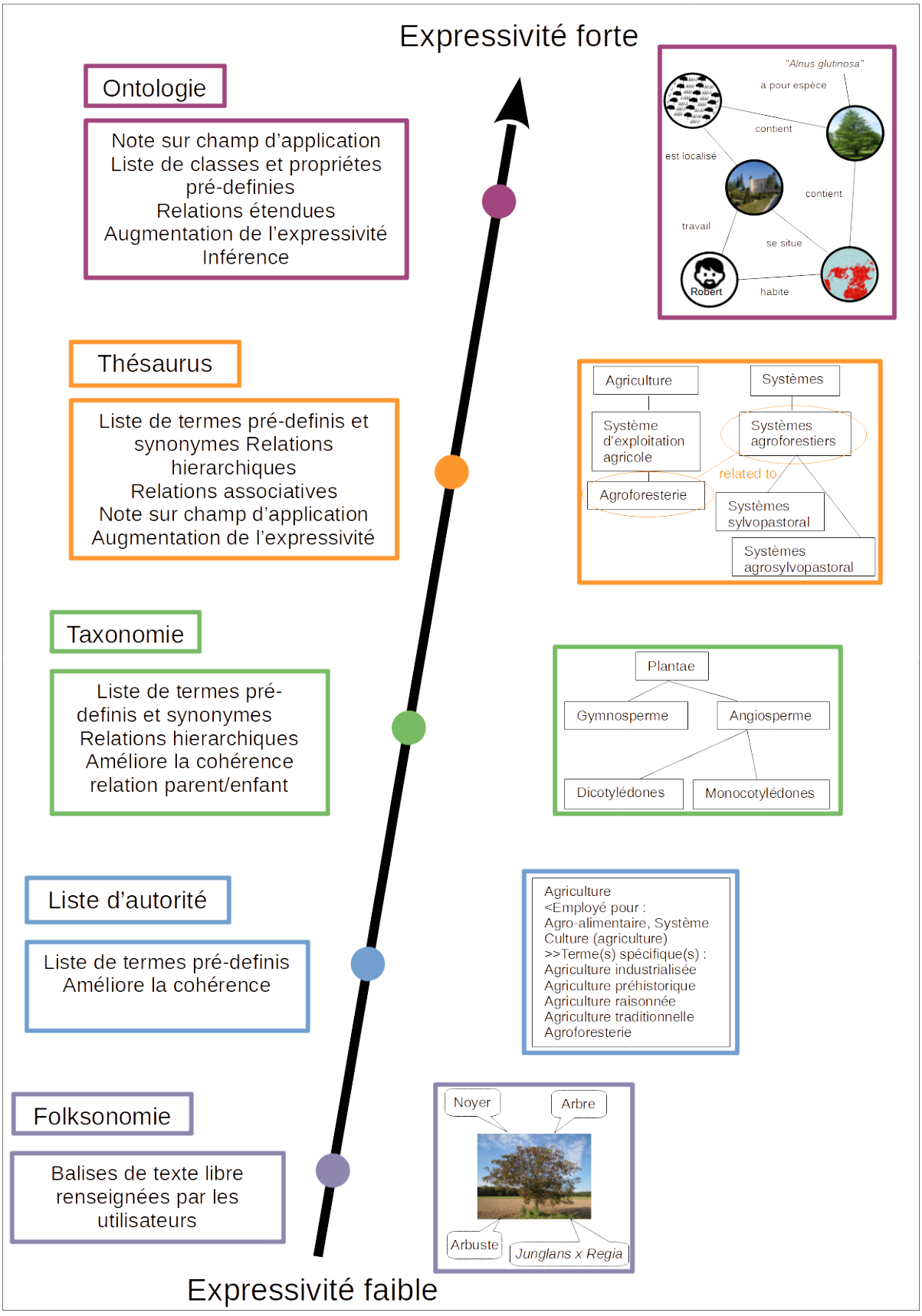


FIGURE 2.5 – *Les différents modes de classification.*

2.1.6 Normalisation des systèmes de gestion des connaissances

Ontologies, thésaurus et folksonomies « répondent à des besoins différents et peuvent très bien être combinés au sein d'une même application pour permettre différentes fonctionnalités, offrant ainsi différents modèles et permettant différentes pratiques pour l'indexation de contenus en ligne. » [GANDON et PASSANT, 2015]. Afin de permettre leur interopérabilité, des normes ont été mises en place pour les encadrer et les homogénéiser [DALBERG, 2017].

La norme ISO 25964

La norme ISO 25964 est publiée en deux partie: La premier partie ISO 25964-1:2011 [3] propose un cadre normatif pour le développement et la maintenance des thésaurus monolingues ou multilingues.
Pour cela elle fournit un modèle sous format UML (voir figure 2.6) à l'intention des concepteurs de thésaurus et des développeurs d'applications pour la gestion de collections d'objet ou de documents. Sur ce modèle, le concept occupe une place centrale avec la classe **ThesaurusConcept**. Ce concept est lié à au moins un élément de la classe **PreferredTerm** par la relation *hasPreferredLabel* et par zéro ou plusieurs élément de la classe **SimpleNonPreferredTerm** par la relation *hasNonPreferredLabel* traduisant le fait que le concept doit être labellisé par au moins un terme descripteur et potentiellement par plusieurs termes non descripteurs. On notera les classes d'association **TopLevelRelationShip**, **HierachicalRelationShip** et **AssociativeRelationShip** et qui traduisent les relations hiérarchiques et associatives possibles entre 2 concepts.
Cette partie recommande aussi un format XML pour l'importation et l'exportation de données de thésaurus afin de renforcer là encore l'interopérabilité au travers des réseaux de communication.
La seconde partie ISO 25964-2:2013 [4] se consacre à l'interopérabilité des thésaurus entre eux et avec d'autres types de vocabulaire couramment utilisés pour la recherche d'informations.
Pour cela elle décrit, distingue et compare les éléments et les caractéristiques des différents vocabulaires et recommande des modalités pour mettre en place et maintenir des correspondances entre plusieurs thésaurus ou entre thésaurus et d'autre vocabulaires.

La norme ISO 25964 a donc un rôle important pour la compatibilité entre les thésaurus et les autres types de vocabulaires, mais elle est insuffisante pour que les thésaurus et autres systèmes d'organisation des connaissances puissent être partagés sur le web. Pour exprimer les correspondances entre ces différents dispositifs sur le web une initiative développée par le W3C a aboutie à la création de SKOS.

SKOS

SKOS fournit un modèle pour exprimer la structure et le contenu de toutes sortes de systèmes d'organisation des connaissances y compris ceux qui ne sont pas conformes à la norme ISO 25964 dans le but de les intégrer au web sémantique. Étant une extension

3. ISO 25964-1:2011 : Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval. https://www.iso.org/standard/53657.html.
4. ISO 25964-2:2013 : Information and documentation — Thesauri and interoperability with other vocabularies — Part 2: Interoperability with other vocabularies. https://www.iso.org/standard/53658.html.

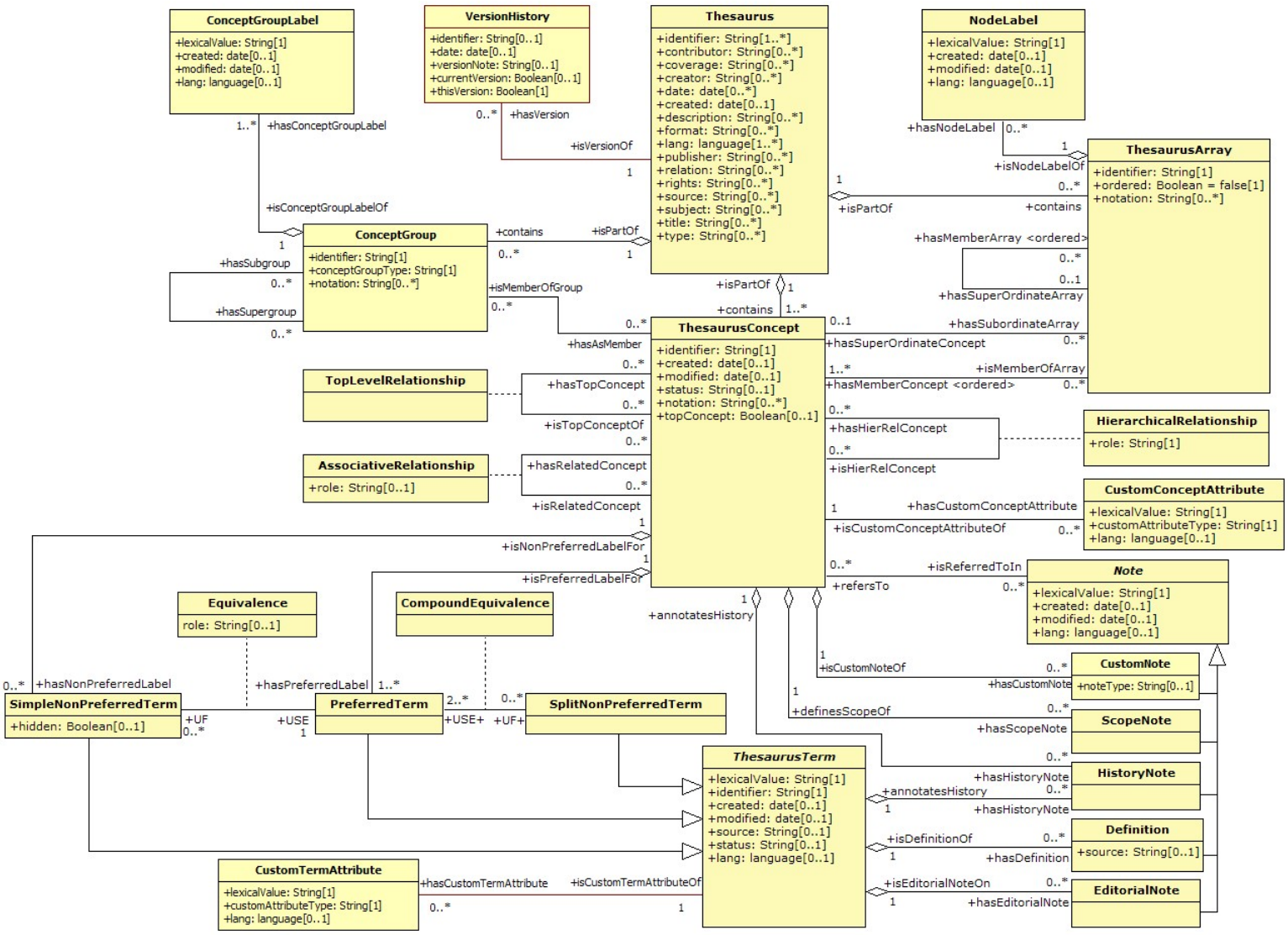


FIGURE 2.6 – *Modèle UML d'un thésaurus selon la norme ISO 25964.*

du RDF (voir chapitre 6 section 6.5.2), SKOS permet la constitution et publication de concepts au travers du web. Ces concepts pourront ainsi être reliés à des données sur le web et être intégrés à d'autres systèmes d'organisation des connaissances. SKOS n'a donc pas pour objectif de remplacer les différents dispositifs en indexation et recherche documentaire, mais de permettre leur partage, faciliter leur réutilisation et augmenter leur interopérabilité. Pour exemple un thésaurus du domaine de l'agroforesterie peut être représenté dans une ontologie grâce à SKOS mais celle ci ne sera pas pour autant une ontologie du domaine de l'agroforesterie et sera simplement une ontologie décrivant ce type de thésaurus. La philosophie de SKOS est d'exploiter au maximum les caractéristiques communes et les similitudes d'usages des différents systèmes d'organisation des connaissances. Dans le module de base de SKOS, l'élément central est là encore le concept. Ce concept est labellisé par des chaînes littérales dans un ou plusieurs langages et peut être documenté par divers types de notes. Différents concepts peuvent être impliqués dans des relations hiérarchiques ou associatives. Les concepts peuvent être regroupés dans des schémas de concepts. En fournissant un langage de modélisation conceptuel délibérément léger et intuitif, SKOS prépare aux développement et au partage de futurs nouveaux systèmes d'organisation des connaissances.

Dans le cas où les spécificités de certains systèmes d'organisation des connaissances ne peuvent être traduites, le modèle de base de SKOS peut se voir doté d'extensions comme par exemple SKOS-XL [5] qui fournit un support supplémentaire pour la description et la liaison des entités lexicales. En effet SKOS-XL opère une réification des chaînes littérales, qui permettent d'étiqueter les concepts, en créant une classe **Label** permettant ainsi aux termes d'être des entités en soi auxquelles on peut attacher des propriétés. Les équipes responsables du développement de la norme ISO 25964 et de SKOS, conscientes de l'importance de l'interopérabilité nécessaire aux échanges d'informations sur le web ont mis en place des correspondances entre les différents éléments et propriétés de leur modèles respectifs [ISAAC, 2013]. Dans cet article, Van Assem et ses collaborateurs [VAN ASSEM et al., 2006] rendent compte de leur méthode pour convertir un thésaurus avec SKOS.

2.2 Quelques exemples de vocabulaires contrôlés

2.2.1 Vocabulaires en Agroforesterie

Au cours des années 80, P.K.R. Nair propose une classification des systèmes agroforestiers (voir Fig.2.7) pour les évaluer et élaborer des stratégies pour leur amélioration [NAIR, 1985]. Son approche se veut simple, pragmatique et basée sur les objectifs. Dans un premier temps le système agroforestier est classé en fonction de la nature de ces composants identifiés au nombre de trois (arbres, plantes herbacées (cultures ou pâturage) et animaux) ce qui aboutit principalement aux catégories de systèmes suivantes : agrosylviculture (cultures et arbres), sylvopastoralisme (animaux et arbres) et agrosylvopastoralisme (cultures, animaux et arbres). Par la suite les rôles ou les fonctions attendus de chacun des composants (en particulier les arbres) sont pris en compte comme la fourniture en bois de chauffage pour les arbres exploités. Ces rôles et fonctions seront placés par la suite en préfixe de la catégorie par exemple système d'agrosylviculture pour la conservation des sols et la production de nourriture.
Par la suite F.L. Sinclair propose un autre système de classification [Fergus L. SINCLAIR, 1999] utilisant la "pratique" comme unité de classification plutôt que le "système" estimant que les expérimentations agroforestières sont très rarement des systèmes agricoles ou forestiers complets.

Plus récemment, un travail de construction de thésaurus [BURRIEL et al., 2017] a été réalisé dans le cadre des projets européens AgroFE (Agroforestry Training in Europe) [*AgroFE : projet Leonardo transfert d'innovation* 2020] et Agrof-MM (Agroforestry – Training – Mediterranean and Mountain Erasmus+) [*Agrof-MM* 2016] visant à mettre en place une banque de données pour le développement de formations et la fourniture de connaissances en agroforesterie. Un extrait de cette classification est visible sur la figure 2.8.

5. SKOS-XL : Simple Knowledge Organization System eXtension for Labels. https://www.w3.org/TR/skos-reference/skos-xl.

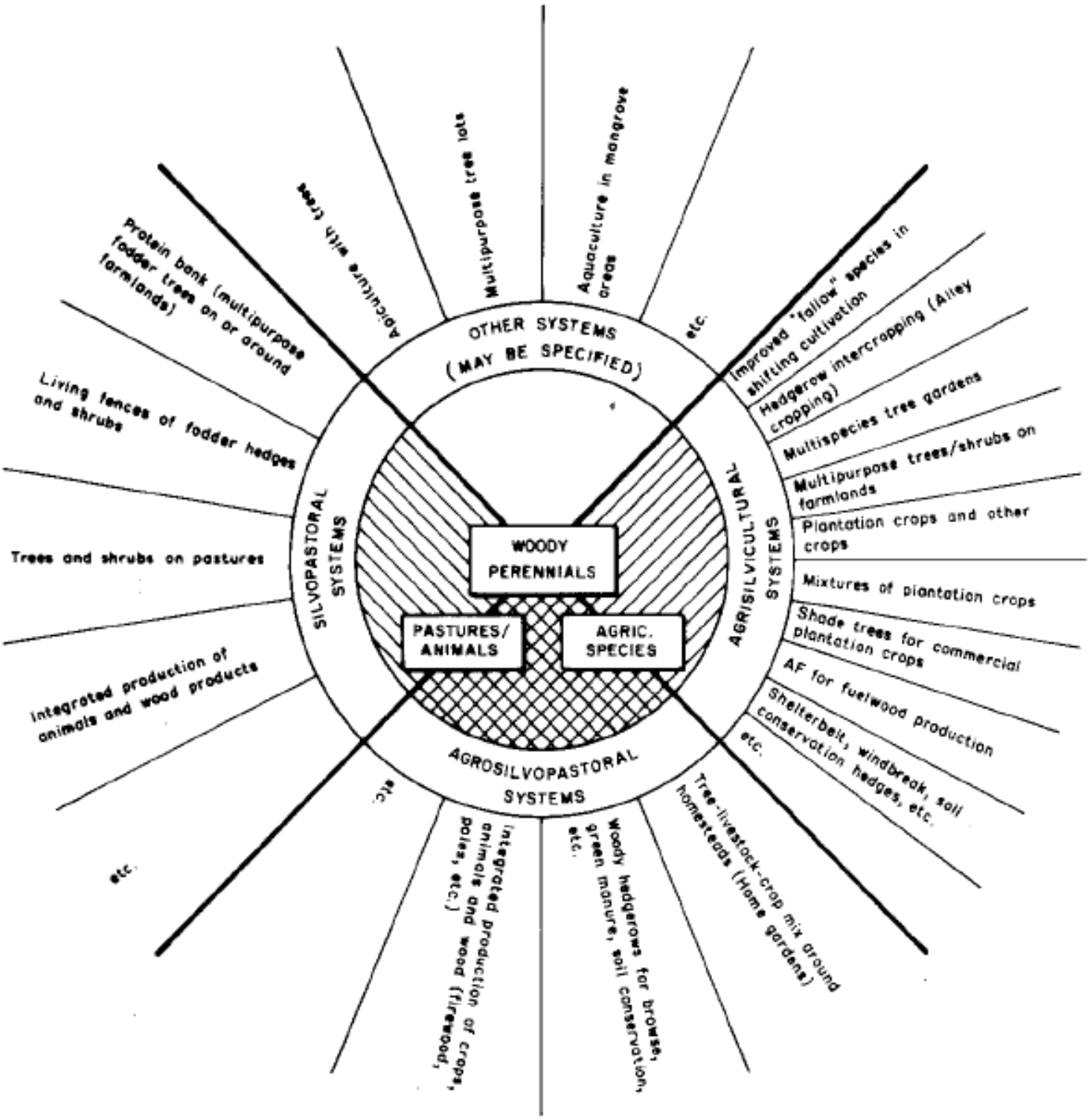


FIGURE 2.7 – *Classification de P.K.R Nair: Catégorisation des systèmes agroforestiers en fonction de la nature des composants (avec des exemples de sous-systèmes/pratiques communs à chaque système).*

Agroforestry : Agrof-MM Thesaurus for Professional and Education-Training Usages

English					Français	Bulgarian e Български език	greek
level 1	level 2	level 3	level 4	Level 5			
Agroforestry					agroforesterie	Агролесовъдство	αγροδασοπονία
1) Agroforestry systems					systèmes agroforestiers	Агролесовъдни системи	αγροδασικά συστήματα
	Agrosilvopastoralism				agro-sylvopastoralisme	Агрогорскипасторализъм	αγροδασογεωργικά συστήματα
		Faidherbia, karité... parklands			parcs à karité, faidherbia...	Агролесовъдна система с участието на дървесния вид Faidherbia sp. (Fabaceae)	παρκα
		Oasis			oasis	Оазис	οάσεις
		High value tree systems			Systèmes à arbres de forte valeur (AGFORWARD)	Системи с висока природна и културна стойност (AGFORWARD)	συστήματα με δέντρα υψηλής αξίας
		Arable and grazed orchards			Vergers cultivés et pâturés	Обработваеми и затревени овощни градини	καλλιεργούμενοι και βοσκόμενοι οπορώνες
			Olive trees systems		Oliveraies cultivées et pâturées	Системи с маслинови дървета (Olea europaea L.)	συστήματα ελιάς

FIGURE 2.8 – *AFAF: Extrait du thésaurus Agrof-MM.*

2.2.2 Vocabulaires en agriculture et écologie

Les exemples suivants ont été choisis pour de multiples raisons dont les suivantes:

- Ils ciblent des domaines connexes au domaine de l'agroforesterie.
- Leurs degrés de formalisation et de standardisation sont élevés et s'appuient sur les standards du web sémantique.
- Pour certains de ces vocabulaires, bon nombre des concepts que nous avons identifiés comme faisant partie du domaine de l'agroforesterie ont été retrouvés (AGROVOC, ENVO).
- Pour d'autres, ils peuvent nous aider à redéfinir certains termes composites de l'agroforesterie dans une approche de description « entité/qualité » équivalentes (PATO).

AGOROVOC :

AGROVOC [6] est à l'origine un thésaurus créé dans les années 80 par la FAO [7](Organisation des Nations unies pour l'alimentation et l'agriculture). Son but est de fournir un vocabulaire contrôlé pour l'indexation des publications scientifiques et technologiques dans le domaine de l'agriculture incluant la foresterie et l'élevage d'animaux [CARACCIOLO et al., 2013]. Numérisé à partir des années 2000, AGROVOC s'oriente vers les standards du Linked Open Data [8] (LOD) et adopte les standards OWL,SKOS et son extension SKOS-XL (extension de SKOS permettant l'expression d'assertions et de relations entre termes ; exemple : <FAO> <est l'acronyme de> <Food and Agriculture Organization>). Ce travail de formalisation consistant à transformer ce thésaurus en ontologie est décrit par Dagobert Soergel et collaborateur [SOERGEL et al., 2004].
Toujours managé par la FAO, AGROVOC est maintenant propriété d'une communauté internationale et est maintenu par des experts et des institutions du domaine de l'agriculture. AGROVOC rassemble 32 000 concepts définis en 20 langues différentes. Ces 32 000 concepts sont organisés hiérarchiquement à partir de 25 « top concepts » (organism, substances, entities, phenomena ...). AGROVOC est aligné avec plus d'une trentaine de vocabulaires à travers le monde (CASS, NALT, DBPEDIA, DNB, BNCF, EARTh, EUROVOC, ...) Les données d'AGROVOC peuvent être visualisées sur le web via plusieurs interfaces. La figure 2.9 montre l'interface web à disposition du grand public. La recherche d'un concept à partir d'un terme (ici agroforesterie) peut se faire en plusieurs langues. Sur le panneau de gauche on visualise le concept au sein de son arbre hiérarchique et sur le panneau de droite, différents renseignements comme sa définition sous forme textuelle, le concept qui le généralise: "Système d'exploitation agricole", celui avec lequel il est en relation: "Systèmes agroforestiers", les différents termes dans d'autres langues avec lesquels il est désigné, ... L'adoption des standards du LOD permet aussi d'accéder directement aux concepts d'AGROVOC via l'URL d'un concept comme le montre la figure 2.10 où grâce à l'outil Loddy [9] spécialisé dans la publication des donnés du LOD on accède au concept "Arbuste" par son URL http://aims.fao.org/aos/agrovoc/c_26823. Les données sont alors présentées sous format RDF. Les 3 panneaux exposent les différents triplets dont le concept

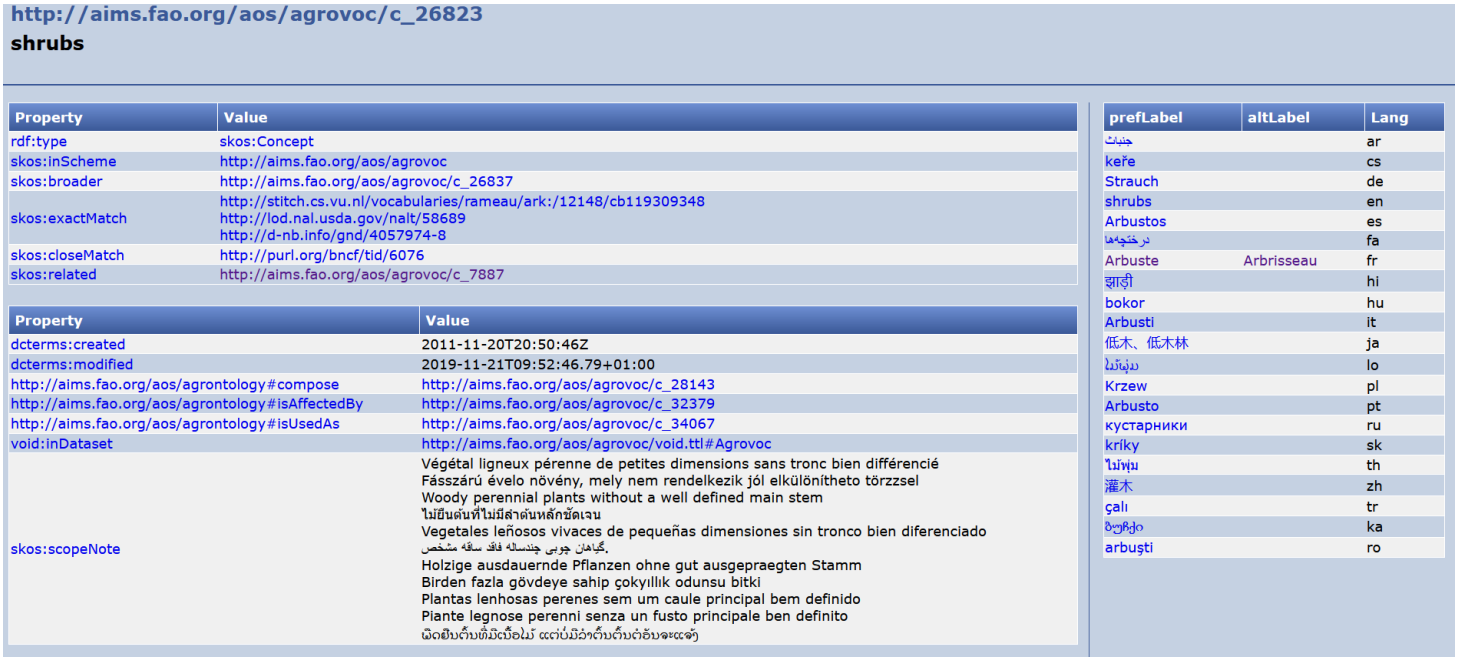
6. AGROVOC : A thesaurus for agriculture. http://www.fao.org/agrovoc/fr.
7. FAO : Food and Agriculture Organisation of the United Nations. http://www.fao.org/home/fr/.
8. LOD : Linked Open Data. https://lod-cloud.net/.
9. Loddy : lightweight system for supporting publication of Linked Open Data. https://bitbucket.org/art-uniroma2/loddy/src/master/README.md.



FIGURE 2.9 – *Description du concept c_207 ("Agroforesterie"), sur le site web d'AGROVOC.*

"Arbuste" identifié par son URI est le sujet.
Le panneau inférieur gauche fournit les triplets décrivant les caractéristiques du concept incluant par exemple sa date de création liée par la relation *dcterms:created* et de modification liée par la relation *dcterms:modified* et des notes d'information textuelle en plusieurs langues relié par la relation *skos:scopeNote* à celui-ci. On retrouve aussi des relations sémantiques appartenant au vocabulaire d'AGROVOC comme *isAffectedBy* liant le concept "Arbuste" au concept http://aims.fao.org/aos/agrovoc/c_32376 (concept "Arboricide" d'AGROVOC traduisant le fait qu'une arbuste puisse être affecté par un arboricide).
Le panneau supérieur gauche nous indique les triplets avec les principales relations SKOS exprimant les relations de hiérarchisation et d'association du concept "Arbuste" avec d'autres concepts comme on le voit avec la relation *skos:broader* exprimant la généralisation de ce concept avec le concept AGROVOC "Plante ligneuse" identifié par l'URI http://aims.fao.org/aos/agrovoc/c_26837. Toujours sur cette même figure, on notera que les relations ne se limitent pas à lier les concepts d'AGROVOC entre eux comme on le voit avec la relation *skos:exactMatch* qui lie notre concept dans une relation d'équivalence avec les concepts http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb119309348 (Concept "Arbuste" de la liste d'autorité française RAMEAU), http://lod.nal.usda.gov/nalt/58689 (Concept "shrubs" du thésaurus américain NAL [10]), et http://d-nb.info/gnd/4057974-8 (Concept "Strauch" du catalogue de la bibliothèque nationale allemande DNB [11])). Cela illustre bien l'utilisation de SKOS pour lier les concepts d'AGROVOC à d'autres vocabulaires. Mais cet alignement est coûteux en terme de réalisation, et si certaines correspondances peuvent être réalisées de manière automatisée (en particulier les relations d'équivalence), d'autres nécessitent le concours de nombreux experts en documentation spécialisés pour chaque domaine abordé par le vocabulaire AGROVOC. Pour pallier cette difficulté, les concepteurs d'AGROVOC ont donc délibérément limité ces correspondances entre concepts à des relations d'équivalences pures *skos:exactMatch* plus facilement établies via des outils informatiques. Le dernier panneau situé à gauche présente les labels en différentes langues permettant d'identifier ce concept "Arbuste". Ce sont des liens qui renvoient vers d'autres pages comme le montre en exemple la figure 2.11 avec l'activation du lien "Arbuste" descripteur de notre concept en français. Cette nouvelle page nous renseigne sur le fait que ce descripteur est plus qu'une simple chaîne de caractères mais l'instance d'une classe du vocabulaire SKOS-XL, la classe **skosxl:Label** possédant de facto sa propre URI http://aims.fao.org/aos/agrovoc/xl_fr_1299493512878. Cette réification d'un littéral en concept constitue le rôle prépondérant de l'utilisation de SKOS-XL car pour rappel un triplet RDF ne peut avoir en sujet un littéral. SKOS-XL va permettre ainsi une modélisation linguistique plus fine du concept "Arbuste" en permettant la mise en place de relations avec des étiquettes en tant que sujet. Parmi ces relations, des relations lexicales peuvent être mises en place entre étiquette comme on le visualise avec le lien créé entre notre étiquette http://aims.fao.org/aos/agrovoc/xl_fr_1299493512878 dont la forme lexicale est "Arbuste" et l'étiquette http://aims.fao.org/aos/agrovoc/xl_fr_1299493513675 dont la forme lexicale est "Arbrisseau" par la relation du vocabulaire AGROVOC *hasRelatedTerm* traduisant que le terme "Arbuste" peut être mis en relation avec le terme "Arbrisseau". Un des avantages de cette association entre termes lexicaux est qu'elle peut être faite sans l'implication des concepts désignés par ces termes ce qui permet d'augmenter l'efficience de la recherche d'un utilisateur en fournissant plus de termes associés aux termes de sa recherche sans affecter la qualité de la définition d'un concept vis à vis d'un autre. D'autres types de relations peuvent aussi être appliqués sur ces étiquettes réifiées apportant plus d'informations

10. NAL : The National Agricultural Library's Agricultural Thesaurus and Glossary. https://agclass.nal.usda.gov/.
11. DNB Katalog : Katalog der Deutschen Nationalbibliothek. https://portal.dnb.de.



http://aims.fao.org/aos/agrovoc/c_26823
shrubs

Property	Value
rdf:type	skos:Concept
skos:inScheme	http://aims.fao.org/aos/agrovoc
skos:broader	http://aims.fao.org/aos/agrovoc/c_26837
skos:exactMatch	http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb119309348 http://lod.nal.usda.gov/nalt/58689 http://d-nb.info/gnd/4057974-8
skos:closeMatch	http://purl.org/bncf/tid/6076
skos:related	http://aims.fao.org/aos/agrovoc/c_7887

Property	Value
dcterms:created	2011-11-20T20:50:46Z
dcterms:modified	2019-11-21T09:52:46.79+01:00
http://aims.fao.org/aos/agrontology#compose	http://aims.fao.org/aos/agrovoc/c_28143
http://aims.fao.org/aos/agrontology#isAffectedBy	http://aims.fao.org/aos/agrovoc/c_32379
http://aims.fao.org/aos/agrontology#isUsedAs	http://aims.fao.org/aos/agrovoc/c_34067
void:inDataset	http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc
skos:scopeNote	Végétal ligneux pérenne de petites dimensions sans tronc bien différencié Fásszárú évelo növény, mely nem rendelkezik jól elkülönítheto törzzsel Woody perennial plants without a well defined main stem ไม้ยืนต้นที่ไม่มีลำต้นหลักชัดเจน Vegetales leñosos vivaces de pequeñas dimensiones sin tronco bien diferenciado گیاهان چوبی چندساله فاقد ساقه مشخص. Holzige ausdauernde Pflanzen ohne gut ausgepraegten Stamm Birden fazla gövdeye sahip çokyıllık odunsu bitki Plantas lenhosas perenes sem um caule principal bem definido Piante legnose perenni senza un fusto principale ben definito ผิวเผินต็นที่มีเนื้อไม้ ตต่ปมีลำต็นต็นต่อับจะตอ

prefLabel	altLabel	Lang
جنبات		ar
keře		cs
Strauch		de
shrubs		en
Arbustos		es
درختچه‌ها		fa
Arbuste	Arbrisseau	fr
झाड़ी		hi
bokor		hu
Arbusti		it
低木、低木林		ja
ไม้พุ่ม		lo
Krzew		pl
Arbusto		pt
кустарники		ru
kríky		sk
ไม้พุ่ม		th
灌木		zh
çalı		tr
ბუჩქი		ka
arbuști		ro

FIGURE 2.10 – *Description du concept c_26823 ("Arbuste"), visualisé sous forme de page HTML par Loddy.*

comme leur date de création ou leur appartenance à un référentiel particulier comme on le visualise avec la relation *void:inDataset* qui lie cette chaîne littérale "Arbrisseau" au référentiel d'AGROVOC. Les machines peuvent elles aussi accéder aux informations



http://aims.fao.org/aos/agrovoc/xl_fr_1299493512878

Property	Value
rdf:type	skosxl:Label

Property	Value
dcterms:created	1989-09-20T00:00:00Z
http://aims.fao.org/aos/agrontology#hasRelatedTerm	http://aims.fao.org/aos/agrovoc/xl_fr_1299493513675
void:inDataset	http://aims.fao.org/aos/agrovoc/void.ttl#Agrovoc
skos:notation	26823
skosxl:literalForm	Arbuste

prefLabel	altLabel	Lang
		fr

FIGURE 2.11 – *Description du label ("Arbuste"), visualisée sous forme de page HTML par Loddy.*

du LOD directement via les URI adéquates, mais il existe aussi d'autres points d'accès aux données d'AGROVOC comme des services web ou un "endpoint SPARQL [12]" permettant aux applications d'intégrer de manière transparente le vocabulaire AGROVOC.

ENVO :

ENVO [13] est une ontologie pour la description concise et contrôlée des environnements. Initié en 2013, la construction de cette ontologie fut axée initialement sur les aires biotiques, les caractéristiques environnementales et les matériaux environnementaux dans les domaines de la recherche génomique et microbiologique.

12. AGROVOC Thesaurus : SPARQL service. https://agrovoc.uniroma2.it/sparql/.
13. ENVO : Environment Ontology - Ontology of environmental features and habitats . http://www.obofoundry.org/ontology/envo.html.

Par la suite, face à son utilisation dans d'autres spécialités, ENVO a été élargie, améliorée et généralisée pour répondre au besoin d'une sémantique environnementale commune à une multitude de domaines [BUTTIGIEG, PAFILIS et al., 2016]. ENVO fournit la description des habitats, des processus écologiques, des milieux anthropiques et des entités impliquées au sein des initiatives en matière de santé environnementale et de développement durable.

PATO :

PATO [14] est une ontologie descriptive de la qualité phénotypique. Elle rassemble 2014 termes conçus pour être utilisés avec des ontologies d'"entités porteuses de qualité". Pour illustrer son utilisation, l'article de Georgios V. Gkoutos et collaborateurs [GKOUTOS et al., s. d.] nous propose une démarche permettant de redéfinir les termes de l'ontologie HPO [15] (Human Phenotype Ontologie), qui fournit un vocabulaire normalisé des anomalies phénotypiques rencontrées dans les maladies humaines, à l'aide d'autres ontologies pour relier et interpréter informatiquement des relations entre des composants anatomiques et des maladies décrites avec des phénotypes anormaux et les gènes qui en sont responsables. Les exemples sont donnés dans le domaine des anomalies phénotypiques musculo-squelettique.
Les termes de HPO utilisent 2 modèles :

- Succession de termes :
  « Mitral valve prolapse » (phénotype) que l'on peut décomposer en « mitral valve » (entité anatomique) et « prolapsed » (qualité).

- Terme unique :
  « Arachnodactyly » (phénotype) que l'on peut décomposer en « finger » (entité anatomique) et « long and slender(mince) » (qualité).

Or les termes de HPO en mêlant entité et qualité ne facilitent pas le traitement informatique et le rapprochement avec d'autres ontologies. Les auteurs proposent donc en utilisant PATO et d'autres ontologies comme FMA (Foundational Model of Anatomy) ,GO (Gene ontology), Cell Type ontology, ... de créer des descriptions « entité/qualité » équivalentes pour les termes composites (phénotypique) de HPO. La figure 2.12 montre cette redéfinition. Les auteurs nous ont donc montré les avantages donnés par

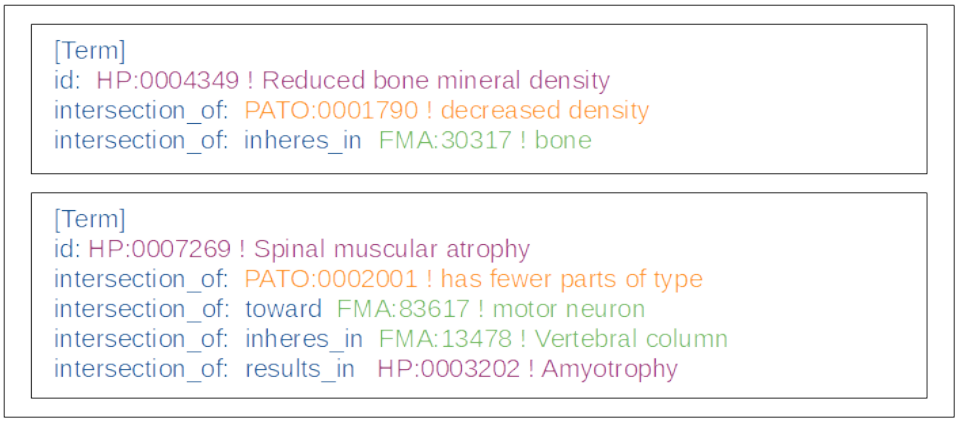


FIGURE 2.12 – *Définition des termes "Reduced bone mineral density" et "Spinal muscular atrophy" dans une approche de description « entité/qualité » avec PATO.*

14. PATO : Phenotype And Trait Ontology. http://www.obofoundry.org/ontology/pato.html.
15. HPO : Human Phenotype Ontologie. https://hpo.jax.org/app/.

le fait de fournir aux termes d'une ontologie une description « entité/qualité » équivalente qui utilise les termes de cette même description. Ainsi les termes propres à un domaine de connaissance sont respectés et leurs descriptions équivalentes formées sur la base d'ontologie de termes simples et moins ambigus permettent un meilleur traitement des données et une plus grande possibilité de croisement des données avec un autre domaine comme par exemple le domaine des humains et celui de animaux. En appliquant ce principe aux termes composites ou porteurs de "qualité" de l'agroforesterie nous pourrions ainsi redéfinir ces termes avec les termes d'autres ontologies augmentant ainsi leur visibilité et l'efficience de leur description. La figure 2.13 montre un exemple de ce que serait la redéfinition des termes "Augmentation de la biodiversité" et "Bande enherbée" de notre vocabulaire sur l'agroforesterie (préfixé par AFV:) avec des termes empruntés aux ontologies PATO et AGROVOC. On note avec la définition du concept "AFV:000010 !Bande enherbée", la réutilisation du terme "AFV:000001 !Augmentation de la biodiversité"

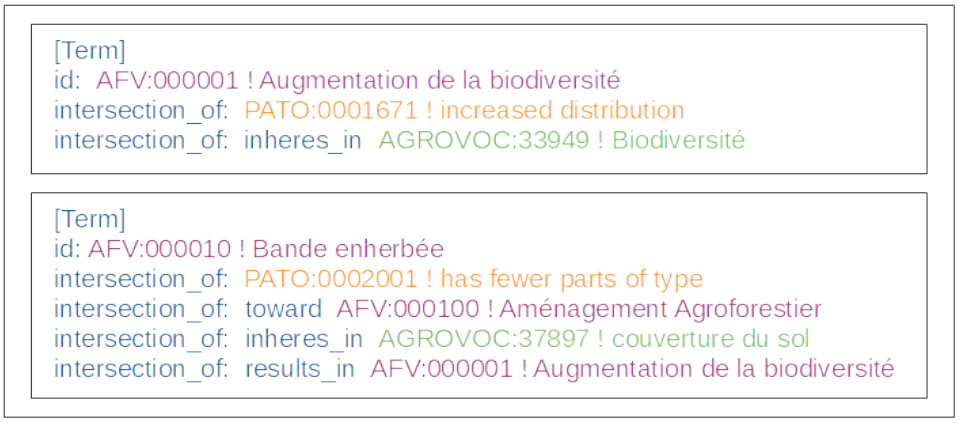


FIGURE 2.13 – *Définition des termes "Augmentation de la biodiversité" et "Bande enherbée" dans une approche de description « entité/qualité » avec PATO.*

2.3 Quelques exemples de systèmes d'information et bases de connaissance

Les systèmes d'informations permettent de stocker et de faire circuler l'information entre différents protagonistes que se soit dans l'univers d'une entreprise ou à un niveau plus large comme pour tout les citoyens d'un pays. Reposant en grande partie sur l'utilisation d'outils informatiques comme des réseaux d'ordinateurs et d'outils de télécommunications, ils facilitent la collecte, le stockage, le traitement et la distribution de l'information de la plupart de nos organisations sociales et économiques. Au sein des systèmes d'informations se trouvent des bases de données collectionnant les informations de façon plus ou moins organisée par exemple sous forme de lignes, de colonnes ou de tableaux facilitant la gestion des données. Mais toutes les informations d'une organisation ne sont pas toujours structurées comme on le voit avec les documents électroniques comme les emails. L'ingénierie des connaissances à pour but de gérer les informations brutes d'une organisation et d'en assurer les échanges entre partenaires. Une des voies utilisée pour transformer les données brutes en véritable connaissance est représentée par l'utilisation de bases de connaissance [GORIA, 2006]. Les bases de connaissance rassemblent les connaissances spécifiques d'un domaine de spécialité pouvant être traitées par des ordinateurs. En opposition aux bases de

données où l'information et organisée et structurée de façon immuable (pour exemple champ "Hauteur" et champ "Circonférence" d'une base de données sur les arbres), les bases de connaissance contiennent en particulier des faits (une plante ligneuse de mon jardin mesure 10 mètres) et des règles (toutes les plantes ligneuses de plus de sept mètres sont des arbres). La présence de règles permet la déduction de nouveaux faits (la plante ligneuse de mon jardin est un arbre) à l'aide d'un moteur d'inférence simulant des déductions logiques [BAADER, CALVANESE et al., 2003]. Une base de connaissance est en fait une ontologie à laquelle on rajouté des individus du monde réel.

2.3.1 Dans les domaines "grand public"

Freebase

En 2007 la société Metaweb lance l'application Freebase « une base de données créée et éditée en collaboration, contenant des informations générales et structurées destinées à un large public » [KURT et al., 2007] marquant le début d'un accès grand public au web sémantique. Au travers de thèmes génériques comme "arts et divertissements", "Société" ou "Sport" des types de ressources sont proposés comme "livres", "film", "auteur", "acteur", "personnage", "ville". A ces types de ressources sont affectées des propriétés comme par exemple pour un livre son auteur, son éditeur ou son nombre de pages qui relie cette ressource à une autre ressource ou à une donnée typée (littéral, nombre, date, ...). Au travers d'une interface centrée sur une ressource particulière Freebase expose les différentes propriétés rattachées à cette ressource (voir figure 2.14), Les contributeurs peuvent aussi créer une nouvelle ressource en profitant du cadre offert par les propriétés déjà affectées au type de ressource concernée et aux ressources déjà existantes mais peuvent aussi créer de nouveaux type de ressources ou de nouvelles propriétés rattachées [BOLLACKER et al., 2008]. En 2010 la société Metaweb est rachetée par Google qui utilisera la base de données de Freebase pour alimenter en 2012 sa propre base de connaissance le "Knowledge Graph" un système basé sur des algorithmes sémantiques proposant des données complémentaires à la requête des utilisateurs de son moteur de recherche. En 2014 les contributions des utilisateurs de Freebase sont stoppées après une production de 1.9 billions de triplets RDF et le site sera définitivement fermé en 2016.

Moteur de recherche Google

Bien que principalement basé sur des mécanismes d'indexation permettant une recherche par mots clés, Le moteur de recherche Google utilise une base de connaissance pour compléter les informations fournies à l'utilisateur sur sa page de résultats. Le "Knowledge Panel", le panneau à droite de la page (voir figure 2.15) montre les propriétés ainsi que les valeurs et ressources qui sont rattachées à l'entité nommée <Tom Gruber> sujet de la recherche. Les informations présentées sont issues du "Knowledge Graph" de Google et ne sont modifiables que par les personnes morales désignées par la ressource.

DBpedia

DBpedia a pour but « de maintenir des données structurées extraites de différents chapitres de Wikipedia » [*DBpedia* 2021]. Des connaissances multilingues sur de nombreux domaines de compétences sont mises à disposition de tous sous format RDF accessible sur le web [LEHMANN et al., 2015] permettant ainsi à des programmeurs informatiques

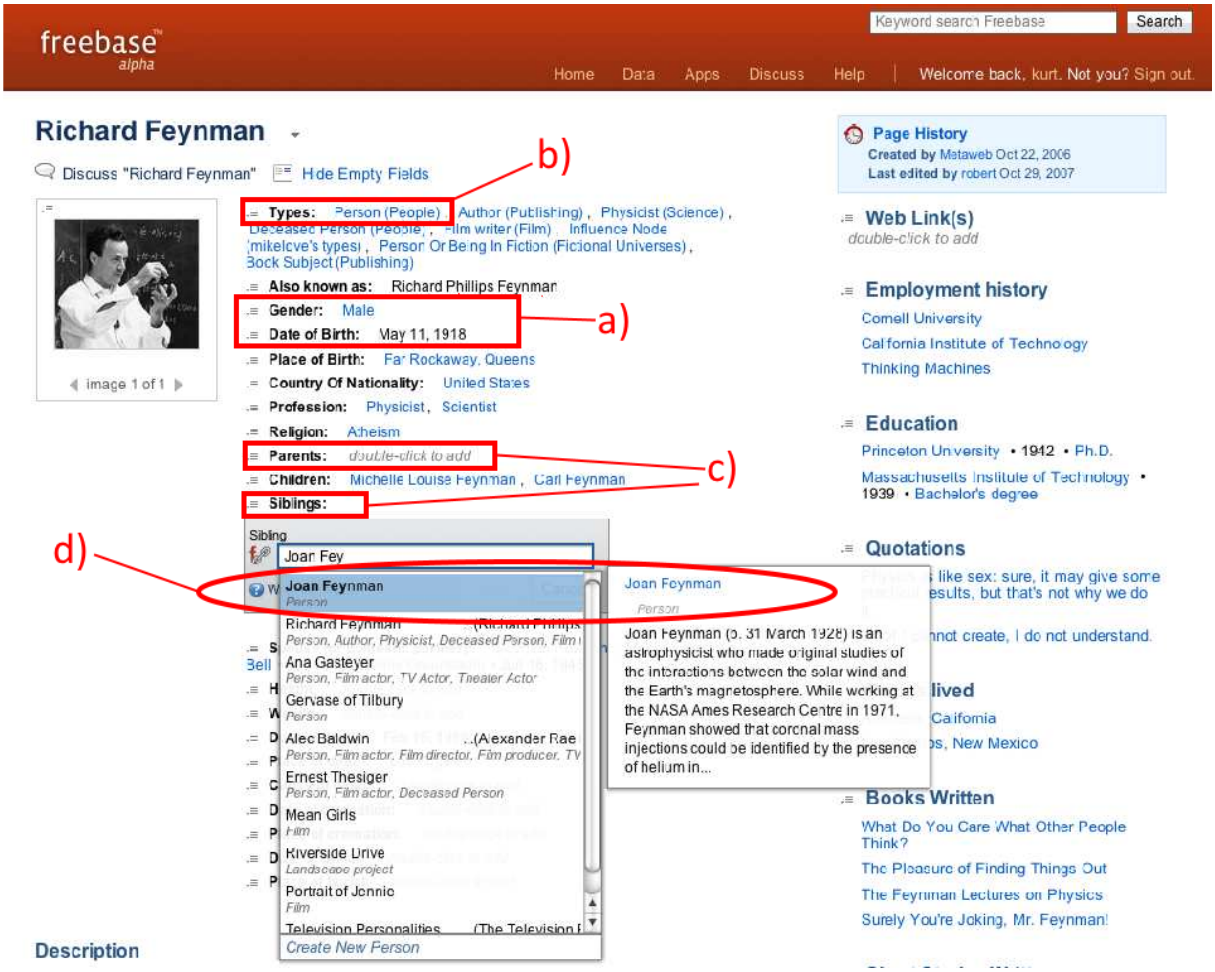


FIGURE 2.14 – *Freebase: Description du physicien Richard Feynman*
Les propriétés ("Gender", "Date of Birth",...) a) sont celles rattachées au type "Person" b) de la ressouce. Les propriétés non renseignées comme "Parents" et "Sibling" c) peuvent complétées par un contributeur. Lors de la saisie du frère de Richard, la ressource Joan Freynman d) qui est déjà saisie dans freebase est proposé au contributeur.

de récupérer ces données via les outils du web sémantique. En se conformant aux standards du réseau "Linked Open Data" (LOD), une initiative du web sémantique pour un accès et une réutilisation des données sans restriction, DBpedia est une base de connaissance libre de toute licence d'exploitation et liée à de nombreuses autres bases. DBpedia est devenue l'élément central de ce réseau par le volume de ses données et les relations qui la lient avec d'autres bases de connaissance ouvertes et liées du web avec neuf milliard et demi de triplets RDF décrivant à peu près cinq milliard de ressources.

2.3.2 Dans le domaine agroforestier

Il est acté [LINIGER et al., 1998] qu'il existe un manque de compréhension quantitative et prévisionnelle au sujet des pratiques agroforestières et de leur importance afin d'en faciliter l'adoption. A aujourd'hui il n'existe pas de base de connaissance dans le domaine de l'agroforesterie.

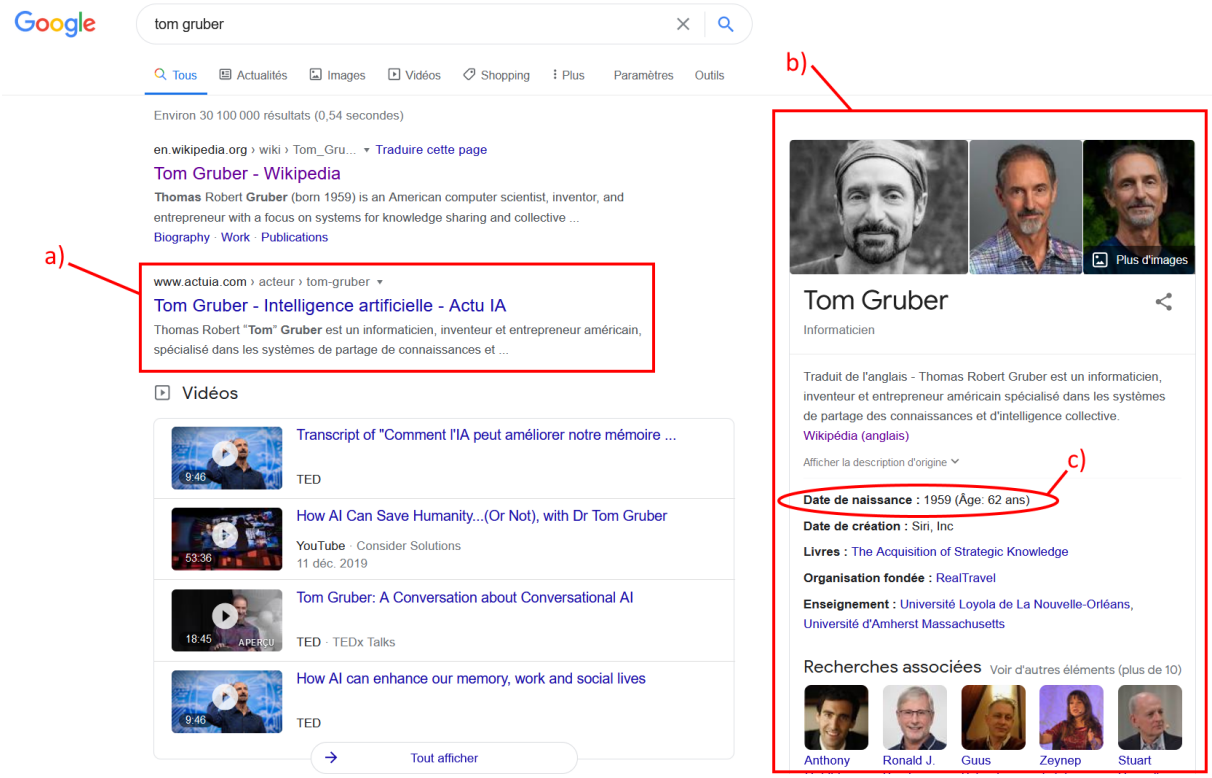


FIGURE 2.15 – *Résultat de la recherche Google sur Tom Gruber: a) Page HTML issue de la recherche par indexation ; b) Le "Knowledge Panel" affiche les données issues de la recherche sémantique: images, définition de la page Wikipédia, propriétés et valeurs ; c) La propriété "Date de naissance" et sa valeur rattachée à l'entité nommée <Tom Gruber>.*

2.3.3 Dans le domaine agricole

Dans le domaine de l'agriculture, différents projets proposent des modèles conceptuels ou des systèmes d'information avec base de connaissance pour collecter, traiter et communiquer les informations.

GIEA :

Le projet GIEA (Gestion des Informations de l'Exploitation Agricole) [DUFY et al., 2006] s'adosse à l'expertise humaine pour identifier les concepts clés relatifs à la communauté de pratique en agriculture et dégager un langage métier commun à l'ensemble des acteurs qui va permettre la compréhension des données échangées. Trois sous-thèmes ont été investis à cet effet, à savoir le sol, l'élevage et l'exploitation.

ICASA :

Le consortium ICASA (International Consortium for Agricultural Systems Applications) [WHITE et al., 2013] fournit des modèles et des standards de données pour l'intégration de données provenant d'expérimentations conduites dans des parcelles agricoles. Les normes proposées se présentent sous la forme d'une liste de variables de références hiérarchisées pour la description des pratiques de gestion, des traitements, des données pédologiques, des données météorologiques et des réponses des céréales

cultivées aux conditions expérimentales. L'objectif est de faciliter les usages autour de la collecte et de l'échange de données dans le cadre de la recherche agronomique, et in fine d'en faciliter l'analyse au travers de systèmes centrés sur de la simulation ou de l'aide à la décision.

FOODIE :
Le projet FOODIE (Farm-Oriented Open Data in Europe) [PALMA et al., 2016] s'appuie sur les directives INSPIRE (Infrastructure pour l'Information Spatiale au niveau Européen) qui portent à la fois sur les standard ISO/OGC visant à cadrer la représentation de l'information géographique, et des spécifications qui viennent s'y ajouter pour l'agriculture et l'aquaculture, pour en décliner une version en langage OWL à partir des diagrammes de classes UML. L'ontologie nommée FOODIE ainsi construite, a été produite par application des règles de conversion définies dans [ISO/TC et S. J. D. COX, 2014]. Si FOODIE a été exploitée pour des besoins en agriculture de précision, il n'en va pas de même pour l'agroforesterie, et certains des concepts définis dans FOODIE ne sont pas pensés pour la structuration d'aménagements mettant en jeu des arbres et des cultures intercalaires. L'exemple du concept "Plot" qui est défini comme étant une "zone agricole continue plantée d'une espèce végétale cultivée" est un exemple de cette difficulté à prendre en charge toute la complexité de l'agroforesterie, et en particulier des interactions pouvant être bénéfiques entre espèces cultivées. FOODIE reste toutefois un travail qui pose des bases très intéressantes pour tout ce qui concerne l'organisation spatiale des parcelles et des zones aménagées au sein de ces parcelles.

2.3.4 Dans les autres domaines scientifiques

L'utilisation des ontologies a pris un essor considérable en biologie moléculaire et dans le domaine biomédical [BARD et RHEE, 2004].

GENE ONTOLOGY :
Partant du principe que la grande majorité des gènes codant pour les protéines impliquées dans les fonctions biologiques essentielles des eucaryotes sont communs aux différentes espèces, "gene ontology" fournit un vocabulaire contrôlé afin que tous les biologistes (biochimistes, généticiens, biologistes cellulaires, biologistes en physiologie animale...) puissent partager leurs connaissances sur les gènes, les protéines qui en découlent et le rôle biologique de ces protéines au sein de différents organismes [ASHBURNER et al., 2000]. Dans ce cadre "gene ontology" se divise en trois catégories : les fonctions biologiques (auxquelles contribue les gènes et les protéines issus de ces gènes), les fonctions moléculaires (rôles des protéines au sein de l'activité cellulaire) et la composition cellulaire (place de la protéine active dans un élément cellulaire).

Les vocabulaires contrôlés gardent leur rôle prépondérant lorsqu'il s'agit d'organiser des connaissances pour faciliter la recherche d'information. Pour cela ils fournissent un lexique composé de termes reconnus et consensuels dans un domaine particulier imposant de fait leur utilisation. Leur évolution nous amène aujourd'hui aux

ontologies qui par leur formalisme accru sont plus aptes à assumer ce rôle au sein d'applications informatiques. De plus les ontologies permettent aussi une interopérabilité renforcée entre des systèmes hétérogènes [ROUSSEY et al., 2010] et donnent accès via des mécanismes d'inférences à la possibilité de création de nouvelles connaissances. Les systèmes de gestion de données basés sur l'utilisation des ontologies sont appelés bases de connaissance. Les avantages offerts par ces bases de connaissance sont exploités aujourd'hui dans de nombreux domaines. Dans certains de ces domaines, comme par exemple pour la recherche biomédicale des initiatives comme "The OBO foundry" (The Open Biological and Biomedical Ontology Foundry) rassemble plus d'une soixantaine d'ontologies [SMITH et al., 2007] montrant le grand intérêt des acteurs de ce domaine pour les ontologies. Cet engouement semble plus limité dans les domaines de la botanique et de l'agriculture et il n'existe pas aujourd'hui d'exemple d'initiative de création d'une base de connaissance en agroforesterie. Néanmoins nous utiliserons les expériences de mise en place d'ontologies et de bases de connaissance réalisées dans d'autres domaines pour la réalisation de notre objectif.

III

L'agroforesterie

3.1 Introduction 35
3.2 Définition 36
3.3 Les aménagements agroforestiers historiques en Europe 37
3.4 Problématique de l'agroforesterie en France et dans les pays de l'UE 38
3.5 Les services fournis par l'agroforesterie 39
3.5.1 Stabilisation des sols 39
3.5.2 Restauration des sols 40
3.5.3 Bien-être animal 41
3.5.4 Retour de la biodiversité 41
3.5.5 Stockage du carbone 42
3.5.6 Limitation de la pollution 42
3.6 Les aménagements agroforestiers 43

3.1 Introduction

De nos jours l'agriculture est devenue un véritable sujet de société. Face aux évidences des différents écueils de l'industrialisation de notre agriculture mis en lumière dans ces dernières décennies, comme la dangerosité de certains pesticides ou la détérioration des terres arables, les consommateurs réclament aujourd'hui une agriculture plus soucieuse de la santé publique, du respect de l'environnement et de l'équité sociale. Face à cette demande d'une agriculture durable et solidaire de multiples initiatives ont vu le jour. Nous revenons sur certaines de ces initiatives :

– L'agriculture biologique
Officiellement reconnue en 1981 par les pouvoirs publics français [*L'agriculture biologique (AB)* 2020], elle vise à préserver la qualité des sols, la qualité de l'eau et de l'air et la biodiversité en prônant des pratiques de culture et d'élevage soucieuses du respect des équilibres naturels. Pour atteindre ces objectifs, elle exclut l'usage de produits chimiques de synthèse, d'organismes génétiquement modifiés (ou OGM) et encourage la limitation des intrants.

- L'agriculture de conservation
  D'après la FAO [*Qu'est-ce que l'agriculture de conservation ?* 2020], « L'agriculture de conservation est un système cultural qui peut empêcher la perte de terres arables tout en régénérant les terres dégradées. ». Elle encourage, tout à la fois, une réduction drastique du travail du sol par des moyens mécaniques, une couverture permanente du sol et la diversification végétale incluant des espèces non productives mais d'intérêts agronomiques comme l'utilisation de couverts riches en azote.
- L'agroécologie
  Le ministère de l'agriculture et de l'environnement en donne la définition suivante : « L'agroécologie est une façon de concevoir des systèmes de production qui s'appuient sur les fonctionnalités offertes par les écosystèmes. » [*Qu'est-ce que l'agroécologie ?* 2020]. Elle a donc pour but d'utiliser au maximum les services offerts par la nature pour les besoins de l'agriculture tout en maintenant sa capacité de régénération.
- La permaculture
  Il s'agit d'un mouvement écologique pour la création de systèmes agricoles stables, initié par Bill Mollison et David Holmgren dans les années 70. Le mouvement milite en faveur de l'écologie et de l'utilisation du savoir faire des sociétés traditionnelles pour la création de systèmes agricoles reproduisant la diversité, la stabilité et la résilience des écosystèmes naturels.

Toutes ces nouvelles agricultures ont pour enjeux la protection de l'environnement, le bien-être animal, une alimentation saine, l'autonomie alimentaire et le travail équitable pour tous les producteurs. L'agroforesterie tout en partageant ces valeurs, se distingue, en se focalisant sur le rôle naturel des arbres pour mener à bien ces défis.
Les sections suivantes ont pour but de décrire l'agroforesterie, ses principes, son histoire, et ses apports en matière de service. Un exemple d'un des multiples aménagements associés à cette pratique, est également présenté. Cet aménagement servira ensuite pour étayer le cas d'étude privilégié dans cette thèse.

3.2 Définition

L'agroforesterie consiste en la réintroduction des arbres dans les métiers de l'agriculture. Plusieurs définitions existent pour l'agroforesterie. Le centre agroforestier mondial (ICRAF) [1] en donne cette définition : « L'agroforesterie est une dénomination collective pour les systèmes et technologies d'utilisation des terres où les plantes vivaces ligneuses (arbres, arbustes, palmiers, bambous, etc.) sont délibérément utilisées dans les mêmes unités de gestion des terres que les cultures agricoles et/ou les animaux, dans une certaine forme d'arrangement spatial ou de séquence temporelle. Dans les systèmes agroforestiers, il existe des interactions à la fois écologiques et économiques entre les différentes composantes. » [LUNDGREN et RAINTREE, 1982], mais il est aussi proposé une définition plus sociale de l'agroforesterie [LEAKEY, 2017]. Dans ce document nous nous limiterons à la définition suivant : « L'agroforesterie est une technique agricole associant des arbres à des cultures ou des animaux ».

1. ICRAF : World Agroforestry. https://www.worldagroforestry.org/.

3.3 Les aménagements agroforestiers historiques en Europe

L'agroforesterie est une pratique ancienne en Europe. De nombreux exemples attestent qu'au fil des siècles,une même unité d'exploitation a combiné avec succès, présence d'arbres et pratique de l'agriculture et/ou de l'élevage. Nous présentons ici quelques exemples parmi les plus notoires :

La dehesa est une structure d'agro-sylvo-pastoralisme dont l'origine remonte au moyen âge et qui est toujours présente notamment en Espagne et au Portugal où elle porte le nom de "montado". Héritée des domaines seigneuriaux de l'époque carolingienne, il s'agit maintenant principalement de domaines municipaux, sur lesquels les habitants d'une commune peuvent faire paître gratuitement leur bétail et profiter de produits forestiers non ligneux. Basé sur le biome « forêts, terres boisées et broussailles méditerranéennes », on y retrouve des arbres clairsemés, le plus souvent des chênes, et des plantes herbacées. Les broussailles sont éliminées par le pâturage des bovins, ovins et porcins et par une mise en culture fourragère occasionnelle d'avoine ou de millet, maintenant ainsi la densité des arbres de 30 à 100 arbres par hectare. Les chênes (*Quercus ilex* ou *Quercus suber*) fournissent des glands pour l'alimentation des porcins (garant du label "pata negra" du fameux jambon ibérique), du bois raméal pour le feu et un fourrage d'appoint en période de sécheresse. Cet environnement a conservé sa biodiversité permettant d'exploiter des produits de sous bois comme les champignons ou le gibier. Avec une superficie estimée à environ quatre millions d'hectares [IBAÑEZ et al., 1981] pour la pointe ibérique, la dehesa représente le plus important système agroforestier d'Europe.

Les huertas se rencontrent sur les plaines du pourtour méditerranéen. Il s'agit à l'origine de petites unités d'agriculture vivrière. Elles mélangent des rangées d'arbres fruitiers sous lesquels sont pratiquées agriculture maraîchère et/ou floriculture. Ce sont des exploitations intensives à l'irrigation contrôlée répondant à des conditions de culture difficiles (manque d'eau et de terres arables). Introduite lors de la domination musulmane du VIIe au XIVe siècle de la péninsule ibérique et du sud de la France, cette pratique placera les régions du sud au cœur de la production de fruits, de légumes et de fleurs.
En France, d'autres aménagements agroforestiers ont façonné notre paysage au cours de ces derniers siècles et sont encore toujours présents souvent à l'échelle régionale :

Les prés-vergers Ils mêlent principalement l'exploitation de pommiers ou de poiriers hautes tiges avec le pâturage de bovins. Couvrant 600 000 hectares au début XXe siècle ils assurent une production de cidre, d'alcool de poire, ou d'eau de vie. Cet alcool sera aussi largement utilisé pour la fabrication d'explosifs et de carburant durant les deux guerres mondiales. Une densité de 50-100 arbres par hectare apporte un ombrage pour le bien être des animaux et prolonge l'exploitation de la prairie de deux à quatre semaines en période de sécheresse. Dans cet écosystème stable, la prairie est entretenue par le pâturage des animaux, ce qui participe à la préparation du sol pour la récolte des fruits. Les déjections du bétail permettent le recyclage de l'azote consommé par les cultures et la prairie contribue à la filtration des nitrates issus de ces mêmes excréments afin de limiter la pollution des aquifères. Aujourd'hui il subsiste plus de 100 000

hectares de prés-vergers principalement en Bretagne, Normandie et Pays de la Loire, ce qui en fait le plus important système agroforestier de France.

Les noyers du Dauphiné Selon les écrits de Pline l'Ancien (premier siècle aprés J.-C.), la pratique de cultures entre des noyers est très ancienne dans le Dauphiné et perdure jusqu'à aujourd'hui. Les cultures intercalaires sont devenues des cultures relais qui permettent, pendant une dizaine d'années, d'assurer un revenu aux nuciculteurs en attente de la capacité de fructification des noyers. En fin d'exploitation les vieux arbres sont vendus, apportant ainsi un revenu supplémentaire à l'exploitant. Toujours en France, d'autres aménagements traditionnels ont également révélé leur efficacité, au point d'être spécifiquement distingués :

- Les hautains
  Décrite au début de notre ère, cette technique est largement répandue au moyen âge. Elle consiste à utiliser des arbres comme tuteurs de la vigne. Les arbres sont plantés de façon régulière, avec à leur pied des plants de vignes dont ils supportent les rameaux. Les espèces utilisées sont les érabliers, les noyers, les ormes ou les chênes
- Les joualles
  Des arbres, comme le pêcher, le cerisier, le noyer, l'amandier ou le chêne truffier, sont insérés parmi les rangées de vignes. Pratiquée avant le XVIIIe siécle, cette technique permettait aux agriculteurs de passer progressivement de la culture de la vigne à une arboriculture fruitière sur une parcelle.
- Les oullières
  Elles furent pratiquées au XVIIIe siècle jusqu'au début de la mécanisation de l'agriculture. Entre des rangées de vignes largement espacées étaient cultivées des céréales ou des plantes sarclées comme la betterave, le maïs ou la pomme de terre. Des lignes d'arbres fruitiers pouvaient y être également introduites.
- Les plantades
  Constituées de rangées de chênes largement espacées, elles permettaient le pâturage. Comme pour la dehesa, sa gestion communale lui apportait une dimension sociale. La pratique de la glandée, par l'éleveur muni d'une perche qui faisait chuter les glands du chêne, permettait de nourrir un élevage porcin.

3.4 Problématique de l'agroforesterie en France et dans les pays de l'UE

Après la fin de la deuxième guerre mondiale, commence pour les pays développés la période des « Trente Glorieuses » qui allie une forte croissance économique et une explosion démographique. Le plan Marshall consistant en des prêts américains accordés aux états de l'Europe, en échange d'importation de produits américains, va accélérer la mécanisation de l'agriculture Européenne.
En France, on assiste à une mécanisation et une intensification de l'agriculture sans précédent. Cette évolution rapide aura de multiples conséquences qui modifieront notre paysage agricole. En effet avec la mécanisation, l'arbre devient un obstacle, car il gêne les déplacements de machines agricoles de plus en plus imposantes. A cela, se rajoute des actions politiques comme "le remembrement" au cours des années soixante et quatre-vingt, qui conduit à la réunification de parcelles agricoles pour une gestion plus efficace des exploitations. La politique de remembrement s'accompagne de la suppression de nombreuses haies et talus bordant les petites parcelles, et nuit en particulier à la biodiversité.

Au fur et à mesure des années, cette conversion massive à une agriculture intensive et les modification paysagères qui en découlent, entraînent une véritable perte du savoir des techniques agroforestières. Certains pays ont réalisé depuis les années soixante-dix des expérimentations agroforestières de grande envergure comme la Chine [DUPRAZ, 1994] et la Nouvelle Zélande [DUPRAZ, AUCLAIR et al., 1992], accompagnées d'études scientifiques mais les conditions climatiques et les espèces utilisées sont différentes de celles rencontrées en Europe réduisant ainsi la possibilité d'un transfert de ces techniques éprouvées et validées.

3.5 Les services fournis par l'agroforesterie

L'agroforesterie ne se résume pas en une simple plantation d'arbres, mais il s'agit d'une association étroite entre les arbres et l'agriculture afin d'obtenir une réelle synergie, bénéfique pour les deux parties [DUPRAZ et LIAGRE, 2008]. Cette synergie permet, entre autres choses, l'amélioration des conditions d'élevage des animaux, le retour de la biodiversité, ainsi que dans une perspective de développement durable dans les pratiques agricoles, la stabilisation et la restauration des sols et la limitation des intrants (eau, engrais et pesticides). Les sections suivantes fournissent une liste non exhaustive des services écosystémiques pouvant résulter de la pratique de l'agroforesterie (voir figure 3.1).

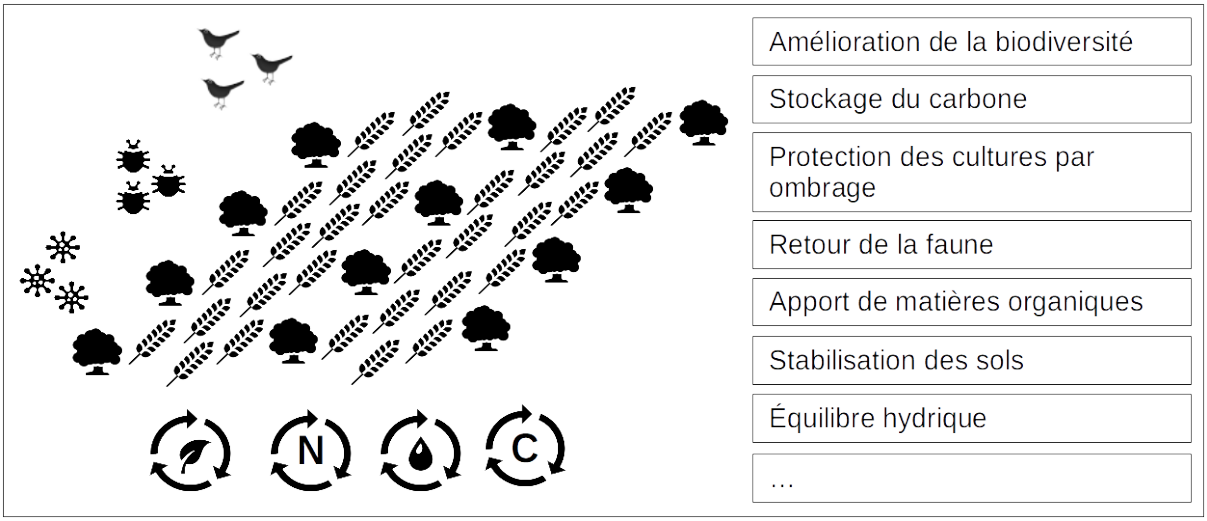


FIGURE 3.1 – *Quelques uns des services écosystémiques apportés par une parcelle agroforestière.*

3.5.1 Stabilisation des sols

Les cultures pratiquées sur de grandes surfaces augmentent le risque d'érosion qui à terme mène à une perte difficilement remédiable des terres arables. Cette perte a été estimée à 12% de la superficie de L'Europe et 10% pour celle de la France entre 1991 et 2005 [ROBERT et CHEVERRY, 2009]. L'érosion des terres arables peut être provoquée par deux mécanismes distincts :

– L'érosion éolienne
  Dans ce cas, la présence des arbres surtout lorsqu'ils sont disposés en ligne de façon

perpendiculaire aux vents dominants, réduit la vitesse des vents. Cette réduction de la vitesse des vents par la présence des houppiers diminue la dispersion des nuages de poussière générés par le vent et les travaux agricoles. On notera par ailleurs que l'exposition de ces arbres au vent est amplifiée par rapport à des conditions de sylviculture classique ce qui entraîne une augmentation de leur résistance aux vents tempétueux.

– L'érosion hydrique
  En ce qui concerne l'érosion due aux écoulements pluviaux, ce sont les bandes enherbées disposées le long des lignes d'arbres qui jouent un rôle essentiel pour la protection des sols. En effet il à été constaté que la présence de ces bandes enherbées entraîne la formation spontanée de talus vis à vis du niveau du sol des cultures intercalaires. Ces talus forment une barrière à l'écoulement des eaux et peuvent dans le cas d'un terrain en pente entraîner un terrassement passif des surfaces cultivables. Au niveau de ces talus, une partie de l'eau est absorbée par les racines des herbacées et son infiltration dans le sol est facilitée par l'augmentation de la porosité du sol due aux racines des arbres.

3.5.2 Restauration des sols

Les arbres ne se contentent pas de stabiliser les sols, ils en assurent le maintien physico-chimique et biologique, ainsi que la restauration. Cette capacité à améliorer la qualité des sol est un des avantages majeurs attribués à l'agroforesterie dans le cadre d'une agriculture durable [YOUNG, 1989], [NAIR, 2011]. Ces fonctions de maintien et de restauration du sol sont communes à tous les arbres que ce soit dans des conditions de sylviculture classique ou dans des conditions de culture agroforestière, mais elles se trouvent amplifiées dans le cas de l'agroforesterie par un enracinement plus profond des arbres [MULIA et DUPRAZ, 2006; CARDINAEL et al., 2015]. Le schéma figure 3.2 montre la distribution racinaire des arbres agroforestiers modifiée horizontalement et verticalement pour échapper à la concurrence induite par les cultures dans les premiers centimètres du sous-sol. Voici quelques-unes de ces fonctions:

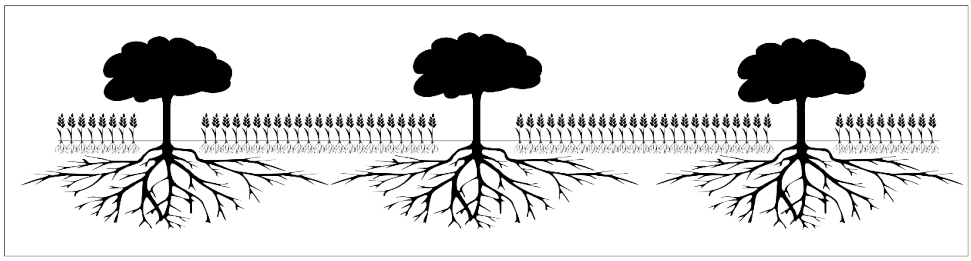


FIGURE 3.2 – *Distribution racinaire des arbres dans des conditions d'agroforesterie. Les arbres s'enracinent plus profondément pour échapper à la concurrence des cultures.*

– Enrichissement en matière organique des terres cultivables.
  La matière organique fournit l'énergie nécessaire aux organismes du sol. De sa présence et de sa qualité, dépendront la biodiversité et les fonctions biologiques associées du sol. Les arbres fournissent de la matière organique en surface et en profondeur via leurs débris végétaux provenant de la chute de feuilles et de branches mortes ainsi que de la décomposition racinaire. Un des indicateurs important pour juger de la bonne santé des sols est le carbone organique du sol (SOC) dont le taux est fonction de la vitesse de décomposition et de remplacement de la matière organique. En réalisant une méta-analyse sur les données de 53 études publiées, Andrea De Stefano et Michael G. Jacobson constatent que la transformation de pâtu-

rages, de parcelles non cultivées ou de parcelles de monoculture en parcelles agroforestières s'accompagne d'une augmentation du stockage du SOC [STEFANO et JACOBSON, 2018].

- Mise à disposition des nutriments pour les cultures.
  On pourrait penser que les arbres et les cultures soient uniquement en concurrence vis à vis des nutriments essentiels présents dans le sol, mais des phénomènes de facilitation peuvent aussi apparaître au sein des aménagements agroforestiers [JOSE et al., 2004]. L'agroforesterie peut jouer sur la disponibilité des nutriments du sol en jouant sur le mélange des cultures et des arbres de façon à minimiser la concurrence. Ainsi la mise en place de cultures d'hiver, lors de la période d'hivernage des arbres limite la concurrence. Mais l'arbre peut aussi jouer un rôle actif grâce à son enracinement profond qui lui permet de remonter des cations provenant de l'altération de la roche mère [ROWE et al., 1999 ; CADISCH et al., 2004].
- Amélioration du microbiote du sol.
  Le sol ne se résume pas à une simple agglomération de matière organique et de particules minérales mais se compose aussi d'une multitude d'organismes diverses contribuant à de nombreux services écosystémiques garants de la santé et de la fertilité du sol [BARRIOS, 2007]. Des études montrent l'influence positive de l'agroforesterie sur certains de ces micro-organismes [ZHANG et al., 2018 ; POSADA et al., 2018].

3.5.3 Bien-être animal

Les périodes de fortes chaleurs et de fort rayonnement solaire peuvent être préjudiciable pour la santé des animaux d'élevage comme le montre cette étude sur les ruminant domestiques [SILANIKOVE, 2000]. La présence d'arbres sur un lieu de pâturage permet de fournir abri et ombrage aux animaux d'élevage afin de réduire leurs stress en cas d'épisodes météorologiques extrêmes. L'ombrage serait même un besoin pour les ovins et participerait au bien-être de ceux-ci même en dehors des périodes estivales [GINANE et al., 2018] En plus de fournir un microclimat bénéfique, les arbres d'ombrage peuvent en même temps fournir une diversification des ressources fourragères comme l'atteste cette étude chez les bovins [KARKI et GOODMAN, 2010]. Cette relation arbre-animal peut être bilatérale, comme on le voit dans cette autre étude ou la présence d'oies, et par voie de conséquence de déjections, augmente les rendements de production des noyers qui leur servent d'abri [DUBOIS et al., 2008]. D'autre effets peuvent être induits comme, par exemple, l'augmentation de l'utilisation des parcours extérieurs chez le poulet élevé en semi-liberté [FANATICO et al., 2016 ; GERMAIN, 2014]. Parmi les bénéfices potentiels, que peut leur apporter l'agroforesterie, le bien-être des animaux est le plus plébiscité par les éleveurs [BÉRAL et al., 2018].

3.5.4 Retour de la biodiversité

Devant l'effondrement de la biodiversité mondiale, des décisions avec effet à court terme doivent être rapidement engagées [RANDS et al., 2010]. Une des causes de la diminution de cette biodiversité est pour de nombreux organismes la perte de leurs habitats sous la pression anthropique. L'agroforesterie peut aider à rétablir cette biodiversité en fournissant de nouveaux habitats au sein de nos surfaces agricoles.
Des études démontrent cette augmentation de biodiversité à l'actif de l'agroforesterie : Ainsi Paul J. Burgess le constate pour des invertébrés, des arthropodes aériens et des petits mammifères, dans les exploitations agroforestières du Royaume Uni [Paul J.

BURGESS, 1999]. Ranjith Udawatta et ses collaborateurs mettent eux en évidence, sur la base de nombreux documents sur la biodiversité des aménagements agroforestiers publiés sur une trentaine d'années, l'augmentation de la diversité des espèces florales, fauniques et de la diversité microbienne du sol de façon significative sur des parcelles agroforestières par rapport à des parcelles de monoculture, des parcelles de cultures adjacentes, et certaines forêts. Cette augmentation de la biodiversité se retrouve amplifiée à proximité des arbres [P. UDAWATTA et al., 2019]. En plus de lutter contre le déclin de la biodiversité, cette réintroduction amène au sein même de l'exploitation agricole des animaux pouvant jouer un rôle d'ennemis naturels dans la lutte contre les parasites des cultures [DIX et al., 1995 ; STAMPS et LINIT, 1997]. Des effets positifs sur la pollinisation ont aussi été mis en évidence [VARAH et al., 2020].

3.5.5 Stockage du carbone

De part les conditions favorables de leur environnement comme le fait de ne pas se concurrencer vis à vis de l'accès à la lumière et aux nutriments du sol, de profiter du travail du sol et parfois de la fertilisation des cultures, les arbres agroforestiers produisent au même âge jusqu'à trois fois plus de biomasse que leurs équivalents forestiers [GAVALAND et BURNEL, 2005]. Cette croissance accélérée place l'agroforesterie comme un outil idéal de séquestration du carbone pour l'agriculture [HAMON et al., 2009]. Une estimation du potentiel de séquestration du carbone par des pratiques agroforestière à l'échelle européenne serait d'environ 1500 millions de tonnes équivalent CO^2 par an, ce qui correspondrait à 37% des émissions de l'union en 2007 [AERTSENS et al., 2013] Le Groupe d'Experts Intergouvernemental sur l'Évolution du Climat (GIEC) a d'ailleurs préconisé la pratique de l'agroforesterie comme une des solutions pour lutter contre le réchauffement climatique [BILAN, 2007].

3.5.6 Limitation de la pollution

Les précipitations induisent un effet de lixiviation des sols ou des éléments solubles comme les nitrates sont entraînés vers le sol profond. Ce phénomène est impliqué dans la pollution des nappes phréatiques, dès lors que l'excédent des traitements azotés destinés aux cultures et non interceptés par celles ci, passe en dessous du sol de culture. Le filet racinaire des arbres agroforestiers placé sous le sol de culture (voir figure 3.2) intercepte ces éléments solubles limitant ainsi la pollution aux nitrates [ROWE et al., 1999 ; BERGERON et al., 2011 ; TULLY et al., 2012].

Nous avons donc entrevu quelques uns des multiples services que peut offrir la pratique de l'agroforesterie. Certains de ces services affecteront l'environnement local de l'exploitation comme c'est le cas pour la protection des sols avec des effets attendus contre l'érosion ou quand il s'agit du bien être des animaux exploités ou tout simplement du paysage. D'autres effets peuvent être considérés à une échelle plus globale comme le maintien de la biodiversité, la lutte contre le changement climatique et la protection des nappes phréatiques. L'étude de ces services à des échelles d'influences différentes doit aussi être prise en compte sur des échelles de temps variées allant d'une saison au cycle de la vie d'un arbre. Ces services sont donc le résultat

d'interactions complexes entre les différents éléments biotiques et abiotiques de l'aménagement agroforestier qu'il faut étudier sur des échelles spatiales et temporelles très diversifiées.

3.6 Les aménagements agroforestiers

L'aménagement est un élément capital pour la réussite d'une parcelle agroforestière il ne suffit pas de tenir compte des conditions météorologiques et pédologiques dans le choix des espèces d'arbres à planter.
C'est un système dynamique en perpétuelle évolution du fait, d'une part des saisonnalités intriquées des arbres, des cultures, et de la présence d'animaux d'élevage et d'autre part de la croissance continue des arbres jusqu'à leur abattage.
Il faut tenir compte des contraintes techniques et économiques de l'exploitation agroforestière sur toute la durée de la vie de l'arbre pour obtenir une bonne estimation de sa rentabilité.
Les aménagements agroforestiers les plus pratiqués sont l'agrisylviculture (arbres destinés à la production de bois avec cultures intercalaires (céréales, maïs, orge...)) et le sylvopastoralisme (association sur un même espace d'activités sylvicoles et pastorales), mais d'autres pratiques agricoles peuvent également être associées comme la viticulture, l'apiculture ou l'aquaculture.

Un exemple d'aménagement agroforestier moderne dans le cadre de l'agrisylviculture

Dans nos travaux nous focaliserons notre attention sur un type d'aménagement particulier de l'agrisylviculture que sont les cultures en allées. Il s'agit d'un aménagement intraparcellaire avec des alignements d'arbres réguliers et discontinus. Sur cette photographie (voir figure 3.3), on aperçoit des bandes de cultures de blé placées entre des lignes parallèles de noyers. Ce système flexible pratiqué dans les pays d'Europe tempérée de façon majoritaire [TORQUEBIAU, 2000] permet une production durable et offre de multiples services écologiques [QUINKENSTEIN et al., 2009].
La conversion d'une parcelle agricole de monoculture classique en ce type d'aménagement est soumise à de multiples paramètres relevant des conditions initiales du terrain et des choix de l'exploitant. Voici en résumé les différent paramètres qu'il est important de prendre en compte:

- Le choix des arbres.
  Le choix des espèces d'arbres à utiliser est bien sûr capital pour s'assurer de la réussite de l'aménagement sur le long terme. Les principaux critères à prendre en compte sont:
  - Les objectifs de production.
    Les attentes de l'exploitant peuvent être variées comme la production de bois d'ornement, la production de fruits ou les deux. D'autres visées peuvent être considérées comme par exemple la mise à disposition de fleurs pour l'apiculture ou la fourniture d'un complément alimentaire pour le bétail. Des mélanges d'espèces peuvent être envisagés pour une diversité de production ou une sécurisation de l'investissement.
  - Les conditions météorologiques.
    Le climat est bien entendu une donnée essentielle pour le choix de l'arbre qui doit être adapté aux conditions de température et de pluviométrie moyennes



FIGURE 3.3 – *Parcelle agroforestière. Site des Eduts (André Gavaland – INRA Toulouse).*

ainsi qu'aux aléas climatiques de la parcelle comme des risques accrus de sécheresse, d'inondation ou de gelées hors saison.
  - Les conditions pédologiques et géologiques.
    Certaines espèces préféreront des sols acides, d'autres des sols calcaire ou argileux et certains arbres ne résisteront pas à un engorgement d'eau d'un sous-sol même temporaire.
  - Les conditions topologiques.
    Certaines espèces sont plus adaptées à des conditions de forte altitude, ou de maintien dans de fortes pentes.
- Le positionnement des arbres.
  Pour une même parcelle, plusieurs aménagements spatiaux peuvent être réalisés et peuvent être influencés par le simple choix esthétique de l'exploitant. Néanmoins des aspects pratiques sont à prendre en compte afin de réduire les risques d'accidents, de minimiser le travail de l'exploitant et de maximiser les rendements. Voici les principales lignes directrices:
  - Les lignes d'arbres doivent être suffisamment espacées pour permettre un passage facilité des engins agricoles. La largeur des tournières en fin de ligne qui permettent de retourner les engins doit être adaptée pour la même raison.
  - Les lignes sont de préférence placées dans une orientation Nord Sud afin de minimiser la compétition pour la lumière entre arbres et culture.
  - Dans les zones fortement exposées aux vent, l'orientation perpendiculaire au vent dominant des lignes d'arbres peut limiter les impacts de celui-ci.
- Le choix des cultures.
  Là aussi les conditions climatiques et pédologiques devront être prises en compte, et il ne serait pas raisonnable d'aller contre nature, dans le choix de la ou des cultures associées. Par contre des associations peuvent être faites pour créer ou renforcer des phénomènes de complémentarité tant au niveau des cultures que de celui des arbres. Pour exemple la mise en place d'une culture d'hiver renforcera par concurrence le phénomène de filet racinaire au niveau des arbres dont la croissance

des racines se fera plus tard. Un autre exemple est de ne pas choisir des arbres et des cultures ayant des prédateurs communs.

Le schéma suivant représente une parcelle agroforestière idéale (voir figure 3.4). Les lignes d'arbres sont parallèles et orientées sur un axe nord-sud, uniformisant ainsi l'ensoleillement des cultures pour une maturation uniforme avant récolte. Les lignes d'arbres sont perpendiculaires au vent dominant et au sens d'une crue potentielle pour limiter l'érosion éolienne et les dégâts qui seraient causés par la crue d'un cours d'eau limitrophe. Dans le rectangle formé par la parcelle, les tournières sont placées sur la largeur du terrain, ce qui fait qu'elles occupent moins de place, augmentant ainsi la densité relative des arbres sur la parcelle et facilitant le travail de l'exploitant en limitant ces manœuvres de retournement.

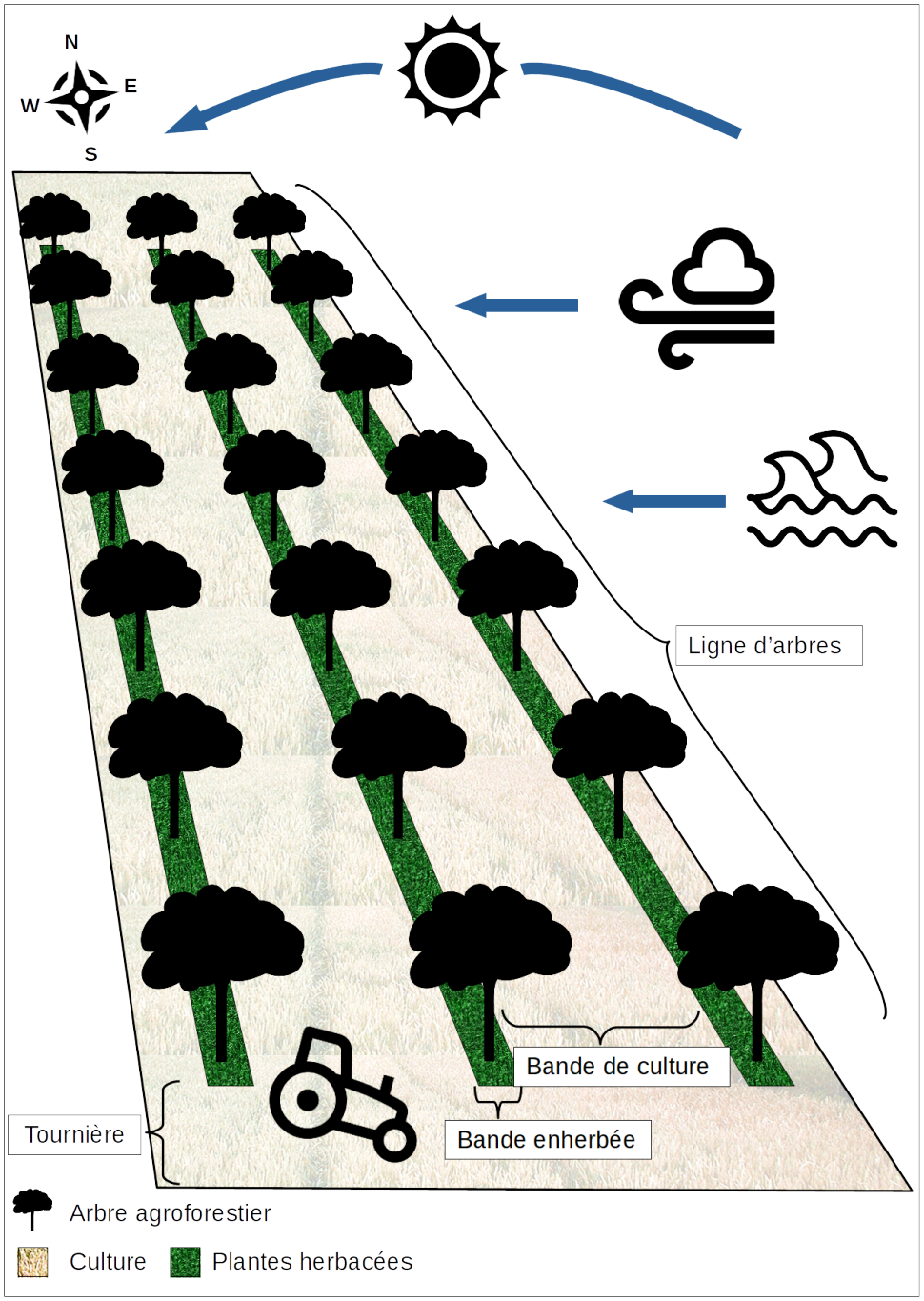


FIGURE 3.4 – *Schéma d'une parcelle agroforestière avec des cultures en bande.*

En Résumé

Les services validés ou pressentis de l'agroforesterie sont nombreux et peuvent être exploités pour nous aider à obtenir des solutions durables face à nos problématiques actuelles et futures en terme de production agricole, de gestions des ressources et d'environnement. Mais valider scientifiquement les bienfaits de la pratique de l'agroforesterie n'est pas une chose simple. Pour vérifier les effets durables de cette technique agricole, ces performances doivent être étudiées d'un point de vue agronomique, économique et social mais aussi environnemental et sanitaire. Au niveau des aménagements, la multitude des interactions entre les éléments biotiques et abiotiques sont à prendre en compte sur des échelles spatiales et temporelles très diverses.
Dans le cadre de ce travail de thèse, nous recherchons une solution afin de faciliter les études sur l'agroforesterie dans une approche systémique afin de vérifier et de valider les bienfaits durables de l'ensemble des pratiques agroforestières.
Pour atteindre notre objectif, nous avons choisi de nous orienter vers une démarche ontologique pour la gestion des bases de données en agroforesterie. Le but de cette démarche est de mettre en place un système pour la gestion des connaissances dans le domaine de l'agroforesterie souple et pérenne, mobilisable via des outils informatiques, pouvant de plus s'interfacer avec les connaissances d'autres domaines concernés par l'approche systémique de l'agroforesterie. Un tel système porte le nom de "base de connaissance" et sa constitution est intimement lié à la notion d'ontologie. Nous commencerons donc par décrire ce qu'est une ontologie dans le chapitre suivant.

IV

Les ontologies

4.1 De la philosophie à l'informatique 48
4.2 Définition de l'ontologie 49
4.3 Contenu des ontologies 50
4.3.1 Les concepts 50
4.3.2 Les relations 52
4.3.3 Les contraintes 52
4.3.4 La hiérarchie des concepts 52
4.3.5 Les différentes informations exprimées par les ontologies 53
4.4 Les différents types d'ontologies 53
4.4.1 Les ontologies terminologiques 54
4.4.2 Les ontologies cadres 54
4.5 Rôle des ontologies 55
4.5.1 Représentation et uniformisation de la connaissance d'un domaine 55
4.5.2 Réutilisation et échange de la connaissance 55
4.6 En résumé 55

L'ontologie, est un terme philosophique provenant de la pensée Aristotélicienne, qui signifie "théorie de l'être". Ce terme a été repris en informatique avec les sens suivants : « ensemble structuré des termes et concepts représentant le sens d'un champ d'informations » ou encore « spécifications formelles explicites de termes d'un domaine et de relations entre elles » [Tom R. GRUBER, 1993]. Une ontologie définit donc les termes employés pour décrire et représenter un domaine de connaissance. Ontologie, thésaurus, et taxonomie ont en commun, de fournir un vocabulaire contrôlé, où termes et concepts sont organisés de façon hiérarchique, mais leur utilisation et leur vocation divergent dans le monde de la représentation des connaissances. En effet, si les taxonomies sont utilisées pour classifier, et les thésaurus pour indexer et rechercher, les ontologies apportent des enrichissements dans la description des connaissances qui permettent de raisonner sur les différentes entités d'un domaine. L'essor du web sémantique a entraîné, en parallèle, l'apparition d'agents logiciels qui tirent parti d'ontologies, afin de faciliter les accès à des bases de connaissance généralistes ou de spécialité. Les exemples les plus connus, sont en particulier Freebase, et plus

récemment Knowledge Graph de Google, ainsi que Wikidata de Wikimedia. Le référentiel Schema.org va, en sa qualité de composant ontologique, fournir les types des entités décrites dans Google Knowledge Graph. Il existe plusieurs types d'ontologies parmi lesquelles, les ontologies terminologiques et les ontologies cadres. Les ontologies terminologiques se concentrent sur les concepts d'un domaine en fournissant les définitions et les termes qui y sont associés. Dans ce cas, les relations entre ces concepts sont majoritairement de type généralisation/spécialisation. Les ontologies cadres sont orientées vers la modélisation de la connaissance d'un domaine en exprimant de façon plus poussée les relations entre concepts et les caractéristiques des individus rattachés à ces concepts.

4.1 De la philosophie à l'informatique

L'étymologie du mot ontologie nous ramène au mot latin *ontologia*, dérivé du grec ancien *ontos*, signifiant « ce qui est », et de *logia* ou *logos*, signifiant « discours, traité », qui fait référence au nom des disciplines. La signification de ce mot serait donc l'étude de l'être. En philosophie, l'Ontologie (avec un « O » majuscule) se rapporte donc à la question de la nature de l'être, question première des penseurs grecs tel que Platon. Aristote, avec l'introduction de la notion de métaphysique, propose, en réponse à cette question, d'étudier les modalités et propriétés de l'être: « Il y a une science qui étudie l'être en tant qu'être, et les attributs qui lui appartiennent essentiellement ». Il illustrera sa démarche, en classant les animaux terrestres et marins, selon leurs attributs communs identifiés par leur fonction première, plutôt que par l'apparence prise au sein de chaque individu.
Il est alors naturel de se demander pourquoi le terme d'ontologie, est repris dans le milieu de l'informatique. A la création du concept d'intelligence artificielle (IA) dans les années 50, l'IA symbolique est l'approche dominante. Il y est question de modéliser les raisonnements humains et de les coder sous forme d'algorithme. Cette approche peut être qualifiée de philosophique dans le sens ou elle implique une démarche de questionnement et de réflexion critique sur la connaissance et nous mêmes. l'IA symbolique repose sur des représentations formelles de la connaissance, accessibles à l'humain comme à la machine. Les symboles de ces représentations peuvent être stockés et manipulés, et l'application de règles de manipulation définies sur ces symboles peut induire lors de leurs interprétations des raisonnements.
Dans le même temps, l'informatique se développe, avec en particulier, le paradigme de la programmation orientée objet (POO). En POO, les informaticiens, avant d'implémenter une application, conçoivent des schémas conceptuels de classes (avec des outils comme par exemple la notation UML). Pour cela, ils réfléchissent aux objets qui vont être manipulés au sein de l'application, aux regroupements de ces objets en classes, aux caractéristiques communes de ces objets au sein d'une classe et aux relations entre ces objets. En résumé ils se questionnent sur la définition existentielle des classes d'objets qu'ils utilisent pour concevoir leur application.
En fait, cette façon de de classifier les choses et leurs caractéristiques, est commune à de nombreux systèmes de représentation des connaissances comme les langages dits cadres, les graphes conceptuels, les logiques de description, ou encore la POO. L'objectif premier vise à définir précisément les concepts que nous manipulons, et se rapproche de la notion philosophique d'Ontologie qui désigne l'étude des propriétés générales de ce qui existe.
Ce parallèle expliquerait l'adoption de ce terme philosophique en informatique, et symboliserait, en quelque sorte, notre rêve de pouvoir dépasser l'intelligence humaine avec des machines.

4.2 Définition de l'ontologie

Bien que le terme ait été adopté par les différentes communautés de l'ingénierie des connaissances, sa définition dans ce domaine n'a toujours pas fait consensus.
La définition la plus répandue, est celle de Thomas R. Gruber en 1993: « une ontologie est une spécification explicite d'une conceptualisation » [Tom R. GRUBER, 1993]. Cette définition rend compte de deux aspects essentiels d'une ontologie. Le premier est explicité par le mot "conceptualisation" que Gruber redéfinit comme étant « une vue abstraite et simplifiée du monde que l'ontologie veut représenter dans un but quelconque ». L'ontologie capture donc une vision simplifiée du monde qui n'existe donc que dans notre esprit et que nous ne créons que dans le but de nous représenter, manipuler ou simuler la réalité. Cette vision nous la réalisons avec l'aide de concepts, qui nous permettent de rassembler avec un même terme les caractères communs à plusieurs individus comme le concept d'arbre sous lequel nous regroupons les individus pommiers, chênes ou sapins, et de relations entre ces concepts. Le deuxième aspect est explicité par les mots "spécification explicite" qui traduise le fait que cette conceptualisation doit être faite de façon précise et complète afin que l'ensemble des concepts et des relations entre ces concepts puissent être exprimée sans ambiguïté. Le but à terme étant d'établir les bases d'une rigueur et d'un formalisme prémices du raisonnement.
Mais cette définition est contestée par Guarino, qui argumente que le terme de conceptualisation n'est pas adapté aux ontologies utilisées dans le cadre d'applications réelles en ingénierie des connaissance, ou il s'agit plutôt d'adaptation technique d'ontologie. « Par conséquent, une ontologie ne peut "spécifier" une conceptualisation que d'une manière très indirecte, puisque i) elle ne peut qu'approcher un ensemble de modèles prévus ; ii) un tel ensemble de modèles prévus n'est qu'une faible caractérisation d'une conceptualisation » [GUARINO, 1998].
Dans le cadre de la construction d'ontologie pour l'ingénierie, Willem Nico Borst propose d'ailleurs cette définition ou la notion de "conceptualisation" est remplacée par la notion de "conceptualisation partagée": « Une ontologie est une spécification formelle d'une conceptualisation partagée » [BORST, 1997]. Il met ainsi l'accent sur le fait qu'une ontologie ne peut être réutilisée qu'à la condition, que la conceptualisation véhiculée soit acceptée de tous. De plus l'adjectif "formelle" renforce l'idée que cette conceptualisation soit accessible aux machines. D'autres propositions suivront mais ne feront pas plus l'unanimité. Si les communautés de l'ingénierie de connaissances s'accordent sur la nécessité d'une formalisation de la connaissance et du partage de celle-ci, les moyens d'y parvenir ne sont pas consensuels.
Dans le cadre de l'ingénierie des connaissances les ontologies sont utilisées dans le but final de spécifier un artefact informatique, dans ce contexte l'ontologie, qui est effectivement une conceptualisation dans le sens ou on y définit des concepts, doit aussi être une théorie logique afin d'être codé en langage informatique. « La théorie logique représentant formellement l'ontologie est l'engagement ontologique de cette dernière. »[CHARLET, 2004].
Pour Guarino la notion d'engagement ontologique se doit d'être formalisée: « formaliser l'engagement ontologique d'un langage logique signifie offrir un moyen de spécifier le sens voulu de son vocabulaire en contraignant l'ensemble de ses modèles, en donnant des informations explicites sur la nature voulue des primitives de modélisation utilisées et leurs relations a priori. »[GUARINO et al., 1994].
Bruno Bachimont résume la construction d'une ontologie en trois étapes correspondant à trois engagements: « un engagement sémantique, fixant le sens linguistique des concepts, un engagement ontologique fixant leur sens formel et enfin un engagement computationnel déterminant leur exploitation effective. »[BACHIMONT, 2000]. Je terminerais avec cette définition donnée en 2009 par Thomas Gruber:« Dans le contexte

des sciences informatiques et de l'information, une ontologie définit un ensemble de primitives de représentation avec lesquelles on peut modéliser un domaine de connaissance ou un discours » en définissant ces primitives de représentation ainsi: « Les primitives de représentation sont généralement des classes (ou des ensembles), des attributs (ou des propriétés) et des relations (ou des relations entre les membres de la classe). Les définitions des primitives de représentation comprennent des informations sur leur signification et les contraintes de leur application logiquement cohérente. Dans le contexte des systèmes de bases de données, l'ontologie peut être considérée comme un niveau d'abstraction des modèles de données, analogue aux modèles hiérarchiques et relationnels, mais destiné à modéliser les connaissances sur les individus, leurs attributs et leurs relations avec d'autres individus. »[T. GRUBER et al., 2009].

4.3 Contenu des ontologies

Nous verrons dans cette section les éléments qui composent une ontologie et la façon dont ils sont organisés afin de retranscrire nos connaissances exprimées de manière linguistique et cognitives.
Les éléments essentiels d'une ontologie sont les concepts et les relations. Les ontologies incluent généralement une organisation hiérarchique des concepts qu'elles contiennent et des relations entre ces concepts, ainsi que des règles et des axiomes qui contraignent ces relations.

4.3.1 Les concepts

Selon la définition du dictionnaire Larousse, un concept est une: « Idée générale et abstraite que se fait l'esprit humain d'un objet de pensée concret ou abstrait, et qui lui permet de rattacher à ce même objet les diverses perceptions qu'il en a, et d'en organiser les connaissances. ». Les concepts peuvent être des catégories d'objets (arbres, agriculteurs), des principes (je pense,donc je suis), des idées (celles qu'on ce fait du métier d'agriculteur), des notions concrètes (plus léger, plus grand) ou abstraites (liberté, équipe de recherche).
Alors comment, dans le cadre des ontologies, rendre ce concept si abstrait en une chose pouvant être appréhendée, échangée et manipulée sans ambiguïté par tous et d'autant plus par des machines ?
Un des premiers éléments de réponse à cette question, nous est apporté par la sémiotique qui étudie les signes et les systèmes de signification. En effet, pour exprimer et partager un concept, nous utilisons des signes.
A la fin du XIXe, Ferdinand de Saussure avance que le signe est la réunion de quelque chose que l'on perçoit (le signifiant), et de la représentation cognitive associée à cette perception (le signifié). Le signifiant est ce qui nous est accessible de façon sensorielle (son, image, mot) alors que le signifié est la face immatérielle qui ne nous est accessible que par l'intellect. Pour Saussure, signifiant et signifié sont indissociables comme les deux faces d'une même pièce qui serait le signe [SAUSSURE, 2011].
Dans sa sémiotique Charles Sanders Pierce distingue trois types de signe, le symbole, l'icône et l'indice. Dans cet exemple (voir figure 4.1 le symbole choisi est le terme « feu » mais nous aurions pu choisir les mots « flamme », « incendie » ou « flambée ». Cela démontre que nous dissocions le concept, de sa manifestation linguistique : le terme n'est pas un concept et le concept n'est pas un terme. En 1923, Richards et Ogden expriment les relations entre conceptualisation, réalité et langage au travers de leur triangle sé-

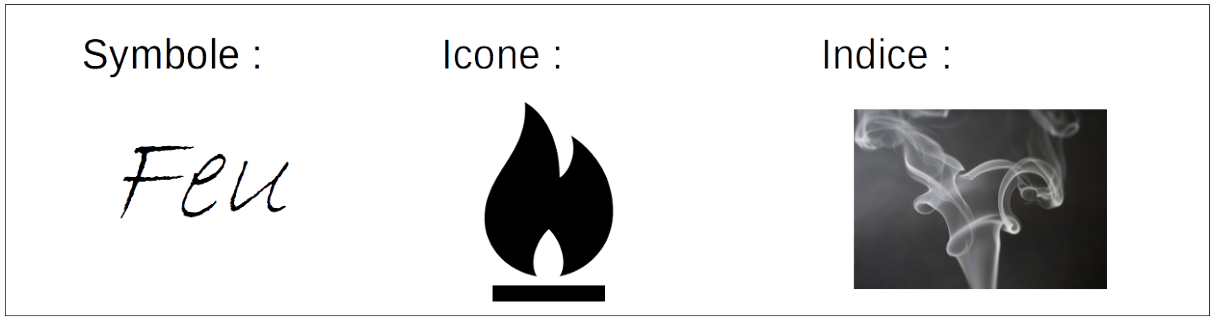


FIGURE 4.1 – *Les trois signes de C. S. Pierce : Le symbole, l'icône et l'indice*

miotique (voir figure 4.2). Ce triangle montre la relation entre le concept, le terme, et la chose nommée. La chose (ici un noyer dans un champ) se rapporte à un concept (la référence de notre intellect). Le mot « arbre » qui signifie (représente) la chose nommée symbolise le concept. Le fait que le concept soit une représentation mentale, pose

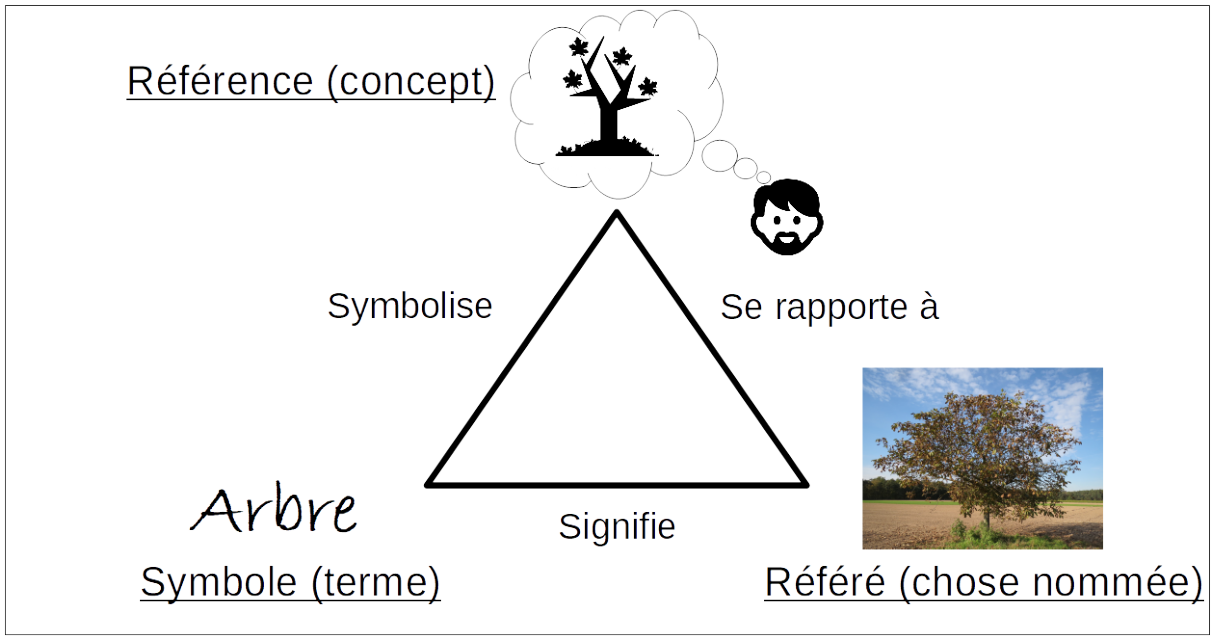


FIGURE 4.2 – *le triangle sémiotique proposé par les linguistes britanniques Charles Kay Ogden et Ivor Armstrong Richards.*

le problème d'obtenir une définition formelle de celui-ci. L'"intension" et l'"extension" sont deux façons de définir un concept. la distinction entre l'intension et l'extension correspond à celle entre "sens" et "référence" soutenue par Friedrich Ludwig Gottlob Frege [FREGE, 1948]. Pour Frege, la "référence" d'une expression linguistique est une partie de la réalité désignée par cette expression alors que le "sens" est ce qui fait le lien entre l'univers du langage et l'univers des choses extra-linguistiques. Les deux notions sont intimement liées, car il considère que le "sens" permet d'accéder à la "référence". L'intension d'un concept est l'ensemble des propriétés communes aux choses auxquelles il s'applique et son extension l'ensemble des choses qu'il englobe.

Pour exemple prenons un concept que nous nommerons A.

Nous pouvons lui associer une intension : « A est une sous catégorie d'arbre qui est cultivé pour obtenir des fruits ayant un intérêt alimentaire et économique pour l'homme ».

Son extension sera l'ensemble des individus de cette catégorie comme par exemple « un pommier, un cerisier, un noyer ... ».

A posteriori, nous pourrions symboliser ce concept A par le terme "arbre fruitier".

Cet exemple nous démontre que le concept se doit de n'avoir qu'un seul sens, qu'une seule définition car c'est cette définition qui conditionnera les individus qu'il englobera, qui en retour rappelons le , apporteront une définition au concept par la présence de leurs caractéristiques communes. Le terme "arbre fruitier" qui dénote le concept n'est nécessaire ni à son intension, ni à son extension et donc à sa définition montrant une fois de plus la distinction entre terme et concept. Le terme pouvant être ambigu, il est donc important de se préoccuper des problèmes de synonymie (un concept signifié par plusieurs termes) et d'homonymie (un terme signifiant plusieurs concepts).

4.3.2 Les relations

Les ontologies définissent aussi des relations qui peuvent exister entre les individus englobés par un concept. Comme pour les concept, nous pouvons associer à ces relations une intension, une extension et une expression linguistique. L'exemple suivant illustre cette démarche : Considérant une relation R.

Son intension est « R est une relation entre une plante et le taxon au cinquième rang de la classification classique des êtres vivants (la famille) »

Son extension est « (pommier,rosaceae),(cerisier,rosaceae),(noyer,juglandaceae) ... »

On peut lui associer la représentation linguistique « est de la famille des ».

Les relations peuvent de plus posséder une « signature », c'est-à-dire la spécification des concepts dont elles vont lier les individus. Dans notre exemple nous pouvons indiquer que la relation R a pour signature (plante,famille) puisqu'elle met en relation des individus du concept plante avec des individus du concept famille.

4.3.3 Les contraintes

Les contraintes sont exprimées au niveau des intensions, elles expriment les lois qui les régissent pour rendre compte de certains aspect de la réalité.
Dans les exemples précédents on voit que notre définition du concept A implique que tous les individus de ce concept "arbres fruitier" sont aussi des individus du concept "arbre" ou encore que la relation R lie des individus de la catégorie des "plantes" avec les individus de la catégorie des "familles" comme l'indique sa signature.

4.3.4 La hiérarchie des concepts

Dans les ontologies les concepts sont généralement organisés sous forme de hiérarchies. Pour exemple si l'on définit les concepts "être vivant", "animal", "plante", "arbre" et "arbre fruitier", nous pouvons les organiser comme présenté dans la figure 4.3. Le fait de placer un concept sous un autre, c'est-à-dire de spécialiser un concept par un autre, est appelé mécanisme de subsomption. Ainsi le concept d'"arbre fruitier" est subsumé

par le concept "arbre", lui-même subsumé par le concept "plante". Dans ce contexte, un individu de la classe "arbre fruitier" sera aussi un individu de la classe "arbre", et par transitivité également un individu de la classe "plante"
Cette organisation rappelle bien, les liens que l'on peut retrouver au sein de ces taxo-

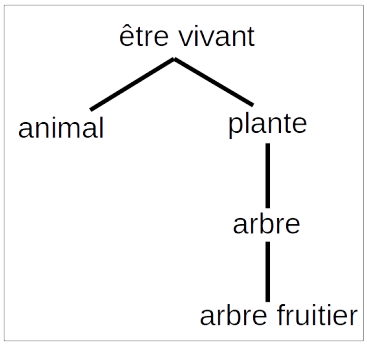


FIGURE 4.3 – *Relations de subsomption entre différents concepts.*

nomies mais il ne doit pas y avoir de confusion. Pour marquer la différence entre taxonomie et ontologie nous aborderons dans la section suivante, les informations que l'on peut trouver dans une ontologie en plus des relations taxonomiques.

4.3.5 Les différentes informations exprimées par les ontologies

Les informations exprimées par les ontologies au travers de leurs intensions sont multiples, comme le montrent les quelques exemples suivants:

- Des relations taxonomiques : Un arbre est une plante.
- Des définitions : Un arbuste est un arbre dont la hauteur est de moins de sept mètres.
- Des contraintes : Un arbre ne peut avoir qu'un seul tronc.
- Des relations de composition: un arbre est composé d'un tronc, de branches et de feuilles.
- Des relations inverses : la relation « partOf » est l'inverse de la relation « hasPart »
- Des algorithmes : Le volume approché d'un tronc est la multiplication de sa circonférence, par sa hauteur.
- ...

Le contenu d'une ontologie peut donc être extrêmement varié, il peut de plus se différencier d'une ontologie à l'autre par le type de l'ontologie considérée, comme nous le verrons dans la section suivante.

4.4 Les différents types d'ontologies

Diverses classifications pour les ontologies ont été proposées. Rafael Valencia Garcia [VALENCIA GARCÍA, 2008] en fait état en les classant principalement selon deux critères qui sont, le type des connaissances quelles contiennent et la motivation de leur création. Selon le premier critère, je propose de retenir dans les sous sections suivantes,

les ontologies terminologiques et les ontologies cadres modélisant la connaissance. Au sein des ontologies modélisant la connaissance et selon le deuxième critère, la motivation, j'aborderais une autre classification.

4.4.1 Les ontologies terminologiques

Selon Gertjan van Heijst [VAN HEIJST et al., 1997] elles sont assimilables à des lexiques dans le sens ou elles précisent les termes utilisés pour représenter la connaissance dans le domaine du discours. Elles sont d'ailleurs souvent issues de vocabulaires contrôles préexistants sous forme de taxonomie ou de thésaurus mais cette transposition ne doit pas faire oublier une différence essentielle qui est que les ontologies terminologiques sont centrées sur le concept et non sur le terme qui n'est utilisé que pour nommer les concepts qu'elles définissent. Comme exemple d'ontologies terminologiques nous pouvons citer WordNet [FELLBAUM, 2012] qui classifie le contenu sémantique et lexical de la langue anglaise, et AGROVOC [CARACCIOLO et al., 2013] le plus grand ensemble de données sur l'agriculture accessible au public.

4.4.2 Les ontologies cadres

Dans le cadre des systèmes d'information basés sur des ontologies, Nicola Guarino propose cette classification [GUARINO, 1998] :

- Les ontologies de haut niveau
  « Les ontologies de haut niveau décrivent des concepts très généraux comme l'espace, le temps, la matière, l'objet, les événements, l'action, ..., qui sont indépendants d'un problème ou d'un domaine particulier : il semble donc raisonnable, au moins en théorie, d'avoir des ontologies de haut niveau unifiées pour de grandes communautés d'utilisateurs. »

- Les ontologies de domaine et de tâches
  « Les ontologies de domaine et les ontologies de tâches décrivent, respectivement, le vocabulaire lié à un domaine générique (comme la médecine ou l'automobile) ou à une tâche ou activité générique (comme le diagnostic ou la vente), en spécialisant les termes introduits dans l'ontologie de premier niveau. »

- Les ontologies d'application
  « Les ontologies d'application décrivent des concepts dépendant à la fois d'un domaine et d'une tâche particuliers, qui sont souvent des spécialisations des deux ontologies connexes. Ces concepts correspondent souvent aux rôles joués par les entités du domaine lors de l'exécution d'une certaine activité, comme une unité remplaçable ou un composant de rechange. »

D'autre critères peuvent être utilisés pour classer les différents types d'ontologies comme par exemple le degré de la formalisation. Celui ci sera intimement lié au langage utilisé pour constituer l'ontologie. En effet on peut utiliser pour décrire une ontologie, le langage naturel, des langages semi-formels ou la logique ce qui aura pour conséquence de modifier ces capacités de traitement en particulier en ingénierie informatique.

4.5 Rôle des ontologies

4.5.1 Représentation et uniformisation de la connaissance d'un domaine

En regroupant les connaissances d'un domaine au sein d'une ontologie, on factorise les connaissances de ce domaine. En effet les connaissances d'une ontologie sont toujours vraies comme par exemple le fait qu'un arbre soit une plante et donc cette connaissance peut être appliquée par la suite à tous les individus de la catégorie des arbres sans qu'il soit nécessaire de ré-exprimer cette connaissance pour chaque individu concerné.

4.5.2 Réutilisation et échange de la connaissance

La modularité des ontologies permet la réutilisation et l'échange des connaissances au sein d'applications différentes. Cela est d'autant plus vrai avec les ontologies de haut niveau (les ontologies « Top level »). Ces capacités de réutilisation et d'échange permettent une interopérabilité entre applications.

4.6 En résumé

Les ontologies sont un moyen de représenter nos connaissances dans un domaine particulier. D'une question philosophique « qu'est ce que l'être », elles sont devenues des objets nous permettant de capitaliser et d'échanger nos connaissances. De formes multiples, elles sont néanmoins toutes centrées sur la notion de concept nous permettant ainsi d'échapper aux pièges de notre langue naturelle comme la synonymie, l'homonymie ou la polysémie. En définissant les concepts qui la composent ; l'ontologie restitue sans ambiguïté les représentations mentales que nous nous faisons des choses de notre univers. Cette formalisation de notre pensée cognitive ouvre la voie à des moyens plus efficaces d'expression et d'échange de notre savoir et à des traitements automatisés de nos connaissances.

Les ontologies terminologiques fournissent des vocabulaires contrôlés pour un domaine de connaissance pouvant être utilisés pour classer et retrouver des informations au travers d'applications informatiques spécialisées comme c'est le cas des thésaurus dans par exemple la gestion des bibliothèques.

Les ontologies de domaine offrent quant à elles un modèle pour la représentation des connaissances dans le cadre d'un domaine, d'une activité ou d'une tache spécifique. Ce modèle pourra être utilisé pour mettre en place une base de connaissance ou des informations factuelles pourront être manipulées par des machines grâce au formalisme poussé de ses langages de représentation. Une base de connaissance peut donc se résumer à une ontologie de domaine à laquelle on ajoute des faits du monde réel. L'adossement de ces langages de représentation aux logiques de description (voir chapitre suivant) permet mon seulement d'exprimer la connaissance terminologique

d'un domaine de façon formelle et structurée mais aussi de réaliser des inférences, mécanismes de notre pensée cognitive qui vont permettent de rendre explicites des données implicites et donc de générer des faits nouveaux dans une base de connaissance.

V

Les formalismes logiques

5.1 Les réseaux sémantiques 57
5.2 Les langages à base de frames 59
5.3 Les graphes conceptuels 61
5.4 Les logiques de description 63
5.4.1 Les constructeurs. 64
5.4.2 Les différentes logiques de description. 64
5.4.3 Raisonnement avec les logiques de description. 66
5.4.4 Les domaines concrets. 67

Le formalisme des ontologies repose sur celui des langages utilisés pour les concevoir. De ce formalisme dépendra l'efficacité de la structure d'une ontologie, et notre capacité à vérifier sa cohérence. De plus ce formalisme sera étroitement lié à la possibilité de réaliser des inférences sur la base des connaissances contenues dans ces ontologies afin de pouvoir créer de nouvelles connaissances. En ingénierie des connaissances, plusieurs initiatives pour permettre une modélisation et une manipulation des connaissances ont été proposées. Il s'agit des réseaux sémantiques, des langages à base de frames (les langages à cadres), des graphes conceptuels et des logiques de description.

5.1 Les réseaux sémantiques

Les réseaux sémantiques sont des notations graphiques pour représenter la connaissance. On peut remonter leurs origines jusqu'à l'antiquité avec l'arbre de Porphyre (voir figure 5.1) avec lequel le philosophe Porphyre de Tyr présente les bases de la logique d'Aristote à propos de ses "catégories animales" sous la forme d'un schéma [PORPHYRE, 1998]. Cette représentation présente des divisions dichotomiques qui montre qu'une espèce est définie par genre et différence, et le processus se répète jusqu'à l'espèce la plus spécialisée formant ainsi une arborescence. Ce type de schématisation sera reprise non seulement en philosophie mais aussi dans les domaines de la psychologie et de la linguistique. Malgré leur utilisation en vogue dans les années soixante aucun formalisme n'est alors encore proposé. Mais suite au travaux de M. Ross Quillian sur un modèle pour le stockage informatique d'information sémantique [QUILLIAN, 1967]

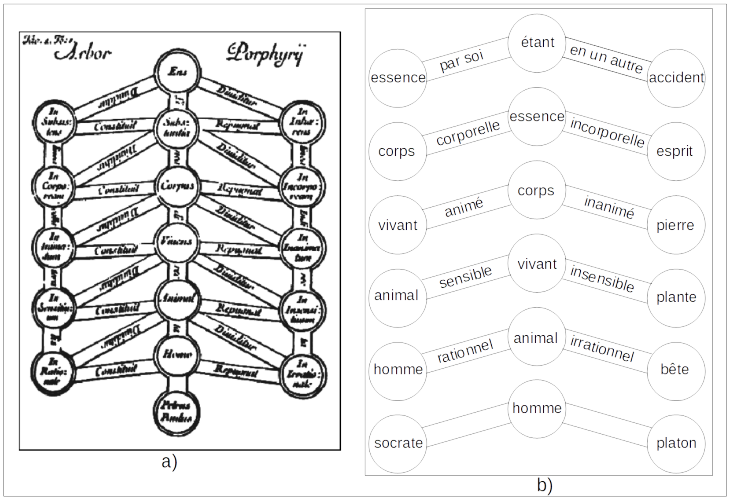


FIGURE 5.1 – *Arbre de Porphyre :*
a) Gravure du livre "Institutiones philosophicae I" de Purchotius, 1730
b) Interprétation de l'arbre de Porphyre

et un programme de lecture automatisé, [QUILLIAN, 1969] et de sa coopération avec Allan M.Collins [COLLINS et QUILLIAN, 1969] des tentatives de formalisation apparaissent pour apporter des capacités computationnelles aux réseaux sémantiques.
Un réseau sémantique se compose d'un graphe constitué de sommets et d'arrêtes interconnectés. Les sommets ou nœuds représentent des concepts et les arrêtes des relations sémantiques entre les concepts. L'exemple suivant (voir figure 5.2) illustre sommairement un réseau sémantique qui représente notre conceptualisation d'un arbre.
Les sommets sont étiquetés soit par des termes désignant des concepts, comme le concept d'arbre ou le concept de plante, soit des termes désignant des objets instance des concepts comme mon noyer, celui de mon jardin, soit des termes désignant des propriétés de ces concepts. Les arcs représentent des relations d'agrégation ("est un"), d'instanciation ("sorte de"), de composition ("a"), ou des relations spécifiques ("stocke").
Pour savoir si un objet représenté par un sommet est membre d'un concept représenté par un autre sommet on suit les arcs d'instanciation et d'agrégation. Ainsi on peut lire sur ce graphe que Mon noyer (sorte de) Arbre fruitier (est un) Arbre et le résumé en Mon noyer (sorte de) Arbre, ce qui nous permet de dire que Mon noyer est composé de branches et stocke du carbone.

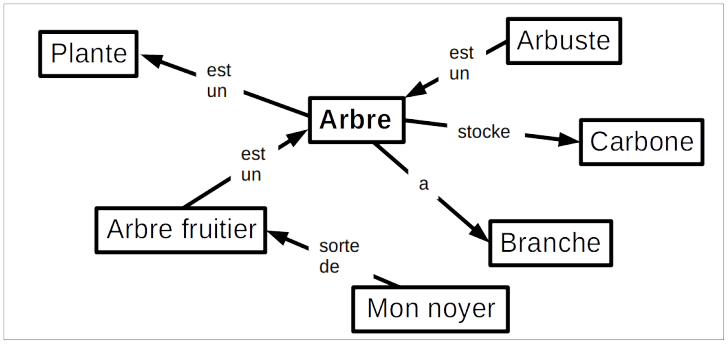


FIGURE 5.2 – *Schéma d'un réseau Sémantique pour le concept d'arbre.*

Ainsi, un réseau sémantique permet de stocker des faits mais aussi les connexions associatives que nous réalisons rendant ainsi certaines informations accessibles à partir d'autres.

5.2 Les langages à base de frames

Marvin Lee Minsky considère qu'à son époque les théories en intelligence artificielle et en psychologie sont axées sur des objets d'une granularité trop fine représentés par exemple par les nœuds d'un réseau sémantique. Il propose pour rendre compte de nos capacités de raisonnement d'étudier des structures plus grandes au sein desquelles des faits et les procédures induites par ces données factuelles sont étroitement liés. Il propose le concept de "frame" (cadre en français) qu'il définit ainsi « Un cadre est une structure de données destinée à représenter une situation stéréotypée, comme le fait de se trouver dans un certain type de salon ou d'aller à l'anniversaire d'un enfant. Certaines de ces informations concernent l'utilisation du cadre, d'autres ce que l'on peut s'attendre à voir se produire, d'autres encore ce qu'il faut faire si ces attentes ne sont pas confirmées. » [MINSKY, 1975]. Ce paradigme repose sur la suggestion que l'être humain utilise des connaissances stéréotypées pour interpréter et agir face aux situations auxquelles il est confronté. Face à une nouvelle situation, il se réfère à une situation qu'il a lui-même déjà vécue n'ayant qu'à ajuster si nécessaire quelques paramètres. Le but de la représentation de ces situations sous forme de cadres est donc de modéliser le contexte d'un problème afin d'en réduire l'espace des solutions potentielles. La conception d'un cadre repose sur trois niveaux emboîtés: Cadre-Attribut-facette (voir figure 5.3).
Les attributs sont des éléments qui décrivent les propriétés d'une situation (le cadre).

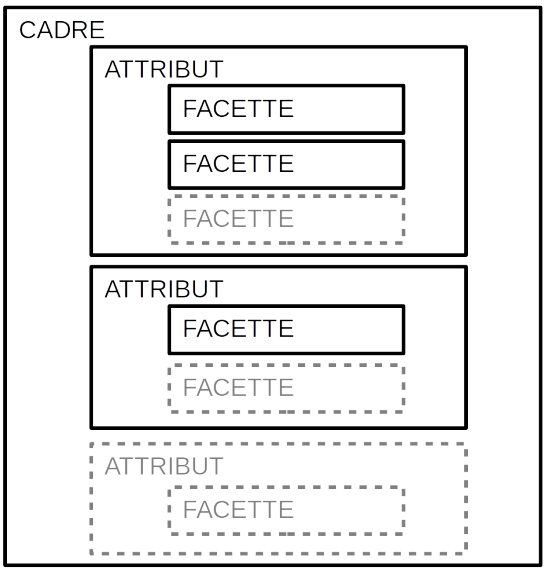


FIGURE 5.3 – *Schéma des trois niveaux emboités d'un cadre.*

Un attribut, aussi nommé "slot" peut être lui-même un cadre. Les facettes sont les formes particulières de description ou de comportement d'un attribut. Elles sont toujours associées à une valeur.
Les facettes peuvent être de plusieurs types :

- Les facettes déclaratives parmi lesquelles :

  Les facettes de typage comme par exemple : $un, $liste-de, $intervalle ...

  Les facette de valeurs comme par exemple :$valeur, $défaut ...

- Les facettes procédurales qui permettent d'exécuter des procédures dans certain cas, comme par exemple $si-besoin, $si-possible, $si-ajout, $si-enlève ...

ARBRE		
Attribut	**Facette**	**Commentaire**
SORTE-DE	$valeur PLANTE	Cadre parent
ESPÈCE	$un CHAÎNE	Procédure de contrainte
FAMILLE	$un CHAÎNE $domaine "Acanthaceae" "Aceraceae" ...	Procédures de contrainte
AGE	$un ENTIER	Procédure de contrainte
HAUTEUR	$un ENTIER	Procédure de contrainte
SEXUALITÉ	$défaut "hermaphrodite"	Valeur par défaut
LOCALISATION	$valeur CORDONNÉES GPS	Cadre

ARBRE FRUITIER		
Attribut	**Facette**	**Commentaire**
SORTE-DE	$valeur ARBRE	Cadre parent
ESPÈCE	$un CHAÎNE	Procédure de contrainte
FAMILLE	$un CHAÎNE $domaine "Acanthaceae" "Aceraceae" ...	Procédures de contrainte
FRUIT	$un CHAÎNE	Procédure de contrainte
AGE	$un ENTIER	Procédure de contrainte
HAUTEUR	$un ENTIER	Procédure de contrainte
SEXUALITÉ	$défaut "hermaphrodite"	Valeur héritée
LOCALISATION	$valeur CORDONNÉES GPS	Cadre

ARBUSTE		
Attribut	**Facette**	**Commentaire**
SORTE-DE	$valeur ARBRE	Cadre parent
ESPÈCE	$un CHAÎNE	Procédure de contrainte
FAMILLE	$un CHAÎNE $domaine "Acanthaceae" "Aceraceae" ...	Procédures de contrainte
AGE	$un ENTIER	Procédure de contrainte
HAUTEUR	$un ENTIER $intervalle [1 700]	Restriction : Les arbustes font moins de 7 mètres
SEXUALITÉ	$défaut "hermaphrodite"	Valeur héritée
LOCALISATION	$valeur CORDONNÉES GPS	Cadre

ARBRE N°10		
Attribut	**Facette**	**Commentaire**
SORTE-DE	$valeur ARBRE FRUITIER	Cadre parent
IDENTIFIANT	$valeur "A10L02"	valeur de l'instance
ESPÈCE	$valeur "Argousier"	valeur de l'instance
FAMILLE	$valeur "Elaeagnaceae"	valeur de l'instance
FRUIT	$valeur "Baies de l'argousier"	valeur de l'instance
AGE	$si-besoin : soustraire(Date du jour, DATE PLANTATION)	Procédure
DATE PLANTATION	$valeur "27/04/2016"	valeur de l'instance
HAUTEUR	$valeur 236	valeur de l'instance
SEXUALITÉ	$valeur "dioïque"	Exception : L'argousier est une plante dioïque
LOCALISATION	$valeur "43.6993, 3.8107"	valeur de l'instance

FIGURE 5.4 – *Exemples de cadres définissant différentes représentations d'arbres.*

Prenons l'exemple suivant (voir figure 5.4) : Le cadre "ARBRE" schématisé ici sous forme de tableau représente le concept d'un arbre. Des attributs (colonne Attribut) qui définissent ce concept sont liés à une ou plusieurs facettes toujours associées à une valeur (colonne Facette). L'attribut "SORTE-DE" lié à la facette déclarative $valeur associée à la valeur "PLANTE" indique le lien hiérarchique entre le concept "ARBRE" et son concept parent "PLANTE". Les attributs "ESPÈCE", "FAMILLE", "AGE" et "HAUTEUR" se voient imposer des contraintes de type avec les facettes $un et $domaine. Les arbres étant en majorité des individus hermaphrodites (dans 90% des cas) l'attribut "SEXUALITÉ" se verra affecté la valeur par défaut "hermaphrodite". La dernière ligne stipule que la propriété "LOCALISATION" du concept "ARBRE" aura pour valeur des individus du concept "COORDONNÉES GPS". Les cadres "ARBRE FRUITIER" et "ARBUSTE" reprennent les attributs du cadre parent "ARBRE" dont ils sont les enfants. D'autres attributs peuvent être ajoutés comme l'Attribut "FRUIT" du cadre "ARBRE FRUITIER" ou être redéfinis comme l'attribut "HAUTEUR" du cadre "ARBUSTE" dont les valeurs seront restreintes à un intervalle de 1 à 700 traduisant le fait qu'un arbuste est défini comme un arbre de moins de 7 mètres. Pour ces deux cadres, l'attribut

"SEXUALITÉ" hérite de la valeur par défaut transmise du cadre "ARBRE" qui les subsume.

Le cadre "ARBRE N°10" représente lui un individu réel, les valeurs des attributs sont renseignées quand elles sont connues et cela conformément aux contraintes imposées par le cadre parent dont il hérite ses attributs. On voit néanmoins que cet héritage n'est pas forcément complet, en effet ici la valeur par défaut "hermaphrodite" de l'attribut "SEXUALITÉ" n'est pas reprise car l'argousier étant une plante dioïque, elle est unisexuée. On notera aussi pour ce dernier cadre la présence du procédure introduite par la facette $si-besoin qui permet le calcul de l'âge de cet arbre fruitier en fonction de sa date de plantation.
Un cadre peut donc contenir des informations qui sont attribuées par défaut lorsque l'on visualise mentalement un concept (comme par exemple la couleur verte du feuillage si vous pensez à un arbre) mais il fournit aussi des informations sur ce à quoi il faut s'attendre en imposant certaines contraintes (comme par exemple l'utilisation de taxons définis pour renseigner la famille taxonomique) et des informations sur la conduite à tenir si ces informations venaient à manquer (comme par exemple le calcul de l'âge de la plante en se référent à sa date de plantation).
On notera aussi sur cet exemple, certains des avantages apportés par ce type de représentation structurelle en plus de la simple description de faits sur certains objets ou évènements. En effet, la représentation de la connaissance sous forme de cadre offre la possibilité d'une organisation hiérarchique comme dans cet exemple avec les concepts de "PLANTE", "ARBRE" et "ARBRE FRUITIER". De plus certains attributs peuvent être réutilisés dans d'autres cadres comme "ESPÈCE" "FAMILLE" ou "AGE" pour décrire des animaux et non plus des plantes. Cette réutilisation des attributs peut faciliter des raisonnements simples comme par exemple la comparaison des valeurs d'un attribut "HAUTEUR" entre un arbre et un immeuble. Pour finir on notera la possibilité de modifier les valeurs par défaut des attributs comme pour l'attribut "SEXUALITÉ" dans cet exemple, ce qui confère à cette forme de représentation une certaine souplesse facilitant la représentation du monde réel.
Les langages à base de frames sont donc un moyen de représentation et de raisonnement des connaissances basés sur la structure. Leur structure simple et souple et la possibilité d'utiliser des procédures rattachées en font les candidats idéaux pour des applications de programmation. Les applications visées par les langages à cadre sont la vision artificielle, l'analyse du discours et la compréhension du langage. Des langages informatiques ont été créés pour la mise en œuvre de représentation sous forme de cadres parmi lesquels KRL [BOBROW et WINOGRAD, 1977], KL-ONE [BRACHMAN et SCHMOLZE, 1989] et LOOM [MACGREGOR et BURSTEIN, 1991].

5.3 Les graphes conceptuels

La théorie des graphes conceptuels a été initialement proposée par John F. Sowa [John F. SOWA, 1976 ; John F. SOWA, 1984 ; John F.. SOWA, 2000] dont l'objectif des travaux était de mettre en place un système de logique simple et efficace pour représenter les structures du langage naturel. Il définit un graphe conceptuel (CG) comme étant « une représentation graphique de la logique basée sur les réseaux sémantiques de l'intelligence artificielle et les graphes existentiels de Charles Sanders Peirce. »[John F. SOWA, 2008]. A la vue des travaux réalisés à son époque sur les réseaux sémantiques, Sowa considère que « certains de ces réseaux ont été explicitement conçus pour mettre en œuvre des hypothèses sur les mécanismes cognitifs humains, tandis que d'autres ont été conçus principalement pour l'efficacité des ordinateurs. Parfois, des

raisons informatiques peuvent conduire aux mêmes conclusions que les preuves psychologiques »[John F. SOWA, 1987]. Il cherche donc en s'appuyant en particulier sur les travaux de Charles Sanders Peirce [PEIRCE, 1958 ; John F. SOWA, 2011] et de Lucien Tesnière [TESNIERE, 1959], un formalisme de représentation basée sur la logique pouvant exprimer la pensée humaine au travers de son langage. Un graphe conceptuel est donc un formalisme de représentation des connaissances et de raisonnements pouvant être traduit en formule logique.
La définition formelle d'un graphe conceptuel est la suivante :

- Un graphe conceptuel est un graphe fini, connexe et biparti
- Les nœuds d'un graphe conceptuel sont de deux types: les concepts et les relations conceptuelles
- Chaque relation conceptuelle possède au moins un arc, qui doit être lui-même obligatoirement relié à un concept.

Selon la notation graphique suivante où les concepts sont des rectangles, les relation conceptuelles des ovales et les arcs du graphe des flèches, la figure 5.5 montre pour exemple le graphe conceptuel de la phrase: "Nicolas est allé à Restinclières en tracteur". Les quatre concepts issus de cette phrase portent les étiquettes "Personne", "Aller", "Lieu" et "Tracteur" et pour deux d'entre eux des instances sont identifiées, les individus Nicolas et Restinclières (domaine agroforestier situé au nord de la ville de Montpellier). Les trois relations conceptuelles ont elles aussi des étiquettes, qui représentent alors le type des relations qui unissent ces concepts, la relation de type "agent" (Agnt), la relation de type "destination" (Dest) et la relation de type "instrument" (Inst). Sur les arcs les flèches sortantes indiquent le premier argument de la relation conceptuelle et celle entrante le second argument de cette même relation. Ce graphe décrit donc que la personne Nicolas est un agent d'une instance du concept aller, le lieu Restinclières étant la destination et tracteur l'instrument. Ce graphe qui repose sur un système de logique, peut être représenté par la logique des prédicats. Si l'on considère les prédicats $Personne(Nicolas)$, $Aller(x)$, $Lieu(Restinclières)$ et $Tracteur(y)$ ainsi que les relations conceptuelles$Agnt(x, Nicolas)$, $Dest(x, Restinclières)$ et $Inst(x, y)$, le graphe peut être transposé en cette formule logique :
$(\exists x)(\exists y)(Aller(x) \wedge Personne(Nicolas) \wedge Lieu(Restinclières) \wedge Tracteur(y) \wedge Agnt(x, Nicolas) \wedge Dest(x, Restinclières) \wedge Inst(x, y))$.
On notera la ressemblance entre les graphes conceptuels et les langage à cadres dé-

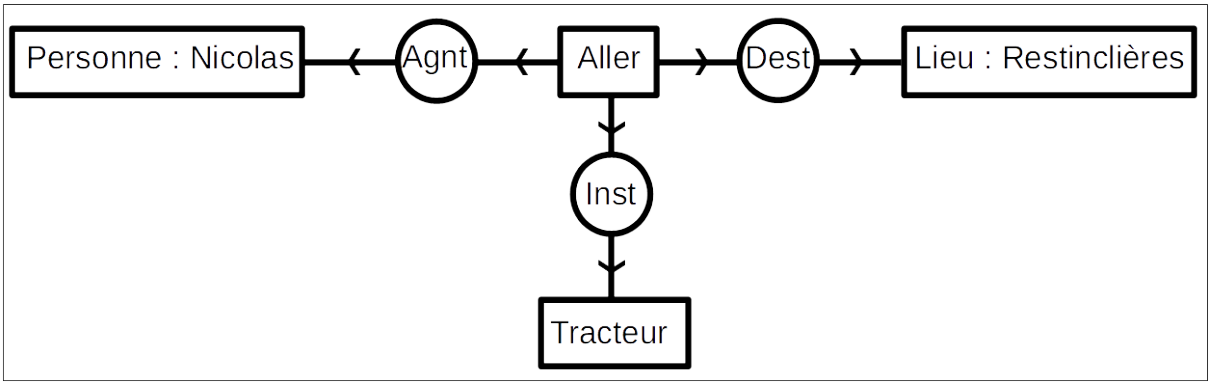


FIGURE 5.5 – *Graphe conceptuel de la phrase: "Nicolas est allé à Restinclières en tracteur".*

crits dans la section précédente (5.2). En effet les relations conceptuelles peuvent être vues comme les attributs d'un cadre, les étiquettes de type comme des contraintes, et les noms Nicolas et Restinclières comme des valeurs de facettes. La différence primor-

diale est que pour les graphes conceptuels la généricité est plus grande (pas de facettes limitées en nombre) et que toute la logique peut être représentée.

5.4 Les logiques de description

Nous aborderons dans cette section les logiques de description en tant qu'outil qui fournira les capacités de raisonnement logique de notre base de connaissance sur l'agroforesterie
Selon la définition de Franz Baader et Werner Nutt les logiques de description sont « une famille de formalismes de représentation de la connaissance qui représentent la connaissance d'un domaine d'application (le "monde") en définissant d'abord les concepts pertinents du domaine (sa terminologie), puis en utilisant ces concepts pour spécifier les propriétés des objets et des individus se produisant dans le domaine (la description du monde). » [BAADER, CALVANESE et al., 2003].
Les logiques de description sont issues des recherches sur le calcul des prédicats du premier ordre encore nommée logique du premier ordre, sur les réseaux sémantiques et sur les langages à cadres [BRACHMAN et SCHMOLZE, 1985]. Elles exploitent des sous-ensembles de la logique du premier ordre et sont utilisées pour représenter la connaissance terminologique d'un domaine d'application d'une façon structurée et formelle. Une base de connaissance, représentée avec les logiques de description est constituée par trois éléments: le concept, le rôle et l'instance. Le concept représente un ensemble d'individus, le rôle une relation binaire entre individus et l'instance un individu particulier faisant partie de l'entité générique que représente le concept. Ces trois entités respectent les principes suivants:

- Concepts et rôles ont une description structurée grâce à l'utilisation de constructeurs qui introduisent les rôles associés aux concepts et les restrictions sur ces rôles. A cette description une sémantique est apportée via une interprétation.

- La description et la manipulation se fait sur deux niveaux:
  - Le niveau terminologique ou TBox pour les concepts et les rôles. C'est à ce niveau que seront décrites les connaissance générales d'un domaine. Ce niveau terminologique peut être scindé en deux parties, une RBox ("Role Box") qui représentera l'ensemble des rôles et une TBox qui désignera dans ce cas uniquement l'ensemble des concepts. L'ensemble composé de cette RBox et de cette TBox sera alors nommé CBox ("Constraint Box").
  - Le niveau factuel ou niveau des assertions ou ABox pour les instances. Ce niveau représente une situation précise dans l'univers du domaine. Cette situation constitués d'individus trouve les moyens de son expression en utilisant les concepts et les rôles de la TBox.

- Des relations de subsomption peuvent s'appliquer sur les concepts et les rôles pour former une hiérarchie de concepts et éventuellement une hiérarchie de rôles.

- Pour obtenir des raisonnements au sein d'une base de connaissance on utilise les opérations de classification et d'instanciation:
  - La classification permet de déterminer la position d'un concept ou d'un rôle dans sa hiérarchie.
  - L'instanciation permet de retrouver les concepts auxquels un individu peut appartenir.

Les logiques de description ont une approche similaire à celle des ontologies telles que nous les avons décrites au chapitre précédent 4 en ce sens où pour représenter un individu elles déterminent une entité générique, le concept et les relations entre individus ou entre concepts. Soit une base de connaissance $\mathcal{K}$, un ensemble fini de concepts $\mathcal{N}_c$, un ensemble fini de rôles $\mathcal{N}_r$ et un ensemble fini d'instances $\mathcal{N}_i$. Si $\mathcal{N}_c$, $\mathcal{N}_r$ et $\mathcal{N}_i$ sont disjoints deux à deux $\mathcal{K}$ est défini ainsi:

$$\mathcal{K} =< \mathcal{N}_c, \mathcal{N}_r, \mathcal{N}_i >$$

Si l'on prend en considération les niveaux terminologiques $\mathcal{T}$ (TBox) et factuel $\mathcal{A}$ (ABox) on obtient :

$$\mathcal{K} =< \mathcal{T}, \mathcal{A} >$$

Les concepts peuvent être atomiques (primitifs) ou complexes (définis). Les concepts complexes ont une définition formée avec des concept atomiques servant de briques élémentaires et de constructeurs qui sont des opérateurs logiques. Il en va de même pour les rôles primitifs ou définis, les rôle définis possédant une description des propriétés associées au rôle.

5.4.1 Les constructeurs.

Il existe deux types de constructeurs:

- Les constructeurs pour les concepts: Ce sont des opérateurs pour combiner et former des concepts complexes comme les opérateurs booléens et les quantificateurs. Par exemple:
  - le constructeur de conjonction $\sqcap$ avec la syntaxe $C \sqcap D$ ou C et D sont des concepts arbitraires décrit un concept complexe qui contient l'ensemble des individus qui sont à la fois membre du concept C et membre du concept D.
  - le constructeur d'équivalence $\equiv$ avec la syntaxe $C \equiv D$ qui sert à énoncer l'équivalence entre les concept C et D.
  - le constructeur d'inclusion $\sqsubseteq$ avec la syntaxe $C \sqsubseteq D$ qui sert à énoncer que l'ensemble des individus membres du concept C sont aussi membre du concept D.
- Les constructeurs pour les rôles: Ce sont des opérateurs pour former des rôles complexes comme l'inversion ou les fermetures transitives.

5.4.2 Les différentes logiques de description.

Des correspondances sont réalisées entre les logiques de description et les formalismes de la logique des prédicats, des langages à cadre et des réseaux sémantiques. Pour les langages à cadres Brachman et Levesque propose la logique $\mathcal{FL}$ (Frame-based Language) et $\mathcal{FL}^-$ [BRACHMAN et LEVESQUE, 1984]. En rajoutant la négation atomique à la logique $\mathcal{FL}^-$ Manfred Schmidt-Schauß et Gert Smolka introduisent la logique $\mathcal{AL}$ (Attribute Language) [SCHMIDT-SCHAUSS et SMOLKA, 1991]. Baader et Nutt qualifient cette logique de "minimale" car pour eux une logique moins expressive serait dénuée d'intérêt [BAADER, CALVANESE et al., 2003]. Le tableau 5.1 montre les constructeurs et leur syntaxe de la logique minimale $\mathcal{AL}$. On notera la définition du concept universel qui définit le concept le plus général et du concept vide qui définit le concept le plus spécifique. Cette logique minimale $\mathcal{AL} = \{\top, \bot, \neg A, C \sqcap D, \forall R.C, \exists R.\top\}$ peut être étendue à ajoutant des constructeurs de concepts ou de rôles. La nomenclature des

Constructeur	Syntaxe
Concept atomique A	A
Concept universel	$\top$
Concept vide	$\bot$
Négation atomique	$\neg A$
Conjonction	$C \sqcap D$
Quantification universelle complète	$\forall R.C$
Quantification existentielle limitée	$\exists R.\top$

TABLE 5.1 – *Tableau des constructeurs et de leur syntaxe pour la logique $\mathcal{AL}$.*

logiques de description impose alors que les nouveaux constructeurs soit représentés par des lettres qui vont êtres concaténées au nom de la logique d'origine. Le tableau 5.2 montre un exemple de ces différents constructeurs. Ainsi la logique $\mathcal{ALC}$ représente la logique $\mathcal{AL}$ à laquelle on a rajouté la négation complète. Certaines extensions peuvent être équivalentes comme par exemple $\mathcal{ALC}$ et $\mathcal{AL\mathcal{U}\mathcal{E}}$ du fait de l'expression possible des constructeurs de disjonction et de quantification existentielle par la négation complète des constructeurs de conjonction et de quantification universelle présents dans la logique $\mathcal{AL}$, $C \sqcup D \equiv \neg(\neg C \sqcap \neg D)$ et $\exists R.C \equiv \neg\forall R.\neg C$ [BAADER, CALVANESE et al., 2003].

Lettre	Constructeur / Contrainte	Syntaxe
$\mathcal{C}$	Négation de concept (complète)	$\neg C$
$\mathcal{U}$	Disjonction	$C \sqcup D$
$\mathcal{E}$	Quantification existentielle typée	$\exists R.C$
$\mathcal{R}^+$	Transitivité des rôles	R^+C
$\mathcal{F}$	Fonctionnalité	$\leq 1R$
$\mathcal{N}$	Restriction de cardinalité	$\leq nR; \geq nR$
$\mathcal{Q}$	Restriction de cardinalité qualifiée	$\leq nR.R; \geq nR.R$
$\mathcal{O}$	Énumération	$\{a_1 \ldots a_n\}$
$\mathcal{H}$	Hiérarchie de rôles atomiques	$R_1 \sqsubseteq R_2$
$\mathcal{R}$	Hiérarchie de rôles non-atomiques	$R_1 \sqcap R_2$
$\mathcal{I}$	Inversion de rôles	R^{-1}

TABLE 5.2 – *Constructeurs de rôles et concepts et contraintes sur l'interprétation des rôles pour L'extension de la logique $\mathcal{AL}$*

La logique minimale $\mathcal{AL}$ peut aussi être étendue en ajoutant des contraintes sur l'interprétation des rôles. Le tableau 5.2 montre la spécification d'un ensemble de rôles transitifs noté $\mathcal{R}^+$, la logique $\mathcal{ALC}$ étendue par $\mathcal{R}^+$ est désigné par la lettre $\mathcal{S}$.
Pour illustrer la création d'une base de connaissance avec les logiques de description dans le cadre de l'agroforesterie prenons l'exemple suivant:
Définissons les concepts atomiques *Arbre*, *Tronc* , *Arbuste* et *Fruit* et les rôles *produit_des* et *possède_un*.
En utilisant le constructeur de négation atomique nous pouvons alors définir le nouveau concept ($\neg$*Arbre*) qui décrira l'ensemble de tout les éléments de notre base de connaissance qui ne sont pas des arbres.
Nous pouvons définir un nouveau concept complexe comme le concept *Arbre fruitier* avec l'aide des constructeurs de quantification universelle et de conjonction. L'expression ($Arbre\ fruitier \equiv Arbre \sqcap \forall produit_des.Fruit$) transpose alors la phrase "Tous les arbres fruitiers sont des arbres qui produisent des fruits".

Avec la restriction de cardinalité $\mathcal{Q}$ nous pouvons exprimer le fait qu'un arbre ne peut posséder qu'un seul tronc avec les expressions (*Arbre* $\sqsubseteq \exists$*possède_un.Tronc*) et (*Arbre* $\sqsubseteq \leq 1$ *possède_un*).

5.4.3 Raisonnement avec les logiques de description.

La sémantique apportée à la description de la connaissance est fournie par des interprétations. Une interprétation $\mathcal{I}$ se compose d'un domaine d'interprétation $\triangle^{\mathcal{I}}$ qui consiste en un ensemble non vide d'individus d'une fonction d'interprétation $.^{\mathcal{I}}$. Cette fonction d'interprétation assigne chaque concept atomique $\mathcal{A}$ à un sous domaine de $\triangle^{\mathcal{I}}$ noté $\mathcal{A}^{\mathcal{I}}$ tel que:

$$\mathcal{A}^{\mathcal{I}} \subseteq \triangle^{\mathcal{I}}$$

Et chaque rôle atomique $\mathcal{R}$ est assigné une relation binaire telle que:

$$\mathcal{R}^{\mathcal{I}} \subseteq \triangle^{\mathcal{I}} \times \triangle^{\mathcal{I}}$$

Pour supporter la notion de concepts composés, la fonction d'interprétation $.^{\mathcal{I}}$ est étendue aux constructeurs comme le montre le tableau 5.3. La séparation entre les catégo-

Lettre	Constructeur	Sémantique
$\mathcal{AL}$	$\top$	$\triangle^{\mathcal{I}}$
$\mathcal{AL}$	$\bot$	$\emptyset$
$\mathcal{AL}$	$\neg A$	$\triangle^{\mathcal{I}} \setminus A^{\mathcal{I}}$
$\mathcal{AL}$	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
$\mathcal{AL}$	$\forall R.C$	$\{a \in \triangle^{\mathcal{I}} \mid \forall b.(a,b) \in R^{\mathcal{I}} \Rightarrow b \in C^{\mathcal{I}}\}$
$\mathcal{AL}$	$\exists R.\top$	$\{a \in \triangle^{\mathcal{I}} \mid \exists b.(a,b) \in R^{\mathcal{I}}\}$
$\mathcal{C}$	$\neg C$	$\triangle^{\mathcal{I}} \setminus C^{\mathcal{I}}$
$\mathcal{U}$	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
$\mathcal{E}$	$\exists R.C$	$\{a \in \triangle^{\mathcal{I}} \mid \exists b.(a,b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\}$
$\mathcal{R}^+$	R^+C	$U_{n\geq 1}(R^{\mathcal{I}})^n$
$\mathcal{F}$	$\leq 1R$	$\{a \in \triangle^{\mathcal{I}} \mid\mid \{b \mid\in R^{\mathcal{I}}\} \mid\leq 1\}$
$\mathcal{N}$	$\leq nR$	$\{a,b \in \triangle^{\mathcal{I}} \mid\mid \{(a,b) \in R^{\mathcal{I}}\} \mid\leq n\}$
	$\geq nR$	$\{a,b \in \triangle^{\mathcal{I}} \mid\mid \{(a,b) \in R^{\mathcal{I}}\} \mid\geq n\}$
	$= nR$	$\{a,b \in \triangle^{\mathcal{I}} \mid\mid \{(a,b) \in R^{\mathcal{I}}\} \mid= n\}$
$\mathcal{Q}$	$\leq nR.C$	$\{a,b \in \triangle^{\mathcal{I}} \mid\mid \{(a,b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\} \mid\leq n\}$
	$\geq nR.C$	$\{a,b \in \triangle^{\mathcal{I}} \mid\mid \{(a,b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\} \mid\geq n\}$
	$= nR.C$	$\{a,b \in \triangle^{\mathcal{I}} \mid\mid \{(a,b) \in R^{\mathcal{I}} \wedge b \in C^{\mathcal{I}}\} \mid= n\}$
$\mathcal{O}$	$\{a_1 \dots\ a_n\}$	$\{a_1^{\mathcal{I}} \dots\ a_n^{\mathcal{I}}\}$
$\mathcal{H}$	$R_1 \sqsubseteq R_2$	$R_1^{\mathcal{I}} \subseteq R_2^{\mathcal{I}}$
$\mathcal{R}$	$R_1 \sqcap R_2$	$R_1^{\mathcal{I}} \cap R_2^{\mathcal{I}}$
$\mathcal{I}$	R^{-1}	$\{(a,b) \in \triangle^{\mathcal{I}} \times \triangle^{\mathcal{I}} \mid (b,a) \in R^{\mathcal{I}}\}$

TABLE 5.3 – *Sémantique des constructeurs*

ries d'individus et les individus représentés par la TBox et la ABox offre une aptitude naturelle pour le raisonnement car si les interactions se font entres les individus, la majeure partie de notre raisonnement se fait sur des catégories d'individus. Un autre avantage des logiques de description pour le raisonnement est que les algorithmes d'inférence sont moins complexes que ceux de la logique du premier ordre [TSARKOV

et HORROCKS, 2003]. L'expressivité des logiques de description est proportionnelle aux nombres de constructeurs qu'elles fournissent, mais plus le nombre de constructeurs augmente plus la complexité des déductions pouvant être réalisées augmente de façon exponentielle [NARDI et BRACHMAN, 2003]. Avec les logiques de description il faut donc toujours être attentif au rapport entre l'expressivité et les capacités de déductions.

5.4.4 Les domaines concrets.

Avec les logiques de description, la connaissance devant être définie à un niveau logique abstrait, le traitement de domaines concrets comme par exemple les nombres entiers ou réels et les opérateurs sur ces domaines comme l'égalité ou d'autre opérateurs numériques ne pouvait être pris en compte. Pour pallier ce manquement, Franz Baader et Philipp Hanschke proposent une extension de la logique $\mathcal{ALC}$ avec un domaine concret $\mathcal{D}$ [BAADER et HANSCHKE, 1991]. Cette extension crée un second domaine d'interprétation $\triangle_{\mathcal{D}}^{\mathcal{I}}$ disjoint de $\triangle^{\mathcal{I}}$ et un nouveau type de rôle $\mathcal{U}$ défini comme une relation binaire sur $\triangle^{\mathcal{I}} \times \triangle_{\mathcal{D}}^{\mathcal{I}}$.
En reprenant notre exemple précédent, si nous voulions définir le fait qu'un arbuste est un arbre d'une hauteur de moins de 7 mètres, on pourra donc spécifier des assertions de rôle comme par exemple $u(hauteur, 650)$ où $u \in \mathcal{U}$ associant une valeur d'un élément du domaine concret (650) à un élément abstrait comme ici la hauteur d'un arbre.

En résumé :

Ces différentes propositions, réseaux sémantiques, langages à base des frames, graphes conceptuels et logiques de description ont pour but de fournir un moyen de représentation de la connaissance humaine. Elles font partie de la famille des systèmes à base de connaissance et leurs capacités de modélisation peuvent être exploitées pour construire des ontologies. Ces propositions et tout particulièrement les logiques de description inspireront les travaux réalisés dans le cadre du web sémantique [BAADER, CALVANESE et al., 2003 ; HORROCKS, PATEL-SCHNEIDER et VAN HARMELEN, 2003 ; ZOU et al., 2004] aboutissant entre autre à la création du langage OWL dont l'aspect formel repose sur les logiques de description.

VI

Le web et les principes du web sémantique

6.1 La naissance du web 70
6.2 Présentation du W3C 71
6.2.1 Les principes du W3C 71
6.2.2 Des outils et standards de qualité 71
6.3 Le web sémantique 72
6.4 Ressources, IRI, URI, URL et URN 72
6.5 Les langages de description 74
6.5.1 XML 74
6.5.2 RDF 75
6.5.3 RDFS 77
6.5.4 OWL 80
6.5.5 SKOS 83
6.6 SPARQL : un langage de requête pour les données du web sémantique 84
6.7 La pyramide du web sémantique 86

Pour mener à bien notre objectif de mettre en place une base de connaissance pour l'agroforesterie, nous avons fait le choix d'utiliser les outils du web sémantique mis à disposition par le World Wide Web Consortium. Dans ce chapitre nous présenterons une vue globale de l'histoire du web, de son évolution en web sémantique et des principaux outils qui ont été développés pour ce dernier. Les premiers réseaux informatiques émergent dans les années 1960, et dans la continuité de l'expansion des réseaux de télécommunications, à l'exemple de la téléphonie mondiale. Ces réseaux basés sur les architectures existantes de la téléphonie permettent pour la première fois, à des utilisateurs d'ordinateurs distincts et éloignés, de communiquer. La vision de J.C.R. Licklider [LICKLIDER et CLARK, 1962], d'une interaction homme machine accessible à tous, dans les domaines de l'enseignement, de l'apprentissage, de la planification et de la conception, se concrétisera en une trentaine d'années. En effet, à l'aube des années 90, le grand public possède des ordinateurs personnels équipés de modems lui permettant d'accéder au réseau internet au travers duquel il peut échanger des informations, comme il peut le faire notamment avec des applications de messagerie électronique.
Les premiers réseaux informatiques sont indépendants les uns des autres, et constitués d'un nombre limité d'ordinateurs partageant une même infrastructure technique. Les

échanges sont alors limités à de courtes distances à cause de ces contraintes matérielles. Internet, ou INTERnational NETwork, est le maillage de tous ces réseaux de natures diverses qui communiquent ensemble grâce au protocole de communication TCP/IP [1]. Ce protocole est développé par Robert Elliot Kahn et Vinton Gray Cerf, et inspiré des travaux de Louis Pouzin sur la commutation de paquets [2], une technique qui permet d'acheminer les informations sur des distances plus importantes. D'autres protocoles se surajoutent à TCP/IP, à l'exemple de FTP [3] pour copier, modifier ou supprimer des fichiers d'un ordinateur à l'autre, ou encore du protocole SMTP [4] pour l'envoi de message d'un ordinateur à l'autre.
C'est donc dans ce contexte, qu'une nouvelle exploitation du réseau Internet apparaît, créant une vraie révolution dans son usage: le web ou 'world wide web' qui désigne un système de documents interconnectés, aussi désigné par la "toile".

6.1 La naissance du web

Au début des années 1990, Tim Berners-Lee, chercheur au centre Européen pour la recherche nucléaire (CERN) se consacre à la problématique de la documentation foisonnante. En effet, le CERN a pour mission l'étude de la physique fondamentale et la découverte des constituants et des lois de l'Univers. Plusieurs milliers de scientifiques de plus d'une centaine de pays, y travaillent sur site, et/ou à partir de leur pays d'origine. La masse de documents générés par leurs recherches, résultats d'expériences, rapports scientifiques, notes de services ..., est à la mesure de l'entreprise. Tim Berners-Lee, pour faciliter l'accès aux documents, propose une solution basée sur un système hypertexte distribué. Le terme hypertexte nous vient du sociologue Ted Nelson en 1965, et est à la base d'un concept en rupture avec la lecture linéaire d'un texte. Un document hypertexte est un ensemble de documents liés entre eux par des hyperliens. Ces liens permettent au lecteur de naviguer à son gré, d'un document à l'autre qui peuvent être des textes, des notes ou des images. Le principe étant, pour Tim Berners-Lee, que chaque scientifique, où qu'il soit dans le monde, puisse accéder aux différentes informations d'un même projet, hébergées sur plusieurs machines. Par la suite, Tim Berners-Lee, et le chercheur Robert Cailliau, qui travaille lui aussi au CERN, mettent au point les différents protocoles, qui en fournissant les règles nécessaires au bon fonctionnement des communications sur internet, permettront l'émergence du 'word wide web'. Parmi ces protocoles, trois sont des plus essentiels pour le fonctionnement du web :

- Les URL (Uniform Resource Locator), communément appelées adresses web, correspondent aux hyperliens permettant d'accéder aux pages web.
- Le langage HTML (HyperText Markup Language) a pour rôle la création et structuration des pages web. Les pages web sont donc des documents hypertexte au travers desquels on peut naviguer grâce aux hyperliens.
- Le protocole de communication HTTP (HyperText Transfer Protocol) est le protocole client-serveur, permettant d'échanger les pages web d'une machine à l'autre via internet.

Les utilisateurs accèdent via un programme dédié, le navigateur, aux pages web qui sont des fichiers HTML stockés sur des serveurs informatiques, les serveurs HTTP. A

1. TCP/IP : (Transmission Control Protocol/Internet Protocol) Ensemble des protocoles de communication sur Internet basé sur la notion adressage IP.

2. commutation de paquets : technique utilisée pour le transfert de données dans les réseaux informatiques.

3. FTP : (File Transfer Protocol) Protocole de partage de fichiers entre deux ordinateurs .

4. SMTP : (Simple Mail Transfer Protocol) Protocole d'échange de messages entre deux ordinateurs .

partir de la première page web consultée, ils peuvent accéder à d'autres pages web, en suivant les liens qui les unissent, les URL. En 1993 le CERN qui n'a pas vocation à conduire de la recherche en informatique, transfère le projet au domaine public. Le web se développe alors de façon exponentielle, pour atteindre 1.8 milliard de sites web dans le monde en 2020 [*Total number of Websites - Internet Live Stats* 2020]. Pour assurer la pérennité du web, Tim Berners-Lee fonde le World Wide Web Consortium (W3C) en 1994 à son départ du CERN.

6.2 Présentation du W3C

Le World Wide Web Consortium est un organisme international à but non lucratif. Il élabore les standards du web pour assurer sa croissance sur le long terme. Il est géré au quotidien par une équipe d'une soixantaine de personne hébergées par quatre institutions, le MIT [5] au USA, l'ERCIM [6] en France, l'université de Keio [7] au Japon et l'université Beihang [8] en Chine. Il est dirigé par son président Tim Berners-Lee et le PDG Jeffrey Jaffe. Le W3C compte, en octobre 2020, 439 membres qui sont des entreprises publiques ou privées parmi lesquelles on note la présence des GAFAM [9]. Le W3C fait aussi appel au grand public par l'entremise de différents blogs, forums de discussion, groupes de travail et organisations d'événements internationaux. Le W3C est financé par la cotisation annuelle de ses membres, par des parrainages et des dons.

6.2.1 Les principes du W3C

Les principes du W3C sont envisagés comme des plus "altruistes" [*W3C Mission* 2017] :

- Un web pour tous. Il s'agit de permettre à tous, sans conditions de moyen matériel ou logiciel, de langage, de culture, de localisation géographique et d'aptitude physique d'accéder au web, pour communiquer, commercer et partager leurs connaissances.
- Un web partout (ubiquitaire). Il s'agit de rendre le web accessible sur tout type de dispositif, non seulement un ordinateur mais aussi un téléphone portable, ou encore des appareils ménagers.

6.2.2 Des outils et standards de qualité

Pour concrétiser sa vision de participation et de partage des connaissances à l'échelle mondiale, le W3C a créé différents groupes qui sont en ligne sur le web. Il s'agit de groupes de travail, de groupes d'intérêts et groupes communautaires et d'entreprises. Ils élaborent ensemble les standards nécessaires au bon fonctionnement du web et à son évolution. le W3C fournit une documentation très complète sur ces standards ainsi

5. MIT : Massachusetts Institute of Technology. http://web.mit.edu.
6. ERCIM : European Research Consortium for Informatics and Mathematics. https://www.ercim.eu.
7. Keio University : établissement privé d'enseignement supérieur. https://www.keio.ac.jp/ja.
8. Beihang University : Université d'aéronautique et d'astronautique de Pékin https://ev.buaa.edu.cn/.
9. Les géants du web : Google, Apple, Facebook, Amazon et Microsoft.

que des guides d'utilisation, tous accessibles gratuitement sur le web. La participation de membres issus du monde académique d'une part et du monde de l'industrie d'autre part, permet une meilleur adéquation entre ces outils et les besoins réels de l'industrie mondiale, ce qui contribue à une meilleurs diffusion de ceux-ci à travers le monde.

En résumé l'architecture du web repose sur des organismes indépendants comme le W3C, avec des standards ouverts et libres, ne faisant l'objet ni de brevet, ni de royalties.

6.3 Le web sémantique

En 2002, Tim Berners-Lee, James Hendler et Ora Lassila rédigent un article intitulé « A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities » [BERNERS-LEE et al., 2002]. Ils y décrivent une fiction, dans laquelle, une sœur et un frère, Lucy et Pete, doivent faire soigner leur mère malade. Dans cette mission, ils sont aidés par des agents logiciels qui communiquent avec eux, et entre eux, pour trouver des solutions adaptées aux contraintes que Lucy et Pete leur imposent. Ces agents logiciels utilisent leurs données personnelles, et des données publiques récoltées sur le web, pour répondre à différentes requêtes en terme de rendez vous médical, de transport de malade, et de choix d'un établissement de santé. Dans cet article, les auteurs révèlent une lacune importante du web, qui a été créé pour les humains, et non pour les machines. Effectivement les pages web sont conçues pour une lecture faite par des humains, et l'information qu'elles contiennent, est difficilement exploitable par les machines. Prenons l'exemple d'une page web, qui serait composée uniquement de la photo d'un grand chêne, avec écrit en dessous, le mot "Arbre" ; pour un être humain il ne fait nul doute que ce mot signifiera un arbre végétal, mais pour une machine, ce terme pourrait aussi bien signifier un arbre généalogique, ou un arbre binaire mathématique, qu'un arbre végétal car elle n'a pas accès au sens de la donnée. Les auteurs proposent une évolution du web vers le web sémantique, un web dans lequel les données mais aussi leurs sens, sont accessibles aux machines.

6.4 Ressources, IRI, URI, URL et URN

L'identification des ressources du web, qui peuvent être, par exemple, des images, des vidéos, des pages web disponibles à partir du réseau internet, a évolué depuis la naissance du web. Afin de normaliser l'identification et l'accès aux ressources de l'internet, l'IETF [10], un organisme de standardisation pour le fonctionnement d'internet, publie des "Request for comments" (RFC) à propos de ces identificateurs. Cette normalisation se fait en coopération avec le W3C. Lors de ce travail de normalisation, la notion même de ressource est modifiée, et devient *tout ce qui peut être identifié par un identifiant de ressource*. La notion de ressource alors n'est plus restreinte au monde de l'informatique, de l'internet, ou du web. Ce changement de paradigme pour la notion de ressource, est capital pour le web sémantique, car il confère à ces identificateurs, une notion d'universalité. En effet à partir de cette définition de la ressource, il est possible de dire qu'un identificateur de ressource peut nommer tout élément de notre univers. Une ressource peut donc être aussi bien une page web, une chose, une personne ou même une idée. Les identificateurs utilisés, pour nommer les ressources, sont:

10. IETF : Internet Engineering Task Force. www.ietf.org.

- L'URI (Uniform Resource Identifier) : la RFC3986 [IETF, 2020a] définit une URI, comme étant une chaîne de caractères de syntaxe restreinte, pouvant identifier une ressource physique ou abstraite. Elle précise qu'une URI identifie une ressource de manière intemporelle, qu'elle soit modifiée, déplacée ou supprimée. La tour Eiffel peinte en vert, déplacée à Montpellier ou démontée, restera ainsi la tour Eiffel. Les URI peuvent être de plusieurs types : les URL et les URN, ou un hybride des deux.
- L'URL (Uniform Resource Locator) (RFC3986): Les URL, sous-ensemble des URI, permettent non seulement d'identifier une ressource, mais en fournissent également les mécanismes d'accès. Par exemple l'URI "http://www.umontpellier.fr/formation/admission-et-inscription/" désigne une page web de l'université de Montpellier pour les inscriptions d'étudiants. "http:" nous indique que l'on va utiliser le protocole HTTP, "//www.umontpellier.fr" que nous accéderons au réseau informatique de l'Université de Montpellier pour atteindre la page "admission-et-inscription.html" via le chemin "/formation". D'autres protocoles peuvent être utilisés pour accéder à une ressource sur internet comme par exemple "ftp:" pour l'échange de fichiers avec le protocole FTP.
- L'URN (Uniform Resource Name) (RFC8141) [IETF, 2020c] : Les URN sont un autre sous-ensemble des URI, qui identifie une ressource par son nom au sein d'un espace de noms, sans fournir d'information sur sa localisation ou le moyen d'y accéder. La notion d'espace de noms se définit comme un ensemble de noms dans lequel ces noms, ont la propriété d'être uniques. Ainsi deux noms identiques, ne peuvent être confondus s'ils n'appartiennent pas au même espace de noms. La méthode d'accès est alors représentée par le motif "urn:" placé en préfixe de l'identificateur. Pour exemple la RFC8141 qui définit le standard des URN, sera identifiée par l'URN suivante : "urn:ietf:rfc:8141", rfc:8141 étant un nom unique, dans l'espace de noms nommé "ietf" de l'IETF. En revanche, pour accéder au document de la RFC8141 sous forme de page web sur internet, on utilisera l'URL "https://tools.ietf.org/html/rfc8141".
- L'IRI (Internationalized Resource Identifier): La RFC3987 [IETF, 2020b] internationalise les URI en permettant l'utilisation des alphabets de toutes les langues internationales en utilisant le jeu de caractère Unicode (ISO 10646). La norme précise qu'une IRI peut remplacer le cas échéant une URI pour identifier une ressource.

La figure 6.1 montre les relations entre ces différente normes d'identification de ressources.

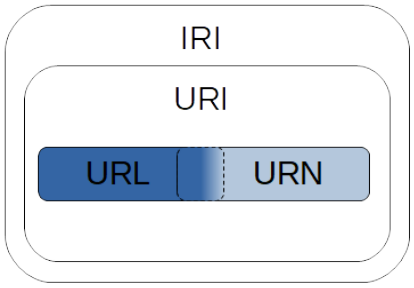


FIGURE 6.1 – *Relation entre IRI, URI, URL et URN : Les URI sont des cas particuliers des IRI. Les URI peuvent être des URL des URN ou bien les deux*

6.5 Les langages de description

6.5.1 XML

Le langage XML (Extensible Markup Language) [W3C, 2017], en français « langage de balisage extensible », est un sous-ensemble de SGML (Standard Generalized Markup Language).
Le SGML, en anglais « Standard Generalized Markup Language », est le fruit des travaux initiés par Charles Goldfarb en 1969, et qui aboutira en 1986 par la publication d'une norme ISO [ISO, 2020b]. Son intention première est de résoudre des problèmes d'interopérabilité, entre la création de documents et leur édition. Pour atteindre cet objectif, la solution retenue est de séparer la structure logique d'un document, comme par exemple son titre, ses chapitres et ses sous-chapitres, des différentes mises en page qui pourrait lui être appliquées en fonction du support d'édition choisi. En effet cette mise en page sera généralement différente si l'on cherche à visionner ce document sur un écran de téléphone mobile, ou sur un feuille A4 en mode paysage. Les différentes structures logiques du document sont identifiées par des balises placées au sein du document. Le nom des balises est délimité par des chevrons (<Nom_de_la_balise>).Ces balises seront interprétées par la suite au moment de l'édition par des programmes pour l'application de styles de mise en page comme par exemple une taille plus grande pour les caractères du titre d'un document. SGML est un métalangage qui décrit donc une grammaire, imposant sa syntaxe. Mais les contraintes syntaxiques mises à part, il n'y a pas de limitations imposées pour le nom des balises et tout un chacun peut donc créer son propre vocabulaire. SGML servira de base pour la création de l'HTML permettant l'édition de pages web. Malgré tout, SGML est considéré comme trop complexe pour être pleinement déployé sur le web, le W3C développe donc XML une version restreinte de SGML. Néanmoins, XML reste entièrement interopérable avec SGML et HTML. La figure 6.2 montre un exemple de code XML. On y observe des balises comme article, titre, chapitre titre-chapitre et paragraphe, identifiables par la présence de chevrons. Les balises sont en doublon avec une balise ouvrante (<...>) et une autre fermante (</...>), elles encadrent des portions de texte du document qui correspondent aux parties logiques de l'article. L'imbrication des balises rend compte de la structure de l'article, on visualise le fait que l'article est composé d'un titre et d'un chapitre qui contient lui-même un titre et un paragraphe. On notera que si l'utilisation des balises, est à la disposition des programmes qui traiteront la mise en page de cet article, ce document reste intelligible à tous.

```
<?xml version="1.0" encoding="UTF-8"?>
<article xmlns="http://monDomaine.fr/ns/monArticle">
  <titre>Mon article sur l'agroforesterie</titre>
  <chapitre>
    <titre-chapitre>Définition de l'agroforesterie</titre-chapitre>
    <paragraphe>L'agroforesterie consiste en la réintroduction...</
       paragraphe>
  </chapitre>
</article>
```

FIGURE 6.2 – *Exemple d'un code XML.*

6.5.2 RDF

Le langage RDF (Resource Description Framework) a été développé par le W3C en 1998 [LASSILA et SWICK, 1998]. C'est un modèle de données pour la description formelle des ressources et métadonnées du web. RDF est considéré de ce fait comme le langage de base du web sémantique. Le fondement du RDF est de décrire pour une ressource, une propriété nommée, et la valeur de celle-ci. Si l'on considère cette propriété, comme un attribut de la ressource, on peut faire le parallèle avec la notion de paire attribut-valeur utilisée communément dans le monde de la représentation des données. Si par contre l'on considère cette propriété comme la relation entre deux ressources, on peut alors voir le modèle RDF comme un diagramme entité-relation (ERD). Un document RDF est un ensemble d'unités atomiques, nommées Triplet ou Statement. Un Triplet RDF consiste en un triplet, composé d'un sujet, d'un prédicat et d'un object (figure 6.3). On peut voir le triplet RDF, comme une phrase simple, composé de son sujet, de son verbe et de son complément.

<Sujet> <Prédicat> <Objet>

FIGURE 6.3 – *structure d'un triplet RDF.*

- Le sujet est la ressource, sur laquelle s'applique la description
- Le prédicat est la propriété applicable à cette ressource
- L'objet est la valeur de la propriété, qui peut être une valeur littérale ou bien une autre ressource.

C'est à partir de cette structuration simple, qu'il est possible de bâtir des déclarations portant sur les ressources du web. Le triplet RDF peut être illustré sous la forme d'un diagramme étiqueté et orienté (figure 6.4). Sujet et objet sont alors représentés par deux nœuds distincts du graphe, et le prédicat qui est la relation entre le sujet et l'objet, est lui représenté par une arête les reliant. Cette relation a un sens, du sujet vers l'objet exprimé par la direction de l'arête.

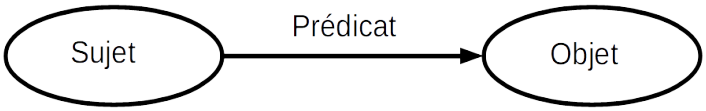


FIGURE 6.4 – *Triplet RDF sous la forme d'un graphe étiqueté et orienté.*

La figure 6.5 montre un exemple de description sous la forme d'un pseudo-code, à partir du texte suivant : Robert est un scientifique qui a pour ami Nicolas, qui est un agroforestier. Robert a pour sujet d'étude l'agroforesterie. L'agroforesterie est pratiquée au domaine de Restinclières. La vidéo intitulée 'Agroforesterie: cultiver entre les arbres' a pour sujet l'agroforesterie. La figure 6.6 traduit sous la forme d'un graphe, le pseudo-code précédemment présenté.

Pour passer de ce modèle abstrait sous forme de graphe, à un modèle concret pouvant être traité par des moyens informatiques, les URI sont utilisés. Les URI, comme nous l'avons vu dans la section précédente (6.4), permettent d'identifier toutes les choses de notre univers. Le RFC3986 stipule que « Likewise, abstract concepts can be

```
<Robert> <est un> <scientifique>
<Robert> <a pour ami> <Nicolas>
<Nicolas> <est un> <agroforestier>
<Robert> <a pour sujet d'étude> <l'agroforesterie>
<l'agroforesterie> <est pratiquée au> <domaine de Restinclières>
<la vidéo 'Agroforesterie: cultiver entre les arbres'> <a pour sujet> <l'agroforesterie>
```

FIGURE 6.5 – *Pseudo code décrivant Robert et l'agroforesterie.*

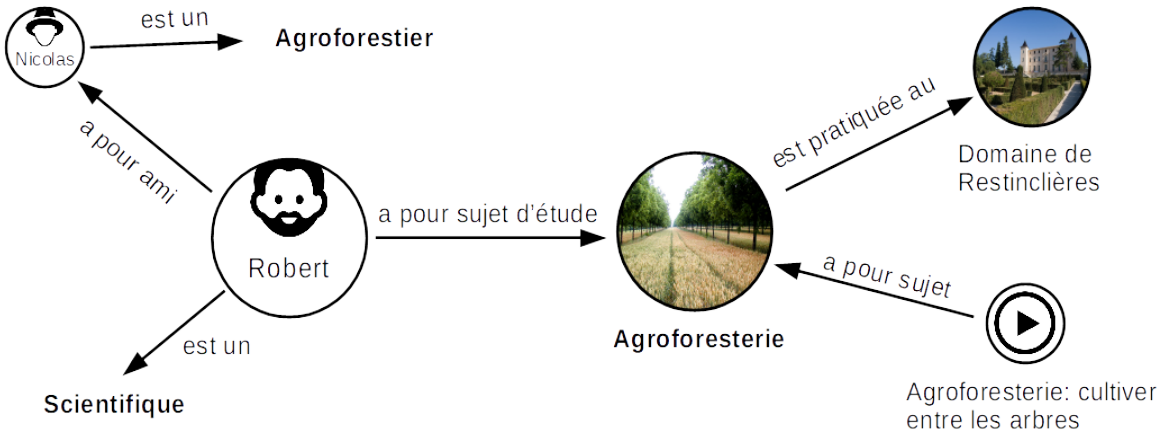


FIGURE 6.6 – *Graphe informel décrivant Robert et l'agroforesterie.*

resources, such as the operators and operands of a mathematical equation, the types of a relationship (e.g., "parent" or "employee"), or numeric values (e.g., zero, one, and infinity). » [IETF, 2020a] indiquant ainsi que les URI ne servent pas seulement à obtenir des ressources à partir d'internet mais identifient aussi des concepts abstraits. Ils sont assimilables aux mots de notre langage. Leur utilisation va permettre de définir sans ambiguïté les différents nœuds et arcs du triplet RDF en les nommant selon les règles suivantes :

- Le sujet est un URI ou un nœud anonyme.
- Le prédicat est toujours un URI.
- L'objet est un URI, un littéral ou un nœud anonyme.

Les nœuds anonymes sont des artéfacts utilisés, lors de l'implémentation du modèle abstrait de RDF. Ils correspondent, en particulier, à l'expression de relations N-aires (figure 6.7), qui ne sont pas prises en compte dans le modèle abstrait du RDF dont les relations sont uniquement binaires. Les littéraux sont des chaînes de caractères, des nombres ou des dates.

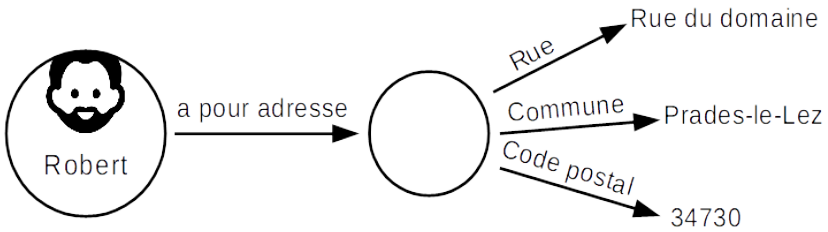


FIGURE 6.7 – *Graphe informel décrivant l'adresse de Robert sous forme de relation N-aire.*

En exploitant des URI déjà connus, comme l'URI "http://www.wikidata.org/entity/Q397350", qui pointe sur le concept d'Agroforesterie utilisé par Wikipédia, ou l'URI "https://youtu.be/GxBvUi5W0R4", qui permet d'accéder à la vidéo "Agroforesterie: cultiver entre les arbres" sur Youtube, notre exemple précédent pourra être retranscrit de cette autre manière visible en figure 6.8.

```
<http://example.org/Robert> <http://example.org/est_un>
   <http://example.org/scientifique>
<http://example.org/Robert> <http://example.org/a_pour_ami>
   <http://example.org/Nicolas>
<http://example.org/Nicolas> <http://example.org/est _un>
   <http://example.org/agroforestier>
<http://example.org/Robert> <http://example.org/a_pour_sujet_d_etude
   >
   <http://www.wikidata.org/entity/Q397350>
<http://www.wikidata.org/entity/Q397350>
   <http://example.org/est\_pratiquee\_au>
      <https://restinclieres.herault.fr/>
<https://youtu.be/GxBvUi5W0R4> <http://example.org/a_pour_sujet>
   <http://www.wikidata.org/entity/Q397350>
```

FIGURE 6.8 – *Pseudo code décrivant Robert et l'agroforesterie avec l'utilisation d'URI.*

Le RDF est donc un modèle de données qui décrit la manière dont sont représentées les données de façon abstraite (pseudo code, graphe). La publication d'un graphe RDF sur le web nécessitera sa transposition sous forme de fichier texte structuré, exploitable au sein des ordinateurs. Cette sérialisation peut être réalisée avec différentes syntaxes. La sérialisation RDF/XML proposée par le W3C à l'origine de la création du RDF est souvent confondue avec ce dernier, mais elle n'en est qu'une syntaxe possible. Par la suite, d'autre syntaxes sont apparues comme la syntaxe N3 (simplification de la lecture par rapport au RDF/XML), Turtle (extension de N-Triples, sous-ensemble de N3) ou JSON-LD (basé sur le format JSON). Les schémas suivants montrent des extraits de sérialisation de notre exemple en XML/RDF (figure 6.9) et Turtle (figure 6.10)

La grande généricité de la structure du langage RDF, permet son utilisation pour la constitution de schémas et de vocabulaires, dans des domaines applicatifs spécialisés. Les langages RDFS et OWL décrits dans les sections suivantes, sont d'ailleurs publiés sous forme de documents RDF.

6.5.3 RDFS

RDFS (RDF Schema) est une extension du vocabulaire RDF, il fournit un vocabulaire de modélisation des données pour les données RDF. Ce vocabulaire va permettre de renforcer l'interprétation des ressources présentes sur le web. Les deux notions les plus importantes qu'il permet d'exprimer, sont d'une part la notion de classe pour les

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 <rdf:Description rdf:about="http://example.org/Robert">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
  <foaf:knows rdf:resource="http://example.org/Nicolas"/>
  <foaf:topic_interest rdf:resource="http://www.wikidata.org/entity/
     Q397350">
 </rdf:Description>
 <rdf:Description rdf:about="http://www.wikidata.org/entity/Q397350"
    >
  <rdf:label>Agroforesterie</rdf:label>
 </rdf:Description>
 <rdf:Description rdf:about="https://youtu.be/GxBvUi5W0R4">
  <dcterms:subject rdf:resource="http://www.wikidata.org/entity/
     Q397350">
 </rdf:Description>
</rdf:RDF>
```

FIGURE 6.9 – *Document RDF sérialisé en RDF/XML.*

sujets et objets des triplets RDF, et d'autre part celle de l'ensemble de définition des prédicats.

La notion de classe : Certaines ressources ont des relations étroites entre elles, formant ainsi des sous-ensembles appelés classes. Par exemple une pomme, une banane et une tomate peuvent être regroupées en une classe fruit. Les ressources faisant partie d'une même classe, sont appelées instance de cette classe. Les classes sont elles-mêmes des ressources identifiables par des URI, et donc a fortiori peuvent être partie constituante d'un triplet RDF. Le prédicat *rdf:type* peut être utilisé pour signifier qu'une instance appartient à une classe comme dans l'exemple suivant :

```
<#Robert> rdf:type <#Personne>.
```

Qui stipule que Robert fait partie de la classe des personnes. L'ensemble des classes de RDFS est la classe *rdfs:Class*. Une classe peut être elle-même un sous-ensemble d'une autre classe, le prédicat *rdfs:subClassOf* permettra de l'indiquer:

```
<#Personne> rdfs:subClassOf <#Humain>.
```

Toute instance de la classe personne, fera alors partie de la classe des humains. Robert qui est une personne fait donc aussi partie de la classe des humains.

L'ensemble de définition des prédicats : Les prédicats RDF font partie de la classe *rdf:Property* qui est elle-même une instance de *rdfs:Class*. Le prédicat *rdfs:subPropertyOf*

```
@base <http://example.org/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<#Robert>
  a foaf:Person ;
  foaf:knows <#Nicolas> ;
  foaf:topic_interest <http://www.wikidata.org/entity/Q397350>.
<http://www.wikidata.org/entity/Q397350>
  rdf:label "Agroforesterie", "Agroforestry"@en.
<https://youtu.be/GxBvUi5W0R4>
  dcterms:subject <http://www.wikidata.org/entity/Q397350>.
```

FIGURE 6.10 – *Document RDF sérialisé en Turttle.*

indique qu'une propriété est la sous-propriété d'une autre propriété ce qui implique qu'un sujet et qu'un objet mis en relation par cette propriété sont aussi mis en relation par cette autre propriété. RDFS enrichit la notion de prédicat RDF en définissant pour les prédicats un ensemble de départ et un ensemble d'arrivée. La propriété *rdfs:domain* indique que tout sujet d'un prédicat donné est une instance d'une ou plusieurs classes. La propriété *rdfs:range* indique elle, que tout objet d'un prédicat donné, est une instance d'une ou plusieurs classes. Dans l'exemple suivant :

```
ex:pratiqueLaTechniqueAgroforestiere rdfs:domain <#Agroforestier>
ex:pratiqueLaTechniqueAgroforestiere rdfs:range <#
   TechniqueAgroforestiere>.
```

les deux triplets indiquent que si le prédicat *ex:pratiqueLaTechniqueAgroforestiere* est utilisé dans un triplet, il aura pour sujet une instance de la classe des agroforestiers et pour objet une instance de la classe des techniques agroforestières. Comme dans l'exemple suivant où Nicolas, fait partie de la classe des agroforestiers, et l'agrosylviculture fait partie des techniques agroforestières.

```
<#Nicolas> ex:pratiqueLaTechniqueAgroforestiere <#AgroSylviculture>
```

RDFS fournit donc des ressources pour déterminer les caractéristiques d'autres ressources comme l'appartenance à une classe, ou les domaines de départ et d'arrivée d'un prédicat. Bien que semblant similaire au concept de la programmation orientée objet dans ses notions de classes et de propriétés, on observe une différence notable. En programmation objet, une classe est définie avec ses attributs (ses propriétés) généralement typés et ce de façon immuable. RDFS, en revanche, décrit ses propriétés en fonction des classes de ressources auxquelles elles s'appliquent, ce qui permet de redéfinir facilement et au besoin de nouvelles propriétés sans avoir à modifier ses classes. Cette approche étend indéfiniment la description de ressources existantes, ce qui est un des principes du web sémantique. Si Nicolas décrit un arbre avec les propriétés de hauteur et de circonférence, Robert peut à postériori rajouter la propriété localisation qui indiquera les coordonnées GPS de cet arbre sans pour autant modifier les descriptions faites par Nicolas.

La recommandation du W3C [*RDF Schema 1.1* 2017] propose d'autres termes RDF en relation avec les notions de classe et de propriétés ainsi que pour des notions non décrites dans ce chapitre de collections, de listes et de réification des triplets RDF. Cette recommandation n'a pas pour vocation d'être exhaustive dans la description de ces notions, elle envisage que d'autres techniques peuvent exister pour renforcer l'expressivité dans la description des ressources du web. Parmi ces techniques proposées, nous noterons le langage OWL (Ontology Web Language), qui est un langage d'ontologie.

6.5.4 OWL

OWL (Ontology Web Language) est un langage de représentation des connaissances. Il a pour but de formuler et d'organiser le savoir humain pour son utilisation et son partage. Le résultat de cette modélisation du savoir est une ontologie. Son usage est préconisé pour des applications dédiées aux traitements de l'information contenue dans les ressources du web.

OWL est une extension de RDFS et utilise la syntaxe RDF pour sa grammaire, de ce fait les ontologies OWL peuvent s'exprimer sous forme de document RDF. On peut donc représenter une ontologies OWL sous la forme d'une structure abstraite ou sous la forme d'un graphe RDF. OWL se distingue néanmoins de RDF et RDFS pour deux raisons. La première est que OWL est un langage dédié à la réalisation d'ontologies, il bénéficie d'une plus grande capacité à exprimer le savoir humain que RDF et RDFS en disposant d'un vocabulaire plus étendu pour exprimer les caractéristiques des propriétés et pour définir des classes dont les individus partagent ces caractéristiques. On pourrait citer comme exemple que s'il est possible avec RDFS de définir une classe *Arbre* dont les individus sont liés par la propriété *a pour espèce* avec des individus d'une classe Espèce taxonomique, il n'est pas possible avec RDFS de limiter l'occurrence de cette relation pour signifier qu'un arbre ne puisse se voir attribuer qu'une seule espèce. La deuxième distinction et que OWL présente une plus grand capacité à fournir des connaissances interprétables par les machines en s'appuyant sur un couche sémantique basée sur une logique du premier ordre.

Nous décrivons ici la dernière version de ce langage nommé "OWL 2" [GRAU et al., 2008] publié en 2012 [*OWL 2 Web Ontology Language Document Overview (Second Edition)* 2017] par le "OWL Working Group", un groupe de travail du W3C. On peut se représenter les classes OWL d'un point de vue ensembliste comme le montre la figure 6.11 ou des ensembles se chevauchent ou contiennent d'autres ensembles. Dans cette figure l'ellipse bleue à gauche représente la classe des *Choses endémiques de Métropole* et l'ellipse vert clair à droite, la classe des *Arbres*. Leur intersection décrit la classe des individus qui sont des arbres et qui sont endémiques de Métropole, c'est-à-dire la classe des *Arbres endémiques de Métropole*. Au sein de la classe des *Arbres* se trouve un sous-ensemble qui est celui représentant la classe des *Arbres utilisés en agroforesterie*. L'intersection de ces trois classe représentera au final l'ensemble des arbres qui sont endémiques de Métropole et qui sont utilisés en agroforesterie c'est-à-dire la classe des *Arbres endémiques de Métropole utilisés en agroforesterie*.
Les ensembles que nous venons de symboliser sont appelés des axiomes, Une Tbox est un ensemble d'axiomes. Dans cet exemple la première ellipse qui représente les *Choses endémiques de Métropole* peut s'exprimer comme l'ensemble de tous les individus liés par la propriété *endémiques de* avec la valeur "Métropole" que l'on peut considérer

comme étant une instance d'une classe *Localisation géographique*. Cette propriété *endémiques de* peut être utilisée avec des arbres, des buissons, des animaux ou des maladies, ainsi on peut définir avec OWL une classe de toutes les choses qui sont endémiques du territoire métropolitain. Cette définition pourra donc permettre, à postériori, de classer n'importe quel individu nouvellement introduit dans la base de connaissance qui répondrait à ce critère. Cette définition d'une classe ou les instances doivent respecter une condition sur une de leur propriété s'appelle une restriction. On voit aussi que l'on peut prendre l'intersection des deux classes *Arbres endémiques de Métropole* et *Arbres utilisés en agroforesterie* pour former la classe *Arbres endémiques de Métropole utilisés en agroforesterie*. Cela est permis par OWL qui fournit dans son vocabulaire des opérateurs logiques pour l'intersection, l'union et la négation de classes. Cet exemple montre un les atouts principaux de OWL qui permet de définir des classes à partir d'expressions dans lesquelles on retrouve des combinaisons de restrictions et de classes. Les résultats de ces expressions peuvent alors être réutilisés dans d'autres expressions. Ces nouvelles classes portent alors leur définition au sein même de leur structure ce qui va permettre à des applications d'y accéder. Au travers des outils qui l'exploitent, OWL permet de

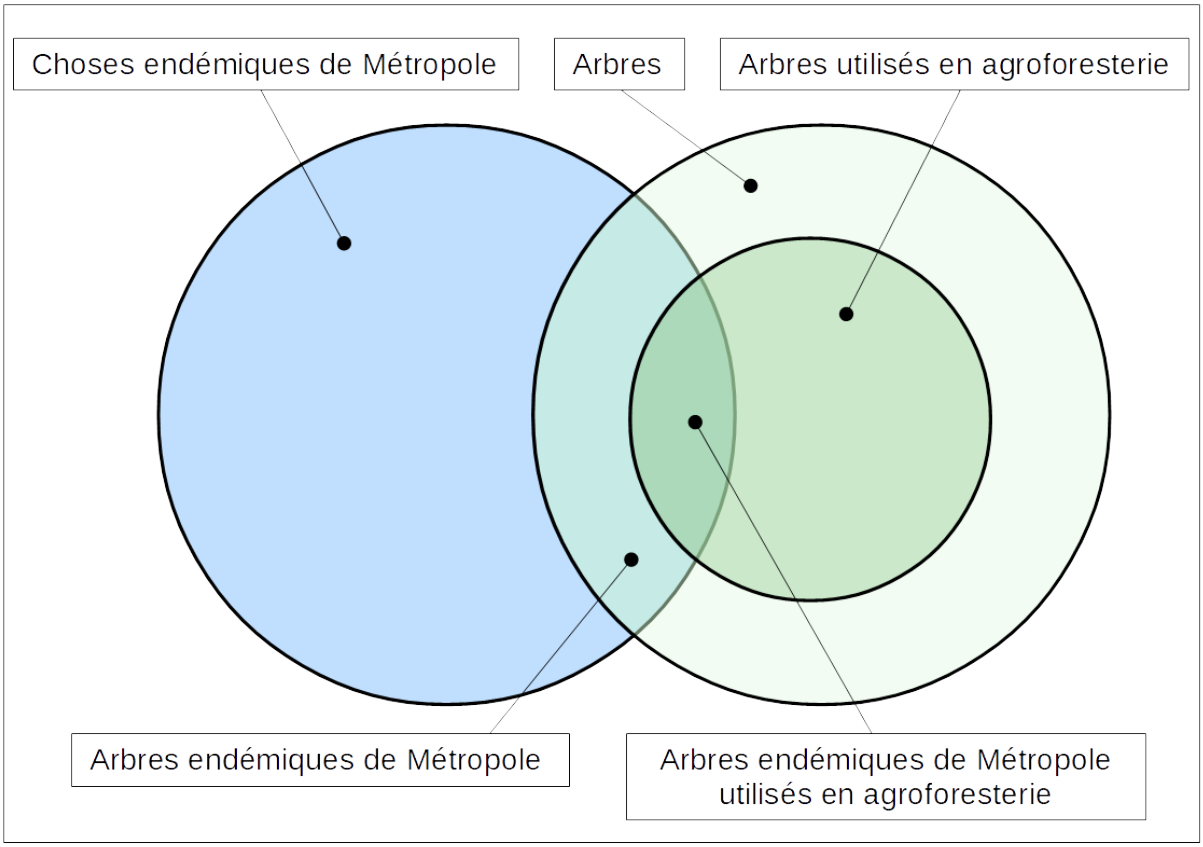


FIGURE 6.11 – *Vision ensembliste des classes OWL.*

vérifier la cohérence des données, de déduire des connaissances nouvelles et de transformer une information implicite en une information explicite. Il est important de comprendre qu'OWL n'est pas un langage de programmation mais OWL décrit une réalité factuelle avec un formalisme logique. C'est cette description sous un formalisme logique qui sera exploitée par des programmes appelés des "moteurs d'inférences" pour permettre de générer des faits nouveaux sur la base de faits déclarés au préalable. Pour acquérir cette signification formelle de ses structures OWL utilise deux sémantiques: la "Sémantique directe" [*OWL 2 Web Ontology Language Direct Semantics (Second Edition)* 2017] et la "Sémantique basée sur le RDF" [*OWL 2 Web Ontology Language RDF-Based Semantics (Second Edition)* 2017]. Comme nous l'avons vu précédemment les ontolo-

gies OWL peuvent être exprimées avec la syntaxe de RDF et donc être représentées sous forme de graphes RDF. La "Sémantique basée sur le RDF" attribue directement une signification aux graphe RDF et donc indirectement aux structures d'une ontologie OWL. Cette sémantique est cent pour cent compatible avec la sémantique de RDF ce qui a pour effet que toutes ontologie OWL interprété avec cette "Sémantique basée sur le RDF" est interprétable avec la sémantique du RDF, on parle alors d'ontologies "OWL 2 Full". La "Sémantique directe" s'appuie sur le principe des logiques de description (voir section 5.4 du chapitre 5) et s'applique directement sur les structures des ontologies OWL. Ce rapprochement avec les logiques de description permet aux créateurs d'outils pour OWL, de bénéficier des nombreuses avancées et des expériences d'implémentations faites dans ce domaine. Cette compatibilité a néanmoins un prix et certaines restrictions syntaxiques doivent être opérées sur les structures d'OWL. Pour être compatible avec la logique de description SROIQ qui offre un bon compromis entre expressivité et performance de calcul on utilise un sous-ensemble des spécification structurelles de OWL nommé "OWL 2 DL" Les logiques de description permettant de représenter la connaissance d'une manière formelle et structurée c'est-à-dire manipulable mathématiquement selon une logique du premier ordre alors un ensemble de propositions considérées comme vraies implique une conclusion vraie. Par exemple si l'on considère les propositions suivantes:
Tous les arbres ont un tronc
le noyer est un arbre
comme vraies, alors la conclusion:
Le noyer a un tronc
est vraie. Donc si l'on fournit au sein d'un document OWL des propositions considérées comme vraies alors une application informatique pourra déduire des conclusions qui seront elles-mêmes considérées comme vraies et qui fourniront des réponses aux questionnements inhérents à la constitution du modèle.

Pour obtenir certaines réponses, il est parfois nécessaires de limiter encore plus l'expressivité d'OWL pour obtenir des facilitations en termes de calcul ou de mise en œuvre. Pour cela OWL proposent des "profils". Ces profils sont au nombre de trois: "OWL 2 EL", "OWL 2 QL", et "OWL 2 RL", ce sont des restrictions syntaxique de OWL qui sont chacune plus restrictif que "OWL 2 DL".
Voici les différents cas d'utilisation de ces profils:

- OWL 2 EL s'appuie sur le principe de la logique de description $\mathcal{EL}^{++}$ [BAADER, BRANDT et al., 2005] ce qui permet aux applications de toujours répondre en un temps raisonnable. Il est utilisé pour des applications utilisant de très grandes ontologies orientées vers l'expression de classe quand l'expressivité peut être limitée au profit des performances. Il est utilisé dans les domaines où l'on manipule des objets complexes comme dans le cas de la recherche en biologie.
- OWL 2 QL convient aux applications manipulant de petites ontologies légères organisant une multitudes d'individus et utilisant des requêtes relationnelles avec des langages comme SQL pour accéder aux données.
- OWL 2 RL est conseillé pour des applications où l'on cherche à préserver un maximum d'expressivité toute en gardant des temps de calcul raisonnable. Cela est réalisé en définissant un sous-ensemble syntaxique de OWL 2 qui peut être mis en œuvre en utilisant des technologies basées sur des règles. Lorsqu'on utilise OWL 2 RL, une implémentation basée sur des règles peut fonctionner directement sur des triplets RDF ce qui rend ce profil idéal lorsque l'on doit remanier des données par des règles supplémentaires.

Ces trois différents profils d'OWL permettent de s'adapter aux besoins des applications informatiques les plus répandues dans le domaine de l'ingénierie des connaissances en offrant différents compromis entre expressivité et capacité de calcul ou des facilités d'intégration avec d'autres systèmes comme les systèmes de gestion de base de données relationnelle, mais d'autres profils peuvent être créés pour s'adapter à d'autres situations particulières.

OWL est donc un outil qui facilite le développement et le partage d'ontologies sur le web et sa sémantique qui repose sur les logiques de descriptions permet un meilleur accès à la connaissance via des machines. Son développement fait suite à d'autres langages de création et de manipulation d'ontologies comme KIF [*Knowledge Interchange Format (KIF)* 2021], LOOM [*Loom* 2007] ou DAML (DARPA Agent Markup Language) [*DAML.org* 2006]. Nous utiliserons OWL pour mettre en place une base de connaissance en aménagement agroforestier en exploitant ses capacités à exprimer des objets complexes comme:
"La catégorie des arbres dont la hauteur est de plus de 10 mètres qui sont endémiques de la zone USDA "Climat méditerranéen franc" et utilisés en agroforesterie." et ses capacités computationnelles qui permettrons de réaliser des inférences afin de compléter les descriptions de nos aménagements.

6.5.5 SKOS

Simple Knowledge Organization System (SKOS) a été créé pour transposer des systèmes d'organisation de la connaissance (SOC) tels que les taxonomies et les thésaurus vers le web sémantique. Il fournit un modèle pour conceptualiser la structure de base et le contenu de vocabulaires contrôlés, en s'appuyant sur les différentes normes ISO déjà existantes pour ces vocabulaires comme ISO 25964-2:2013 [ISO, 2020a]. Étant une extension du modèle RDF, SKOS permet de publier et de partager ces vocabulaires sur le web.

L'élément fondamental de SKOS n'est pas le terme linguistique utilisé dans les taxonomies mais le concept sous-jacent à ce terme. Ce concept existe comme une entité abstraite indépendamment des termes utilisés pour le nommer. Sa description en tant que concept dans un triplet RDF se fera en le plaçant en tant que sujet de ce triplet qui aura pour prédicat *rdf:type* et comme objet *skos:concept*. Le concept sera alors nommé grâce à l'association *skos:prefLabel*. Les concepts peuvent être rassemblés au sein d'une entité désignant le SOC dont ils sont issus. Cette entité sera du type *skos:conceptScheme* et les concepts lui seront rattachés par la relation *skos:inScheme*.

Les concept sont liés entre eux par des relations sémantiques hiérarchisées, indiquant que l'un est un concept plus large *skos:broader* ou au contraire, plus restreint *skos:narrower* que l'autre, ou non hiérarchisées, indiquant une simple association de concepts *skos:related*.

Du fait de sa spécialisation SKOS est l'outil idéal pour exposer des vocabulaires contrôlés sur le web, sa structure plus légère que celle d'OWL permet de mettre en place plus facilement et plus rapidement des ontologies terminologiques nouvellement

créées ou issues de taxonomies ou de thésaurus déjà existants. Au cours de ce travail de thèse il sera exploité pour mettre en place une ontologie terminologique pour l'agroforesterie.

6.6 SPARQL : un langage de requête pour les données du web sémantique

SPARQL (SPARQL Protocole And RDF Query Language) [*SPARQL 1.1 Overview* 2018] permet de requêter des données RDF. C'est un ensemble de spécifications permettant d'interroger et de manipuler le contenu d'un graphe RDF ou d'une base de données RDF au travers de langages et de protocoles. Son mode de fonctionnement consiste à exprimer des motifs de graphe dans lesquels les données recherchées sont exprimées sous forme de variables. L'unité atomique de ces motifs est la structure Sujet-Prédicat-Objet du triplet RDF. SPARQL va ainsi permettre d'extraire des sous-ensembles de graphes RDF en exprimant des contraintes sous forme d'équations où les variables sont les éléments recherchés parmi les sujets, prédicats ou objets des triplets RDF qui composent le graphe.

Pour illustrer ce fonctionnement considérons le graphe RDF suivant (voir figure 6.12). Si l'on cherche à extraire les ressources liées à la ressource "Parcelle A3" avec le prédicat "contient" et plus particulièrement les libellés lié à ces ressources par le prédicat "a pour espèce" on procèdera ainsi:

- Les ressources liées à "Parcelle A3" par le prédicat "contient" sont:

```
(<Parcelle A3> <contient> ?ressources)
```

- Le libellé lié à ces ressources par le prédicat "a pour espèce" sont:

```
(?ressources <a pour espèce> ?libellé)
```

Les variables recherchées (à lier à des valeurs) "?ressources" et "?libellé" sont indiquées par le point d'interrogation qui les précédent. En rapprochant ces deux équations on obtiendra avec la syntaxe de SPARQL :

```
SELECT ?ressources ?libellé
WHERE {
<Parcelle A3> <contient> ?ressources.
?ressources <a pour espèce> ?libellé
}
```

Le résultat obtenu sera :

ressources	libellé
<Culture>	Triticum durum
<Arbre 1>	Juglans regia
<Arbre 2>	Alnus glutinosa

Comme on le voit, cette syntaxe est proche du langage SQL, utilisée pour interroger les bases de données relationnelles, dont il est fortement inspiré. Leur fonctionnalités sont elles aussi très proches puisque SPARQL peut rechercher, ajouter, modifier ou supprimer des données RDF comme le fait SQL avec les données d'une base relationnelle.

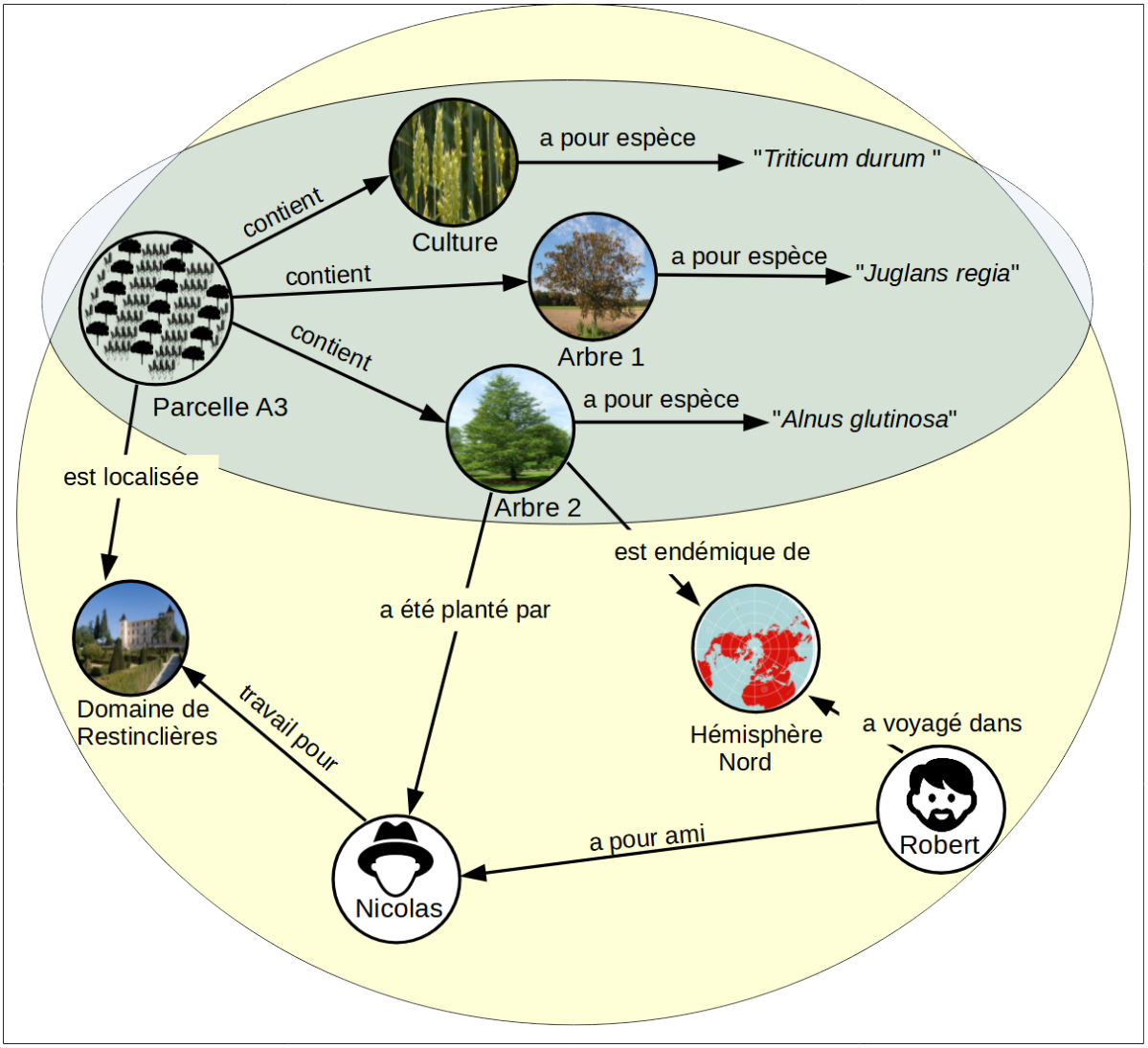


FIGURE 6.12 – *Un exemple de graphe RDF.*

Des services SPARQL ("SPARQL endpoints") permettent d'accéder aux données du web. Ce sont des point d'accès fonctionnant sous le protocole HTTP permettant de traiter des requêtes SPARQL. Ces données font généralement parties des données ouverte et liées du web ("Linked Open Data") comme c'est le cas avec DBpedia [*DBpedia* 2021] qui expose les données de Wikipédia [*Wikipédia, l'encyclopédie libre* 2021] sous format RDF. Un des avantages de ces terminaux de services SPARQL par rapport aux services web classique c'est qu'il n'est pas nécessaire de connaitre au préalable la description du service pour obtenir des données particulières puisque des requêtes SPARQL peuvent être construites directement sur les données.

6.7 La pyramide du web sémantique

La pile du web sémantique ou « Semantic Web Stack » [11] est un schéma qui résume les différentes technologies et leur contribution pour un bon fonctionnement du web sémantique. La figure 6.13 s'inspire de ce schéma, et montre l'organisation des technologies déjà opérationnelles comme celles qui ont été décrites dans les chapitres précédents ou qui doivent être développées. L'empilement de bas en haut traduit la dépendance des techniques les une par rapport aux autres. La figure 6.14 reprend le schéma précédent en rassemblant ces technologies par blocs fonctionnels.

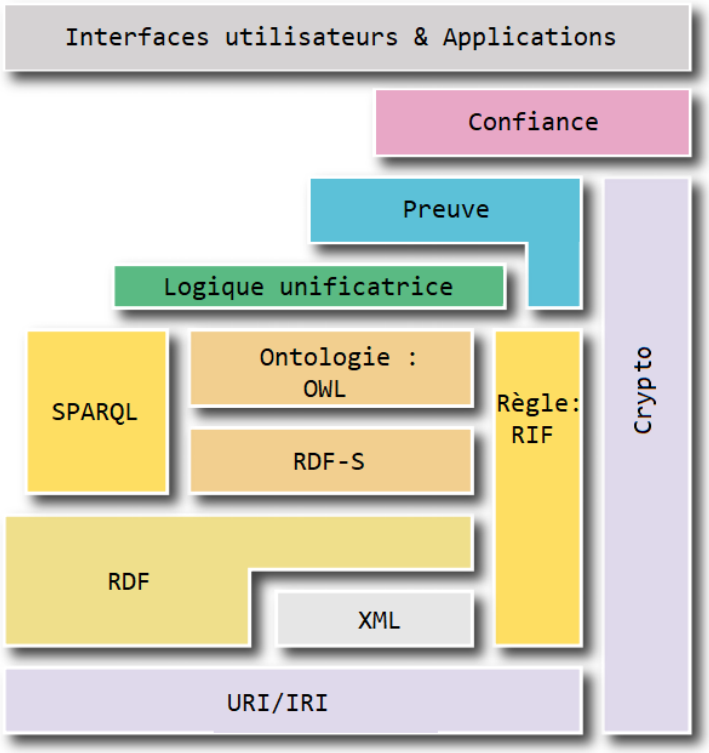


FIGURE 6.13 – *La pyramide du web sémantique.*

On peut donc résumer ainsi les grandes fonctionnalités attendues du web sémantique :

– Description :

  Pour la description des ressources, à la fois les URI, et les langages XML et RDF sont mobilisés pour fournir un socle minimal à la représentation des données et ainsi faciliter une large interopérabilité entre systèmes se partageant les ressources.

– Raisonnement :

  RDFS, OWL, la famille des logiques terminologiques et les règles sont utilisés pour le raisonnement au travers de mécanismes d'inférences réalisés par des agents logiciels spécialisés, désignés par moteurs d'inférence. Le *Rule Interchange Format* (RIF) est un standard pour l'échange de règles entre les différents moteurs d'inférences.

11. Semantic Web Stack : Terme et schéma de l'architecture du web sémantique proposé par Tim Berner-Lee lors d'une conférence [BERNERS-LEE, 2000].

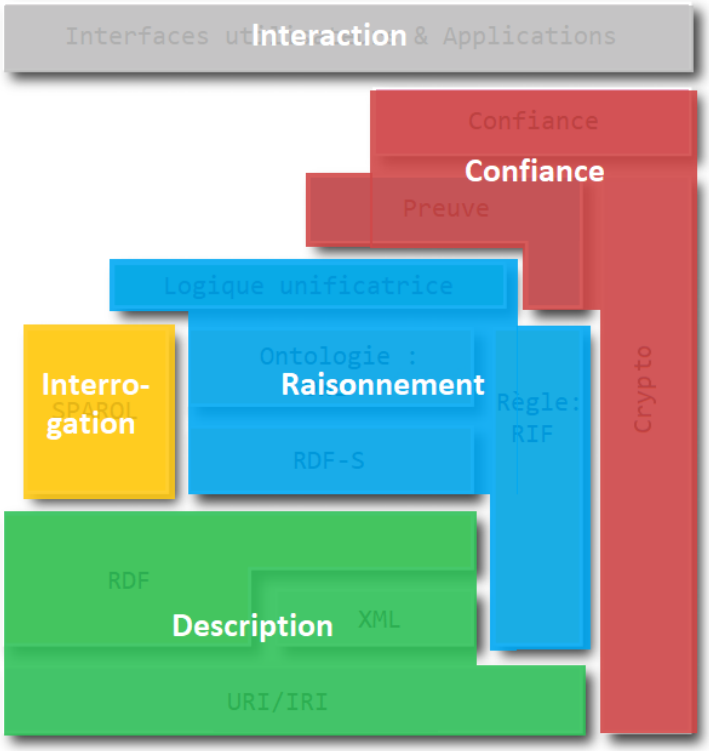


FIGURE 6.14 – *Blocs fonctionnels de la pyramide du web sémantique.*

– Interrogation :

  SPARQL est l'équivalent du langage SQL pour les bases de données relationnelles. Spécialisé dans le traitement des triplets RDF, il permet d'interroger, d'ajouter, de modifier ou de supprimer des triplets contenus dans des triplestores qui sont des bases de données spécialisées dans le stockage de graphes RDF.

– Confiance :

  Le réseau internet n'impose aucun contrôle sur la qualité des données qui y circulent. Les informations contenues dans les pages web sont laissées à la responsabilité de ceux qui les créent et certains dérapages se produisent comme pour exemple le phénomène des *fake news*. Le web sémantique doit donc être capable de s'assurer de la bonne qualité des données qu'il véhicule. Dans ce but, des techniques pour la cryptographie, l'apport de preuve de véracité et d'indices de confiance sur les données collectées doivent être développées.

– Interaction :

  Par le biais d'interfaces utilisateur et d'applications informatiques, la connaissance contenue au sein des graphes RDF est accessible via le web. Certains langages de programmation possèdent des librairies spécialisées pour la traitement de graphes RDF, comme la librairie JENA développée en langage JAVA et qui permet la manipulation de langages tel que RDFS, OWL ou SPARQL. Des programmes permettent de créer et de maintenir des ontologies comme le logiciel Protégé de l'université de Stanford [RESEARCH, STANFORD CENTER FOR BIOMEDICAL INFORMATICS, 2021].

Si certains blocs fonctionnels n'ont pas encore été totalement développés comme pour la cryptographie, la preuve et la confiance, les autres fonctionnalités attendues du web sémantique sont maintenant matures et sont utilisées dans de nombreux domaines.

En Résumé : le langage RDF permet donc de décrire les entités du monde qui nous entoure de manière simple avec un formalisme qui permet aux machines d'accéder à cette description tout en restant lisible pour l'être humain. La sémantique d'un document RDF peut être exprimée en théorie des ensembles ce qui confère au RDF la généricité et l'universalité de la notion d'ensemble. De plus, cette sémantique peut être aussi exprimée en théorie des modèles moyennant certaines contraintes ce qui lui permet d'être traduite en formule de logique du premier ordre ouvrant ainsi des possibilités de traitements complexes par des outils informatiques. Parmi ces traitements complexes on trouve la capacité de réaliser des inférences permettant de créer des données nouvelles sur la base de faits déjà connus.
Pour reprendre la métaphore linguistique de Bruno Menon à propos du web de données, « il s'agit bien d'un système d'expression et de communication. Les URI/IRI en sont les mots et les ontologies conçues en RDF Schema ou en OWL sont les dictionnaires qui permettent d'en élucider la sémantique. RDF, qui régit la construction des assertions, en est la syntaxe. Et XML est l'un des systèmes d'écriture avec lesquels cette langue peut être notée. Les jeux de données RDF sont les textes ou les ouvrages de référence rédigés dans cette langue, que SPARQL permet de consulter et de compiler. La strate des règles d'inférence et celle des interfaces et applications pourraient tenir lieu d'une pragmatique et compléter ainsi l'image d'ensemble. » [MENON, 2016]
Les structures en charge du développement et de l'évolution des outils du web sémantique semblent structurellement pérennes et efficaces et reposent sur l'intervention de nombreux acteurs. Elles apportent de plus une documentation très complète.
Les outils du web sémantique semblent donc idéaux pour la constitution d'une base de connaissance. Nous les avons utilisés tout au long de ce travail de thèse afin de réaliser notre objectif de mettre en place un système de gestion de données efficace et pérenne dans le cadre de l'agroforesterie.

Introduction aux contributions de la thèse

Les trois chapitres suivants (chapitres 7, 8, 9) décrivent dans le détail l'ensemble des travaux réalisés pendant la thèse. Les deux premiers (chapitres 7 et 8), sont entièrement consacrés à la présentation de la démarche ontologique qui a été retenue pour la gestion des données et connaissances en agroforesterie. Le troisième chapitre (chapitre 9) est plus orienté vers des considérations technologiques, et la mise en place d'une application permettant de tirer parti du modèle de connaissances. Il s'agit en effet d'inciter les acteurs de l'agroforesterie à exploiter une telle démarche sémantique basée sur la mobilisation de l'existant.

Plusieurs activités sont détaillées successivement tout au long des ces trois contributions :

- l'analyse de l'existant qui porte à la fois sur l'analyse du vocabulaire de spécialité, et sur l'appropriation de données issues d'un jeu de données de l'agroforesterie pris en exemple que nous exposerons au chapitre 7
- la modélisation conceptuelle et les processus de dérivation/formalisation qui ont permis d'aboutir à une ontologie de domaine pour les aménagements agroforestiers que nous présenterons au chapitre 8
- le développement d'une brique logicielle offrant à la fois un cadre opérationnel à l'ontologie et permettant d'évaluer la démarche que nous détaillerons au chapitre 9

Le chapitre 9 veut également promouvoir les approches sémantiques et souligner l'importance de la mouvance des sources de données ouvertes et liées auprès des acteurs de l'agroforesterie.

VII

Première contribution : un référentiel terminologique dédié

7.1 Introduction à la modélisation des connaissances liées à l'agroforesterie 91
7.2 Les concepts et termes de l'agroforesterie 92
7.3 Acquisition des jeux de données d'Agroof 93
7.4 Les données issues des expérimentations du domaine de Restinclières 97
7.5 Concepts obtenus à partir d'un travail d'extraction de textes 98
7.6 Utilisation de référentiels terminologiques existants 100
7.6.1 La classification des systèmes agroforestiers selon P.K.R. Nair . 100
7.6.2 La classification des systèmes agroforestiers selon E. Torquebiau 100
7.6.3 Le thésaurus de l'Agence Française d'AgroForesterie 102
7.6.4 Les vocabulaires du web sémantique 103
7.7 Une ontologie terminologique pour l'agroforesterie 104
7.8 Démarche retenue pour la construction d'un référentiel terminologique dédié 104
7.8.1 Description générale 104
7.8.2 Résultats 109
7.8.3 Justifications concernant la construction d'un référentiel 116
7.8.4 Extraits du référentiel en cours de construction 118
7.8.5 Discussion autour de la version actuelle 122
7.8.6 Les éléments d'importance retenus pour l'élaboration d'un cadre général de modélisation des connaissances 123

7.1 Introduction à la modélisation des connaissances liées à l'agroforesterie

La section 4.4 du chapitre 4 a fait état de différents types d'ontologies. Nous nous intéresserons principalement aux ontologies terminologiques (ou ressources onto-terminologiques) et aux ontologies de domaine qui seront au centre des activités de modélisation.

Nous les considérons en effet comme des entités complémentaires. La construction d'une ontologie terminologique dédiée à l'agroforesterie est envisagée en premier lieu, pour définir et organiser l'ensemble des concepts, et des termes qui s'y rattachent. Le besoin d'un vocabulaire de spécialité pour l'agroforesterie a d'ores et déjà été à l'origine de différents travaux [NAIR, 1985; TORQUEBIAU, 2000; HYVÖNEN et al., 2008; BURRIEL et al., 2017] pour identifier et préciser les notions d'importance au cœur de l'agroforesterie, et en partager les significations. Une ontologie de domaine peut ensuite s'adosser au vocabulaire de spécialité (voir chapitre 8) pour promouvoir la construction de bases de connaissance à la fois dédiées à des contextes d'étude spécifiques, mais également exploitables à d'autres fins. Les interfaces et les applications présentées ensuite au chapitre 9 vont en conséquence mettre à contribution l'ensemble des composants ontologiques construits au premier abord.

Ce chapitre décrit étape après étape, les moyens mis en œuvre pour identifier dans un premier temps, puis structurer au sein d'un référentiel dans un second temps, les termes et concepts clés de l'agroforesterie.

7.2 Les concepts et termes de l'agroforesterie

Nous avons procédé avec méthode, et avons initié un travail de veille bibliographique traitant de la classification des systèmes agroforestiers du début des années 80 à nos jours. Ce travail a été complété par la compilation de sources de données relatives à la description de parcelles agroforestières.

Les concepts qui se révèlent pertinents dans un contexte spécifique lié à un domaine d'expertise, peuvent être identifiés de plusieurs façons :

- Analyse de l'univers du discours :
  Lorsque des agroforestiers parlent de leur travail, ils utilisent des termes précis qui, bien souvent, désignent les concepts en relation avec leur métier. Ainsi, par exemple, les termes de "ligne d'arbre" et de "bande enherbée" désignent des concepts qui sont propres à l'agroforesterie. Durant mon travail de thèse, je suis resté très attentif à ces termes spécifiques entendus lors de présentations et discussions avec des agroforestiers à l'occasion de colloques, séminaires ou lors de simple visites.

- Analyse de documents de travail:
  Les divers documents édités par les agroforestiers au cours de leur activité, à l'exemple de feuilles tabulées, contiennent des données sur les concepts cibles. Au travers des différents documents qu'ils nous ont partagés, nous avons extrait des concepts en relation avec les expériences menées, comme par exemple la notion de "diamètre d'arbre à hauteur de poitrine".

- Analyse par fouille de textes :
  Dans le cadre de la construction et du peuplement d'ontologies à partir de textes, P. Ciamino [CIMIANO et al., 2006] considère l'extraction de la terminologie d'un domaine spécialisé, et l'identification des synonymes comme des activités essentielles. À cet effet, il préconise l'usage de logiciels de fouille de texte performants et capables d'extraire les mots les plus spécifiques d'un corpus spécialisé.

- Veille bibliographique autour des classifications en agroforesterie :
  Différents travaux visant à décrire l'essence même des systèmes agroforestiers ont

été conduits depuis l'introduction du néologisme agroforesterie en 1977 (d'après P.K.R. Nair). Ces travaux offrent différentes perceptions sur la manière de structurer et de définir les concepts terminologiques liés à l'agroforesterie, et se révèlent des "matières à penser" pour la définition d'un référentiel dédié.

– Portails de référentiels :
Les ontologies et référentiels terminologiques sont légions dans les sciences du vivant et de l'environnement. Il est donc difficile de tous les connaître. Les portails vont dans se sens héberger les ontologies associées à de vastes champs disciplinaires, et proposer toute une série de services autour de la manipulation de ces ontologies et de leurs concepts. Il est ainsi rendu possible de naviguer visuellement dans l'arborescence de concepts d'une ontologie, mais également de rechercher tous les concepts toutes ontologies confondues associés à un terme d'intérêt. Parmi les portails dédiés aux sciences de l'environnement et du vivant, nous pouvons citer : AgroPortal [1] [Clement JONQUET et al., 2018] (agriculture, agronomie), BioPortal [NOY et al., 2009] (sciences du vivant et de la santé) et EcoPortal [FIORE et al., 2017] (écologie, biodiversité) qui sont tous trois adossés à la technologie OntoPortal Appliance ; ainsi qu'OLS [2] [CÔTÉ et al., 2006] (sciences du vivant et de la santé). Nous exploiterons en priorité AgroPortal en raison des champs thématiques couverts, et de ses fonctionnalités autour de l'aide à la découverte de concepts au sein de 163 ontologies.

7.3 Acquisition des jeux de données d'Agroof

L'objectif est de tirer parti de l'expertise acquise par la Société Coopérative et Participative Agroof qui est impliquée dans la mise en pratique de différents systèmes agroforestiers, en particulier dans le sud de la France, et dans différents projets de recherche participatifs (en agroforesterie maraîchère par exemple).
Le bureau d'études Agroof a, de manière intéressante, constitué diverses banques de données portant sur différents sites agricoles situés en France à partir de ses activités de promotions, de formations et d'accompagnements aux agroforestiers. Ces sources de données se présentent bien souvent sous forme de classeurs contenant des feuilles tabulées (voir Fig.7.1) et fournissent des renseignements divers sur la nature des sites et des différents aménagements et travaux agroforestiers qui y sont pratiqués. Il est à remarquer que l'organisation des données est en lien étroit avec l'aménagement des sites agroforestiers. Ainsi l'organisation topologique interne aux sites prédomine. Dans cet ordre, un site contient des parcelles (parcelles cadastrales), qui sont elles-mêmes composées d'aménagements (voir Fig.7.2) qui sont amenés à varier selon le système visé.
La figure 7.2 illustre une organisation spatiale fréquemment retrouvée en Europe relevant de l'agroforesterie en disposition linéaire. Un exemple d'aménagement illustre un alignement continu (sans espace) de type haie brise-vent. Un autre alignement est constitué d'alignements discontinus (lignes d'arbres plantés à intervalle régulier), de bandes enherbées (privilégiant la biodiversité) et tournières (espaces laissés libres pour le retournement des engins mécaniques). L'aménagement est donc l'unité spatiale élémentaire qui va être à la base de tout aménagement agroforestier. Un aménagement peut par exemple être constitué d'une simple haie séparant deux parcelles, d'arbres distribués de façon éparse comme dans le cadre d'un pâturage ombragé ou bien d'un parcours de volailles, ou d'arbres en ligne comme dans le cas de l'agrisylviculture. Dans ce dernier type d'aménagement, les agroforestiers plantent habituellement des

1. AgroPortal : http://agroportal.lirmm.fr/.
2. OLS : The Ontology Lookup Service https://www.ebi.ac.uk/ols/index.

SITE	AMENAGEMENT								
NOM	NOM	BLOCS	Année de début	Localisation de l'aménagement	Nature de l'aménagement	Largeur de la bande	Densité	Distance sur la ligne	Distance entre les lignes/bandes
Berneuil	AMENAGEMENT 1	BLOC 1	2008	Intraparcellaire	Forestiers haut jet Intraparcellaires	2	50	7	26
St Jouin De Marnes	AMENAGEMENT 1	BLOC 1	2006	Intraparcellaire	Forestiers haut jet Intraparcellaires	3	35	10	27
Challignac	AMENAGEMENT 1	BLOC 1	2008	Intraparcellaire	Forestiers haut jet Intraparcellaires	3	75	5	27
Vézénobres	AMENAGEMENT 1	BLOC 1	1995	Intraparcellaire	Forestiers haut jet Intraparcellaires	1 ;5	100	10	10
Saint Aulais la Chapelle	AMENAGEMENT 1	BLOC 1	2007	Intraparcellaire	Forestiers haut jet Intraparcellaires	5	65	6	25
Chateaudun	AMENAGEMENT 1	BLOC 1	2008	Intraparcellaire	Forestiers haut jet Intraparcellaires	2	34	10	26
Bonnes	AMENAGEMENT 1	BLOC 1	2008	Intraparcellaire	Forestiers haut jet Intraparcellaires	3	48	7	30
Parcé sur Sarthe	AMENAGEMENT 1	BLOC 1	2007	Intraparcellaire	Forestiers haut jet Intraparcellaires	3	30	10	30
Parcé sur Sarthe	AMENAGEMENT 1	BLOC 2	2007	Intraparcellaire	Forestiers haut jet Intraparcellaires	3	300	3	3

FIGURE 7.1 – *Extrait d'un fichier tabulé fourni par le bureau d'études Agroof.*

lignes d'arbres placées parallèlement et séparées par des cultures intercalaires (de céréales ou de légumineuses). Sur ces lignes, aux abords et entre les arbres, l'exploitation mécanique des cultures est rendue difficile ce qui implique que la ligne soit généralement inutilisée pour la culture principale. La ligne est alors laissée à l'état naturel, sur une certaine largeur, ou semée de spécimens pouvant apporter un service écosystémique, et forme ainsi ce qui est appelé une bande enherbée.

Dans les sources de données en notre possession, nous allons donc retrouver par exemple :

- Au niveau du site:
  - Nom du site
  - Commentaires
  - ...
- Au niveau de la parcelle:
  - Nom de la parcelle
  - Nom Prénom Propriétaire
  - Nom Prénom Exploitant
  - Nom Prénom Contact Recherche
  - Surface
  - Type de système agricole
  - Coordonnées GPS
  - ...
- Au niveau du bloc:
  - Nom du bloc
  - Vocation expérimentale
  - Éléments paysagers environnants
  - Surface
  - Coordonnées GPS
  - Topographie
  - Classe de PH
  - Humidité
  - Irrigation
  - Production(s) végétale(s) annuelle(s)
  - Production(s) végétale(s) pérenne(s)
  - Production(s) animale(s)

    - Traitement(s) phytosanitaires(s)
    - ...
- Pour les aménagements:
    - Nom aménagement
    - Bloc(s) concerné(s)
    - Année de début
    - Localisation de l'aménagement
    - Nature de l'aménagement
    - Objectifs initiaux
    - Essences
    - Conduite des arbres
    - Année de plantation des arbres
    - Largeur de la bande
    - Distance sur la ligne
    - Distance entre les lignes/bandes
    - Densité
    - Protection
    - Paillage
    - ...

L'étude des données collectées a permis de dresser un inventaire des concepts mobilisés dans la description d'aménagements agroforestiers, et ce pour une entreprise spécialisée dans la conception de ce type d'aménagement. L'organisation spatiale des sites est au centre des démarches, et illustre de l'importance des relations topologiques qui se nouent entre les éléments du système. Nous voyons apparaître notamment les notions de lignes d'arbres, de bandes enherbées et d'allées de culture qui viennent en support de la notion d'aménagement dans un système agro-sylvicole.

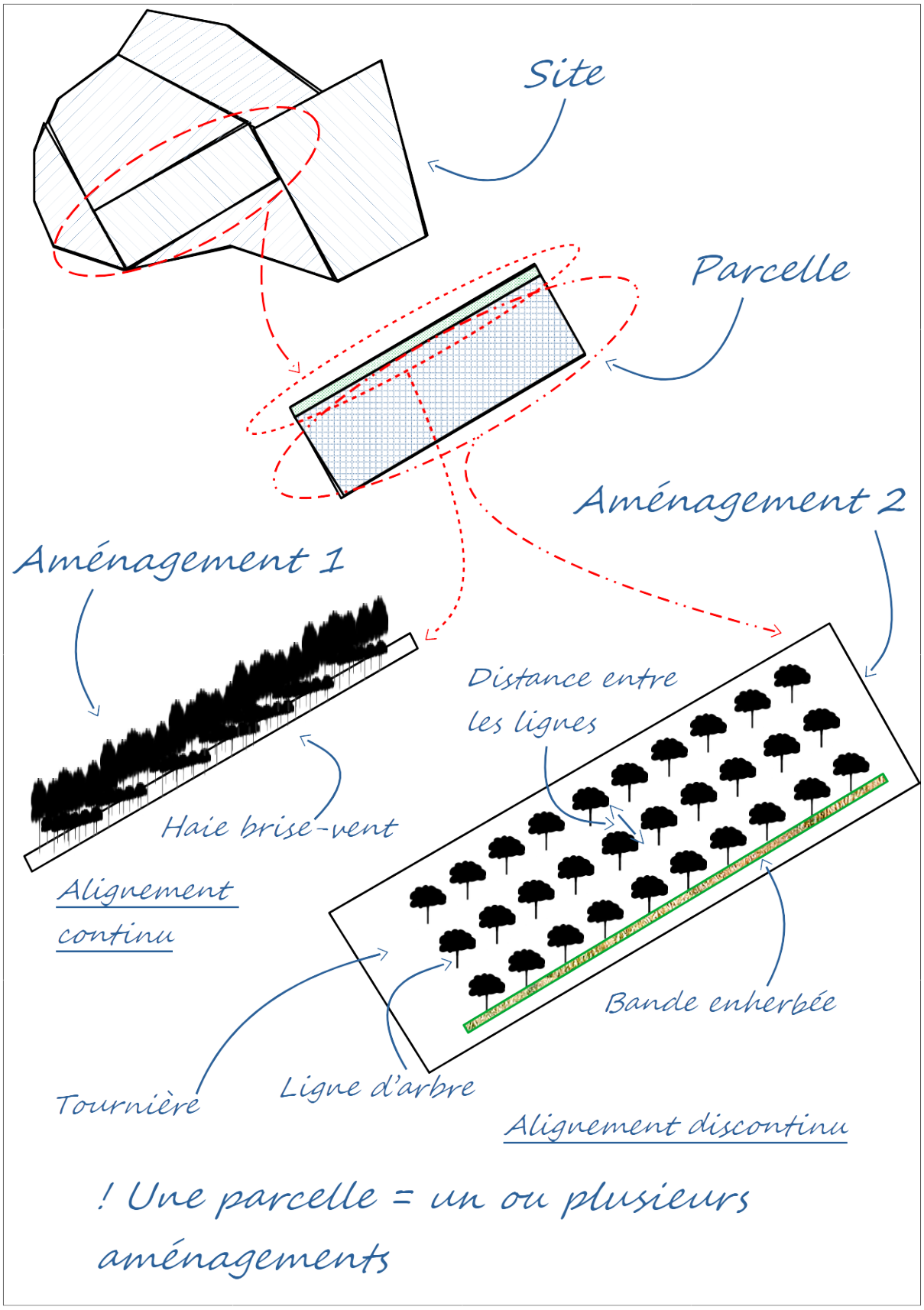


FIGURE 7.2 – *Relations d'inclusion topologique et exemples d'aménagement*

7.4 Les données issues des expérimentations du domaine de Restinclières

Le domaine de Restinclières est un site expérimental dédié à la pratique de l'agroforesterie et en particulier de l'agrisylviculture, depuis plus de vingt ans maintenant, et a fait l'objet de nombreux travaux de recherche sur les intérêts comparés de l'agroforesterie. Une description détaillée de ce domaine est proposée à la section 9.1.1. La littérature scientifique relativement abondante [DUPRAZ, LECOMTE et al., 2002 ; FABIEN LIAGRE et al., 2004] et les jeux de données associés au domaine de Restinclières nous ont permis de disposer d'un cas d'étude riche et documenté à même de nous permettre d'évaluer, et de valider l'approche de modélisation symbolique mise en place. Nous avons pu ainsi travailler sur des mesures faites sur le terrain qui se révèlent être des mesures habituellement exploitées dans le contexte de l'agroforesterie, mais aussi de l'agriculture et de la foresterie de manière séparée. Les relations de complémentarité entre arbres et cultures y ont été particulièrement étudiées avec des objectifs de préservation de la biodiversité et dans un contexte de développement durable, sans remettre en question pour autant les rendements de production. Nous donnons ici un rapide aperçu de ces mesures qui sont ici surtout associées au rendement de production et à la qualité du sol :

- Pour la foresterie:
  - Le diamètre d'un arbre à hauteur de poitrine
  - La hauteur totale d'un arbre
  - La hauteur d'élagage d'un arbre (hauteur du tronc)
  - La date de plantation d'un arbre
  - ...
- Pour les cultures agricoles:
  - poids total des grains/m²
  - Nombre de grains/m²
  - Nombre d'épis/m²
  - Nombre de grains/épi
  - poids gousses vides sèches (g)
  - rendement (qx/ha)
  - ...
- Pour la pédologie:
  - pH
  - Teneur en carbone organique
  - Teneur en Azote total
  - Teneur en Azote minéral
  - Teneur en eau
  - ...
- Pour les conditions météorologiques:
  - Pluviométrie
  - Rayonnement global
  - Température de couvert

- Température de sol
- ...

– ...

Il est à noter que nous retrouverons bon nombre de ces notions au sein de référentiels dédiés aux sciences de l'environnement et du vivant.

7.5 Concepts obtenus à partir d'un travail d'extraction de textes

L'outil Gargantext [*https://gargantext.org* 2021] de l'ICF-PIF [3], est un logiciel conçu pour l'exploration interactive de grands corpus numériques, et nous a permis d'extraire les termes liés à un corpus numérique de l'agroforesterie voulu exhaustif. La capture d'écran (voir Fig.7.3) fait état des termes les plus fréquemment retrouvés dans le corpus sur un total de 5573 termes. Le nombre d'occurrences de chaque terme est également affiché. La figure suivante (voir Fig.7.4) montre que l'on peut aussi disposer des locutions composées à partir du terme "soil", et présentes dans le corpus. Ces locutions sont également accompagnées de leur nombre d'occurrences. Ainsi nous retrouvons la locution "soil conservation" ("conservation du sol") qui fait référence à une fonction de type "rôle de protection" dans les travaux de P.K.R. Nair [NAIR, 1985].



FIGURE 7.3 – *Gargantext : interface listant les résultats de la recherche des mots les plus fréquents extraits d'un corpus dédié à l'agroforesterie*

Le nuage de mots présenté en figure 7.5 est constitué des soixante premiers termes simples extraits par Gargantext, à partir d'un corpus de mille articles en relation avec l'agroforesterie obtenus auprès de 183 sources d'édition numérique de 1979 à 2016. Ces mots dont la taille dans le nuage est représentative de leurs occurrences dans le corpus sont susceptibles d'aider à la constitution d'une ontologie terminologique de l'agroforesterie.

3. ICF-PIF : Institut des systèmes complexes de Paris Île-de-France. https://iscpif.fr/.



FIGURE 7.4 – *Gargantext : interface illustrant la recherche du terme sol ("soil") et des locutions comprenant ce terme dans le corpus dédié à l'agroforesterie*



FIGURE 7.5 – *Nuage de mots relatifs à l'agroforesterie extraits par l'application Gargantext.*

7.6 Utilisation de référentiels terminologiques existants

7.6.1 La classification des systèmes agroforestiers selon P.K.R. Nair

Le centre international pour la recherche en agroforesterie (ICRAF[4]) a initié de 1982 à 1987 un vaste travail d'inventaire des systèmes agroforestiers pour mieux en appréhender les pratiques, intérêts et limites. De ce travail a découlé une réflexion sur les classifications possibles de ces systèmes. Les critères, à même de permettre de définir une ossature propice à une classification, sont au nombre de quatre : la structure interne du système lui-même (en d'autres termes, les arrangements spatio-temporels de ses composants), les fonctions jouées par le système (en lien avec les produits en sortie du système), la nature socio-économique du système, et encore la dimension écologique et environnementale du système. P.K.R. Nair [NAIR, 1985] propose une synthèse sur les éléments clés des systèmes agroforestiers au regard de ces quatre grands critères. Il distingue par exemple l'agrisylviculture (arbres/arbustes et cultures), du sylvopastoralisme (arbres et pâtures/animaux) ou encore de l'agro-sylvopastoralisme (arbres, cultures et pâtures/animaux) sur la base des composants présents dans le système ciblé (voir figure 7.6). Il va aussi typer le caractère dense ou éparse des arrangements spatiaux, ou encore la présence simultanée ou séquentielle de différents composants dans le système (aménagements temporels). Pour ce qui concerne les aspects fonctionnels de production, deux aspects sont considérés, avec d'une part la production du système, et d'autre part sa durabilité. Parmi les exemples de production, nous retrouvons les notions de "nourriture", "bois de chauffage", "bois d'ornement" et "fourrage". La durabilité se traduit souvent par des rôles visant à préserver l'environnement, à l'exemple de "conservation de l'humidité", "conservation du sol" ou encore "restauration du sol". Les aspects socio-économiques sont également le reflet de deux considérations avec d'une part les entrées du système qui peuvent se révéler coûteuses (coût financier, impact écologique) pour l'agroforestier avec des concepts de gestion des intrants ("fort niveau d'intrants" ou "faible niveau d'intrants") et d'autre part des objectifs socio-économiques très hauts niveaux ("subsistance" , "commercial" ou intermédiaire"). Le dernier critère aborde les aspects d'adaptabilité à l'environnement. Ainsi certains systèmes agro-forestiers seront particulièrement adaptés à des "terres arides" et "semi-arides" ou bien des "terres tropicales".
Les différents systèmes agroforestiers peuvent donc être classés sur la base d'un ou plusieurs de ces critères.

7.6.2 La classification des systèmes agroforestiers selon E. Torquebiau

Une classification radicalement différente est proposée par E. Torquebiau dans [TORQUEBIAU, 2000] sur la base d'une relecture de la définition de l'agroforesterie. L'agroforesterie est alors envisagée comme la mise en valeur du sol, au travers d'associations, entre arbres et agriculture, sur une même surface à même d'offrir produits et services utiles

4. ICRAF : World Agroforestry. https://worldagroforestry.org/.

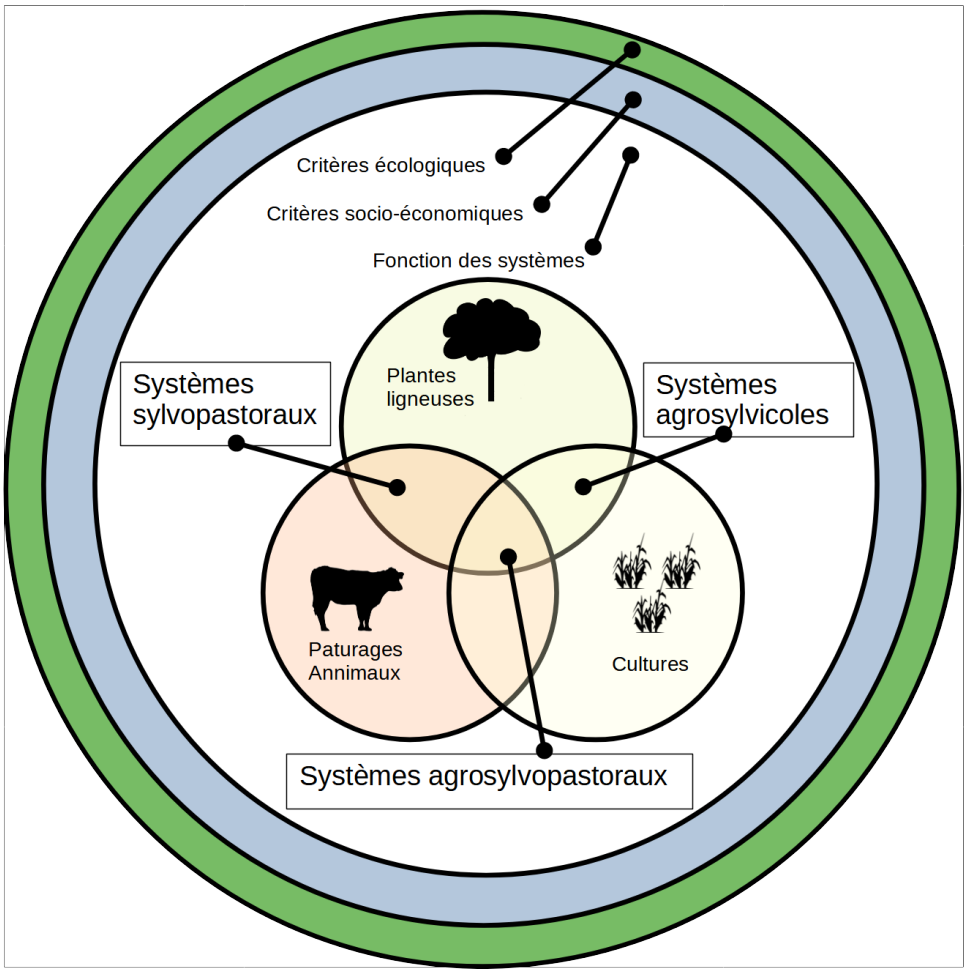


FIGURE 7.6 – *Classification des systèmes agroforestiers selon P.K.R. Nair [*NAIR*, 1985].*

à l'homme. Les systèmes agroforestiers sont alors plutôt classés sur la nature de ces associations, et en six catégories fondées sur des critères physionomiques aisément observables sur le terrain, et à même de façonner le paysage.

Nous listons ici ces six catégories qui viennent se substituer aux trois techniques agroforestières habituelles (à savoir agrisylviculture, sylvopastoralisme, et agro-sylvopastoralisme), et qui sont le reflet des pratiques liées à l'agroforesterie :

1. culture sous couvert arboré
   Il peut s'agir d'arbres dispersés dans des parcelles agricoles, d'arbres d'ombrage associés à des cultures telles que le thé ou le cacao, ou encore de vergers associés à des cultures ;
2. agroforêt
   Les arbres constituent dans ce contexte, un ensemble dense, multi-étagé, et associé à des cultures en sous-bois ;
3. agroforesterie en disposition linéaire
   Les techniques qui exploitent des alignements d'arbres et/ou d'arbustes (haie, brise-vent, culture en couloirs) sont alors de mise ;
4. agroforesterie animale
   L'élément déterminant est alors la présence simultanée d'arbres et animaux au sein de parcelles. Les animaux peuvent alors brouter l'herbe, ou les feuilles des arbres et des arbustes. Il est donc question d'une production fourragère ;

5. agroforesterie séquentielle
   Les arbres et les cultures s'y succèdent dans le temps. Les exemples les plus habituels sont l'agriculture sur brûlis, ou bien les jachères améliorées (jachères arbustives et arborées) ;
6. techniques agroforestières mineures
   Elles consistent en des associations d'arbres et d'animaux particulières à l'exemple de l'aquaforesterie (pour exemple pêcherie de mangrove), ou de l'entomoforesterie (pour exemples mûrier et vers à soie ou apiculture).

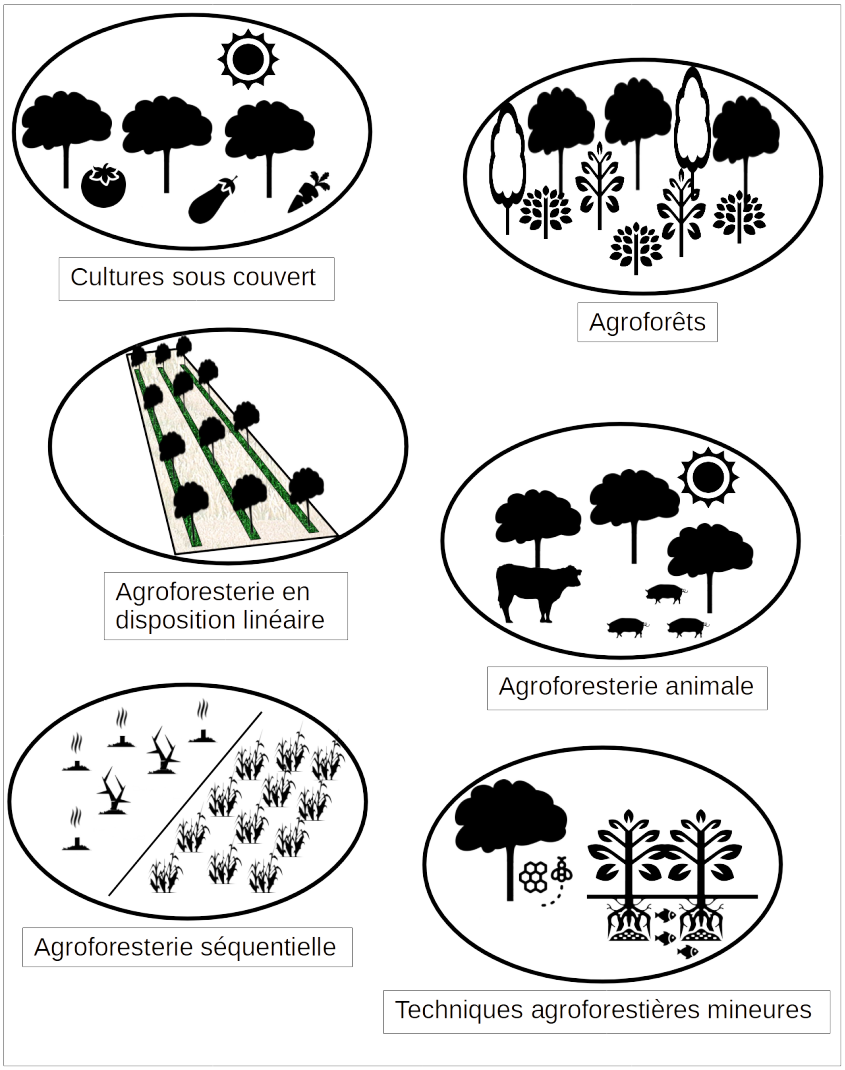


FIGURE 7.7 – *Classification des systèmes agroforestiers selon E. Torquebiau [*TORQUEBIAU, *2000].*

7.6.3 Le thésaurus de l'Agence Française d'AgroForesterie

L'Agence Française d'AgroForesterie (AFAF[5]), et ses partenaires européens, ont également proposé de manière plus récente une typologie de l'agroforesterie basée sur cinq critères et offrant trois niveaux de détail [BURRIEL et al., 2017] et reprenant en partie les critères définis dans [NAIR, 1985](voir figure 7.8). Un critère nouveau concerne en

5. AFAF : Agence Française d'AgroForesterie. https://www.agroforesterie.fr.

particulier les pratiques de culture à l'exemple de la technique de paillage (ou mulching) et les fonctions sont revues en partie sous la forme de services écosystémiques à l'exemple du bois de chauffage (fuelwood). Le référentiel prend la forme d'une taxonomie de descripteurs déclinés dans plusieurs langues (anglais, français, hongrois, bulgare, albanais, grec et libanais). Un fichier pdf reprenant l'ensemble de la taxonomie est disponible sur le site Agrof-MM Knowledge Data Bank.

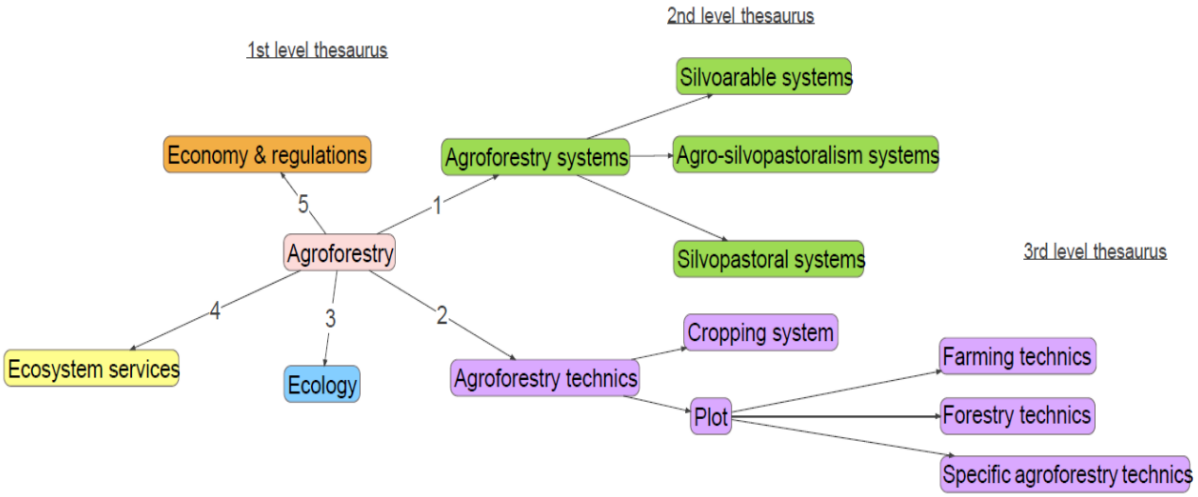


FIGURE 7.8 – *The levels of the thesaurus system for the Agrof-MM KDB – Les cinq critères et les trois différents niveaux du thésaurus de l'AFAF [*BURRIEL *et al., 2017].*

7.6.4 Les vocabulaires du web sémantique

Dans le périmètre du web sémantique et des sources de données ouvertes et liées, il est à remarquer l'existence de l'ontologie terminologique AFO (Pour Agriculture and Forestry Ontology). AFO est un vocabulaire au format SKOS, proposé par l'Université d'Helsinki dans un contexte général de mise à disposition d'un portail sémantique sur le web adossé à une ontologie de haut niveau (YSO General Finnish ontology) qui est étendue ensuite par plusieurs vocabulaires de spécialité [HYVÖNEN, 2010], à l'exemple des sciences du vivant et de l'environnement mais aussi des domaines artistiques. Par ailleurs, AFO n'est pas réellement dédié à l'agroforesterie, mais contient plutôt les descripteurs spécifiques de l'agriculture d'une part, et de la foresterie d'autre part. En ce sens, nous pouvons retrouver certains descripteurs clés de l'agroforesterie comme "soil conservation", "silviculture" ou "agroforestry" à la fois en anglais et en finlandais. Nous ne retrouverons toutefois pas la description des différents systèmes agroforestiers à l'exemple du système sylvo-pastoral "silvopastoral system". En ce sens, les ontologies terminologiques très généralistes de l'agriculture que sont AGROVOC [RAJBHANDARI et KEIZER, 2012] ou NALT (reprises dans GACS) contiendront le plus grand nombre des termes dédiés à l'agroforesterie. Nous retrouverons également de nombreux concepts d'intérêt dans le thésaurus AnaEE de l'INRAE disponible depuis le portail AgroPortal [Clément JONQUET et al., 2018].

7.7 Une ontologie terminologique pour l'agroforesterie

Il nous semble important de poser la question de l'utilité de disposer d'un référentiel terminologique dédié à l'agroforesterie dans le contexte du web sémantique. De nombreux travaux ont cherché à définir l'agroforesterie, à en dégager les principaux concepts et une classification qui pourrait faire consensus. Ces différentes classifications peuvent servir de points de départ à la construction d'un thésaurus. À ce titre, un thésaurus, nommé Agrof-MM, a d'ores et déjà été construit pour les besoins d'indexation d'une base de connaissance dédiée à de la formation en agroforesterie [BURRIEL et al., 2017]. Ce thésaurus ne répond toutefois pas aux principes de publication FAIR (Faciles à trouver, Accessibles, Interopérables et Réutilisables) préconisés dans le contexte du web sémantique. A contrario, des référentiels intégrant de nombreux concepts terminologiques relevant de l'agroforesterie sont bien retrouvés dans la sphère du web sémantique, à l'exemple d'AGROVOC ou d'AnaEE Thésaurus. La première difficulté porte sur le besoin d'exhaustivité. Le domaine de l'agroforesterie est au croisement de diverses disciplines connexes à l'exemple de l'agriculture, et de la foresterie, mais aussi de l'écologie ou de la climatologie. Il est donc difficile d'en cerner les contours. L'agroforesterie est de plus définie de différentes manières, et a été explorée au travers de différents critères classificatoires. Le choix parmi tous ces critères, des critères sur lesquels asseoir l'ossature du thésaurus, s'avère complexe. L'objectif est donc de définir une ossature hiérarchique qui puisse être complétée par d'autres structures si nécessaires en fonction des points de vue d'intérêt.
Le standard du W3C SKOS, introduit dans le chapitre 6 section 6.5.5, propose un méta-modèle pour l'organisation de systèmes d'organisation de la connaissance dits légers [BAKER, BECHHOFER et al., 2013 ; MARTÍNEZ-GONZÁLEZ et ALVITE-DÍEZ, 2019], et s'impose naturellement pour la construction d'une ontologie terminologique pour l'agroforesterie qui puisse aussi satisfaire les principes FAIR [JANOWICZ, HITZLER et al., 2014]. Le langage SKOS a contribué à la construction de nombreux référentiels dans les sciences du vivant et de l'environnement, dont AGROVOC mais aussi SDGO [JOSHI et al., 2021]. Nous avons adopté deux démarches différentes. La première de ces démarches consiste à proposer une version sémantisée du vocabulaire Agrof-MM. La seconde démarche consiste à définir une ossature de thésaurus plus orientée vers les concepts clés de la gestion d'un projet mené en agroforesterie. Nous cherchons en effet à construire un système de connaissances qui tire parti des jeux de données obtenus sur les parcelles d'agroforesterie.

7.8 Démarche retenue pour la construction d'un référentiel terminologique dédié

7.8.1 Description générale

Nous avons défini une chaîne de traitement (voir figure 7.9) qui va nous permettre d'évaluer si il y a lieu de construire un référentiel terminologique pour l'agroforesterie. L'évaluation est volontairement simple et porte à la fois sur le nombre de concepts terminologiques déjà pris en charge par d'autres référentiels, et sur la nature de ces concepts. Si la décision de construire un référentiel dédié est prise, la chaîne de traitement vient également en support de l'activité de construction en fournissant d'une part la collection de termes qui correspondent à des concepts terminologiques retrou-

vés dans différents référentiels accessibles sur le web, et d'autre part la collection de termes par trop spécifiques et non associés à des concepts appartenant à des ontologies terminologiques connexes.

Les différentes étapes de notre chaîne de traitement

Nous revenons ici sur les grandes activités mobilisées dans la chaîne de traitement. En entrée de chaîne, deux tâches sont menées en parallèle :

- **Élaboration d'une liste de termes issus d'une fouille de textes**
  La première ce ces tâches tire parti de l'outil de fouille de textes Gargantext et d'un corpus dédié à l'agroforesterie (en anglais) construit à partir de la plateforme ISTEX [COLLIGNON et CUXAC, 2017] [6], pour construire une liste des termes anglais les plus fréquemment exploités en agroforesterie.

- **Élaboration d'une liste de termes issus d'un référentiel**
  La seconde de ces tâches a consisté à définir, à partir de la terminologie Agrof-MM proposée par l'AFAF [7], un système d'organisation de connaissances au format SKOS et à le rendre disponible sur le web en s'aidant du logiciel OpenTheso [OPENTHESO, 2016]. Ce travail est disponible à l'adresse [8]. La figure 7.10 montre l'exemple de la description du concept terminologique ayant pour label oasis (en anglais comme en français) qui a pour concept de plus haut niveau agro-sylvopastoralisme. Par suite, l'ensemble des labels anglais des concepts retrouvés dans le vocabulaire Agrof-MM a fait l'objet d'une nouvelle liste.

6. https://www.istex.fr/
7. AFAF : http://agrofmm.eu/.
8. Amapvoc : http://www.amapvoc.fr.

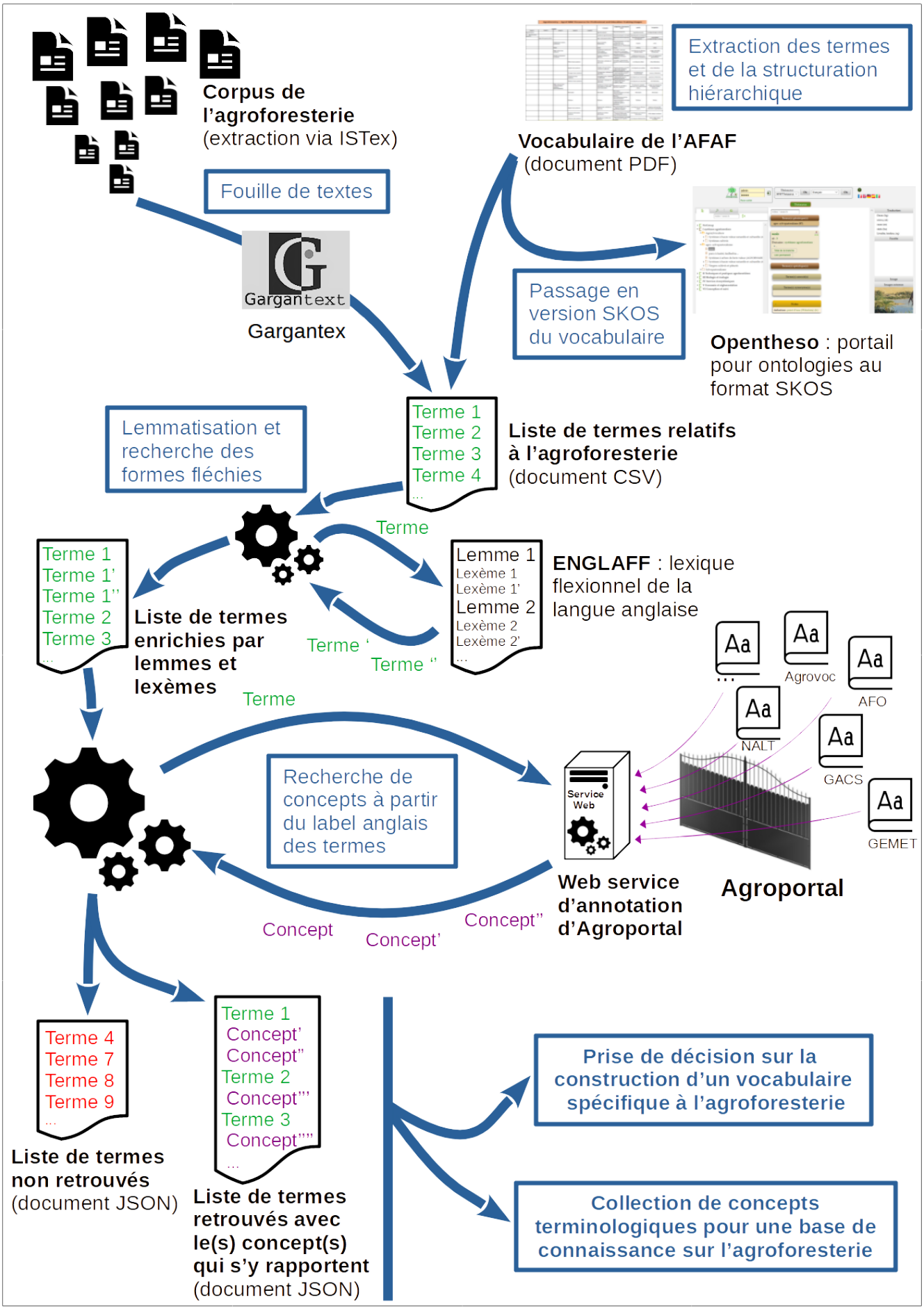


FIGURE 7.9 – *Ensemble des activités de la chaîne de traitement*

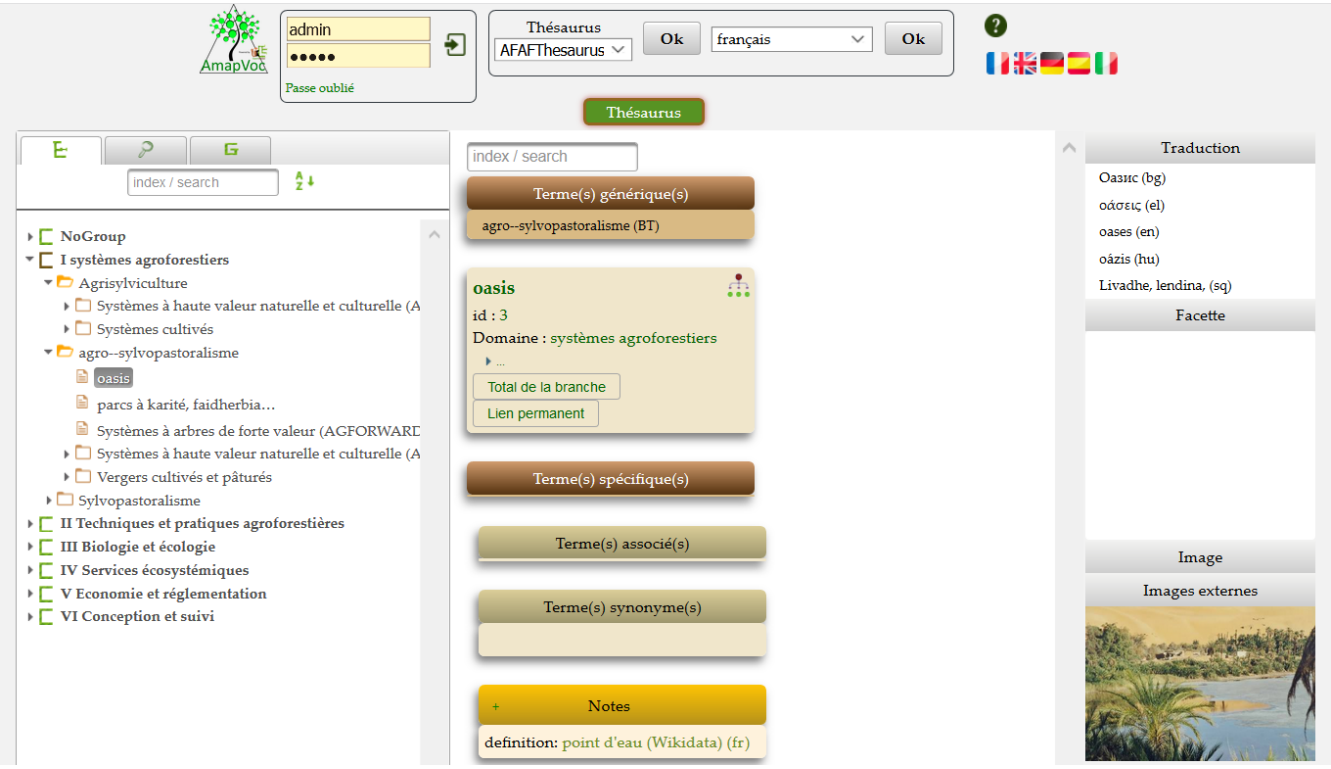


FIGURE 7.10 – *Site web AmapVoc présentant le thésaurus sur l'agroforesterie de L'AFAF.*

Après constitution des deux listes de termes nous avons constaté que pour chaque terme, seule une forme fléchie particulière était représentée dans la liste. Pour exemple dans notre liste issue de la fouille de données réalisée avec Gargantex nous retrouvons le terme "trees" mais pas le terme "tree", qui est son singulier. Afin de renforcer le caractère exhaustif de notre recherche de concepts à partir des termes de nos listes, nous avons rajouté une étape de traitement lexical pour obtenir les différentes formes fléchies des termes trouvés initialement dans le but de les rajouter à leur liste respective :

– **Lemmatisation et recherche des formes fléchies**

  Dans le cadre de la linguistique, la lemmatisation d'un texte consiste à associer aux différentes flexions d'un mot, une unité lexicale commune nommée "lemme". Ce traitement est généralement réalisé pour faciliter les analyses textuelles en réduisant le nombre de graphies d'un mot en un seul. Pour exemple la lemmatisation de "arbres" renverrait au lemme "arbre". Ces traitements de lemmatisation nécessitent la constitution de lexiques de forme fléchies pour chaque langue étudiée. Pour notre traitement nous avons utilisé ENGLAFF [9] un lexique des formes fléchies de la langue anglaise constitué par le CLLE [10] [SAJOUS et al., 2020]. Ce lexique contient plus de 1,1 millions d'entrées avec une forme fléchie, son lemme et une étiquette morphosyntaxique indiquant l'origine grammaticale de sa flexion (déclinaison ou conjugaison) au format GRACE [RAJMAN et al., 1997]. Nous avons donc utilisé le lexique ENGLAFF pour collecter l'ensemble des formes fléchies de chacun des termes anglais extrait initialement et de les incorporer au sein des listes traitées.

Nos listes de termes ainsi constituées et enrichies sont alors utilisées pour cette dernière étape de notre chaîne de traitement :

9. ENGLAFF : ENglish Great Lexicon Available For Free http://redac.univ-tlse2.fr/lexiques/englaff.html.
10. CLLE : Laboratoire Cognition, Langues, Langage, Ergonomie https://clle.univ-tlse2.fr/.

– **Recherche des concepts avec un web service d'annotation**

Les deux collections de termes ont été soumises de manière séparée au service REST NCBO Annotator disponible depuis AgroPortal [SHAH et al., 2009], le portail de référentiels en agriculture et en agronomie. Nous cherchons à établir des correspondances entre ces termes et les concepts organisés au sein des 163 ontologies gérées au sein d'AgroPortal [Clément JONQUET et al., 2018]. Il est à noter que le service Annotator ne met en œuvre qu'une recherche exacte sur la base du terme confronté, ce qui a rendu le travail de recherche de forme fléchies en amont nécessaire.
Le web service NCBO Annotator fourni par AgroPortal (service natif d'OntoPortal Appliance qui est la technologie exploitée par AgroPortal mais aussi BioPortal) est pensé à l'origine pour annoter des jeux de données à partir de concepts organisés au sein d'ontologies terminologiques. Une illustration de l'interface web utilisateur et du fonctionnement de NCBO Annotator est donnée ci dessous (voir les figures Fig.7.11 et 7.12). Annotator traite ici le texte correspondant à la page wikipedia dédiée à l'agroforesterie et en retourne les concepts désignés par les termes utilisés dans le texte. Ces concepts proviennent d'ontologies hébergées par le portail. Nous voyons ainsi que les ontologies préférentiellement sollicitées sont AgroVoc [RAJBHANDARI et KEIZER, 2012], GACS [BAKER et SUOMINEN, 2014; BAKER, CARACCIOLO et al., 2016], NALT [ALONSO, 2004; ALLIANCE OF AGRICULTURAL INFORMATION AND DOCUMENTATION SERVICES OF THE AMERICAS et al., 2019] et AFO pour des concepts associés à des termes très généraux tels que "agroforestry" ou encore "land use". La chaîne de traitement utilise Annotator un peu différemment en injectant les termes simples (agroforestry, ...) et composés (agroforestry system, ...) des listes précédentes afin d'étudier la présence/absence des concepts associés aux termes au sein des différentes ontologies accessibles depuis AgroPortal. Certains termes sont bien représentés et couverts par différents vocabulaires. Cette représentation est également retranscrite.
La chaîne de traitement fournit donc en dernier lieu un document au format JSON intégrant en particulier une liste des termes non retrouvés et une liste des termes retrouvés avec la collection des concepts associés et des ontologies de provenance.
La chaîne de traitement a été outillée en langage Java. Elle est suffisamment générique pour être réutilisée dans un autre contexte disciplinaire. Nos deux objectifs principaux sont d'évaluer la pertinence de définir un référentiel dédié à l'agroforesterie et de disposer des concepts d'intérêt et des alignements possibles avec d'autres référentiels de domaines connexes.

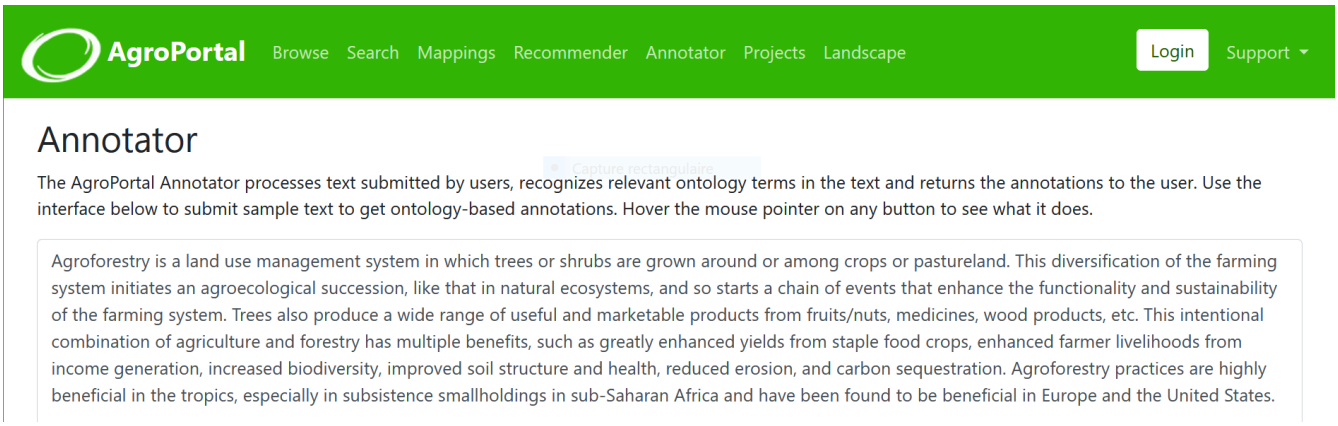


FIGURE 7.11 – *Interface web de l'annotateur du site AgroPortal.*

agroforestry	National Agricultural Library Thesaurus	direct	... carbon sequestration. **Agroforestry** practices are highly ...	agroforestry	National Agricultural Library Thesaurus
agroforestry	AGROVOC	direct	**Agroforestry** is a land ...	agroforestry	AGROVOC
agroforestry	AGROVOC	direct	... carbon sequestration. **Agroforestry** practices are highly ...	agroforestry	AGROVOC
land use planning	Global Agricultural Concept Scheme	direct	... is a **land use management** system in which ...	land use planning	Global Agricultural Concept Scheme
land use planning	National Agricultural Library Thesaurus	direct	... is a **land use management** system in which ...	land use planning	National Agricultural Library Thesaurus
land use	GEneral Multilingual Environmental Thesaurus	direct	... is a **land use** management system in ...	land use	GEneral Multilingual Environmental Thesaurus
land use	Agriculture and Forestry Ontology	direct	... is a **land use** management system in ...	land use	Agriculture and Forestry Ontology
land use	Biodiversity Thesaurus	direct	... is a **land use** management system in ...	land use	Biodiversity Thesaurus

FIGURE 7.12 – *Exemples des différents concepts retrouvés pour les termes de la page de description de l'agroforesterie sur le site Wikipédia via le service Annotator*

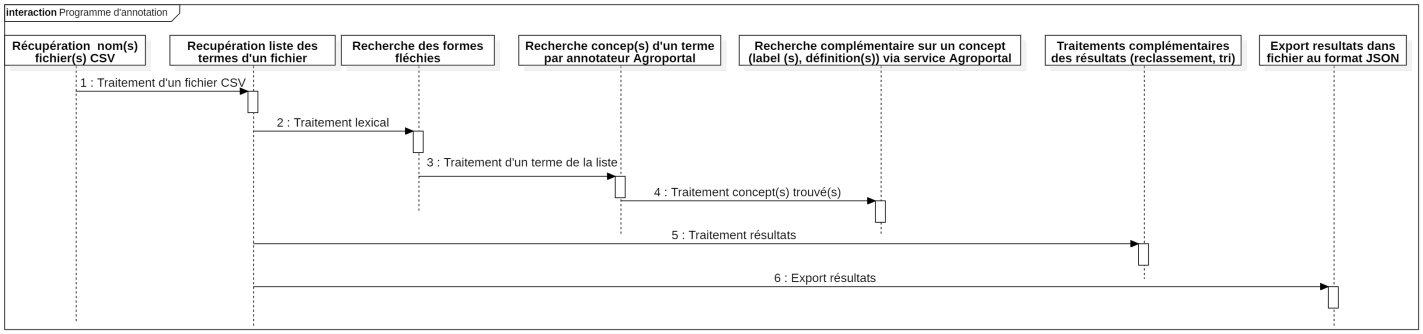


FIGURE 7.13 – *Diagramme de séquence de la chaîne de traitement pour l'annotation*

7.8.2 Résultats

Nous présentons les différents résultats obtenus en sortie de chaîne. Un fichier (un fichier pour chaque liste soumise) donne une vision à la fois générale et détaillée sur les résultats obtenus. Ici les résultats décrits sont associés à la liste de 350 termes obtenue après constitution d'un corpus ISTEX et extraction des termes les plus fréquents avec GarganText. La structuration au format JSON fait ressortir les éléments suivants :

- Nombre de termes simples ou composés issus des premières extractions (« numberOfOrginalTerms »)
- Nombre de termes d'origine exploités par le traitement lexical (« numberOfOriginalTermsWithFlexedForm »)
- Nombre de termes traités provenant du traitement des formes fléchies (« numberOfTermsFlexedForm »)
- Nombre de termes ayant été annotés par au moins un concept (« numberOfAnnotatedTerms »)
- Nombre de termes provenant du traitement des formes fléchies ayant été annotés par au moins un concept (« numberOfAnnotatedTermsFlexedForm »)
- Nombre de termes de la liste d'origine dont seul le traitement des formes fléchies

a permis une annotation (« numberOfOriginalTermsAnnotatedOnlyWithFlexedForm »)

– Nombre de termes n'ayant pas été annotés (« numberOfNotAnnotatedTerms »)

– Nombre de termes provenant du traitement des formes fléchies n'ayant pas été annotés (« numberOfNotAnnotatedTermsFlexedForm »)

– Nombre d'ontologies ayant fourni au moins un concept pour un terme de la liste (« numberOfOntologiesUsed »)

– Liste détaillée des termes annotés (« annotatedTerms »)

– Liste des termes annotés classés par nombre de concepts trouvés dans les différentes ontologies exploitées(« annotatedTermSortByNumberOfConcept »)

– Liste des termes non annotés (« notAnnotatedTerms »)

– Liste des ontologies exploitées ayant fourni au moins un concept pour un terme de la liste initiale (« ontologyAndConceptsForTerms »)

Synthèse

Nous présentons dans la Fig.7.14 un exemple de résultat obtenu. Ainsi sur les 350 termes d'intérêt fréquemment retrouvés dans les documents traitant d'agroforesterie 87 (25 %) ont été retrouvés dans le lexique ENGLAFF, ce qui nous a permis de collecter 173 termes supplémentaires notre recherche de concepts. Sur ce total de 523 termes, 260 ont été annotés (50 %) en étant associés à au moins un concept. Parmi ces 260 termes annotés, 49 sont issus du traitement des formes fléchies et l'on notera que pour 10 des termes provenant de l'extraction d'origine, seul le traitement des formes fléchies a permis une annotation.
Au sein des 263 termes non annotés, 124 proviennent du traitement lexical et l'on constate donc que sur les 139 termes originaux non annotés, 129 (92 %) n'ont pas non plus été retenus par le traitement lexical.
Les concepts identifiés sont distribués dans 98 ontologies.
Deux constats peuvent donc être tirés. Le premier est que l'agroforesterie entretient des liens de proximité avec de nombreux autres domaines expliquant le grand nombre d'ontologies (98 sur 163 ontologies maintenues) associées via leurs concepts. Le second est qu'un nombre important de termes est spécifique à l'agroforesterie, et reste à décrire et organiser.

```
numberOfOriginalTerms: 350
numberOfOriginalTermsWithFlexedForm:87
numberOfTermsFlexedForm: 173
numberOfAnnotatedTerms: 260
numberOfAnnotatedTermsFlexedForm: 49
numberOfOriginalTermsAnnotatedOnlyWithFlexedForm:10
numberOfNotAnnotatedTerms: 263
numberOfNotAnnotatedTermsFlexedForm: 124
numberOfOriginalTermsNotAnnotatedAndWithoutFlexedForm:129
numberOfOntologiesUsed: 98
▶ annotatedTerms: [...]
▶ annotatedTermSortByNumberOfConcept: [...]
▶ notAnnotatedTerms: [...]
▶ ontologyAndConceptsForTerms: [...]
```


FIGURE 7.14 – *Structure/résultats en sortie de chaîne*

Détail par terme retrouvé

Nous présentons maintenant un extrait de la liste détaillée des termes annotés (voir Fig.7.15) : Pour chaque terme qui est valué par la propriété « termLabel », l'organisation en est alors la suivante : :

– un tag « flexedTerm » pour indiquer que le terme provient du traitement des formes fléchies.

– le nombre de concepts trouvés pour ce terme « numberOfConceptsForTerm »

– différents détails concernant ces concepts (liste « conceptsOfOntolgy ») avec leur libellé préférentiel « prefLabel », leur URI « uri » et leur(s) définition(s) « definition », accompagnés des espaces de noms des ontologies dont ils ont été extraits (« ontologyUri »). Dans l'illustration, nous voyons que le terme "trees" est associé à 5 concepts, parmi lesquels le concept c_7887 d'AGROVOC ayant pour label préféré "trees" en anglais. A la suite le terme "tree" provenant du traitement des formes fléchies que nous avons réalisé avec le lexique ENGLAFF (tag flexedTerm = true) est associé à cinq autres concepts dont le concept c2_250 d'AnaEE avec pour label préféré "tree".De même, le terme "agroforestry" est associé à 6 concepts différents.

```
  numberOfOriginalTerms: 350
  numberOfOriginalTermsWithFlexedForm:87
  numberOfTermsFlexedForm: 173
  numberOfAnnotatedTerms: 260
  numberOfAnnotatedTermsFlexedForm: 49
  ...
  numberOfOntologiesUsed: 98
▼ annotatedTerms:
  ▼ 0:
    ▼ term:
        termLabel: "trees"
        flexedTerm: false
      numberOfConceptsForTerm: 5
    ▼ conceptsOfOntolgy:
      ▼ 0:
        ▼ concept:
            prefLabel: "trees"
            uri: "http://aims.fao.org/aos/agrovoc/c_7887"
          ▼ definition:
              0: "http://aims.fao.org/aos/agrovoc/
    xDef_e5549baa"
        ▼ ontology:
            ontologyUri: "http://data.agroportal.lirmm.fr/
    ontologies/AGROVOC"
      ▶ 1: {...}
      ▶ 2: {...}
      ▶ 3: {...}
      ▶ 4: {...}
  ▼ 1:
    ▼ term:
        termLabel: "tree"
        flexedTerm: true
      numberOfConceptsForTerm: 5
    ▼ conceptsOfOntolgy:
      ▼ 0:
        ▼ concept:
            prefLabel: "Tree"
            uri: "http://top-thesaurus.org/trait#TOP139"
          ▼ definition:
              0: "plants (PO:whole plant) that have secondary
    growth (GO:secondary growth) and have a main ..."
        ▼ ontology:
            ontologyUri: "http://data.agroportal.lirmm.fr/
    ontologies/TOP"
      ▼ 1:
        ▼ concept:
            prefLabel: "tree"
            uri: "http://opendata.inra.fr/anaeeThes/c2_250"
          ▼ definition:
              0: "Any large woody perennial plant with a
    distinct trunk giving rise to branches or leaves at some
    distance from the ground."
        ▼ ontology:
            ontologyUri: "http://data.agroportal.lirmm.fr/
    ontologies/ANAEETHES"
      ▶ 2: {...}
      ▶ 3: {...}
      ▶ 4: {...}
  ▼ 2:
    ▼ term:
        termLabel: "agroforestry"
        flexedTerm: false
      numberOfConceptsForTerm: 6
  :::
```


FIGURE 7.15 – *Détails pour les termes annotés.*

Termes retrouvés triés par ordre d'"importance"

L'information décrite ici reste sensiblement la même, mais il s'agit de retrouver les termes et concepts associés qui sont très souvent retrouvés dans les ontologies hébergées dans AgroPortal. Ainsi les termes annotés sont triés par nombre de concepts trouvés dans les différentes ontologies exploitées par ordre décroissant (voir Fig.7.16). Nous avons la structure pour chaque terme :

- la propriété « termLabel » renseigne le libellé du terme.
- la propriété « numberOfConceptsForTerm » donne le nombre de concepts trouvés dans les différentes ontologies.

Les termes les plus représentés sont logiquement les termes très généraux à l'exemple de "organization" (organisation) lié à 22 concept, de "soil" (sol) lié à 18 concepts ou encore "system" (système) lié à 14 concepts.


```
  numberOfOriginalTerms: 350
  numberOfOriginalTermsWithFlexedForm:87
  numberOfTermsFlexedForm: 173
  numberOfAnnotatedTerms: 260
  numberOfAnnotatedTermsFlexedForm: 49
  numberOfOriginalTermsAnnotatedOnlyWithFlexedForm:10
  ...
  numberOfOntologiesUsed: 98
▶ annotatedTerms: [...]
▼ annotatedTermSortByNumberOfConcept: [...]
  ▶ 0:
  ▶ 1:
  ▶ 2:
  ▼ 3:
    ▼ term:
        termLabel: "organization"
        flexedTerm: false
      numberOfConceptsForTerm: 22
  ▼ 4:
    ▼ term:
        termLabel: "soil"
        flexedTerm: false
      numberOfConceptsForTerm: 18
  ▼ 5:
    ▼ term:
        termLabel: "system"
        flexedTerm: false
      numberOfConceptsForTerm: 14
  ▶ 6:
  ▶ 7:
  ...
  ...
  ...
▶ notAnnotatedTerms: [...]
▶ ontologyAndConceptsForTerms: [...]
```


FIGURE 7.16 – *Termes annotés triés par nombre de concepts retrouvés*

Termes non retrouvés

Il s'agit ici juste d'une liste à plat, avec le libellé des termes non retrouvés (voir Fig.7.17) :

– La valeur de la propriété « termLabel » renseigne le libellé du terme.

– Le tag « flexedTerm » indique si le terme provient du traitement des formes fléchies.

Pour exemple le terme "agroforests" (agroforêt en français) qui désigne une pratique culturale spécifique de l'agroforesterie, ou encore la locution de plus haut niveau "agroforestry practices" ne sont pas retrouvés. Ce n'est pas une surprise, les termes très spécifiques sont d'ordinaire dans un vocabulaire de spécialité, et ce vocabulaire de spécialité est absent d'AgroPortal pour ce qui concerne l'agroforesterie.



FIGURE 7.17 – *Exemple de termes non annotés.*

Référentiels apparentés

Les ontologies, considérées comme étant les plus proches, sont les ontologies qui comprennent le plus grand nombre de concepts en correspondance avec des termes de la liste soumise en entrée. Les ontologies qui ont des concepts associés sont donc listées et triées au regard du nombre de concepts retrouvés comme associés.
Pour chaque ontologie ayant au moins un concept associé, le nombre de concepts « numberOfConcepts » et la liste des correspondances concept - terme annoté « conceptsOfTerm » sont retournés (voir Fig.7.18). La visualisation de ces résultats est présentée

en figure 7.19 pour les douze premiers vocabulaires en terme de concepts retrouvés .On observe que les référentiels les plus connexes sont les référentiels les plus généralistes dans les domaines de l'alimentation et de l'agriculture. Le premier référentiel retrouvé en tête de résultat est AFO basé sur le thésaurus Agriforest de l'université d'Helsinki couvrant entre autres les thèmes de l'agriculture, de la sylviculture et de l'alimentation. En deuxième position, nous retrouvons GACS (Global Agricultural Concept Scheme) qui s'avère une mise en commun des vocabulaires dédiés à l'agriculture et à l'alimentation : AGROVOC, CABT et NALT. Par la suite on retrouvera des référentiels spécialisés dans d'autres domaines comme l'environnement (GEMET, ENVO), la biodiversité (AnaEE Thésaurus, BIODIVTHES), l'agronomie (Agronomy Ontology), ou l'administration des sols (LANDVOC).


```
  numberOfOriginalTerms: 350
  numberOfOriginalTermsWithFlexedForm:87
  numberOfTermsFlexedForm: 173
  numberOfAnnotatedTerms: 260
  numberOfAnnotatedTermsFlexedForm: 49
  numberOfOriginalTermsAnnotatedOnlyWithFlexedForm:10
  ...
  numberOfOntologiesUsed: 98
▶ annotatedTerms: [...]
▶ annotatedTermSortByNumberOfConcept: [...]
▶ notAnnotatedTerms: [...]
▼ ontologyAndConceptsForTerms:
  ▼ 0:
    ▼ ontology:
       ontologyUri: "http://data.agroportal.lirmm.fr/ontologies
      /GACS"
    numberOfConcepts: 136
    ▼ conceptsOfTerm:
      ▼ 0:
        ▼ term:
           termLabel: "trees"
        ▼ concept:
            prefLabel: "trees"
            uri: "http://id.agrisemantics.org/gacs/C386"
          ▼ definition: []
      ▼ 1:
        ▼ term:
           termLabel: "agroforestry"
        ▼ concept:
            prefLabel: "agroforestry"
            uri: "http://id.agrisemantics.org/gacs/C3865"
          ▼ definition:
              0: "http://id.agrisemantics.org/gacs/
     def_e786ceb4"
              1: "http://id.agrisemantics.org/gacs/
     def_2353caf4"
       ▶ 2: {...}
       ▶ 3: {...}
  ...
  :::
  ...
```


FIGURE 7.18 – *Détails pour les ontologies connexes*

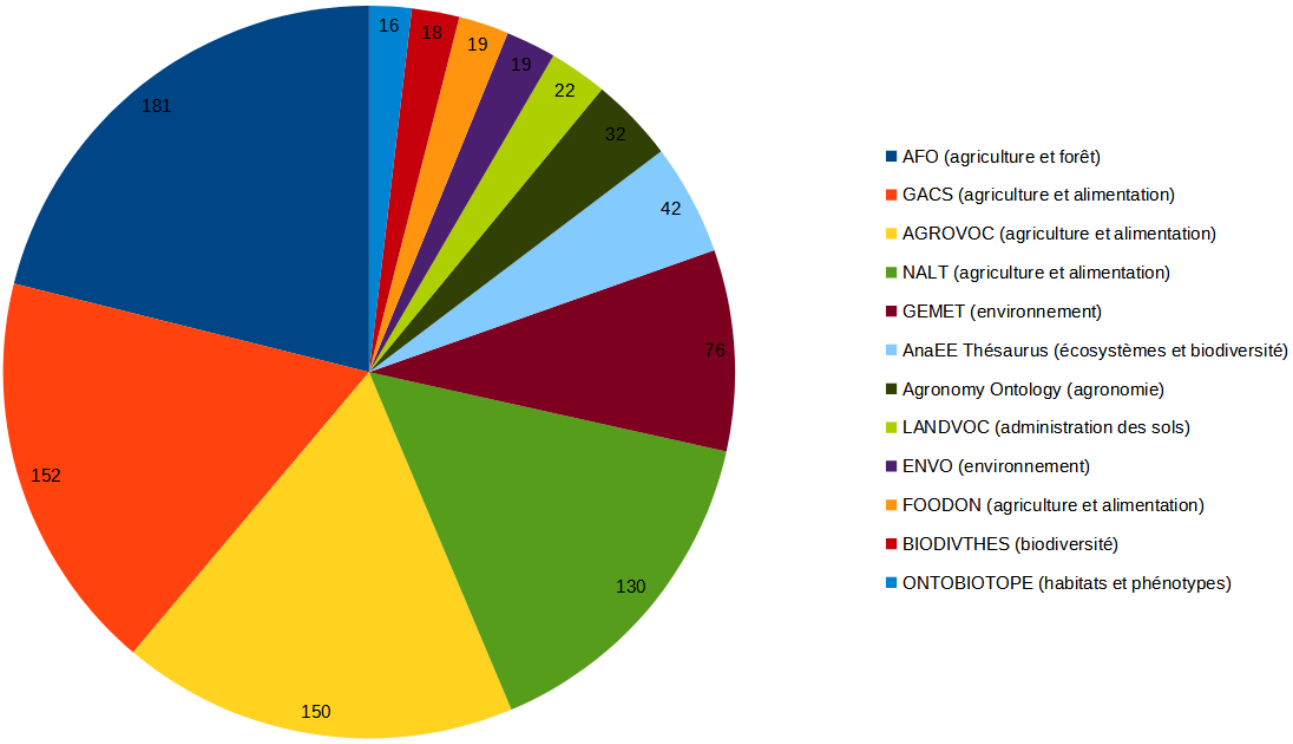


FIGURE 7.19 – *Les différentes ontologies contenant les concepts retrouvés avec les termes issus de l'extraction avec GarganText (classées par nombre de concepts retrouvés).*

7.8.3 Justifications concernant la construction d'un référentiel

L'analyse des résultats en sortie de chaîne de traitement fait ressortir deux grandes tendances. Les termes les plus généraux de l'agroforesterie sont partagés par de nombreuses disciplines, à l'exemple de l'agriculture, de l'agronomie, de l'écologie, de la biodiversité, de la taxonomie du vivant, de la biologie végétale, de la gouvernance des territoires, ou encore de l'environnement. Nous retrouvons en effet de nombreux concepts terminologiques transverses à ces domaines comme "soil", "system", "plant", "tree". À l'inverse, les termes les plus spécifiques de l'agroforesterie ne trouvent pas de réelle correspondance parmi les concepts terminologiques organisés au sein des ontologies hébergées par AgroPortal.
Nous nous concentrons ici sur les concepts organisés au sein du thésaurus Agrof-MM, et de la liste de termes associée. Les ontologies qui ont le plus de liens en commun sont dans l'ordre (voir Fig.7.20) : AFO (agriculture et forêt), AGROVOC (agriculture et alimentation), GACS (agriculture et alimentation), NALT (agriculture et alimentation), GEMET (environnement), AnaEE Thésaurus (écosystèmes et biodiversité), Agronomy Ontology (agronomie), et dans une moindre mesure : ENVO (environnement), LANDVOC (administration des sols), BIODIVTHES (biodiversité), ONTOBIOTOPE (habitats et phénotypes), FLOPO (phénotype et plante), PATO (phénotypes et traits) et PO (anatomie, morphologie et développement des plantes). Ce résultat est fidèle à ce qui était attendu, puisque l'agroforesterie est en lien très étroit avec l'agriculture, la forêt, l'environnement, les sols, les organismes vivants, et l'écologie. Sur les 253 termes extraits du thésaurus, 91 ont permis de dériver 186 termes supplémentaires avec la recherche de formes fléchies. Sur un total donc de 439 termes 160 sont retrouvés comme étant partagés, et 279 ne le sont pas. Nous avons donc près de 63% des concepts de l'agroforesterie qui ne trouvent aucune correspondance sur AgroPortal. Nous listons ci-dessous quelques exemples parmi ces 279 termes. Nous retrouvons des

pratiques agroforestières spécifiques avec les "forest gardens" (ou jardins forêt), ou les "huertas" (systèmes agraires propres aux milieux méditerranéens). Il est à noter que certains systèmes agroforestiers conservent leur désignation dans la langue du pays qui a vu croître ce système (pour "huertas", l'Espagne). Nous retrouvons de même le terme français "joualles" (vignes poussant sur des arbres), ou encore le terme allemand streuobst (arbres fruitiers disséminés sur des terres agricoles convenant aux latitudes tempérées d'Europe centrale). Il va en conséquence être très difficile de retrouver des concepts associés à ces pratiques en dehors d'un vocabulaire de spécialité dédié à l'agroforesterie. Nous retrouvons également des principes d'aménagement avec "border systems" (systèmes de bordure), "trees in plot" (arbres dans la parcelle), ainsi que des éléments d'aménagement avec "alignments" (alignements) et "hedgerow" (haie). L'organisation des concepts associés à ces derniers termes nous intéresse tout particulièrement, puisque nous cherchons à modéliser les techniques agroforestières sur la base des aménagements pratiqués sur les sites agroforestiers. À ce titre, il nous semble justifié de construire un référentiel dédié, et nous allons nous orienter vers la construction d'un premier module de référentiel terminologique pour une définition et une organisation des concepts liés aux aménagements de parcelles agroforestières. Nous nous aidons en ce sens des résultats, et en particulier des termes non annotés en sortie de la chaîne de traitement.

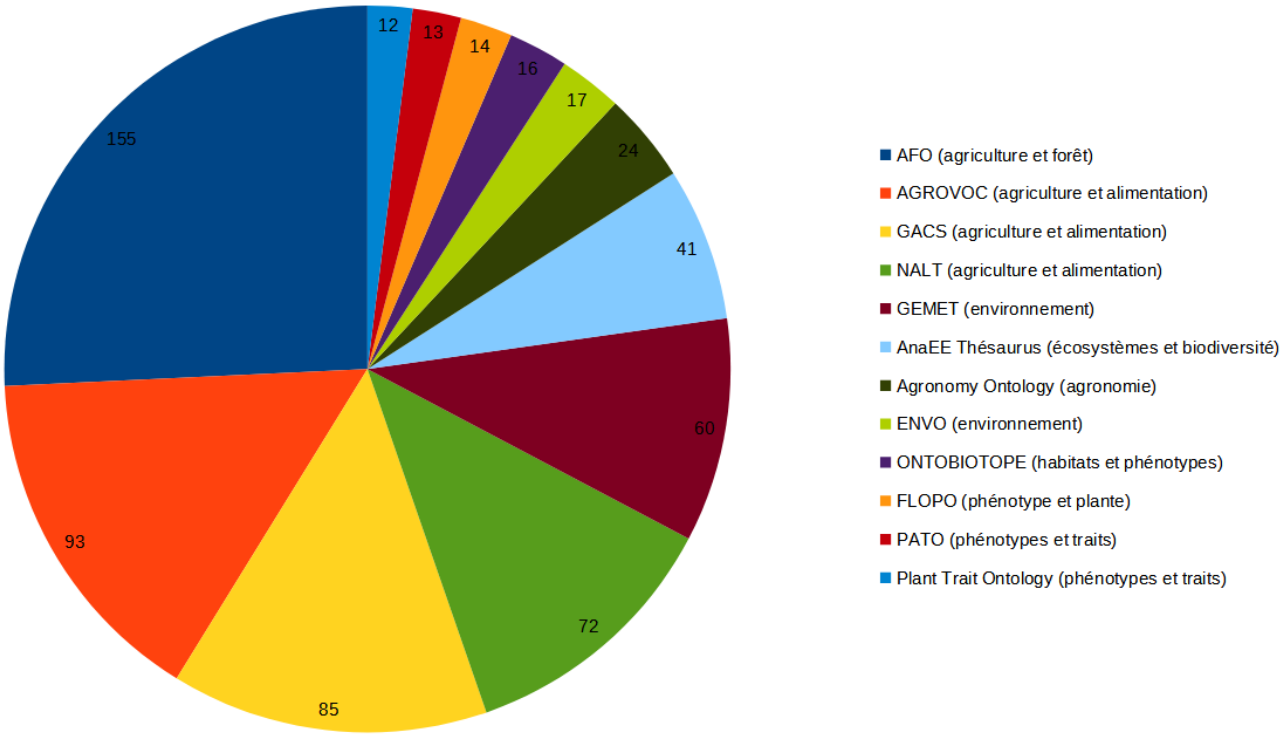


FIGURE 7.20 – *Les différentes ontologies contenant les concepts retrouvés avec les termes issus du thésaurus Agrof-MM (classées par nombre de concepts retrouvés).*

```
▼ notAnnotatedTerms: [...]
      ▼ 0:
          termLabel: "homegardens"
      ▼ 1:
          termLabel: "food orchards"
      ▼ 2:
          termLabel: "urban agroforestry"
      ▼ 3:
          termLabel: "forest gardens"
      ▼ 4:
          termLabel: "huertas"
      ▼ 5:
          termLabel: "vineyard systems"
      ▼ 6:
          termLabel: "silvoarables"
      ▼ 7:
          termLabel: "trees in plot"
      ▼ 8:
          termLabel: "bioenergy systems"
      ▼ 9:
          termLabel: "hautains"
      ▼ 10:
          termLabel: "coltura promiscua"
      ▼ 11:
          termLabel: "joualles"
      ▼ 12:
          termLabel: "silvopastoralism"
      ▼ 13:
          termLabel: "grazed forests"
      ▼ 14:
          termLabel: "wood pastures"
      ▼ 15:
          termLabel: "border systems"
      ▼ 16:
          termLabel: "alignments"
      ▼ 17:
          termLabel: "hedgerow"
      ▶ 18: {...}
   ...
   ...
```


FIGURE 7.21 – *Termes du vocabulaire Agrof-MM sans correspondance*

7.8.4 Extraits du référentiel en cours de construction

Nous présentons dans cette section un extrait de notre référentiel en nous focalisant sur les différents types d'aménagements agroforestiers relevés au cours de notre analyse du corpus agroforestier, illustrés par la figure 7.22.

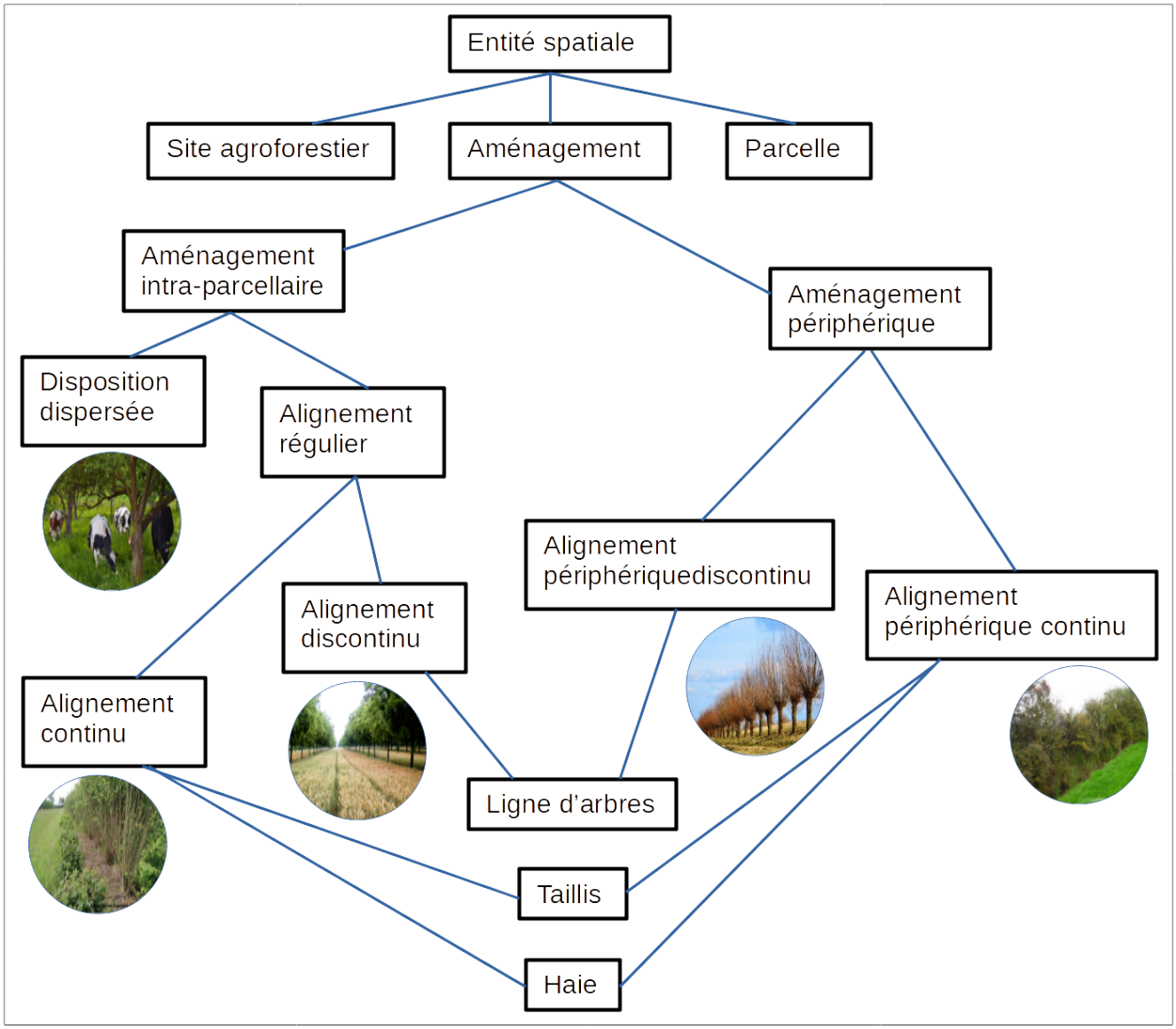


FIGURE 7.22 – *AgroforestryVoc: Schéma des différents types d'aménagements agroforestiers.*

La figure 7.23 montre un extrait de cette ontologie "AgroforestryVoc" en syntaxe Turtle. Les deux démarches se complètent, et nous alignons également les termes définis avec les termes retrouvés dans d'autres vocabulaires, à l'exemple d'AGROVOC. Dans cet extrait, différents concepts et propriétés de SKOS sont mobilisés pour structurer l'ontologie :

- Le triplet "afv:caf_cs_1 a skos:ConceptScheme" indique que l'entité "afv:caf_cs_1" est un individu de la classe **ConceptScheme** de SKOS. Il s'agit donc de l'instance de type schéma de concepts, à laquelle seront rattachés tous les individus de type concept terminologique. "afv:caf_cs_1" sera donc le point d'entrée du vocabulaire mis en place.

- "afv:caf_tccs_3" est un individu de la classe **Concept** et les termes qui le nomment en français et en anglais sont donnés avec la propriété *prefLabel*. Il est donc nommé en français "Entité spatiale"

- Ce concept "Entité spatiale" est lié au point d'entrée du vocabulaire par la propriété *inScheme* et la propriété *hasTopconcept* indique qu'il s'agit d'un concept de plus haut

niveau qui sera mis en avant lors du parcours de ce vocabulaire nommé "AgroforestryVoc".

- La propriété *definition* permet de relier un concept à une description de celui sous forme textuelle.
- La propriété *broader* permet d'indiquer une relation directe de subsomption entre deux concepts. On voit dans cet exemple que le concept "Thème expérimental" ("afv:caf_16") est subsumé par le concept "Propriétés d'éléments agroforestier" ("afv:caf_tccs_1").
- La propriété *narrower* indique elle aussi une relation directe de subsomption entre deux concepts mais cette fois du sujet vers l'objet. Le concept "Aménagement" ("afv:caf_48") subsume les concept "Aménagement intra-parcellaire" et "Aménagement périphérique".
- La propriété *exactMatch* relie quant à elle deux concepts considérés comme entièrement similaires. Dans cet exemple notre concept "Parcelle" semble être en tout point identique au concept de l'ontologie Agrovoc agrovoc:c_fdfbb37f. Elle permet ainsi de rapprocher des concepts provenant de vocabulaires différents.

Le vocabulaire de SKOS n'est bien sûr pas limité aux seuls concepts et propriétés présentés dans cet exemple [*SKOS Simple Knowledge Organization System Reference* 2018]. La figure 7.24 visualise cet extrait avec le logiciel Protégé.

```
@prefix afv:   <http://www.afy.fr/AgroforestryVoc/> .
@prefix skos:  <http://www.w3.org/2004/02/skos/core#> .
@prefix agrovoc: <http://aims.fao.org/aos/agrovoc/> .

afv:caf_cs_1    a    skos:ConceptScheme ;
     skos:prefLabel   "AgroForestryVoc"@fr , "AgroForestryVoc"@en ;
     skos:hasTopConcept  afv:caf_tccs_3 .

afv:caf_tccs_3  a   skos:Concept ;
     skos:prefLabel    "Entité spatiale"@fr , "spatial entity"@en .
     skos:definition   "Entité ponctuelle, linéaire ou surfacique."@fr ,
                       "point, linear or surface entity."@en ;
     skos:inScheme     afv:caf_cs_1 .

afv:caf_31    a        skos:Concept ;
     skos:prefLabel    "Parcelle"@fr , "plot"@en .
     skos:definition    "Une parcelle de terrain est une zone de terrain ayant une propriété
                          particuliére, une utilisation du sol ou une autre caractéristique particulière.
                          Une parcelle est souvent utilisée comme base pour un cadastre ou un
                        système d'enregistrement des terres"@fr ,
                         "A parcel (or plot) of land is an area of land with a particular ownership,
                          land use, or other characteristic. A parcel is frequently used as the basis for
                          a cadastre or land registration system."@en ;
     skos:broader      afv:caf_tccs_3
     skos:exactMatch agrovoc:c_fdfbb37f .

afv:caf_48    a       skos:Concept ;
     skos:prefLabel    "Aménagement"@fr , "management"@en .
     skos:definition    "Surface de terrain où l'on mène une activité et un ensemble de méthodes et
                         de pratiques pour une production de matières premières agricoles ou
                         ligneuses"@fr ,
                        "area of land where an activity and a set of methods and practices are
                         carried out for the production of agricultural or timber raw materials"@en ;
     skos:broader      afv:caf_tccs_3
     skos:narrower     afv:caf_35 , afv:caf_36 .

afv:caf_35    a       skos:Concept ;
     skos:prefLabel    "Aménagement intra-parcellaire"@fr , "intra-plot development"@en .
     skos:definition    "Aménagement agroforestier au coeur même d'une parcelle"@fr ,
                         "Agroforestry development in the heart of a plot"@en ;
     skos:broader      afv:caf_48 .

afv:caf_36    a       skos:Concept ;
     skos:prefLabel    "Aménagement périphérique"@fr , "peripheral development"@en .
     skos:definition    "Aménagement agroforestier à la périphérie d'une parcelle"@fr ,
                       "Agroforestry development at the edge of a plot"@en ;
     skos:broader      afv:caf_48 .

afv:caf_37    a       skos:Concept ;
      skos:prefLabel  "Alignement régulier"@fr , "regular alignment"@en .
      skos:definition  "Aménagement intra-parcellaire régulier"@fr ,
                      "Regular intra-plot development"@en  ;
      skos:broader     afv:caf_35 ;
      skos:narrower    afv:caf_38 , afv:caf_39 .

afv:caf_38    a       skos:Concept ;
      skos:prefLabel  "Alignement discontinu"@fr , "discontinuous alignment"@en  .
      skos:definition  "Aménagement intra-parcellaire régulier discontinu"@fr ,
                      "Regular discontinuous intra-plot development"@en ;
      skos:broader     afv:caf_37 .

afv:caf_39    a       skos:Concept ;
      skos:prefLabel  "Alignement continu"@fr , "continuous alignment"@en  .
      skos:definition  "Aménagement intra-parcellaire régulier continu"@fr ,
                      "Regular continuous intra-plot development"@en ;
      skos:broader     afv:caf_37 .
```

FIGURE 7.23 – *AgroforestryVoc: Extrait en langage Turtle.*

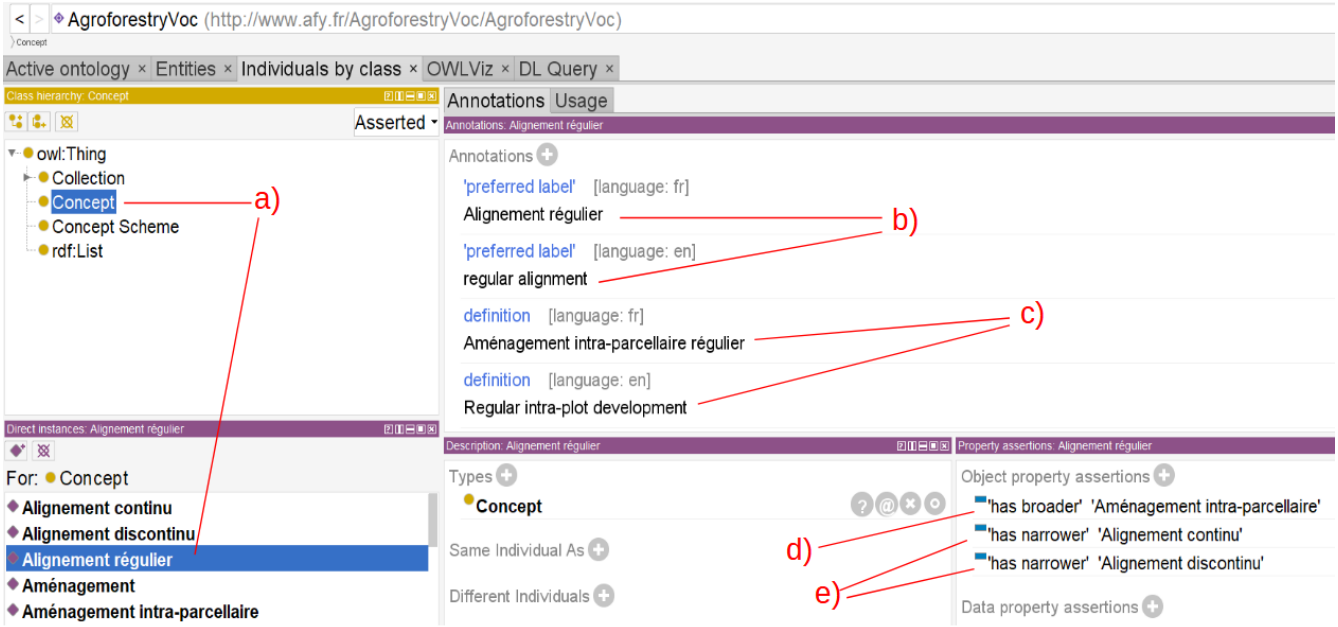


FIGURE 7.24 – *AgroforestryVoc: Exemple d'une ontologie terminologique pour l'agroforesterie visualisé avec le logiciel Protégé: a) Le concept "Alignement régulier" est sélectionné ; b) Les termes utilisés pour le nommer sont fournis en français et en anglais ; c) Une définition est donnée pour ce concept ; d) Ce concept est subsumé par le concept "Aménagement intra-parcellaire" ; e) Ce concept subsume entre autres les concepts "Alignement continu" et "Alignement discontinu" .*

7.8.5 Discussion autour de la version actuelle

Une version bêta de l'ontologie terminologique est construite. Nous sommes conscients d'un certain nombre de limites du travail réalisé. Nous nous sommes surtout efforcés de préciser les concepts clés à même de normaliser la description des aménagements agroforestiers. Pour ce faire, nous avons exploité un corpus de documents relatifs à l'agroforesterie, extrait les termes les plus fréquents et consulté les ontologies terminologiques du monde de l'agriculture, de l'agronomie et de l'écologie. Nous n'avons pas pour l'instant stabilisé toute la terminologie relative à l'agroforesterie. De plus, il faut garder en tête qu'une ontologie ne peut être mise en place sans la participation active de la communauté d'expertise associée. Si l'on souhaite que l'ontologie soit à la fois partagée et le reflet de l'expertise de la communauté, il faut pouvoir impliquer les différents acteurs du domaine dans le processus de construction. Cette implication permettrait de s'assurer de l'engagement sémantique [BACHIMONT, 2000] du référentiel construit en fixant le sens linguistique des concepts. Un logiciel comme Opentheso vient ainsi faciliter la construction collective de référentiels, en permettant à une communauté d'interagir lors du processus de construction. Nous n'avons pas pour l'instant réussi à fédérer suffisamment les acteurs de l'agroforesterie pour pouvoir prétendre avoir construit un référentiel qui les rassemble. Néanmoins, cette première brique de référentiel nous sera utile pour la définition d'un cadre de base de connaissance général à l'agroforesterie, en dotant les concepts spécifiques de l'agroforesterie de définitions, quand ils sont absents dans les ontologies terminologiques disponibles sur le web.
L'utilisation d'une ontologie terminologique va permettre d'enrichir sémantiquement le contenu d'une base de connaissance. Il convient maintenant de définir le cadre nécessaire à la construction de bases de connaissance pouvant être mutualisées en agroforesterie.

7.8.6 Les éléments d'importance retenus pour l'élaboration d'un cadre général de modélisation des connaissances

L'observation

La plupart des données brutes issues des bases de données analysées proviennent d'observations et de mesures faites par des capteurs ou des personnes au sein des parcelles agroforestières. Ces observations diverses et variées portent sur des sujets et des propriétés de ces sujets comme pour exemple la hauteur et le diamètre des arbres à hauteur de poitrine pour les arbres, l'humidité et le pH pour les sols ou un rendement par hectare d'une culture. La notion d'observation, avec en particulier l'observation des valeurs des propriétés d'un élément constitutif d'un aménagement agroforestier, est un élément clé pour la constitution du modèle.

Les aspects spatiaux et temporels

Un aménagement agroforestier est par nature spatial, et il est par exemple difficile d'en modifier l'organisation spatiale générale depuis la plantation des arbres jusqu'à leur abattage. De même, la proximité dans l'espace va influencer les éléments notamment biotiques de la parcelle. Ainsi des contraintes de proximité entre arbres et cultures vont conduire à des mécanismes de compétition pour capter différentes ressources comme la lumière ou encore l'eau.
L'aspect temporel revêt également une grande importance au sein des aménagements agroforestiers, car ces systèmes combinent parfois des cultures annuelles, avec l'exploitation d'arbres pouvant se développer sur la parcelle sur une trentaine d'années. Les compétitions dans l'utilisation de ressources communes peuvent être minimisées en prenant en compte les séquences temporelles de l'utilisation des ces ressources par les différentes espèces. Ainsi, un exploitant peut se voir préconiser une culture d'hiver qui profitera d'un ensoleillement accru en l'absence de feuilles sur les arbres pendant la période hivernale.
Les observations elles-mêmes, peuvent également nécessiter la prise en compte de ces aspects spatiaux et temporels, comme par exemple dans le cas de la mesure de l'humidité du sol d'une parcelle en pente, en haut et en bas de cette pente, en hiver et en été.

La taxonomie du vivant

Les principes de l'agroforesterie mettent en exergue les mécanismes d'interaction entre les différents constituants biotiques d'une parcelle afin de créer des synergies entre ces constituants. Ces interactions peuvent varier d'une espèce à l'autre, comme par exemple dans le contexte de deux espèces d'une même famille mobilisées pour fournir un couvert végétal, dont l'une sera capable de fixer l'azote de l'air dans le sol au profit de la culture suivante.
La taxonomie qui est déjà en soi un outil essentiel dans les sciences du vivant pour identifier l'éventail des organismes, nous permettra de décrire, nommer et classer les plantes et les animaux impliqués en tant que constituants d'un aménagement agroforestier.

VIII

Deuxième contribution : construction d'une ontologie de domaine pour les aménagements agroforestiers

8.1 Introduction 125
8.2 Définition d'un modèle conceptuel 128
8.2.1 Définition du modèle conceptuel : Agroforestry 128
8.2.2 Une démarche mobilisant l'existant 129
8.2.3 Intégration de l'ontologie SOSA 130
8.2.4 Extension de l'ontologie GeoSPARQL et de l'ontologie OWL-Time 131
8.3 Passage au modèle formel 140
8.3.1 Utilisation des logiques de descriptions 140
8.3.2 Les fondements logiques d'Agroforestry 141
8.4 Passage au modèle opérationnel 143
8.4.1 Premiers essais d'inférence 144

8.1 Introduction

Nous souhaitons proposer un modèle de connaissances qui offrira un cadre unificateur à la structuration de données et de connaissances en agroforesterie, et qui simplifiera ainsi la construction de bases de connaissance du domaine. Les données issues d'observations jouent un rôle clé dans la compréhension des mécanismes régissant les systèmes agroforestiers, et peuvent également asseoir leur justification par comparaison avec les systèmes agricoles d'un côté, et forestiers de l'autre. L'objectif est donc aussi de pouvoir mettre en commun l'information provenant de plusieurs bases de connaissance construites sur la base de systèmes agroforestiers de même type, ou bien de types différents. Un exploitant qui veut tirer parti de l'agroforesterie va devoir se poser de nombreuses questions sur les arbres et les cultures à exploiter en synergie, les aménagements les plus adaptés au regard de la nature du sol et du climat, ou encore les rotations de culture à respecter. De nombreux paramètres sont en effet à prendre en considération en fonction des services socio-économiques et écosystémiques attendus.

Certaines des réponses peuvent être guidées par la consultation d'une base de connaissance et l'analyse des connaissances obtenues. La construction d'un modèle qui sera au cœur d'une telle base de connaissance est l'objectif de cette section.
Nous exploitons pour ce faire une démarche en trois temps, comptant un premier travail de conceptualisation à l'aide de la notation UML, une étape de formalisation, ou d'"ontologisation" qui s'adosse à la famille des logiques de description et enfin une phase d'opérationnalisation qui mobilise les langages du web sémantique, et en particulier OWL 2. La démarche ainsi constituée s'inspire grandement de la proposition de [DIENG et al., 2001] pour ce qui concerne le cycle de vie d'une ontologie. Une ontologie est envisagée sous l'angle du génie logiciel, comme un composant technique évolutif et est régie par un cycle de vie qui prend en charge différentes phases allant de l'analyse des besoins jusqu'à l'évaluation du composant et à ses usages. Parmi ces phases, la construction comprend donc les phases successives de conceptualisation, ontologisation et opérationnalisation, mais peut aussi inclure une phase dite d'intégration, qui permet de tirer parti d'autres composants ontologiques déjà existants. Il est à noter que la méthodologie Neon [HAASE et al., 2008] préconise globalement les mêmes étapes de conceptualisation, de réutilisation, de formalisation et d'implémentation pour la construction d'une ontologie.
La figure 8.1 offre une synthèse sur le cycle de vie adopté. Il est à noter que la phase de conceptualisation a également compris une phase de rétro-conception en UML des ontologies cadres mobilisées de manière à ancrer l'ontologie Agroforestry à ces ontologies cadres dès la conception.

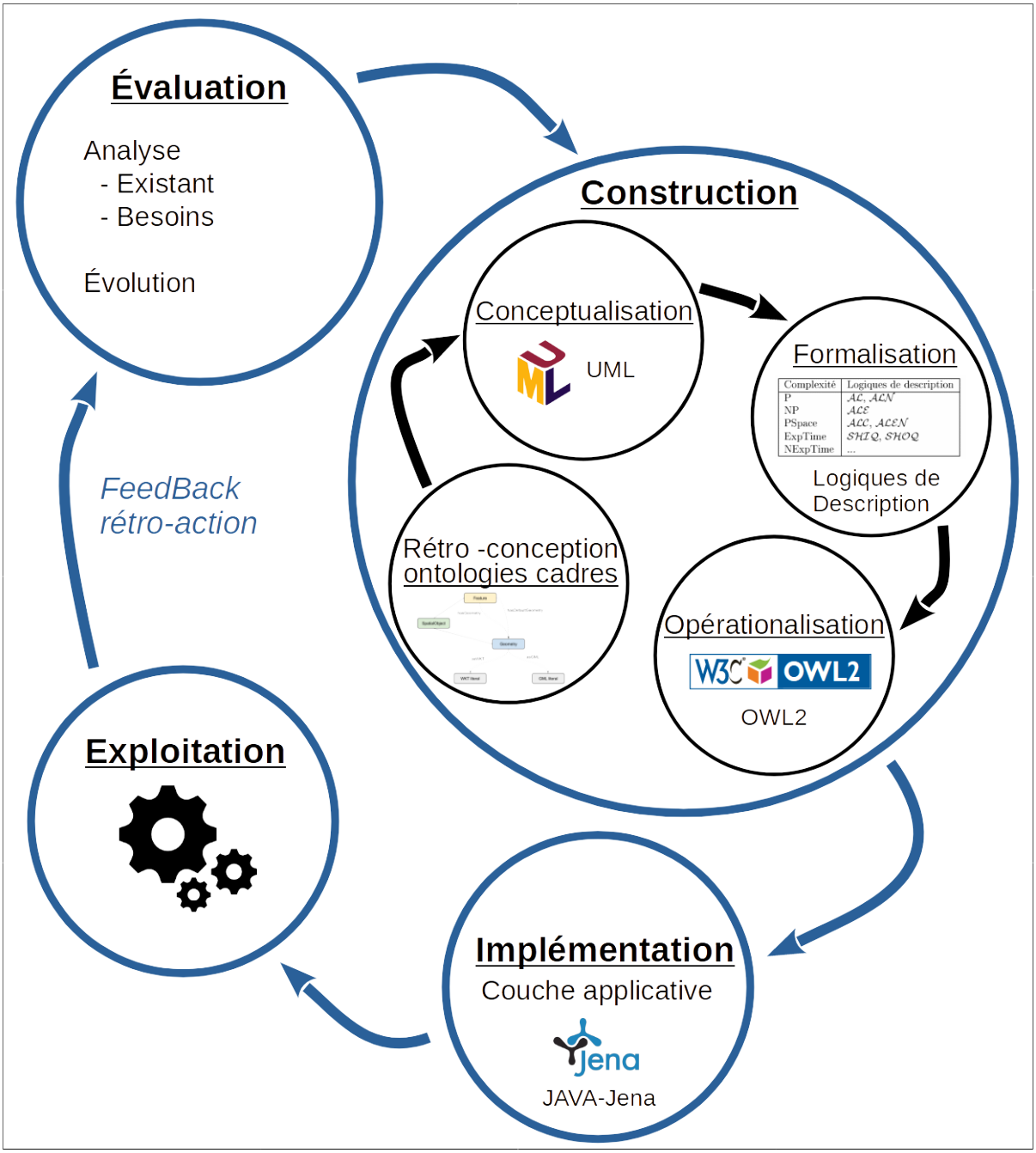


FIGURE 8.1 – *Démarche retenue pour la construction et la maintenance de l'ontologie "Agroforestry".*

8.2 Définition d'un modèle conceptuel

Il nous semblait naturel de mobiliser la notation objet UML [1], et en particulier les diagrammes structurels, pour tout ce concerne la phase de conceptualisation. Le langage de modélisation UML est en effet un standard préconisé par l'OMG [2] pour tout ce qui concerne les activités de conception orientée objet lors de développements logiciels. La notation UML a d'ailleurs fait l'objet de nombreux travaux en ingénierie des connaissances, pour soit faciliter la construction de modules ontologiques [VO et HOANG, 2020], soit guider des développements logiciels à l'aide d'un composant ontologique [KEET et FILLOTTRANI, 2015].
Le langage UML est dit semi-formel [BREU et al., 1997] dans le sens où la sémantique associée à ses éléments de modélisation reste parfois définie de manière incomplète et parfois ambiguë, et ne peut pas être prise en charge de manière totalement aboutie par des agents logiciels. Le langage OCL [OMG, 2012] (Object Constraint Language) vient compléter UML pour tendre vers plus de rigueur. Nous sommes donc conscients que tout ne pourra être modélisé à l'aide de diagrammes de classes.
Nous avons fait le choix de la modularité, et nous présentons de fait différents diagrammes de classes. Lors de la phase préliminaire d'évaluation des besoins, nous avons cherché à identifier les éléments clés pour la constitution d'un modèle pour une représentation volontairement ouverte des aménagements agroforestiers. De fait, nous avons défini un premier modèle pensé pour être réutilisable. Nous avons ensuite opté pour la réutilisation de modules ontologiques existants et prenant en charge des aspects aussi essentiels que l'espace, le temps ou l'observation.

8.2.1 Définition du modèle conceptuel : Agroforestry

Pour qu'un système sémantique, dédié aux aménagements agroforestiers, soit partagé par l'ensemble des acteurs de l'agroforesterie, voire par les acteurs d'autres domaines si l'on considère la nature systémique de l'agroforesterie, il semble essentiel de maintenir un haut niveau de généricité dans la définition de l'ontologie. La vision retenue est la suivante : une parcelle agroforestière évolue dans le temps et dans l'espace, et a pour sous-parties, des éléments biotiques et abiotiques, qui sont envisagés comme des éléments simples ou décomposables, qui possèdent des propriétés intrinsèques (caractéristiques propres), et qui nouent des interrelations de différentes natures. Nous nous sommes intéressés en particulier, aux relations spatiales qui sous-tendent les différents aménagements propres à chaque parcelle.
Tout élément, propriété ou relation de l'ontologie est potentiellement enrichi par un ou plusieurs termes/concepts issus d'ontologies terminologiques. Ces termes utilisés pour qualifier chaque élément d'un aménagement agroforestier, ainsi que leurs propriétés et leurs inter-relations, impliquent soit la mobilisation de l'ontologie terminologique présentée dans la section précédente, soit la réutilisation d'ontologies terminologiques existantes (AGROVOC [CARACCIOLO et al., 2013], PO (Plant Ontology) [JAISWAL et al., 2005], ou encore ENVO (the Environment Ontology) [BUTTIGIEG, MORRISON et al., 2013] et vont dans le sens d'une ouverture vers la sphère des sources de données ouvertes et liées, ou encore LOD (Linked Open Data) [JANOWICZ, HITZLER et al., 2014]. En effet, pour faciliter partage et analyse des données, il est impératif de réutiliser autant que possible, les concepts terminologiques déjà définis par des communautés d'expertise. L'avantage pour nous est double, d'une part la définition

1. UML : Unified Modelling Language. https://www.uml.org/.
2. OMG : Object Management Group. https://www.omg.org/.

de nouveaux concepts, possiblement redondants avec l'existant, est rendue inutile, et d'autre part la mobilisation de concepts faisant consensus offre une meilleure visibilité sur les données.

8.2.2 Une démarche mobilisant l'existant

Pour fournir à ce système les capacités de capitalisation et de partage des connaissances requises, nous avons opté pour une démarche privilégiant l'adossement à différentes ontologies cadres. D'après Thomas Gruber [Thomas R. GRUBER, 1995], l'extensibilité est une qualité importante des ontologies, qui doivent pouvoir se voir ajouter de nouveaux concepts sans que cela ne vienne remettre en question les modélisations antérieures. Dans cette optique, un petit nombre de concepts clés sont proposés, et subsumés par des concepts de plus haut niveau provenant d'ontologies cadres. Ce mécanisme d'extension nous permet de tirer parti de l'expertise d'autres domaines et de règles de bonne pratique pour la résolution de problèmes bien identifiés à un niveau plus général, et qui peuvent impacter la gestion des connaissances des aménagements agroforestiers.
Le choix est donc fait de présenter les efforts de conceptualisation associés à la construction de l'ontologie Agroforestry, au travers de différents diagrammes UML [RUMBAUGH et al., 2004]. Un premier diagramme de classes (Fig.8.2) retranscrit notre volonté de généralisation, et décrit les entités clés du modèle Agroforestry, et leurs inter-relations. Les diagrammes suivants restituent graduellement les notions d'importance empruntées aux ontologies cadre modélisant respectivement les notions d'espace, d'observation et de temps GeoSPARQL [BATTLE et KOLAS, 2012], SOSA [JANOWICZ, HALLER et al., 2019] et OWL-Time [HOBBS et PAN, 2004]. Un dernier diagramme (Fig.8.6) explicite l'adossement d'Agroforestry, aux ontologies de plus haut niveau.
Agroforestry est dédiée à une meilleure compréhension des aménagements agroforestiers et à ce titre, se consacre principalement à la description structurelle des entités présentes au sein d'une parcelle. La classe **StructuralElement** occupe en conséquence, une position privilégiée dans cette modélisation. Nous distinguerons les classes filles disjointes **SimpleElement** et **CompositeElement** qui permettent d'énoncer qu'un élément va admettre d'autres éléments structuraux en son sein, ou au contraire être un élément "feuille" du modèle. La classe **CompositeElement** est ainsi enrichie d'une association *hasElement* qui pointe vers la classe **StructuralElement**. L'association *hasElement* pourra être affinée dans l'ontologie (en premier lieu via des relations méréotopologiques). La relation *hasElement* est inspirée des travaux autour de GeoSPARQL et de BOT (Building Ontology Topology) [MADS RASMUSSEN et al., 2017], et respectivement des différents relations réflexives de geo:SpatialObject vers geo:SpatialObject et de la relation bot:hasElement. Ces éléments de modélisation sont déconnectés de toute considération temporelle. Prenons l'exemple d'un noyer hybride planté dans une parcelle agroforestière donnée. Ce noyer a pour propriété intrinsèque d'être un élément de la parcelle tout au long de son existence. Nous sommes conscients que cette vision simplificatrice peut être sujette à caution dans certains cas de figures. La classe **StructuralElement** se voit associer, via l'association *hasProperty*, à la classe **PropertyOfElement** qui va permettre d'enrichir les éléments structuraux de diverses caractéristiques amenées à évoluer dans le temps. La propriété *PropertyOfElement* est cette fois-ci réifiée et va être spécialisée en exploitant différents référentiels terminologiques à l'exemple de la taille totale de la plante dans Trait Ontology (TO:1000012 de nom "whole plant size"). À ce titre, la modélisation s'inspire de SOSA. Nous avons également défini une super-classe nommée **Resource** venant non seulement généraliser **StructuralElement** mais aussi **PropertyOfElement**. L'intérêt de cette super-classe est de pouvoir enrichir sémantiquement toute ressource du modèle avec des termes empruntés aux ressources

onto-terminologiques rendues disponibles sur le web. À cet effet, nous avons introduit la classe **Term** qui vient également spécialiser la classe **Resource** et dont les instances sont attendues de venir compléter de manière normalisée la description des autres ressources du modèle. Ainsi, les notes, présentes au sein du diagramme de classes UML de la figure 8.2, font état de termes tirés d'ontologies terminologiques qui peuvent être utilisés pour qualifier les éléments d'un aménagement ainsi que leurs propriétés et inter-relations. La propriété **dct:type** est empruntée au standard de métadonnées Dublin Core Terms [DCMI USAGE BOARD, 2006] afin de typer toute ressource, quelle qu'elle soit, y compris une ressource terminologique, décrite dans le modèle.

Le parti-pris dans la définition du modèle de connaissances d'Agroforestry est la réutilisation des ontologies et terminologies existantes. Seul un socle minimum d'organisation de la connaissance vient organiser l'existant afin de rendre cet existant exploitable dans le domaine de l'agroforesterie. Les éléments de modélisation des ontologies cadre mobilisées sont détaillés dans les sous-sections suivantes.

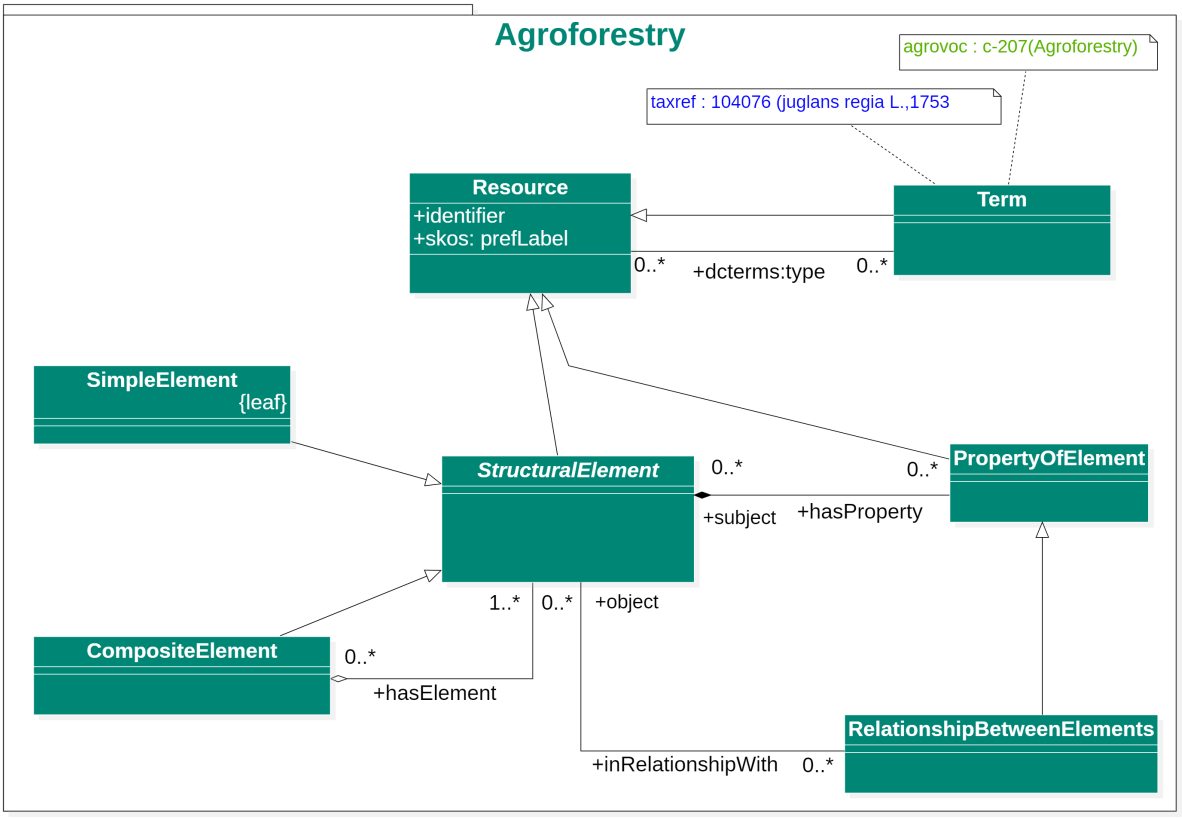


FIGURE 8.2 – *Conceptualisation de l'ontologie "Agroforestry".*

8.2.3 Intégration de l'ontologie SOSA

La première nécessité était de permettre aux agroforestiers de capitaliser les différentes observations faites sur le terrain. Si certaines propriétés de nos éléments agroforestiers possèdent des valeurs immuables (comme par exemple la localisation d'un site), d'autres propriétés (comme l'acidité d'un sol au travers de la mesure de son PH), sont quantifiées ou qualifiées à l'aide d'une procédure d'observation spécifique. Les valeurs collectées au travers de ces observations peuvent varier au cours du temps et sont dépendantes de la méthode de mesure. Cette variabilité doit absolument être

prise en compte pour obtenir un résultat rigoureux en cas de comparaison et d'analyse de données issues de multiples observations. À cet effet, l'ontologie cadre SOSA [JANOWICZ, HALLER et al., 2019], qui a pour élément central l'observation (Fig.8.3), est retenue. SOSA est le module central (mais néanmoins autonome) de l'ontologie SSN (Semantic Sensor Network) [HALLER et al., 2018] qui permet de décrire les capteurs (y compris humains) et les observations acquises par ces capteurs. Sa simplicité et son autonomie servent de pivot d'interopérabilité pour SSN mais aussi pour d'autres ontologies dédiées à l'observation telles que O&M [S. J. COX, 2016] et OBOE [MADIN et al., 2007].
SOSA place l'observation (**sosa:Observation**) au centre de son modèle : Une observation permet ainsi de renseigner la valeur d'une propriété descriptive pour un élément d'intérêt, à un instant ou intervalle de temps donné. L'observation est liée à un individu de la classe **sosa:FeatureOfInterest** par la relation *sosa:hasFeatureOfInterest*. La propriété observée de l'élément (classe **sosa:ObservableProperty**) est liée à l'observation par la relation *sosa:observedProperty* et la valeur de cette propriété est liée à l'observation par la relation *sosa:hasResult*. Dans le modèle Agroforestry, la classe **StructuralElement** vient spécialiser **sosa:FeatureOfInterest**, et va naturellement pouvoir bénéficier de toute la modélisation autour de la notion d'observation. Il est ainsi aisé d'organiser l'information autour de la croissance d'un arbre, en le considérant comme une instance de la classe **sosa:FeatureOfInterest** et en renseignant sa taille, à intervalles réguliers, au travers de plusieurs observations. La classe **ssn:Property** (super-classe de **sosa:ObservableProperty**) a également retenu notre attention et est étendue par la classe **PropertyOfElement** de manière à pouvoir décrire de manière complémentaire les propriétés engagées dans la description des éléments clés de l'agroforesterie. L'idée est de pouvoir disposer à plus long terme de couples éléments/collection de propriétés pertinents pour l'agroforesterie et ainsi en faciliter la réutilisation par la communauté. La classe **PropertyOfElement** est spécialisée à son tour par **RelationshipBetweenElements** afin de prendre en charge des relations autres que des relations structurelles entre éléments d'un aménagement agroforestier. Ces relations peuvent être valuées et datées dans le temps, et sont en capacité de capturer les interactions entre éléments biotiques retrouvés sur les mêmes parcelles. Les sites pilotes en agroforesterie permettront de mettre en exergue les effets de pratiques agroforestières, au moyen d'expérimentations menées sur des parcelles de test. Les classes **sosa:Sample** et **sosa:Sampling** vont dès lors, nous être également utiles ; une parcelle expérimentale pouvant être vue comme un individu **sosa:Sample**. Le diagramme de classes UML présenté en figure 8.3 reprend les principales classes de SOSA dont certaines font l'objet d'extensions dans le modèle Agroforestry.

8.2.4 Extension de l'ontologie GeoSPARQL et de l'ontologie OWL-Time

Une double dimension spatiale et temporelle est nécessaire à toute entité du modèle si l'on veut rendre compte de l'évolution d'un système où les relations de proximité dans le temps sont des plus importantes. En effet les aménagements agroforestiers et en particulier ceux de l'agrosylviculture sont de par leur nature des aménagements spatiaux, où la place de chaque élément (arbre, plante céréalière, ligne d'arbres dans son ensemble, . . .) a une importance primordiale dans le cadre des interactions plante-plante.
La localisation géographique des données d'intérêt est à même de rapprocher et de lier des données issues de domaines différents à l'exemple de l'hydrologie ou de la pédologie. Nous pouvons prendre pour exemple le cas d'une rivière qui pourrait jouxter une parcelle agroforestière et ainsi limiter grandement la compétition en eau entre

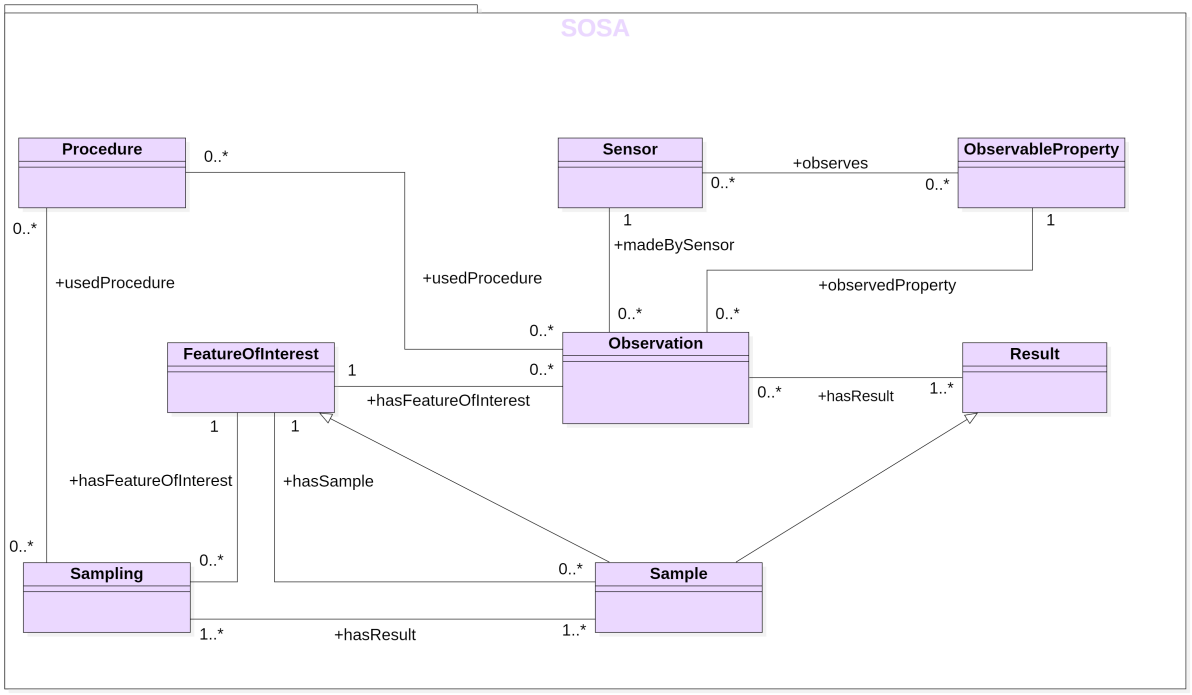


FIGURE 8.3 – *Diagramme de classes autour du module ontologique SOSA.*

arbres et plantes cultivées. GeoSPARQL [BATTLE et KOLAS, 2012] (Fig.8.4) est proposé par le consortium international de l'OGC (Open Geospatial Consortium), pour fournir les éléments nécessaires à la représentation et à l'interrogation de données spatialisées dans le contexte du web sémantique. GeoSPARQL est ainsi organisé de façon modulaire, et peut offrir des services aussi bien aux systèmes basés sur un raisonnement spatial qualitatif, qu'à ceux basés sur des calculs spatiaux quantitatifs. Les calculs spatiaux quantitatifs, à l'exemple d'un calcul de distance, nécessitent de connaître précisément la géométrie des éléments étudiés (géométrie Euclidienne), alors que les systèmes basés sur un raisonnement spatial qualitatif s'appuient sur des relations le plus souvent topologiques entre éléments. Ces dernières relations (à l'exemple de l'adjacence, de l'intersection ou de l'inclusion) sont, par exemple, décrites au travers du formalisme "region connection calculus" (RCC8) [RANDELL et al., 1992]. GeoSPARQL offre également la possibilité d'associer une à plusieurs géométries, à l'exemple de point ou de polygone, à toute entité géographique au travers de la classe **geo:Geometry** qui représente la super classe de toutes les géométries possibles. Dans le modèle Agroforestry, la classe **StructuralElement** est subsumée par **geo:Feature** et tire parti de toute la modélisation proposée par GeoSPARQL autour des géométries. La classe **geo:Geometry** du module cœur de GeoSPARQL est par la suite étendue par différentes préconisations dont le standard OGC, nommé Simple Features Access (sfa), dont certaines géométries sont illustrées dans le diagramme de classes en figure 8.4.

Le temps est une notion tout aussi importante dans le contexte des aménagements agroforestiers dans lequel, par exemple, la saisonnalité annuelle des cultures est confrontée à la vie pluriannuelle des arbres. Afin de mieux appréhender les interactions entre arbres et cultures dans le temps, nous devons définir de façon précise les dates de plantation et les périodes de présence spécifique de chaque élément agroforestier. Ainsi la classe **StructuralElement** subsumée par **sosa:FeatureOfInterest** va pouvoir mobiliser la modélisation retenue dans SOSA, à savoir entretenir une relation de type *sosa:phenomenonTime* avec une entité temporelle **time:TemporalEntity** de l'ontologie OWL-Time visible en

figure 8.5 [HOBBS et PAN, 2004]. La classe **time:TemporalEntity** est ensuite spécialisée en **time:Instant** et **time:Interval**, et il est ainsi aisé d'inscrire un élément d'aménagement agroforestier dans une durée basée sur un intervalle de temps précis ; ou bien de définir une estampille temporelle pour une observation spécifique définie avec le modèle SOSA pour ce même aménagement agroforestier (Fig.8.6).

Nous avons présenté le socle minimal de structuration des connaissances propres aux aménagements forestiers, ainsi que des portions de modèles conceptuels des ontologies cadres véhiculant trois notions qui nous intéressent en tout premier lieu, à savoir l'observation, l'espace et le temps. Il convient maintenant d'expliciter les articulations définies entre les différents modèles au travers de deux nouvelles figures. La figure 8.6 est un nouveau diagramme de classes intégrant SOSA, GeoSPARQL, OWL-Time et Agroforestry pour les besoins propres des aménagements agroforestiers. Le code couleur permet de repérer visuellement la provenance de chaque élément modélisé. La figure 8.7 est un diagramme de paquetages qui exprime les liens de dépendance du paquetage "Agroforestry" vis à vis des paquetages "SOSA", "OWL-Time" ou encore "GeoSPARQL". Ce diagramme facilitera par la suite la définition des liens d"import" de composants ontologiques dans Agroforestry.

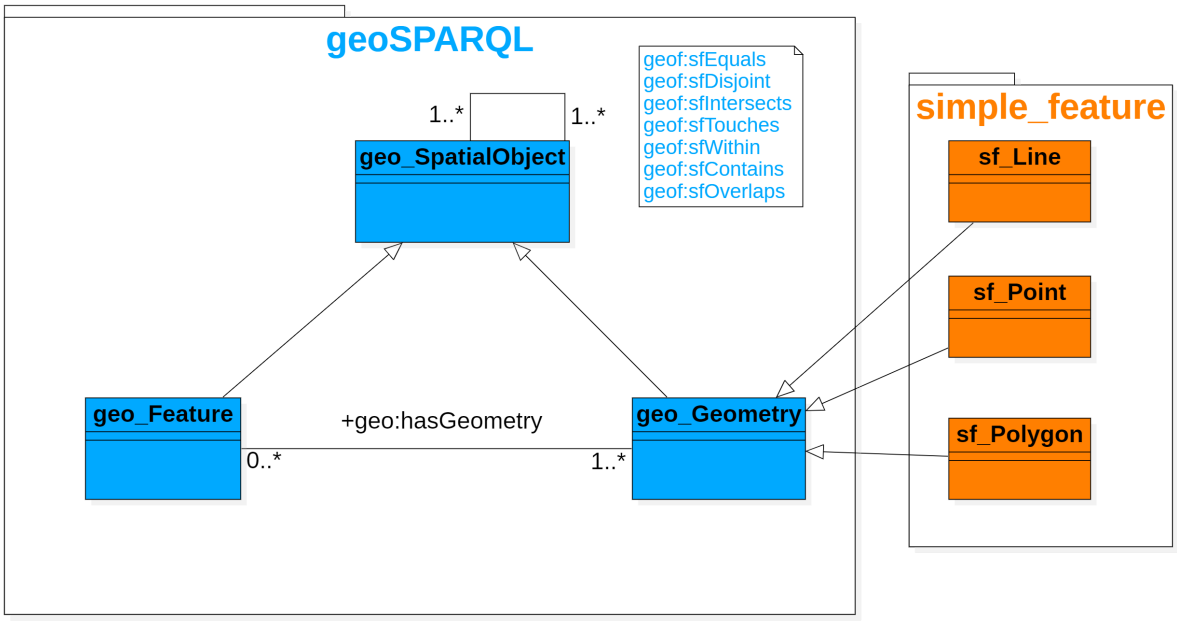


FIGURE 8.4 – *Diagramme de classes autour du module ontologique cœur de GeoSPARQL.*

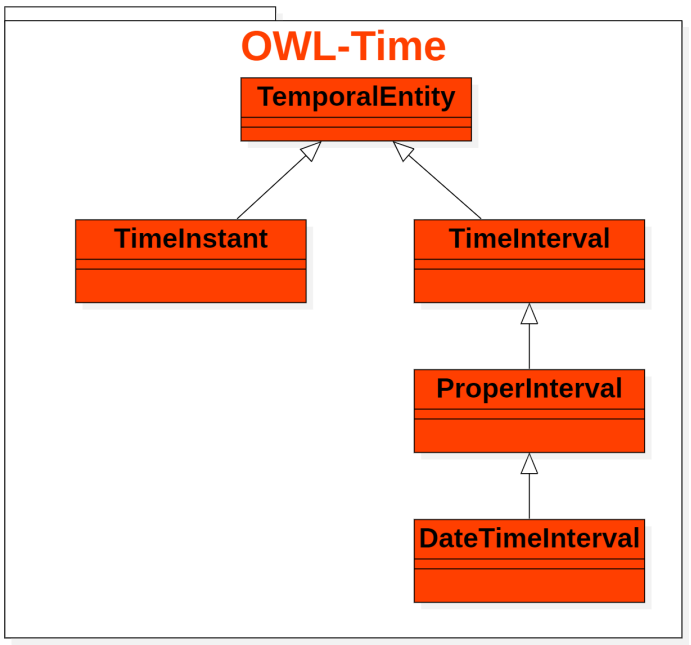


FIGURE 8.5 – *Diagramme de classes autour du module ontologique OWL-Time.*

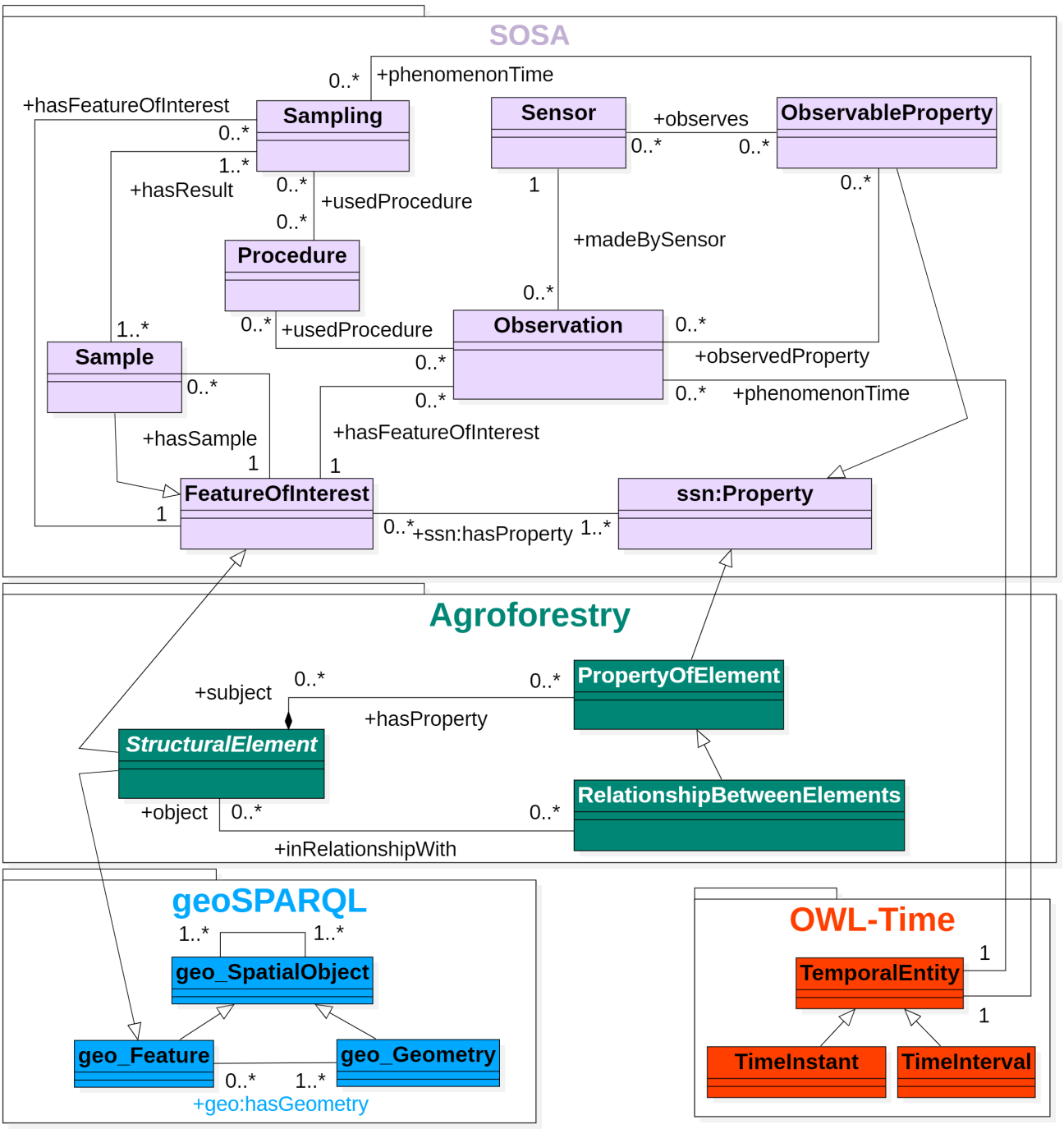


FIGURE 8.6 – *Diagramme de classes intégré.*

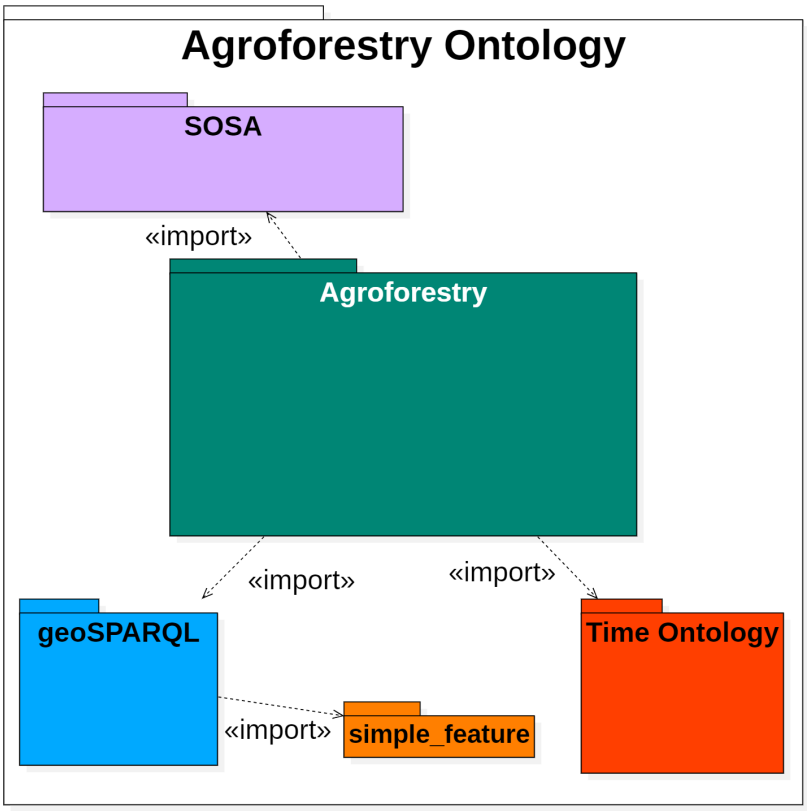


FIGURE 8.7 – *Diagramme de paquetages autour de l'ontologique cadre " Agroforestry ".*

Afin de confronter les différents diagrammes de classes définis aux objets concrets manipulés en agroforesterie, nous présentons trois diagrammes d'instances qui illustrent successivement les apports de notre modèle cadre, de SOSA et de GeoSPARQL. Ces diagrammes auraient pu être regroupés en un seul et même diagramme, puisqu'ils concernent les mêmes objets exemples. Nous avons cependant préféré présenter les trois illustrations séparément pour les rendre plus faciles d'abord, et distinguer ainsi les différentes perceptions prises en considération : organisation des aménagements, acquisition des observations, dimensions spatiale et temporelle.

Le diagramme de la figure 8.8 est une instanciation spécifique des différentes classes du modèle cadre avec un exemple correspondant à la description réelle de la parcelle agroforestière A3 du domaine de Restinclières. La parcelle A3 (PA3AF), instance de la classe **CompositeElement**, est liée par la relation *hasElement* à deux autres individus qui sont respectivement une plantation en ligne (PA3AFL01), et une culture en couloir (PA3AFCS01). La culture PA3AFCS01 est un objet de la classe **SimpleElement** alors que la plantation en ligne PA3AFL01 est un objet de la classe **CompositeElement**, dont un des individus composites est l'arbre numéro 14 (PA3AFL01A14) qui appartient à la classe **SimpleElement**. Les instances PA3AFL01 et PA3AFCS01 sont liées respectivement par les relations *hasProperty* et *inRelationshipWith* à l'instance I1 membre de la classe **RelationshipBetweenElements** qui décrit une relation entre la ligne d'arbres et la culture intercalaire qui est à sa proximité immédiate. L'arbre PA3AFL01A14 est lui lié par la propriété *hasProperty* à l'instance P1 membre de la classe **propertyOfElement** qui décrit une propriété de l'arbre. Tous ces éléments, qu'ils soient membres des classes **CompositeElement**, **SimpleElement**, **RelationshipBetweenElements** ou **propertyOfElement**, sont qualifiés par des membres de la classe **Term** auxquels ils sont liés par la relation *type* empruntée au standard de métadonnées Dublin Core. Les instances de la classe **Term** sont majoritairement des concepts SKOS provenant d'ontologies terminologiques comme Agrovoc, Nalt, AnaEE Thésaurus, ou bien provenant de notre composant terminologique AgroforestryVoc.

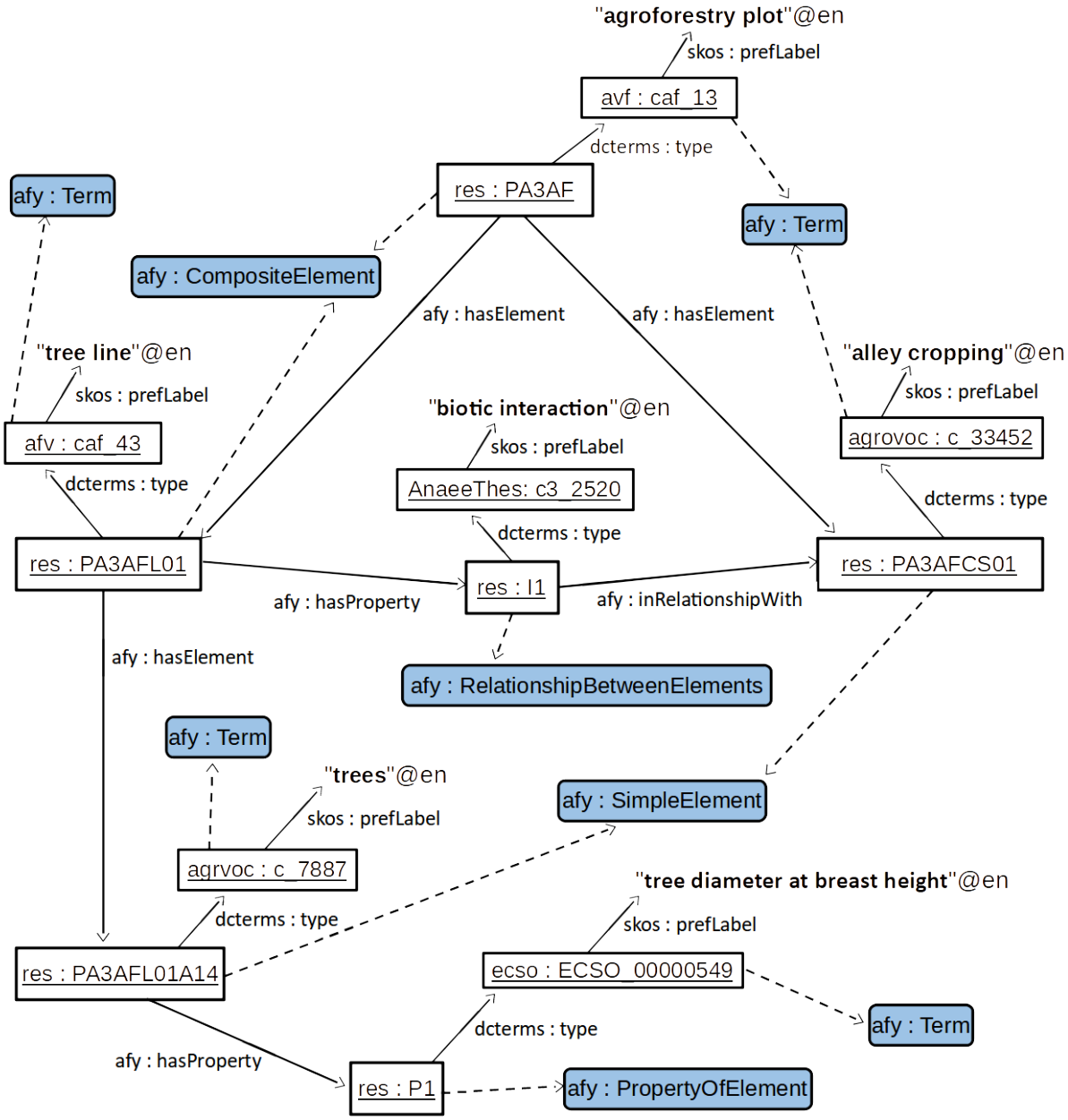


FIGURE 8.8 – *Diagramme d'instances illustrant le modèle cœur de l'ontologie cadre.*

Le diagramme de la figure 8.9 est une instanciation du modèle SOSA. Une observation (PA3AFL01A14_OBS_2015) instance de la classe **sosa:Observation** est réalisée sur l'arbre PA3AFL01A14, instance de la classe **SimpleElement**, mais aussi de la classe **sosa:FeatureOfInterest**, et concerne la valeur de la propriété P1, décrite comme étant le diamètre de l'arbre à hauteur de poitrine, et qui est membre des classes **propertyOfElement** et **sosa:ObervableProperty**. Cette observation est liée à l'individu PA3AFL01A14 par la relation *sosa:hasFeatureOfInterest* et à sa propriété P1 par la relation *sosa:observedProperty*. Elle est aussi liée par la relation *sosa:hasresult* à l'instance PA3AFL01A14_OBS_2015_Measure membre de la classe **sosa:Measure** qui exprime la mesure physique en centimètres de cette propriété ; par la relation *sosa:phenomenomTime* à l'instance PA3AFL01A14_OBS_2015_Time, membre de la classe **time:Instant**, qui exprime la temporalité de cette prise de mesure ; par la relation *sosa:usedProcedure* à l'instance forest_mensuration, membre de la classe **sosa:Procedure**, qui exprime la méthode utilisée pour cette prise de mesure ; et par la relation *sosa:madeBySensor* à l'ins-

tance Lydie_DUFOUR, membre de la classe **sosa:Sensor**, qui exprime le capteur ayant réalisé cette prise de mesure. En l'occurrence le capteur est ici une personne car l'instance Lydie_DUFOUR est aussi membre de la classe **foaf:Person**, ensemble de toutes les personnes défini dans l'ontologie FOAF [3] [GRAVES et al., 2007]

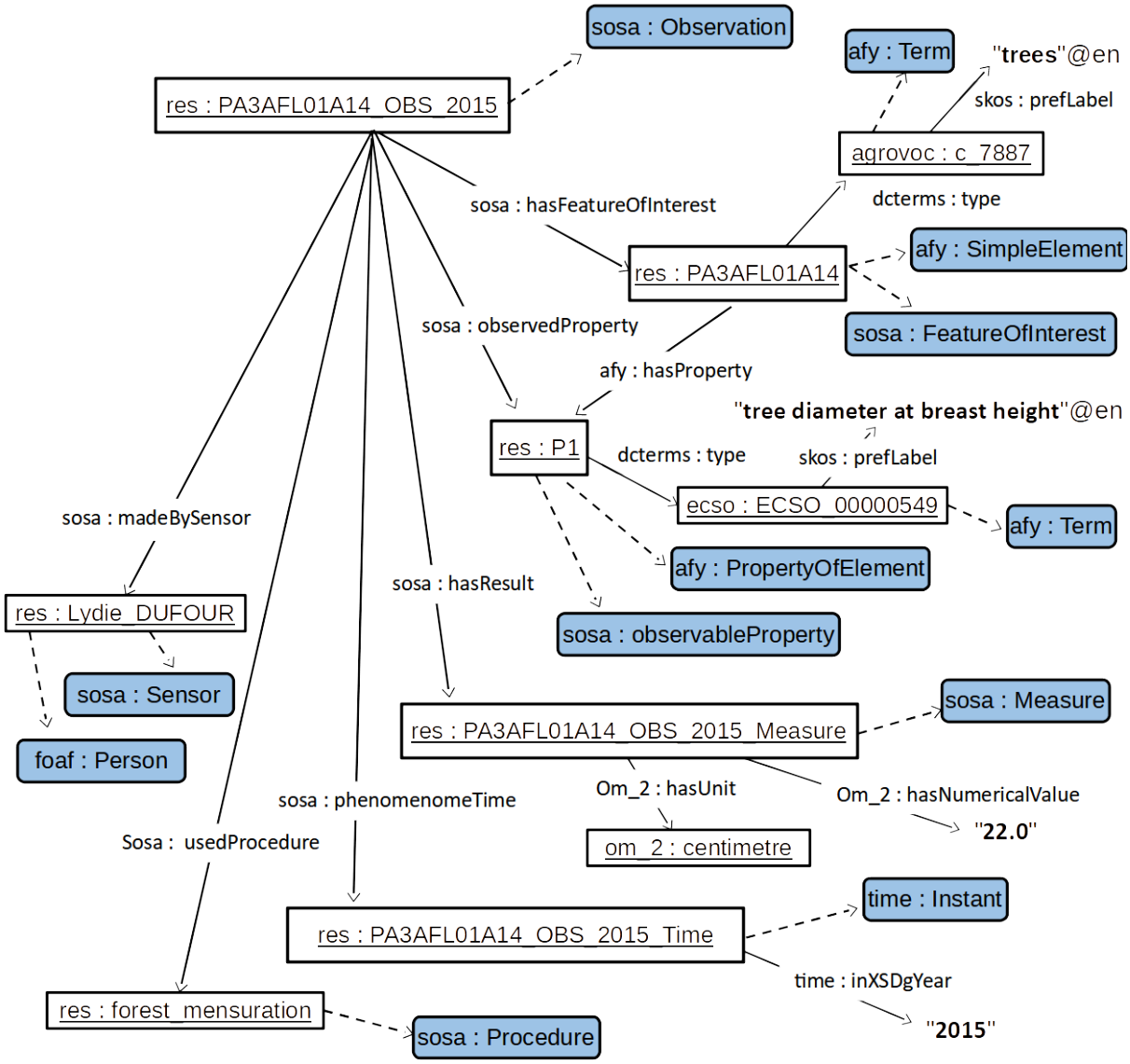


FIGURE 8.9 – *Diagramme d'instances illustrant l'utilisation du modèle SOSA.*

Le diagramme en figure 8.10 est une instanciation du modèle GeoSPARQL. La ligne de plantation PA3AFL01 et L'arbre PA3AFL01A14 sont tous deux membres de la classe **geo:Feature**. PA3AFL01 est lié par la relation *geo:hasDefaultGeometry* à l'instance PA3AFL01_GEOM membre de la classe **sf:Polygon** et PA3AFL01A14 est lié par cette même relation *geo:hasDefaultGeometry* à l'instance PA3AFL01A14_GEOM membre de la classe **sf:Point**, exprimant ainsi pour l'un la localisation d'une surface, et pour l'autre une localisation ponctuelle.

3. FOAF : Friend of a friend ontology. http://www.foaf-project.org.

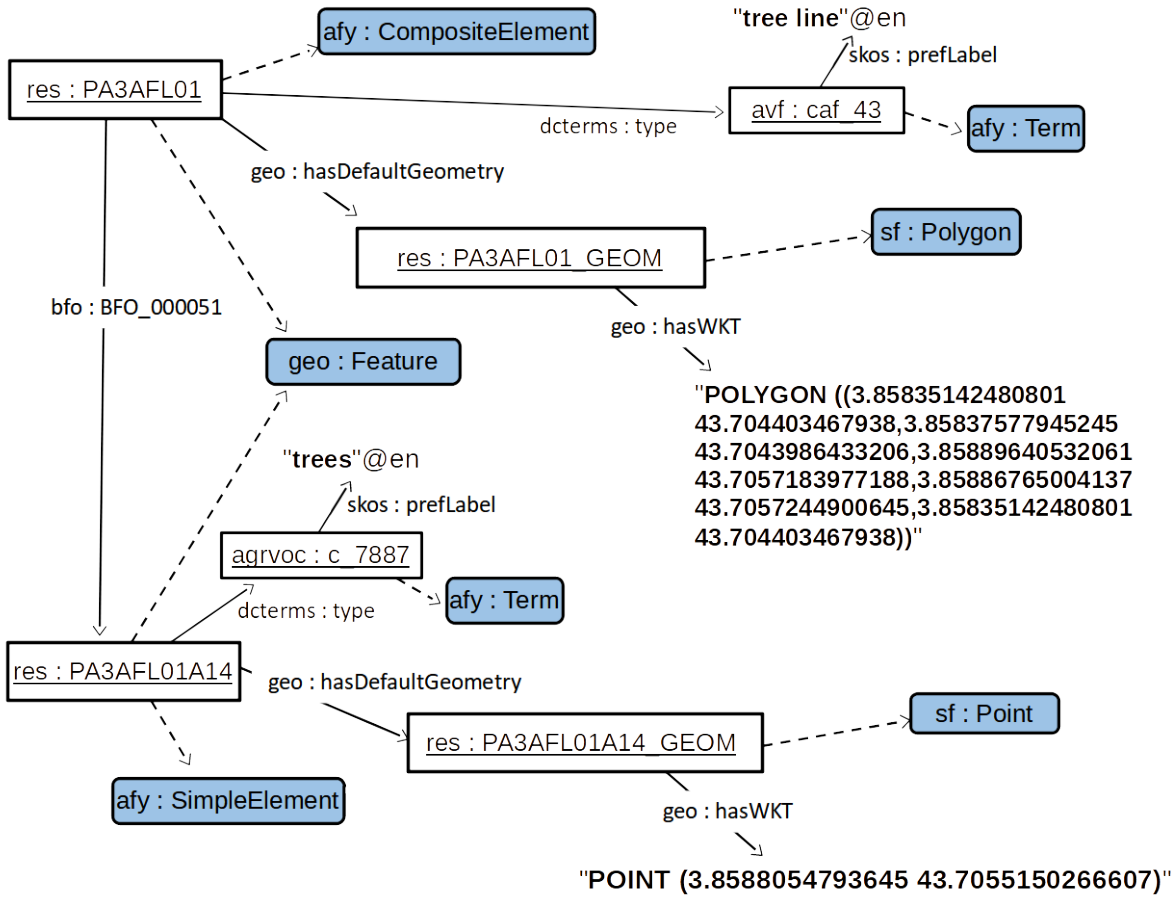


FIGURE 8.10 – *Diagramme d'instances illustrant l'utilisation du modèle GeoSPARQL.*

8.3 Passage au modèle formel

Le langage UML, a facilité la construction de modèles conceptuels pouvant être compris de tous. L'objectif est d'impliquer la communauté experte de l'agroforesterie. Produire des diagrammes de classes et d'objets accessibles aux acteurs du domaine est un élément important dans un contexte de partage des connaissances. Il s'agit maintenant de faire en sorte que les modèles soient aussi compris par les machines. À ce titre, il faut en préciser la sémantique et en permettre tout à la fois la vérification formelle, et l'application de mécanismes de raisonnement. Dans l'absolu, ces modèles pourraient aussi être rendus de plus en plus expressifs.

8.3.1 Utilisation des logiques de descriptions

Pour passer d'un diagramme de classes à un modèle formel, il est donc attendu de mobiliser un formalisme adapté. Notre travail fait le choix des logiques de description décrites précédemment au chapitre 5 à la section 5.4. Les logiques de description correspondent à une famille de formalismes logiques très exploitée actuellement. La section suivante décrit les fondements logiques de l'ontologie Agroforestry.

8.3.2 Les fondements logiques d'Agroforestry

Le socle de connaissances défini spécifiquement pour les aménagements agroforestiers est donc repris dans le formalisme habituellement dévolu aux logiques de description [BAADER, CALVANESE et al., 2003 ; KRÖTZSCH et al., 2014]. Le préfixe "afy" (ontology for AgroForestrY) sera le préfixe retenu pour l'ontologie Agroforestry. Les principaux axiomes d'inclusion et d'équivalence de la TBox (boîte terminologique qui contient l'arborescence de concepts de l'ontologie) sont listés en figure Fig.8.11. Agroforestry se compose essentiellement d'une arborescence de concepts clés, avec **afy:Resource** comme concept le plus général. Les concepts **afy:PropertyOfElement**, **afy:Term** et **afy:StructuralElement** viennent raffiner **afy:Resource**, et sont disjoints deux à deux. Le concept **afy:Term** découle d'une double motivation : il s'agit d'enrichir les éléments décrits dans une base de connaissance en agroforesterie avec des savoirs terminologiques complémentaires et normalisés, mais aussi et dans un second temps de dégager (en faisant évoluer le composant terminologique au besoin) un nouvel ensemble de concepts terminologiques pouvant s'appliquer avec succès à l'agroforesterie. Le concept défini **afy:CompositeElement** représente l'ensemble des éléments structuraux comptant au moins une relation de type tout-partie (*afy:hasElement*) avec tout autre élément structural.

afy:Resource $\sqsubseteq \forall$ dcterms:type.afy:Term $\sqcap \exists$ afy:identifier.xsd:anyURI

afy:PropertyOfElement $\sqsubseteq$ afy:Resource $\sqcap \neg$ afy:StructuralElement

afy:Term $\sqsubseteq$ afy:Resource $\sqcap \neg$ afy:StructuralElement

afy:StructuralElement $\sqsubseteq$ afy:Resource $\sqcap \forall$ afy:hasProperty.afy:PropertyOfElement

afy:CompositeElement $\equiv$ afy:StructuralElement $\sqcap \exists$ afy:hasElement.afy:StructuralElement $\sqcap \forall$ afy:hasElement.afy:StructuralElement

afy:SimpleElement $\equiv$ afy:StructuralElement $\sqcap \neg$ afy:CompositeElement

afy:RelationBetweenElements $\equiv$ afy:PropertyOfElement $\sqcap \exists$ afy:inRelationshipWith.afy:StructuralElement

FIGURE 8.11 – *Concepts d'Agroforestry.*

Pour ce qui concerne la RBox (conteneur des rôles de l'ontologie) présentée en figure Fig.8.12, la propriété *afy:hasElement* a une propriété inverse nommée *afy:elementOf*. Par ailleurs, les axiomes d'inclusion présentés en figure 8.13 qui viennent adosser Agroforestry aux ontologies GeoSPARQL et SOSA sont également présentés. Ainsi **afy:StructuralElement** , en tant qu'entité spatialisée d'intérêt, spécialise à la fois le concept **geo:Feature** et le concept **sosa:FeatureOfInterest**. Le concept **afy:PropertyOfElement** dérive du concept **ssn:Property**. Le concept **afy:Term** est à rapprocher de **skos:Concept**.

Nous présentons également en figure 8.14 un jeu d'assertions possibles (ABox) pour la TBox ainsi construite. Ce jeu d'assertions est dérivé du diagramme d'instances présenté en figure 8.8.

Le standard de description de connaissances OWL 2 s'adossant à la famille des logiques de description, il s'est avéré aisé de dériver le modèle ontologique en un modèle opérationnel.

afy:hasElement $\equiv$ afy:elementOf^{-}
afy:hasProperty $\equiv$ afy:propertyOf^{-}
$\top \sqsubseteq \leq 1$ afy:propertyOf
afy:hasRelationship $\equiv$ afy:inRelationshipWith^{-}
$\top \sqsubseteq \leq 1$ afy:inRelationshipWith
$\top \sqsubseteq \leq 1$ afy:identifier

FIGURE 8.12 – *Propriétés d'Agroforestry.*

afy:StructuralElement $\sqsubseteq$ geo:Feature
afy:StructuralElement $\sqsubseteq$ sosa:FeatureOfInterest
afy:PropertyOfElement $\sqsubseteq$ ssn:Property
afy:Term $\sqsubseteq$ skos:Concept

FIGURE 8.13 – *Adossement aux ontologies de plus haut niveau.*

afy:StructuralElement(res:PA3AF)
afy:StructuralElement(res:PA3AFL01)
afy:SimpleElement(res:PA3AFL01A14)
afy:SimpleElement(res:PA3AFCS01)
afy:RelationshipBetweenElements(res:I1)
afy:hasElement(res:PA3AF,res:PA3AFL01)
afy:hasElement(res:PA3AFL01,res:PA3AFL01A14)
afy:hasProperty(res:PA3AFL01,res:I1)
afy:inRelationshipWith(res:PA3AFCS01,res:I1)
dcterms:type(res:PA3AF,afv:caf_13)
dcterms:type(res:PA3AFL01,afv:caf_43)

FIGURE 8.14 – *Exemple de peuplement de la ABox*

8.4 Passage au modèle opérationnel

Pour faire de l'ontologie Agroforestry, un composant sémantique opérationnel, les langages du web sémantique préconisés par le W3C, RDF, RDFS et OWL 2 sont mis à contribution. Comme il a été vu au chapitre 6, ce choix repose sur les avantages de ces langages de données. Il s'agit de standards préconisés par le W3C, et leur utilisation est rendue sans cesse plus fréquente dans le monde de l'ingénierie des connaissances. Nous pouvons également penser que ces langages vont évoluer mais rester relativement pérennes dans le temps. Ils offrent la possibilité d'accéder à des composants ontologiques déjà existants et de les réutiliser aisément sous la forme de modules interconnectés sur le web.

Nous avons également tiré parti de l'API Jena pour une construction programmatique de l'ontologie en OWL. Jena permet de travailler sur diverses fonctionnalités associées à la manipulation de la base de connaissance (moteur de requêtage ARQ, moteur de règles RDF, classifieurs RDFS et OWL internes, classifieurs externes à l'exemple de Openllet (successeur open source de Pellet)), facilités pour définir des fonctions utilisateurs ; et nous avons donc privilégié l'usage de cette librairie. Le langage de requête SPARQL est également mobilisé pour la manipulation du contenu de l'ontologie. Nous avons construit un premier jeu de requêtes en guise de premier test du modèle de la base de connaissance (contenant une ABox très réduite définie à partir des assertions présentées au dessus). L'ontologie définie a été vérifiée formellement à l'aide du raisonneur Pellet [SIRIN et PARSIA, 2004 ; SIRIN, PARSIA et al., 2007 ; MISHRA et KUMAR, 2011] et a été exploitée pour tirer parti de premiers mécanismes de raisonnement. Nous avons par exemple retrouvé, au travers de l'activité de "réalisation" des instances, les individus venant matérialiser la classe afy:CompositeElement (res:PA3AF, res:PA3AFL01). Nous avons aussi fait jouer la propriété algébrique de transitivité s'appliquant à hasElement, qui a permis de retrouver que res:PA3AF admet aussi comme élément res:PA3AFL01A14 puisque res:PA3AF a pour élément res:PA3AFL01 et que res:PA3AFL01 a pour élément res:PA3AFL01A14.
La figure 8.15 montre un extrait de l'ontologie OWL Agroforestry donnée dans la sérialisation RDF/XML à partir des formules définies en FOL (logique de premier ordre). Des axiomes d'inclusion décrivent respectivement afy:Resource, afy:PropertyOfElement, afy:Term et afy:StructuralElement.

```
<rdf:RDF
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
                                        afy:Resource ⊑ ∀ dcterms:type.afy:Term ⊓ ∃ afy:identifier.xsd:anyURI
<owl:Class rdf:about="http://www.afy.fr/Agroforestry/Resource">
   <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="http://purl.org/dc/terms/type"/>
      <owl:allValuesFrom rdf:resource="http://www.afy.fr/Agroforestry/Term"/>
    </owl:Restriction>
   </rdfs:subClassOf>
   <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty>
        <owl:DatatypeProperty rdf:about="http://www.afy.fr/Agroforestry/identifier"/>
      </owl:onProperty>
      <owl:someValuesFrom rdf:resource="http://www.w3.org/2001/XMLSchema#anyURI"/>
    </owl:Restriction>
   </rdfs:subClassOf>
  </owl:Class>                  afy:PropertyOfElement ⊑ afy:Resource ⊓ ¬ afy:StructuralElement

  <owl:Class rdf:about="http://www.afy.fr/Agroforestry/PropertyOfElement">
   <rdfs:subClassOf>
    <owl:Class rdf:about="http://www.afy.fr/Agroforestry/Resource"/>
   </rdfs:subClassOf>
   <owl:disjointWith>
    <owl:Class rdf:about="http://www.afy.fr/Agroforestry/StructuralElement"/>
   </owl:disjointWith>
   <rdfs:subClassOf rdf:resource="http://www.w3.org/ns/ssn/Property"/>   afy:PropertyOfElement ⊑ ssn:Property
  </owl:Class>

  <owl:Class rdf:about="http://www.afy.fr/Agroforestry/Term">   afy:Term ⊑ afy:Resource ⊓ ¬ afy:StructuralElement
   <rdfs:subClassOf>
    <owl:Class rdf:about="http://www.afy.fr/Agroforestry/Resource"/>
   </rdfs:subClassOf>
   <owl:disjointWith>
    <owl:Class rdf:about="http://www.afy.fr/Agroforestry/StructuralElement"/>
   </owl:disjointWith>
   <rdfs:subClassOf>                                                  afy:Term ⊑ skos:Concept
    <owl:Class rdf:about="http://www.w3.org/2004/02/skos/core#Concept"/>
   </rdfs:subClassOf>
  </owl:Class>
</rdf:RDF>
```

FIGURE 8.15 – *Extrait pour la constitution de l'ontologie cadre Agroforestry avec les langages RDF et OWL. les formules logiques transcrites sont présentées en encadré*

8.4.1 Premiers essais d'inférence

Le socle ontologique construit est volontairement générique, et nous pouvons l'enrichir par de nouveaux concepts définis de manière aboutie (axiomes d'équivalence) en fonction des besoins. L'objectif est de tester en tout premier lieu la capacité du modèle ontologique à prendre en charge les perceptions des agroforestiers. Nous cherchons donc à exprimer les concepts empruntés aux classifications introduites par P.K.R. Nair et E. Torquebiau, qui font sens pour la communauté. Si nous considérons l'exemple du concept "Agroforesterie en disposition linéaire" introduit par E. Torquebiau, nous pouvons retrouver toutes les parcelles tombant sous cette dénomination à l'aide de l'axiome qui pourrait s'énoncer ainsi : une parcelle agroforestière adoptant un arrangement linéaire (LinearArrangementAgroforestryPlot) est un élément structurel étiqueté par le terme afv:caf_13 (parcelle agroforestière ou encore agroforestry plot) et qui possède au moins un élément étiqueté par le terme afv:caf_37 (alignement régulier ou encore regular alignment). Les nouveaux concepts introduits sont définis au sein d'un module ontologique à part entière qui vient étendre l'ontologie AgroForestrY, et

qui prend le nom AgroForestrY-Ext (qualificatif afy-ext). Nous définissons de manière aboutie deux concepts : AgroforestryPlot et LinearArrangementAgroforestryPlot. Ces déinitions mobilisent des restrictions sur des concepts nominaux (ici respectivement {afv:caf_13} et {afv:caf_37} (mise à contribution de la restriction owl:hasValue dans OWL 2) sur lequels vont s'appuyer les mécanismes d'inférence [SIRIN, GRAU et al., 2006]

afy-ext:AgroforestryPlot ≡ afy:StructuralElement ⊓
∃ dcterms:type.{afv:caf_13}

FIGURE 8.16 – *Concept défini pour classer les parcelles agroforestières*

afy-ext:LinearArrangementAgroforestryPlot ≡ afy-ext:AgroforestryPlot⊓
∃ afy:hasElement.(afy:StructuralElement ⊓ ∃ dcterms:type.{afv:caf_37)}

FIGURE 8.17 – *Concept défini pour classer les parcelles agroforestières en disposition linéaire*

La représentation OWL 2 en syntaxe turtle du concept *LinearArrangementAgroforestryPlot* est également présentée ci-dessous (voir figure 8.18). *LinearArrangementAgroforestryPlot* a pour classe équivalente, la classe correspondant à l'intersection de la classe *AgroforestryPlot* et d'une classe construite sur une restriction qui correspond à l'ensemble des individus qui sont associés via la propriété *hasElement* à des individus qui sont membres de la classe *StructuralElement* et qui sont décrits par le concept nominal {afv:caf_37} (regular alignment). La vision très générale d'une agroforesterie centrée

```
afy-ext:LinearArrangementAgroforestryPlot
a                       owl:Class ;
owl:equivalentClass     [ a                     owl:Class ;
                          owl:intersectionOf    ( afy-ext:AgroforestryPlot
                                                  [ a                   owl:Restriction ;
                                                    owl:onProperty      afy:hasElement ;
                                                    owl:someValuesFrom  [ a                   owl:Class ;
                                                                          owl:intersectionOf  ( afy:StructuralElement
                                                                                                [ a                 owl:Restriction ;
                                                                                                  owl:hasValue      afv:caf_37 ;
                                                                                                  owl:onProperty    dcterms:type
                                                                                                ]
                                                                                              )
                                                                        ]
                                                  ]
                                                )
                        ] .
```

FIGURE 8.18 – *Equivalence de classe pour le concept* LinearArrangementAgroforestryPlot *en syntaxe Turtle*

sur des éléments structurels qui peuvent être typés de différentes manières, permet de n'avoir à exprimer qu'un socle de connaissances restreint qui peut ensuite être précisé à l'aide de référentiels terminologiques. Ce découplage est à la fois léger en terme de représentation et adaptable à tous les systèmes agroforestiers. Différentes extensions en fonction des besoins peuvent venir se surajouter au socle de connaissances.
Les logiques de description présentent certaines limites dans leur capacité à représenter certains concepts. C'est le cas notamment lorsqu'il s'agit d'énoncer des axiomes qui introduisent des cycles entre les individus ciblés par la définition. Pour donner un exemple, il ne va pas être possible de définir le concept *Tree_Line*, en indiquant qu'une ligne d'arbres contient au moins trois arbres et que tous les arbres, inclus dans la

ligne, partagent une intersection avec un même segment de droite. Il est juste possible d'énoncer que la ligne d'arbres contient au moins trois arbres, mais rien ne permet de contraindre ces individus à être positionnés sur un même segment. Nous allons retrouver ce problème avec OWL 2. Le problème peut cependant être contourné par l'ajout de règles d'inférence déductive au modèle de connaissances. Plusieurs langages de règles sont disponibles à l'exemple de SWRL [O'CONNOR et al., 2005 ; HASSANPOUR et al., 2009] qui combinent avec profit à la fois OWL et RuleML. Les règles SWRL qui viennent étendre les axiomes OWL, s'apparentent à des clauses de Horn et prennent la forme d'une implication entre un corps et une tête.

La plateforme Jena, que nous utilisons dans ce travail, dispose d'un moteur hybride en charge de l'exécution des règles Jena (ou Jena rules) qui lui sont propres et qui sont reconnus pour leur performance [4]. Il est composé de deux moteurs distinct. Le premier et un moteur de chaînage avant de type RETE [FORGY, 1989] et le second un moteur de chaînage arrière de type LP (Logic Programming) [KOWALSKI, 1988 ; BARAL et GELFOND, 1994].
Les règles Jena s'apparentent également aux clauses de Horn (tête et corps) [HORN, 1951] et ont une syntaxe proche des règles SWRL. L'écriture des règles Jena est plus concise que celle des règles SWRL. Les règles Jena peuvent aussi être construites sur la base de mécanismes de chaînage avant (forward chaining) [DELAHAYE, 1995] comme de chaînage arrière (backward chaining) [RUSSELL et al., 2010] [AL-AJLAN, 2015] et de combinaison des deux (hybrid) (voir figure 8.19). L'usage de fonctions définies par l'utilisateur y est également facilitée. Nous avons donc préféré exploiter les règles Jena pour le large éventail de règles pouvant être définies dans une base de règles Jena.
Les clauses de Horn peuvent être considérées comme des implications ou des déductions énoncées sous la forme de règles "SI ... ALORS ... " qui permettent de déduire de nouveaux faits à partir de faits existants. Un fait est défini comme un littéral qui est une proposition élémentaire étant soit vraie, soit fausse. Nous reprenons la définition des parcelles agroforestières au moyen d'une règle Jena : nous indiquons alors que si un individu quelconque noté ?af satisfait les faits suivants "?af appartient à la classe StructuralElement" et "?af a pour type caf_13" alors ?af est un individu de la classe AgroforestryPlot.

Il est à noter, qu'au regard de notre base de triplets exemple, le moteur d'inférence de Jena permettrait de déduire un nouveau triplet: (res:PA3AF rdf:type afy-ext:AgroforestryPlot). Il en irait de même pour un classifieur comme Pellet et l'axiome afy-ext:AgroforestryPlot.
La règle donnée en exemple exploite du chaînage avant (orienté données).

4. http://jena.apache.org/documentation/inference/index.html

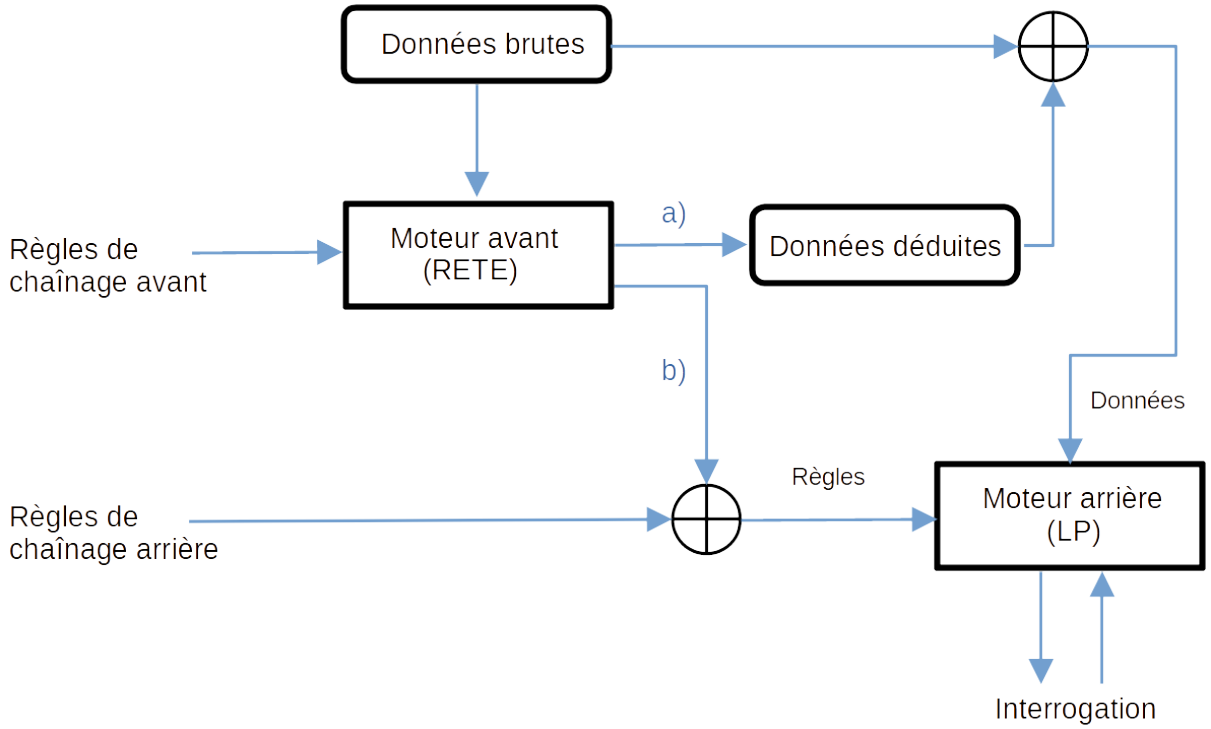


FIGURE 8.19 – ***Moteur hybride Jena*** *: a) Le moteur de chaînage avant s'exécute et maintient un pool de données déduites . b) Si les règles de chaînage avant affirment des règles en chaînage arrière, celles ci seront instanciées préalablement par le moteur avant, puis transmises au moteur arrière.*

```
[AgroforestryPlot_Rule:
        (?af rdf:type afy:StructuralElement)
        (?af dcterms:type afv:caf_13)
        ->
        (?af rdf:type afy-ext:AgroforestryPlot)
]
```

FIGURE 8.20 – *Règle Jena apparentée à l'axiome OWL 2 AgroforestryPlot*

L'objectif de cette section était de dérouler les trois phases sous-tendant la construction de l'ontologie de domaine. Nous avons privilégié une démarche s'inspirant du génie logiciel pour aboutir à la définition d'un module ontologique dédié à l'agroforesterie et qui réutilise des ontologies cadres existantes. Il nous semble important de réutiliser, mais aussi de rendre possible la réutilisation de notre composant ontologique. À cet effet, les standards du web de données sont mobilisés. Il s'agit maintenant d'évaluer la qualité de l'ontologie construite. La section suivant est entièrement consacrée à un cas d'étude grandeur nature. L'ontologie sera confrontée à la réalité du terrain, et pourra faire l'objet d'une maintenance évolutive si nécessaire.

IX

Troisième contribution : Une application pour la validation de notre démarche

9.1 Validation et confrontation du modèle avec la réalité 149
9.1.1 Le domaine de Restinclières 151
9.1.2 Description de l'étude 155
9.1.3 Diagrammes de cas d'utilisation 157
9.1.4 Diagrammes d'instances du modèle Agroforestry 160
9.1.5 Implémentation du modèle Agroforestry 165
9.1.6 L'application AOBRA 173
9.1.7 Mode opératoire avec les axiomes OWL 175
9.1.8 Mode opératoire avec l'application de règles Jena 180
9.1.9 Études réalisées avec AOBRA et leurs résultats 199
9.2 En ce qui concerne la validation et la confrontation de notre modèle avec la réalité 210

9.1 Validation et confrontation du modèle avec la réalité

Pour valider le modèle Agroforestry dans son ensemble, il n'y avait rien de mieux que la confrontation avec un cas d'usage réel. Nous avons donc réalisé un cas d'étude portant sur des aménagements agroforestiers du domaine de Restinclières situé sur la commune de Prades le Lez, dans le département de l'Hérault en France. En règle générale, des modèles de simulation numérique, agricoles ou forestiers, voire agroforestiers [LUEDELING et al., 2016 ; P. J. BURGESS et al., 2019] viennent faciliter les prises de décisions dans la gestion des systèmes de production agricole et sylvicole. Les besoins couverts vont de la gestion quotidienne des exploitations [HOCHMAN et al., 2009] à la simulation de conséquences sur la production face à des aléas pressentis à l'exemple des changements climatiques [HOLZWORTH et al., 2014 ; CHENU et al., 2017]. Les attendus face aux modèles portent en particulier sur :

- La synthèse des informations sur le fonctionnement des systèmes. Les sorties des modèles pourront être des vues agrégées sur les données expérimentales et faciliteront l'identification des interactions qui se nouent entre les différents éléments modélisés.

- La simulation d'effets relatifs aux changements de nombreux facteurs interagissant au sein de ces systèmes comme la qualité des sols, les espèces utilisées, le climat ou les modes de gestion des cultures (labour, émondage, irrigation, ...).

- Le partage des connaissances entre les chercheurs et les professionnels de terrain pour identifier les principaux facteurs de production.

- La mise à disposition d'une aide à la décision pour les acteurs politiques impliqués dans les politiques territoriales.

Il existe différents modèles de simulation à l'exemple de Hi-sAFE [DUPRAZ, WOLZ et al., 2019] qui est un modèle numérique qui simule la la croissance des arbres et des cultures intercalaires implantés sur une parcelle agroforestière. Hi-sAFE modélise également les mécanismes de compétition, qui se jouent entre les arbres et les cultures, en ce qui concerne notamment l'apport en lumière, en azote et en eau. Hi-sAFE cherche par le biais de simulations régulières sur une plage temporelle dépassant les 10 ans, à extrapoler l'évolution des rendements des cultures intercalaires et de la croissance des arbres. Le modèle Hi-sAFE a été expérimenté sur le domaine de Restinclières, et a permis de tirer différents enseignements sur les intérêts comparés des parcelles agroforestières, agricoles et forestières, en terme de services rendus.
Il n'existe pas à notre connaissance de modèles qualitatifs qui aient été exploités pour mieux comprendre les avantages et/ou inconvénients inhérents aux parcelles agroforestières.
L'objectif est donc de démontrer qu'en intégrant de la connaissance experte dans un système, ce système peut permettre de résoudre des problèmes qui usuellement nécessitent de l'expertise humaine. Un modèle qualitatif n'est pas forcément concurrent à un modèle quantitatif comme Hi-sAFE, mais peut venir en complémentarité, en dégageant des connaissances qui peuvent être différentes et/ou expliquées.
Nous cherchons donc à valider notre modèle de connaissances en le confrontant notamment à ce qu'est attendu de réaliser Hi-sAFE. Pour ce faire, nous l'avons testé au travers d'une couche applicative pour observer dans quelles mesures, il pouvait répondre aux objectifs énoncés ci-dessus. L'intérêt de l'approche est aussi de disposer d'une volumétrie de données importante qui sera à même de permettre d'évaluer la performance et la robustesse du système de connaissances mis en place.
Nous avons tout d'abord constitué une base de connaissance conforme au modèle Agroforestry, et alimentée avec tout un jeu de relevés de mesures dendrométriques concernant les arbres, et de rendements de cultures intercalaires obtenus sur différentes parcelles de ce site et sur plusieurs années consécutives. Puis dans une seconde étape, nous avons construit une couche applicative nommée AOBRA (**A**pplication layer with **O**ntology-**B**ased **R**easoning for **A**groforestry) qui vient étendre et outiller la base de connaissance. La façon d'exploiter la base de connaissance au travers de règles faisant jouer des fonctions statistiques pourra paraître décalée, mais il s'agissait pour nous de montrer que notre démarche est adaptable à toute problématique relevant de l'agroforesterie.
Nous présentons très brièvement les expérimentations conduites sur le domaine départemental de Restinclières qui ont eu pour résultat le jeu de données exploité. Nous revenons sur les différentes étapes de conception qui nous ont amenés à constituer le jeu de données "sémantisé". Nous présentons en dernier lieu la couche applicative AOBRA, ainsi que des exemples de sortie du système.

9.1.1 Le domaine de Restinclières

Le site de Restinclières s'étend sur une superficie de 215 hectares, et appartient au Département de l'Hérault, qui en a fait l'acquisition en 1990. Il est bordé par deux cours d'eau : le Lez et le Lirou. Ce site est utilisé pour des expérimentations agroforestières (12000 arbres plantés pendant les hivers 94-95 et 95-96) avec le concours de l'INRA, des Ecoles d'Agronomie de Montpellier et de Toulouse, de la Chambre d'Agriculture de l'Hérault, du Centre Régional de la Propriété Forestière, d'associations et de bureaux d'études indépendants. Les expérimentations sur ce domaine ont pour but d'étudier différents aménagements agroforestiers pour valider scientifiquement certaines pratiques de l'agroforesterie notamment pour l'agrisylviculture. Plusieurs parcelles agroforestières sont aménagées avec des aménagements spatiaux différents (variation de la distance entre les lignes d'arbres et entre les arbres au sein d'une ligne) et diverses essences d'arbres (noyer hybride, cornier, paulownia, peuplier,...). Des parcelles témoins sont également présentes pour comparer la pousse des arbres sans cultures (témoins forestiers) et celle des cultures sans les arbres (témoins agricoles).

Les différentes parcelles et cultures de l'étude

La parcelle agroforestière A3 et son témoin forestier, la parcelle A4, sont au centre de l'étude. Les données des parcelles A3 et A4, ont été collectées entre 2015 et 2018, par Lydie Dufour ingénieur de recherche à l'INRA et Serge-Stéphane Kouakou assistant de recherche au CIRAD. Les résultats nous ont été fournis au moyen d'un classeur Excel. Les arbres étant visibles depuis le logiciel de cartographie "Google Earth" [1], il a ainsi été rendu possible d'extraire les coordonnées GPS de chaque arbre ainsi que celles des différentes subdivisions du site (parcelle agroforestière (Fig. 9.1) et témoin forestier (Fig. 9.2)) . Les coordonnées géographiques du Lez ont été obtenues auprès du Services d'administration nationale des données et référentiel sur l'eau (SANDRE) [2].



FIGURE 9.1 – *Plan de la parcelle agroforestière A3 du domaine de Restinclières.*

1. Google Earth : Le globe terrestre dit le plus détaillé au monde. https://www.google.fr/intl/fr/earth.

2. SANDRE : Service d'aministration nationale des données et référentiels sur l'eau. http://www.sandre.eaufrance.fr.



FIGURE 9.2 – *Plan de la parcelle témoin forestier A4 du domaine de Restinclières.*

La Parcelle agroforestière

La parcelle est constituée de 8 lignes d'arbres plantées parallèlement au Lez orientées Nord Sud, et espacées entre elles de 13 mètres. Le sol est de type alluvial profond.
Un plan de répartition des arbres est visible en annexe A sur la figure A.1 qui montre la position relative de chaque arbre.
Les arbres étudiés dans cette étude sont des noyers communs (*Juglans regia*). Sur chaque ligne, les noyers ont été plantés en hiver 1994-1995, et au moins la moitié d'entre eux ont survécu sur chaque ligne, jusqu'à la prise des mesures entre 2015 et 2017.

Un tableau récapitulatif (Fig. 9.3) fournit pour chaque arbre, l'âge de l'arbre (colonne "Année") , l'année de la mesure (colonne "An civil") , le numéro de la ligne (colonne "Ligne") , la position dans la ligne (colonne "Colonne"), le diamètre de l'arbre pris à hauteur de poitrine (1,3 mètres) (colonne "D130"), la hauteur d'élagage (hauteur des premières branches) (colonne "HE en cm"), et la hauteur totale de l'arbre (colonne "H en cm"). Les coordonnées GPS au format WKT (Well-Known Text) des arbres ont été extraites de la plateforme de visualisation spatiale Google Earth, et rajoutées dans la colonne "WKT".

Le témoin forestier

Le témoin forestier jouxte la parcelle agroforestière et est composé de onze lignes d'arbres, elles aussi parallèles au Lez. Le plan de répartition de la parcelle A4 est présenté en annexe A sur la figure A.2. Un tableau récapitulatif fournit des valeurs de mesure pour des propriétés identiques à celles déjà observées pour les arbres de la parcelle agroforestière. L'intérêt du témoin forestier est de permettre de comparer l'évolution de la croissance de la population d'arbres dans des conditions de sylviculture classique, par opposition avec des conditions d'aménagement agroforestier.

Le témoin agricole

Pour le témoin agricole, une coupe planifiée des arbres a été opérée au sein de la parcelle A3 du rang 45 au rang 50 (zone délimitée en rouge)(Fig. 9.4). Les mesures de

Mesures des noyers de la parcelle A3 (Agroforesterie)

Année	2015	2016	2017
n° année	21	22	23

				cm	cm	cm	
Année	**An civil**	**Ligne**	**Colonne**	H en cm	HE en cm	D130	WKT
21	2015	1	10	1370	435	32,5	POINT (3.85885733993818 43.705661294259)
21	2015	1	11				
21	2015	1	12				
21	2015	1	13	1280	405	23,3	POINT (3.8588206191825 43.7055512425061)
21	2015	1	14	1210	405	22,0	POINT (3.8588054793645 43.7055150266607)
21	2015	1	15	1210	430	24,7	POINT (3.85879417528231 43.7054798314582)
21	2015	1	16				
21	2015	1	17				
21	2015	1	18				
21	2015	1	19	1360	440	26,1	POINT (3.85873531270688 43.7053479583104)
21	2015	1	20				
21	2015	1	21	1190	405	23,9	POINT (3.85870958345631 43.7052706865678)
21	2015	1	22	1360	440	29,6	POINT (3.85869795837973 43.7052345451259)
21	2015	1	23	1370	500	29,9	POINT (3.85868337318098 43.7052043992288)
21	2015	1	24				
21	2015	1	25	1070	430	15,6	POINT (3.8586548204222 43.7051312631374)
21	2015	1	26				
21	2015	1	27	1370	440	27,7	POINT (3.85862759753867 43.7050627240374)
21	2015	1	28				
.....	.....	.....	.....	.....	.....	.....	.....
....	....	....	....	....	....	....	....
...	...	...	...	...	...	...	...
..	..	..	..	..	..	..	..
.	.	.	.	.	.	.	.
22	2016	1	37				
22	2016	1	38	1180	420	31,0	POINT (3.85847432879448 43.7046795454629)
22	2016	1	39				
22	2016	1	40				
22	2016	1	41	1030	430	26,6	POINT (3.85842952993517 43.7045799571879)
22	2016	1	42	690	315	13,4	POINT (3.85841837973453 43.7045434646175)
22	2016	1	43				
22	2016	1	44	990	410	24,2	POINT (3.85839250240375 43.7044662859103)
22	2016	1	45				
22	2016	1	46				
22	2016	1	47				
22	2016	1	48				
22	2016	1	49				
...	...	...	...	...	...	...	...
..	..	..	..	..	..	..	..
.	.	.	.	.	.	.	.

FIGURE 9.3 – *Extrait du tableau des caractéristiques des arbres de la parcelle agroforestière A3.*

rendement des cultures en conditions d'agriculture classique ont été réalisées sur cette parcelle témoin.

Les cultures

Trois types de culture ont fait l'objet d'expérimentations entre 2016 et 2018 :

- La culture de pois (*Pisum sativum*) en 2016
- La culture de blé (*Triticum durum*) en 2017
- La culture d'orge (*Hordeum vulgare*) 2018

Pour exemple, le tableau (Fig. 9.5) montre les caractéristiques relevées a partir des différents échantillons prélevés sur la parcelle agroforestière et le témoin agricole lors de la récolte de blé dur en 2017. Deux tableaux similaires sont fournis pour le pois en 2016 et l'orge en 2018. Les caractéristiques étudiées sont différentes en fonction des cultures (nombre de gousses au mètre carré pour le pois, nombre de grains par épi pour le blé, ...) mais l'extrapolation du rendement à l'hectare a été harmonisée pour chacune des trois cultures (colonne "rendement (qx/ha)" dans notre exemple).

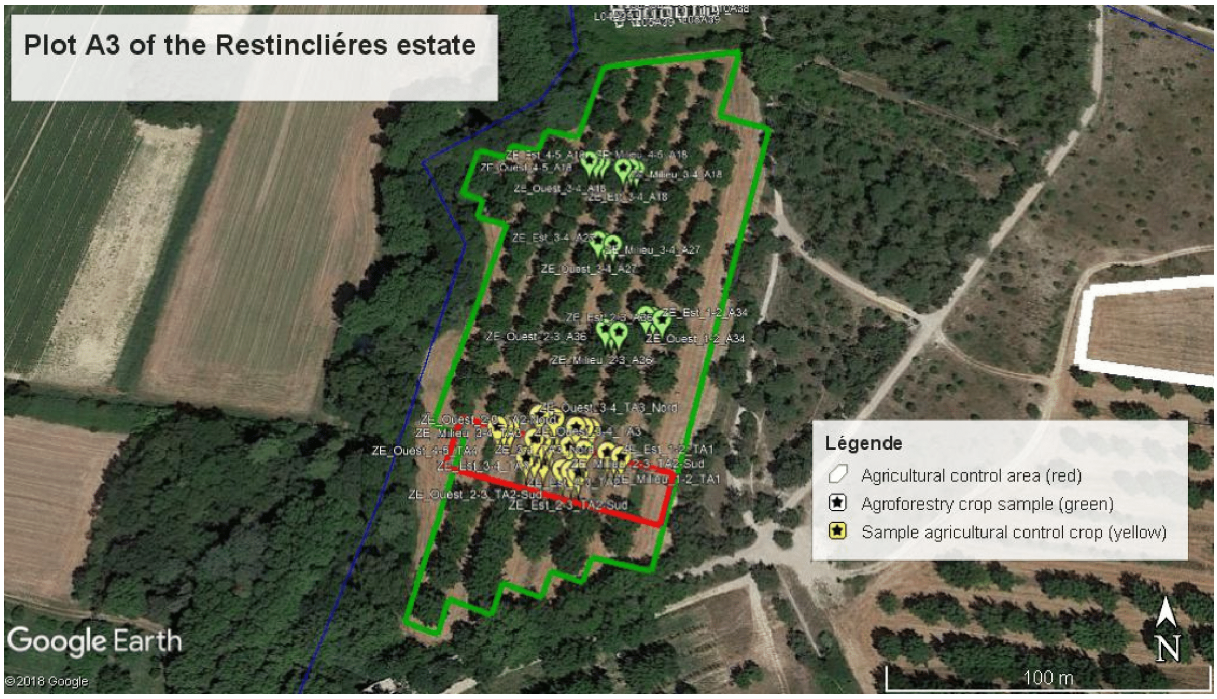


FIGURE 9.4 – *Plan des zone d'échantillonnage de la parcelle agroforestière A3 du domaine de Restinclières.*

D'autres observations sur la parcelle agroforestière B17 et sur son témoin forestier, toutes deux plantées de peupliers (*Populus x canadensis*) ont été traitées. Cependant, les mesures dendrométriques disponibles ne concernaient que l'année 2015 et nous ne disposions pas de données sur les cultures réalisées sur cette parcelle. Pour ces deux parcelles, les coordonnées GPS des arbres ont pu là encore être extraites à partir du logiciel Google Earth.

Rendements et Composantes du rendements - Blé dur récolté le 15 juin 2017
Restincliéres parcelles A3

TA = Témoin agricole
AF = Agroforesterie

Parcelle	Traitement	Emplacement	Rangee	Num Arbre	poids total des grains/m²	PMG	Nbre de grains /m²	Nbre d'épis/m²	Nbre de grains/épi	rendement (qx/ha)
A3	AF	Est	1 - 2	34	251,5	51,31	4901,58	184	26,64	25,15
A3	AF	Est	2 - 3	36	251,17	37,72	6658,80	428	15,56	25,12
A3	AF	Est	3 - 4	18	153,36	50,72	3023,66	253	11,95	15,34
A3	AF	Est	3 - 4	27	336,54	50,39	6678,71	254	26,29	33,65
A3	AF	Est	4 - 5	18	164,48	36,02	4566,35	190	24,03	16,45
A3	AF	Milieu	1 - 2	34	387,67	45,56	8509,00	265	32,11	38,77
A3	AF	Milieu	2 - 3	36	232,22	40,74	5700,05	240	23,75	23,22
A3	AF	Milieu	3 - 4	18	227,66	37,63	6049,96	292	20,72	22,77
A3	AF	Milieu	3 - 4	27	296,17	38,47	7698,73	287	26,82	29,62
A3	AF	Milieu	4 - 5	18	210,98	38,48	5482,85	241	22,75	21,10
A3	AF	Ouest	1 - 2	34	199,82	32,97	6060,66	274	22,12	19,98
A3	AF	Ouest	2 - 3	36	252,13	49,29	5115,24	227	22,53	25,21
A3	AF	Ouest	3 - 4	18	126,73	43,63	5097,45	251	20,31	18,36
A3	AF	Ouest	3 - 4	27	183,61	36,02	5097,45	202	25,23	16,55
A3	AF	Ouest	4 - 5	18	188,84	32,46	5817,62	287	20,27	18,88
A3	TA	Est	1 - 2	TA1	256,96	52,18	4924,49	193	25,52	25,70
A3	TA	Est	2 - 3	TA2	425,28	49,57	8579,38	281	30,53	42,53
A3	TA	Est	3 - 4	TA3 Nord	464,9	48,74	9538,37	306	31,17	46,49
A3	TA	Est	3 - 4	TA3 Sud	324,21	52,57	7548,47	698	10,81	37,77
A3	TA	Est	4 - 5	TA4	377,65	50,03	6167,21	287	21,49	32,42
A3	TA	Milieu	1 - 2	TA1	345,46	52,3	6605,35	307	21,52	34,55
A3	TA	Milieu	2 - 3	TA2	341,7	49,34	6925,42	251	27,59	34,17
A3	TA	Milieu	3 - 4	TA3 Nord	475,02	48,74	9746,00	346	28,17	47,50
A3	TA	Milieu	3 - 4	TA3 Sud	124,91	50,22	2487,26	324	7,68	12,49
A3	TA	Milieu	4 - 5	TA4	351,87	51,1	6885,91	333	20,68	35,19
A3	TA	Ouest	1 - 2	TA1	246,73	50,15	4919,84	80	61,50	24,67
A3	TA	Ouest	2 - 3	TA2	445,71	45,77	9738,04	187	52,08	44,57
A3	TA	Ouest	3 - 4	TA3 Nord	418,74	48,42	8648,08	274	31,56	41,87
A3	TA	Ouest	3 - 4	TA3 Sud	394,46	55,11	7157,68	294	24,35	39,45
A3	TA	Ouest	4 - 5	TA4	306,99	53,32	5757,50	273	21,09	30,70

FIGURE 9.5 – *Tableau des caractéristiques des échantillons issus de la récolte de blé dur de la parcelle A3 en 2017.*

9.1.2 Description de l'étude

Objectif principal de l'étude

L'ensemble du dispositif a été mis en place pour étudier la rentabilité des cultures (rendement à l'hectare) et la production de bois (volume de bois exploitable à la vente) d'une parcelle agroforestière, et de comparer les productions avec celles obtenues dans des conditions de culture classiques en agriculture d'une part et en foresterie d'autre part. En effet, les dispositifs agroforestiers ne pourront être adoptés par une vaste communauté d'agriculteurs et de forestiers, que si il s'avère que ces dispositifs sont rentables. De notre côté, nous cherchons à montrer que notre modèle sémantique est approprié pour faciliter les comparaisons. Nous avons fait le choix de la généricité et de la réutilisation. Chaque cas d'étude nécessite l'extension du modèle par de nouveaux concepts, de nouvelles règles ou de nouveaux traitements. Il nous faut cependant vérifier que les concepts de haut niveau ont tous été introduits. Il nous faut également tester la capacité du système à s'adapter à toute problématique relevant de l'agrofo-

resterie. De plus, il faut nous assurer de l'efficacité du système, et de sa capacité à répondre et à fixer des interprétations, en présence d'une forte volumétrie de données.

Hypothèse de départ

Nous considérons que les conditions de culture sont identiques pour la parcelle agroforestière et la parcelle du témoin forestier ainsi que pour la parcelle du témoin agricole. La composition des sols et les conditions climatiques sont considérées comme étant les mêmes, en raison de la proximité spatiale de ces trois parcelles.

Les paramètres pris en compte pour l'étude

- Au sein de la parcelle forestière et agroforestière
  - la hauteur des arbres
  - la hauteur d'élagage des arbres
  - le diamètre des troncs à hauteur de poitrine
- Au sein de la parcelle du témoin agricole et de la parcelle agroforestière
  - le rendement des cultures (sur échantillons de récolte de 1 mètre carré)

Étude de la rentabilité des cultures

Cette étude est conduite à partir de la comparaison de 2 productions de céréales (blé, orge) et d'une production de légumineuse (pois) dans les conditions de culture en agroforesterie et d'agriculture classique s'étalant sur trois années consécutives (une culture différente par année).
L'indicateur retenu sera la production à l'hectare au sein d'une parcelle.
Pour chacune des trois cultures, quinze échantillons de récolte ont été prélevés à la fois dans la parcelle agroforestière, et dans le témoin agricole (cinq échantillons à l'est, cinq au milieu et cinq à l'ouest de chaque parcelle).
Notre démarche s'efforcera donc :

- D'identifier les échantillons de récolte de la parcelle agroforestière et de son témoin agricole pour une culture et une année données.
- Après analyse statistique des données acquises, de calculer la moyenne du rendement obtenu dans les deux conditions de cultures.

Nous pourrons ainsi comparer le rendement à l'hectare d'une culture pour les deux catégories de parcelles.

Étude de la rentabilité de la production de bois destiné à la vente

Afin d'étudier la rentabilité de la production de bois destiné à la vente, nous nous focaliserons sur l'évolution de la croissance des arbres.
L'indicateur retenu sera l'évolution du ratio entre arbres exploitables et arbres non exploitables sur trois années consécutives dans les conditions de culture en agroforesterie et en foresterie classique.
La notion d'arbre exploitable sera basée sur des critères utilisés dans le monde de l'exploitation forestière, en l'occurrence pour notre étude, une hauteur de tronc (hauteur

d'élagage) et un diamètre de tronc (diamètre à hauteur de poitrine) requis pour l'autorisation à la vente du bois dans le cadre de l'ébénisterie ou de la scierie.
Pour atteindre cet objectif, nous procéderons comme suit :

- Identification des arbres de la parcelle agroforestière et de ceux du témoin forestier.
- Analyse statistique des données acquises sur les différents caractères mesurés (hauteur totale, hauteur du tronc, diamètre, biomasse) sur les différentes années considérées.
- Calcul de la biomasse de chaque arbre au fil des années.
- Identification des arbres éligibles à la vente pour les différentes années selon des critères d'éligibilité en usage dans le négoce du bois (hauteur minimale du tronc de 3 mètres et diamètre minimal du tronc de 36 centimètres).
- Pour les deux types de parcelle (agroforestière et forestière), calcul de la biomasse de tous les arbres de la parcelle et de la biomasse des arbres exploitables.
- Calcul du ratio entre la biomasse de tous les arbres et celle des arbres exploitables pour chacune des parcelles.
- Comparaison du ratio entre les conditions agroforestières et les conditions de foresterie classique.

9.1.3 Diagrammes de cas d'utilisation

Pour faciliter la prise en charge des différents traitements par notre système, nous avons défini différents diagrammes de cas d'utilisation, qui chacun précisent les fonctionnalités attendues. Un premier diagramme de cas d'utilisation (Fig. 9.6) permet de modéliser les besoins des différents acteurs dans le cadre de ce cas d'étude.

Deux grandes fonctionnalités sont mises en évidence :

- La description d'un aménagement agroforestier au travers du modèle Agroforestry ("Describe a layout").
- L'interrogation de la base de connaissance à propos de cet aménagement ("Query the data of a layout").

Différents acteurs sont susceptibles d'alimenter la description d'une parcelle agroforestière : les agriculteurs ("farmer"), les forestiers ("forester") et les scientifiques ("scientist"). Ils sont généralisés par l'acteur fournisseur de données ("data provider").
Ces mêmes acteurs et des décisionnaires ("decision maker") sont susceptibles d'alimenter et d'interroger la base de connaissance. Ils sont généralisés par l'acteur consultant de données ("data consultant").

Le cas d'usage "Describe a layout" généralise les cas d'usages suivants :

- "Describe a physical element" : il s'agit pour le fournisseur de données de pouvoir saisir n'importe quel élément physique simple ou composite d'une parcelle agroforestière comme par exemple une ligne constituée d'arbres. Pour atteindre son objectif ce cas d'utilisation inclut les fonctionnalités suivantes :
  - "Describe the location of an element" : le fournisseur de données doit pouvoir indiquer la localisation géographique de l'élément.
  - "Describe the Temporality of an element" : le fournisseur de données doit pouvoir renseigner la temporalité de l'élément comme par exemple l'intervalle de temps de présence d'un arbre dans une parcelle.

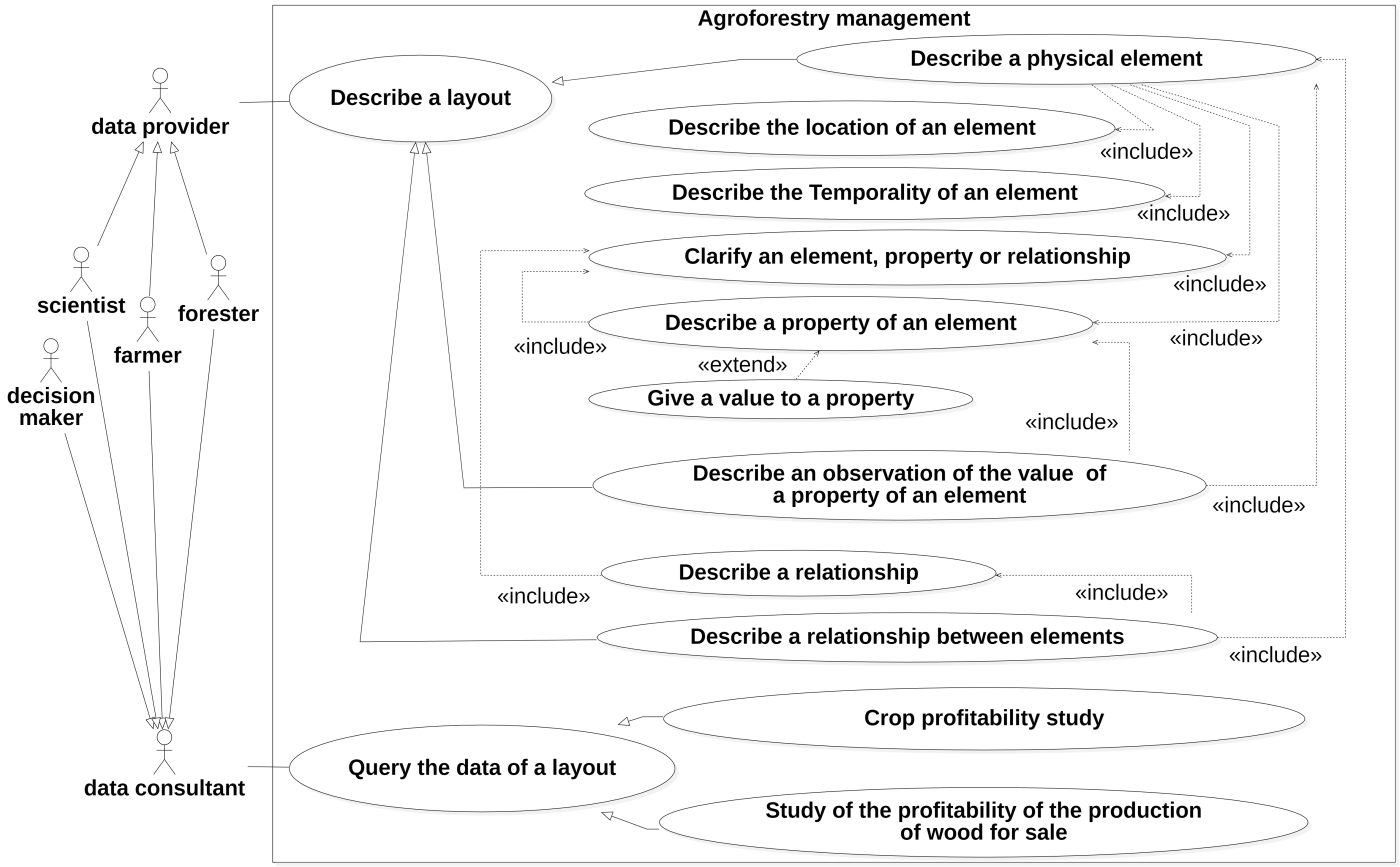


FIGURE 9.6 – *Diagramme de cas d'utilisation UML pour les aménagements agroforestiers.*

  - "Clarify an element, property or relationship" : un élément, une propriété de cet élément et une relation entre deux éléments doivent pouvoir être enrichis par un terme provenant d'une ontologie.
  - "Describe a property of an element" : le fournisseur de données doit pouvoir indiquer une caractéristique de l'élément, comme le fait qu'un arbre appartient à une espèce particulière, ou le fait qu'il possède une hauteur.

– "Describe an observation of the value of a property of an element" : pour des caractéristiques dont les valeurs peuvent varier dans le temps comme le diamètre de l'arbre, on consignera une observation comme étant la valeur d'une propriété observée d'un élément avec son contexte de mesure (temps de l'observation, procédure,...).

– "Describe a relationship between elements" : une description d'une relation entre deux éléments doit pouvoir être saisie.

Le cas d'usage "Query the data of a layout" généralise les cas d'usages suivants :

– "Crop profitability study" : le consultant de données doit pouvoir comparer le rendement des cultures d'une parcelle pour des aménagements distincts.

– "Study of the profitability of the production of wood for sale" : le consultant de données doit pouvoir comparer l'évolution de la croissance des arbres d'une parcelle pour des aménagements distincts.

Les figures suivantes exposent en détail les différentes fonctionnalités de ces deux derniers cas d'usage pour l'étude du rendement des cultures et de la production de bois.

La figure 9.7 introduit des éléments fonctionnels détaillés pour l'étude du rendement des cultures :

- "Compare the crop profitability of an agroforestry plot with that of an agricultural plot" : l'utilisateur doit pouvoir comparer les rendements d'une même culture dans des conditions de culture classique et agroforestière.
- "Comparison of the average crop yield between two plots" : comparaison du rendement d'une même culture sur deux parcelles distinctes.
- "Calculation of the average crop yield on a plot" : calcul de la moyenne du rendement d'une culture sur une parcelle.
- "Identification of crop sampling for a given crop and year on a plot" : identification d'échantillons de récoltes pour une culture et une date donnée.

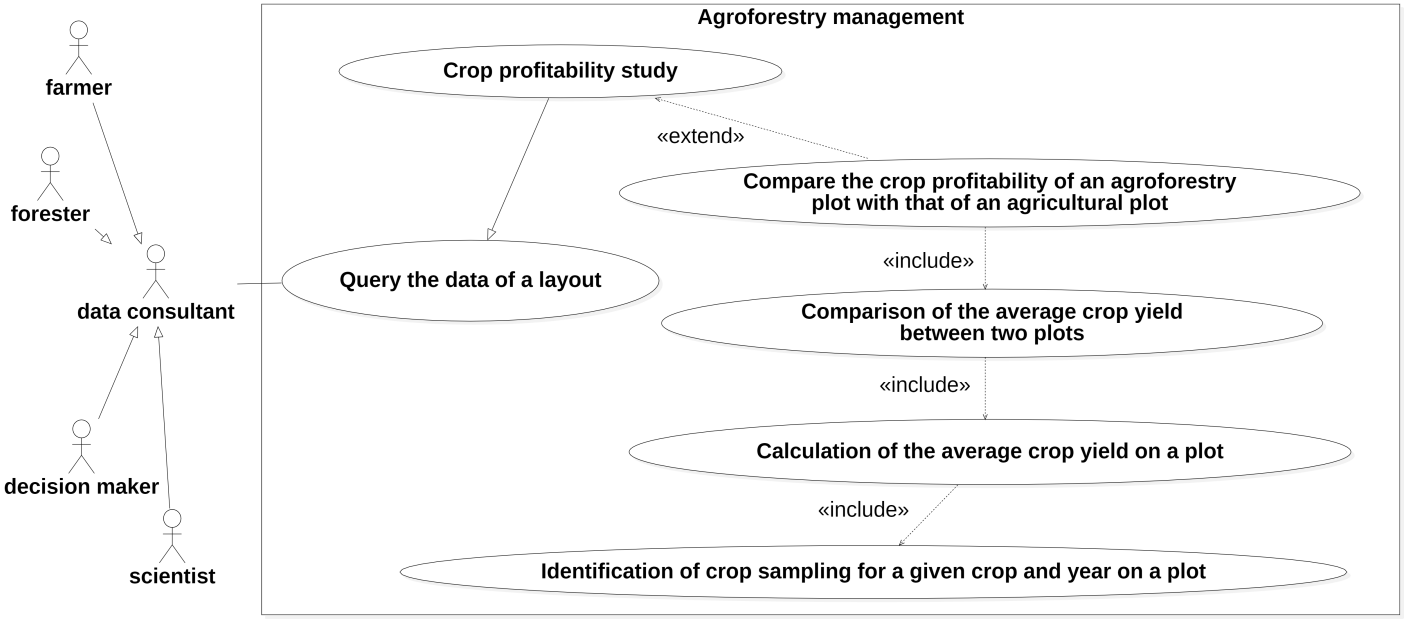


FIGURE 9.7 – *Diagramme de cas d'utilisation UML pour les aménagements agroforestiers : Étude des cultures.*

La figure 9.8 introduit des éléments fonctionnels détaillés pour l'étude de l'évolution de la croissance des arbres :

- "Compare the profitability of an agroforestry plot with that of a forest plot" : comparaison de l'évolution de la croissance des arbres dans des conditions de foresterie classique et agroforestière.
- "Calculation of the ratio of saleable tree biomass to total tree biomass" : calcul du ratio entre la biomasse de bois éligible au négoce du bois sur la biomasse totale des arbres pour une parcelle.
- "Identification of trees in a plot" : identification des arbres d'une parcelle.
- "Identification of the trees in a plot suitable for timber trading" : identification des arbres d'une parcelle éligible au négoce du bois .
- "Calculation of the biomass of a tree" : calcul de la biomasse d'un arbre.

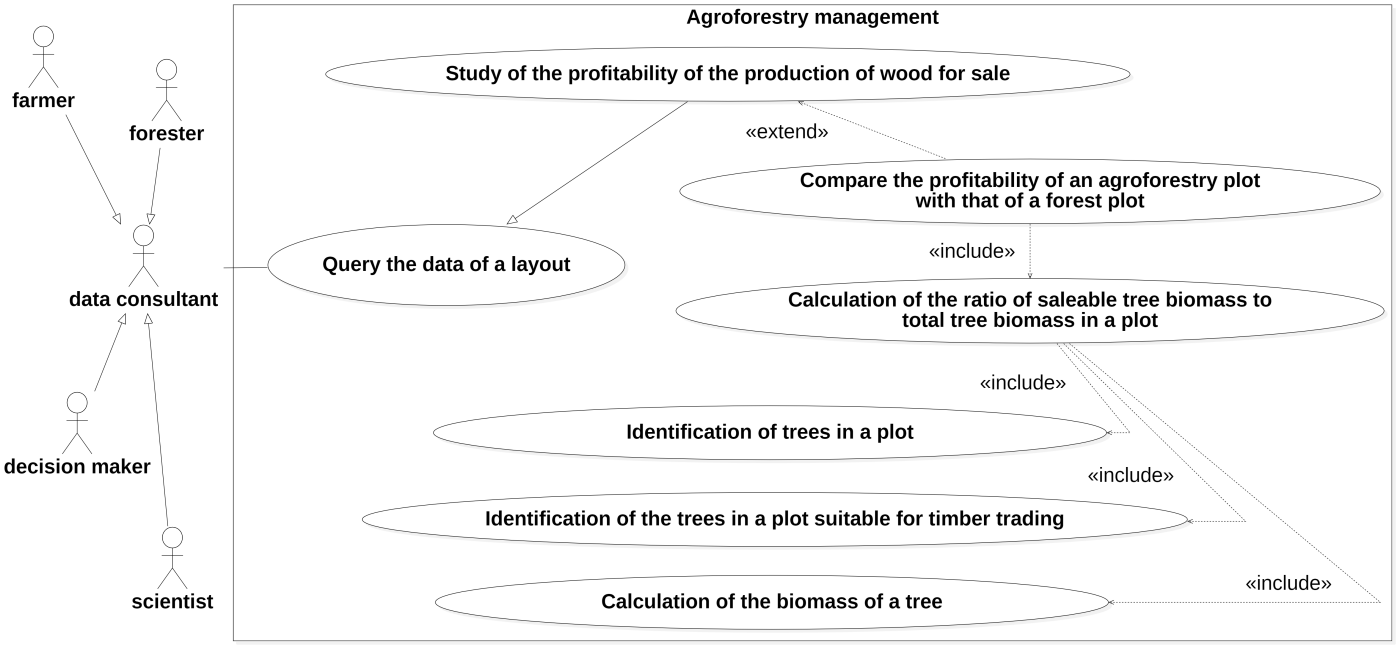


FIGURE 9.8 – *Diagramme de cas d'utilisation UML pour les aménagements agroforestiers : Étude de la croissance des arbres.*

9.1.4 Diagrammes d'instances du modèle Agroforestry

Des diagrammes d'instances UML, conformes au modèle Agroforestry, aident à la compréhension des objets de l'étude. L'idée est aussi de valider les choix opérés lors des activités de la modélisation conceptuelle, par confrontation avec des données réelles.

Structure des éléments

Comme déjà introduit, les éléments décrits peuvent être simples ou composés (Fig. 9.9). La propriété objet *afy:hasElement* est spécialisée par la propriété *BFO_0000051* (has part) pour tendre encore plus vers de la réutilisation de l'existant. Les individus peuvent avoir une désignation qui leur est propre, qui est notée à l'aide de la propriété *skos:prefLabel*. L'individu PA3AF a ainsi pour label local "Agroforestry Plot A3". Le domaine de Restinclières se compose de trois parcelles (une parcelle agroforestière, une parcelle témoin forestière et une parcelle témoin agricole. La parcelle agroforestière est elle-même composée d'un alignement régulier d'arbres, à son tour composé de plusieurs lignes d'arbres.

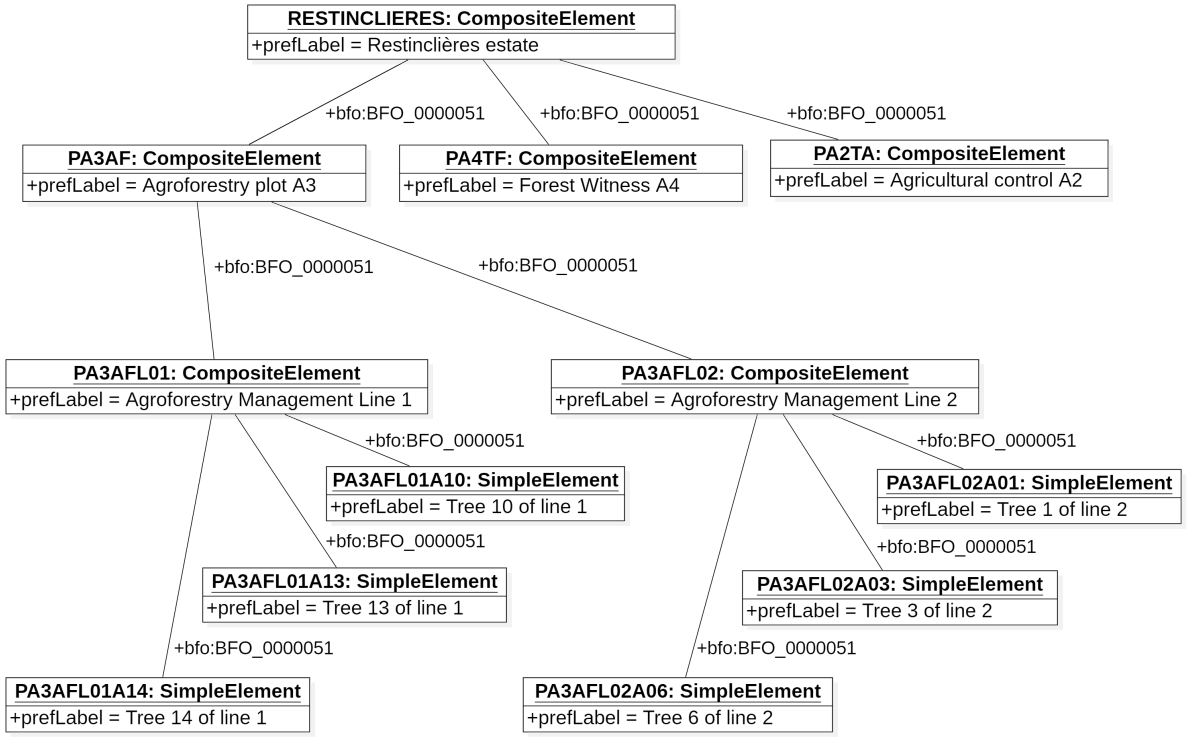


FIGURE 9.9 – *Diagramme d'instances : Description de la composition d'éléments du domaine de Restinclières.*

Typage des éléments, de leurs propriétés intrinsèques, et de leurs relations

Les éléments sont enrichis sémantiquement par des termes issus de vocabulaires partagés (Fig. 9.10).
Ici, l'arbre numéro 10 de la première ligne de la parcelle A3 a pour identifiant la valeur clé PA3AFL01A10, et est qualifié par le concept SKOS c_3052 qui a pour label en anglais "forest trees", et qui provient du vocabulaire contrôlé AGROVOC de la FAO. Ce terme est lui-même généralisé par le concept c_7887 "trees" provenant également d'AGROVOC.

Les caractéristiques (propriétés) des éléments de l'exploitation agroforestière sont des instances de la classe **PropertyOfElement**. Elles sont, elles aussi, enrichies par des termes de vocabulaire. Dans notre jeu de données, la colonne "D-1,30" correspond au diamètre à 1,30 mètres de hauteur (ici pour l'arbre numéro 10). Ce relevé de mesure est utilisé en sylviculture sous le vocable de diamètre à hauteur de poitrine (ou diamètre à hauteur d'homme). Nous indiquons donc que l'arbre numéro 10 de la première ligne possède une propriété *diametreDuTronc*. Cette propriété est à son tour précisée par les les concepts c_34790 d'AGROVOC ("breast height diameter") et ECSO_00000549 ("tree diameter at breast height") de l'ontology ECSO . Nous avons délibérément choisi de pas utiliser directement un terme de référentiel en matière de propriété pour pouvoir justement élargir le choix des étiquettes sémantiques.

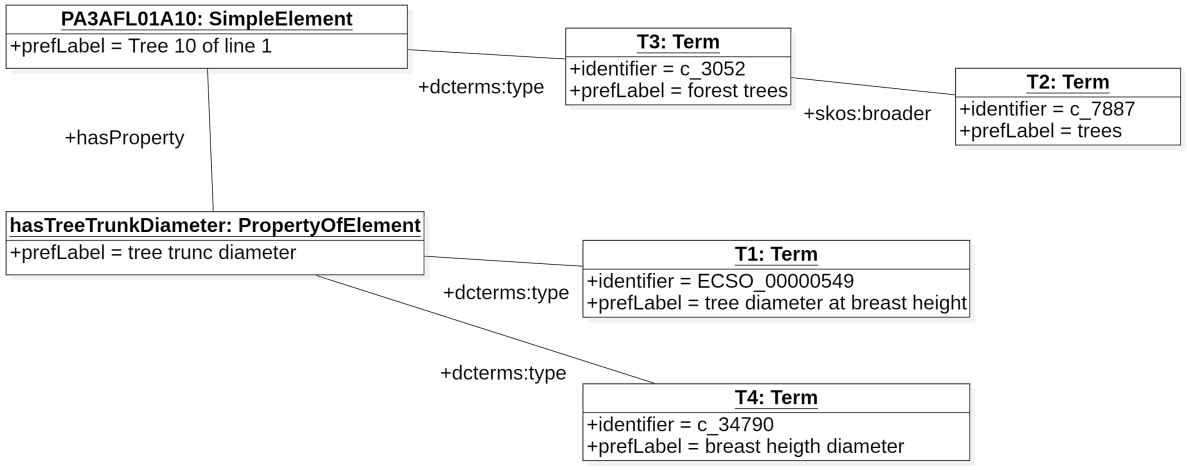


FIGURE 9.10 – *Diagramme d'instances : Description des caractéristiques des éléments et typage de ces éléments et caractéristiques.*

Expression des observations de terrain

Les valeurs des différentes caractéristiques (propriétés) des éléments de l'exploitation agroforestière sont organisées au travers de l'ontologie SOSA (Sensor, Observation, Sample, and Actuator). SOSA permet de décrire les capteurs (y compris humains) et les observations acquises par ces capteurs. SOSA place l'observation au centre de son modèle : Une observation permet ainsi de renseigner la valeur d'une propriété descriptive d'un élément d'intérêt. L'observation est liée à un élément (instance de la classe **StructuralElement**) par la relation *hasFeatureOfInterest*. La propriété observée de l'élément est liée à l'observation par la relation *observedProperty* et la valeur de cette propriété est liée à l'observation par la relation *hasResult*. Les couples propriété et valeur ainsi que propriété et élément observés doivent être cohérents. La relation *hasProperty*, entre un élément (instance de la classe **SimpleElement**) et une propriété (instance de la classe **PropertyOfElement**), sera utilisée pour renforcer cette cohérence (Fig.9.11).

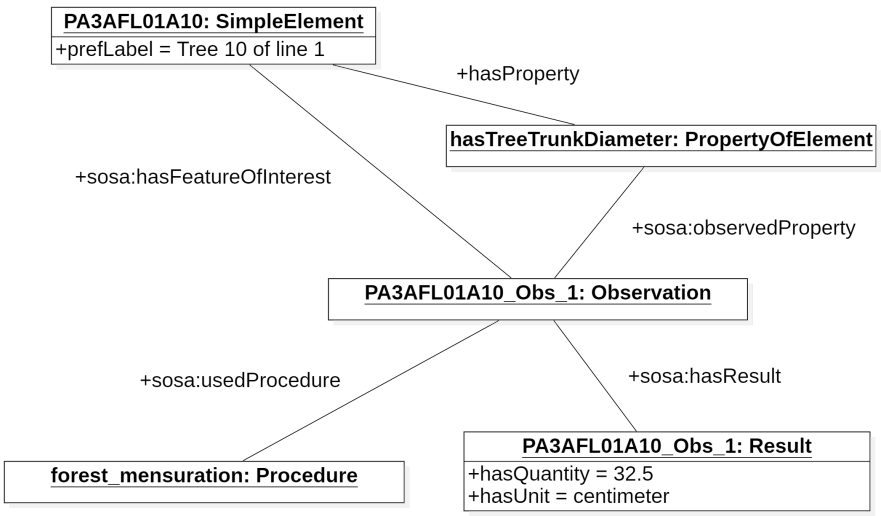


FIGURE 9.11 – *Diagramme d'instances : Expression d'une observation de terrain avec l'ontologie SOSA. Ici le diamètre d'un arbre*

Relation entres les éléments

Des relations peuvent exister entre éléments d'une exploitation agroforestière Elles sont représentées au moyen de la classe **RelationshipBetweenElements**. Ces relations sont normalisées, elles aussi, par des termes de vocabulaire (Fig. 9.12). Ici il est énoncé que la ligne d'arbres PA3AFL01 et la culture intercalaire PA3AFCS01 de la parcelle agroforestière A3 entretiennent une interaction de facilitation typée par le concept "Biotic interaction" (AnaeeThes:c3_2520).

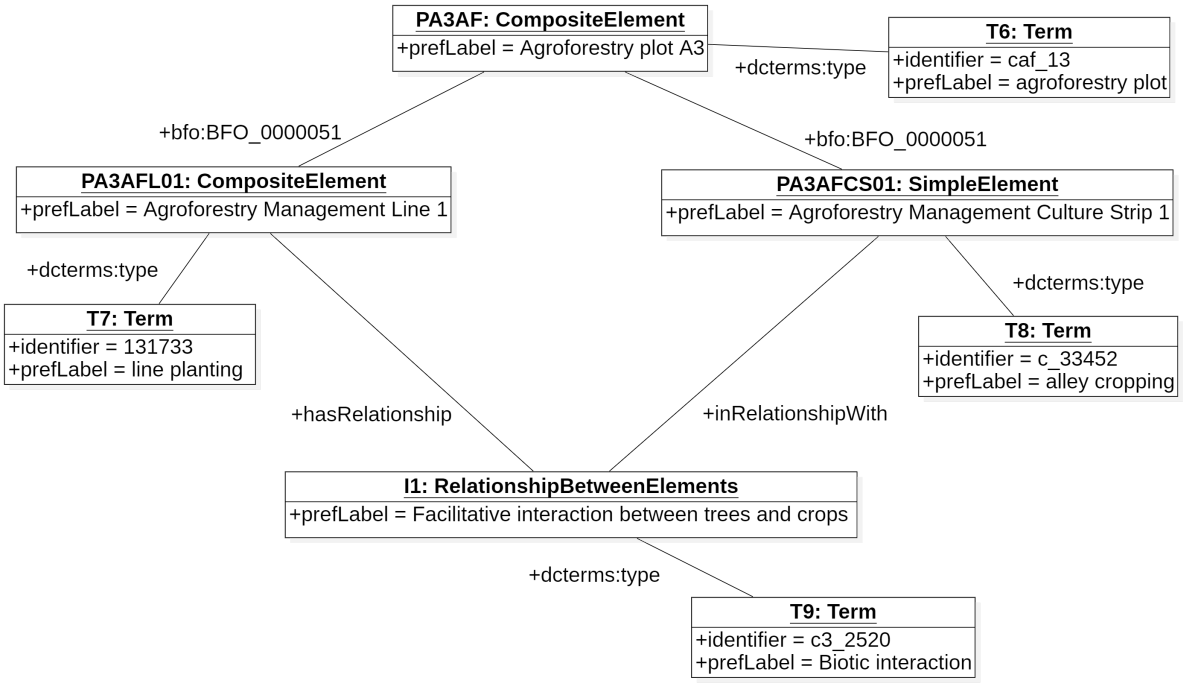


FIGURE 9.12 – *Diagramme d'instances : Description d'une relation sémantique «Facilitive interaction between trees and crops» entre une ligne d'arbres et une culture intercalaire de la parcelle A3.*

Expression de la géolocalisation et de la temporalité

Les aspects de spatialisation sont traités via l'utilisation de l'ontologie GeoSPARQL. Chaque élément de l'exploitation agroforestière peut se voir attribuer une géométrie (point, ligne ou polygone) avec ses coordonnées GPS (Fig. 9.13). Dans cet exemple on attribue à l'arbre numéro 10 de la première ligne d'arbres de la parcelle A3, une géométrie ponctuelle sous la forme d'un point de coordonnées GPS.

Sur cette même figure on observe que les aspects temporels sont traités via l'utilisation de l'ontologie Time Ontology. Chaque élément de l'exploitation peut se voir attribuer un moment et/ou un intervalle de temps. Ici on attribue à l'arbre 10, une date de plantation en décembre 1994.

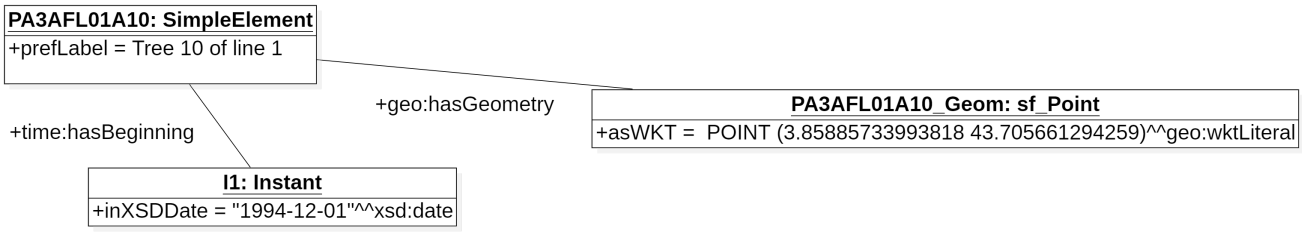


FIGURE 9.13 – *Diagramme d'instances : Expression de la géolocalisation et de la date de plantation d'un arbre de la parcelle A3.*

Réutilisation d'ontologies existantes

De manière complémentaire aux ontologies citées précédemment (en particulier SOSA, GeoSPARQL, Time, Agrovoc), et à notre ontologie terminologique dédiée, d'autres référentiels sont mis à contribution. L'objectif est toujours de réutiliser plutôt que de définir de nouveaux termes/concepts.
Nous avons en conséquence également exploité :

- **OM** (Ontology of units of Measure and related concepts) : OM est une ontologie du domaine des quantités et des unités de mesure permettant d'exploiter au mieux des descriptions sur des données quantitatives [RIJGERSBERG et al., 2013]. Sa dernière version OM-2[3] modélisé avec le langage OWL 2 permet une bonne intégration en particulier dans des modèles approchant les domaines de l'expérimentation scientifique et de l'ingénierie, autorisant des conversions d'unités réalisées par les moteurs d'inférences [KEIL et SCHINDLER, 2019]. Nous utilisons ce vocabulaire pour nous aider à décrire les mesures quantifiés de notre cas d'étude (hauteur des arbres en centimètre, rendement en kilogramme par hectare, ...).

- **AnaEE thesaurus** : thésaurus pour l'étude des écosystèmes et de leur biodiversité. AnaEE France est une infrastructure de recherche Française fournissant des plateformes expérimentales pour l'étude des écosystèmes [CLOBERT et al., 2018]. Afin de capitalisé les données produites par les expérimentations réalisées au sein de leurs installations, ils ont mis en place un vocabulaire partagé pour l'étude des écosystèmes continentaux terrestre et aquatique et de leur biodiversité du nom d'AnaEE Thésaurus [CLASTRE et al., 2018]. Ce vocabulaire a pour but de faciliter la réutilisation des données, de généraliser les résultats et d'améliorer les modèles prédictifs en définissant des concepts tel que celui défini par le terme « soil moisture » (http://opendata.inra.fr/anaeeThes/c2_2574) qui nous permet de clarifier les propriétés des éléments de notre cas d'étude en l'occurrence ici l'humidité du sol relevée pour les parcelles.

- **Environment Ontology** (ENVO) : est une ontologie de termes dédiée à l'environnement [BUTTIGIEG, PAFILIS et al., 2016]. Nous l'avons utilisé par exemple pour définir le Lez comme étant un fleuve avec le concept « river » (http://purl.obolibrary.org/obo/ENVO_00000022) de ce vocabulaire.

- **TAXREF_LD** : mise en place par le Muséum National d'Histoire Naturelle, cette ontologie est un référentiel national taxonomique pour tout les organismes vivant en France métropolitaine et Outre-mer. Il nous permet de définir la taxonomie des plantes et animaux pouvant être présents sur un parcelle agroforestière

3. OM-2 : Ontology of units of Measure. https://github.com/HajoRijgersberg/OM.

[GARGOMINY et al., 2016 ; MICHEL et al., 2017]. Nous avons utilisé par exemple le concept « Juglans regia » (http://taxref.mnhn.fr/lod/taxon/104076/10.0) pour préciser l'espèce des noyers de la parcelle A4 de notre cas d'étude.

– **Woody Plant Ontology** : cette ontologie liste les variables nécessaires pour les observations des plantes ligneuses [MICHOTEY et al., 2019]. Elle nous permet par exemple de disposer des concepts de circonférence d'un arbre « Tree circumference » ('http://www.cropontology.org/rdf/CO_357:1000012') et de hauteur totale d'un arbre « Tree height » (http://www.cropontology.org/rdf/CO_357:1000037).

9.1.5 Implémentation du modèle Agroforestry

Pour une bonne intégration, une évolution facilitée et un partage simplifié de l'ontologie Agroforestry, OWL est retenu comme le langage de représentation des connaissances privilégié.

Diagrammes RDF du modèle Agroforestry

Les figures suivantes montrent les diagrammes RDF du modèle issus des diagrammes d'instances présentés précédemment :

– Le diagramme RDF (dérivé du diagramme d'instances (Fig. 9.9)) décrit la composition des éléments physiques de la parcelle. La parcelle A3 est constituée d'un témoin forestier, d'un témoin agricole et d'une parcelle agroforestière. La parcelle agroforestière est constituée de lignes d'arbres qui sont séparées par des bandes de culture. Les lignes d'arbres sont constituées d'arbres individualisés (Fig.9.14).

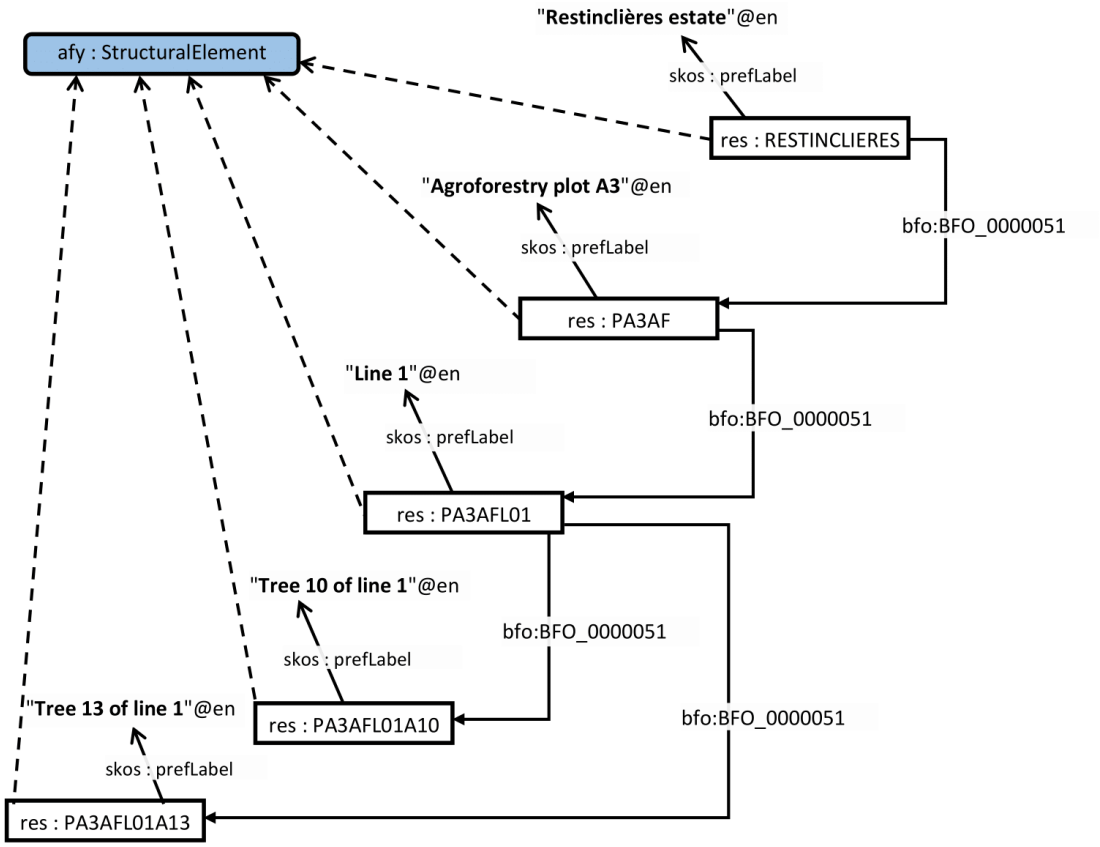


FIGURE 9.14 – *Diagramme RDF : Description et composition d'éléments d'une exploitation agroforestière.*

– Le diagramme RDF (dérivé du diagramme d'instances (Fig. 9.10)) : le terme "forest trees" du vocabulaire contrôlé Agrovoc vient typer l'arbre numéro 2 de la première ligne d'arbre (il indique que cet arbre est dédié à l'exploitation de bois d'œuvre). Ce terme est lui-même généralisé par le terme "trees". Le diamètre du tronc qui est une caractéristique de cet arbre est enrichi par le terme "breast height diameter" qui indique que le diamètre est pris à hauteur de poitrine (Fig.9.15).

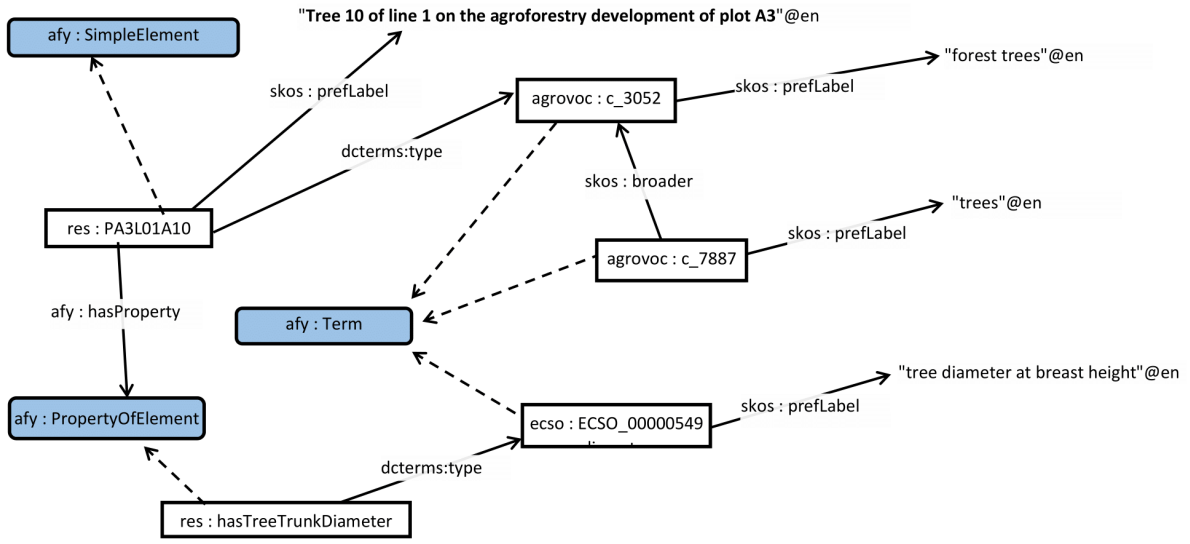


FIGURE 9.15 – *Diagramme RDF : étiquetage d'un élément et d'une propriété par des termes d'une ontologie terminologique.*

– Le diagramme RDF (dérivé du diagramme d'instances (Fig.9.11)) : avec l'aide de l'ontologie SOSA, l'observation de la valeur du diamètre d'un arbre, relevée par Lydie Dufour (capteur humain), en date du 28 juillet 2015 (Fig.9.16) est représentée.

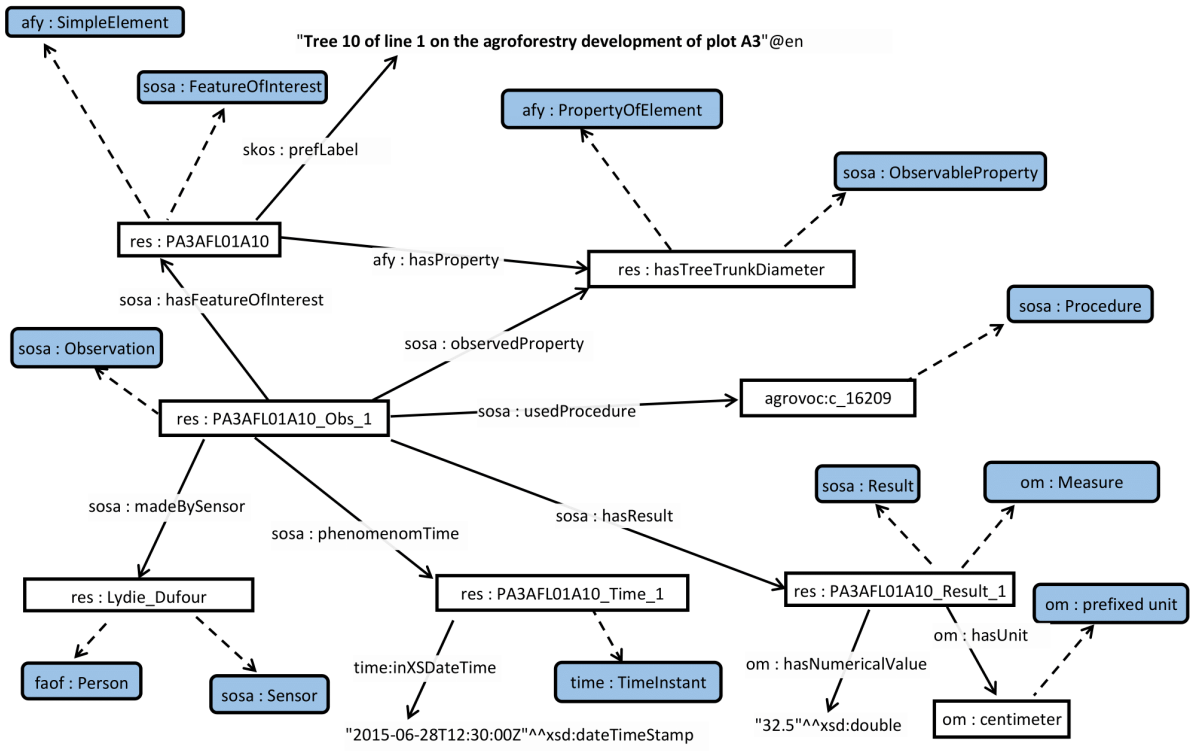


FIGURE 9.16 – *Diagramme RDF : Observation du diamètre d'un arbre avec SOSA.*

– SOSA permet aussi de décrire des échantillonnages. L'échantillonnage effectué sur la parcelle PA3AF durant la récolte de blé est représenté ici. Les échantillons "sosa:Sample" sont le résultat par la relation "sosa:hasResult" de l'échantillonnage

"sosa:Sampling". L'échantillonnage est conduit sur une récolte de blé en 2017 au sein de la parcelle PA3AF et les observations portent sur les échantillons prélevés lors de cette récolte (Fig.9.17).

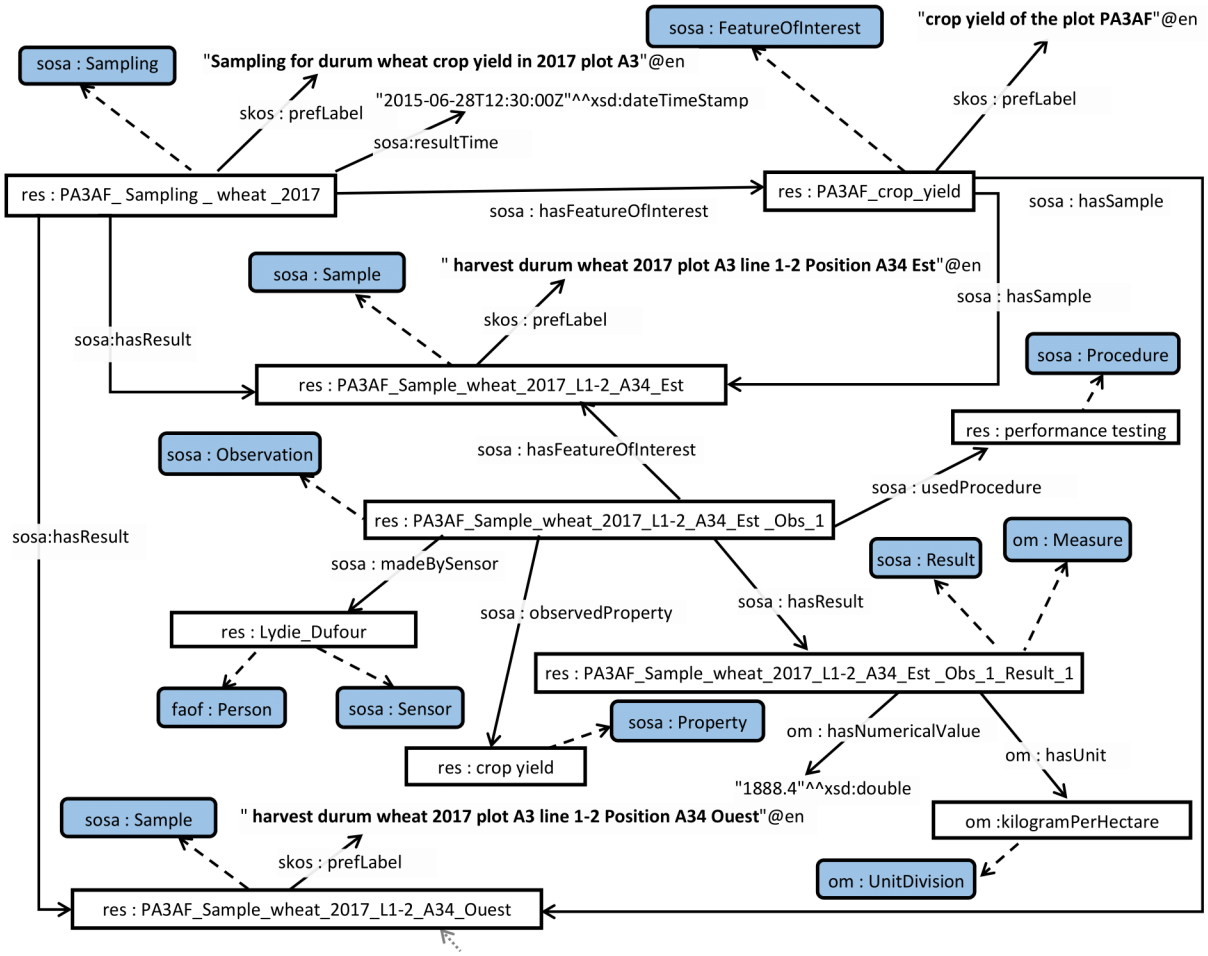


FIGURE 9.17 – *Diagramme RDF : Exemple d'un échantillonnage avec SOSA.*

– Le diagramme RDF (dérivé du diagramme d'instances (Fig. 9.12)) : exprime le fait qu'une parcelle est une sous partie d'un site au travers de la relation sémantique "has part/part of" (Fig.9.18).

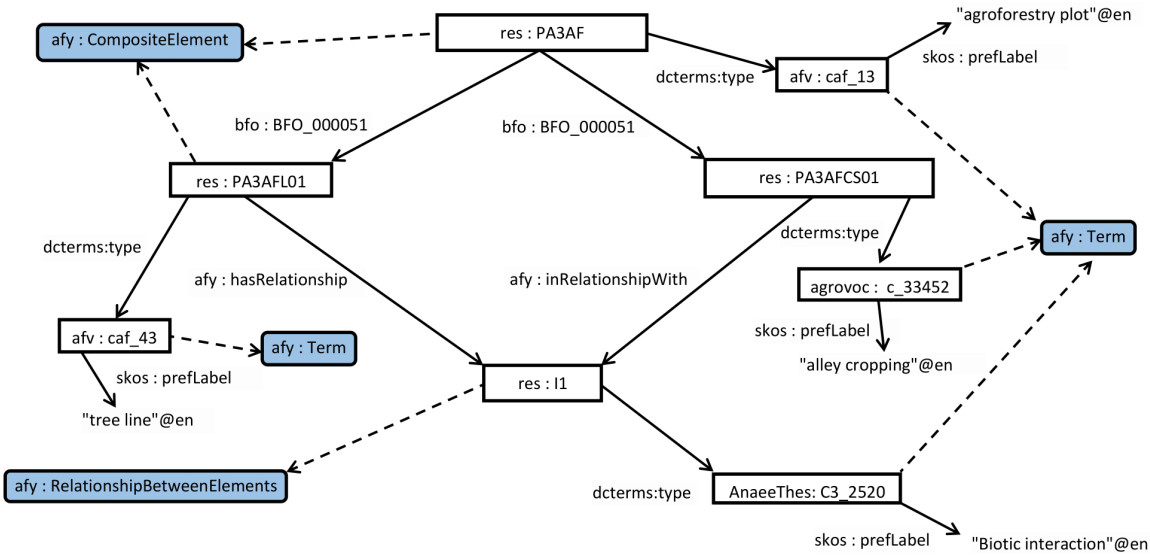


FIGURE 9.18 – *Diagramme RDF : Relation entre la parcelle et le site.*

– Le diagramme RDF (dérivé du diagramme d'instances (Fig. 9.13)) : restitue la prise en charge de données de géolocalisation. GeoSPARQL nous permet par exemple de rendre compte de la localisation d'un arbre avec une géométrie de type point et d'une ligne d'arbres avec une géométrie de type polygone (Fig.9.19).

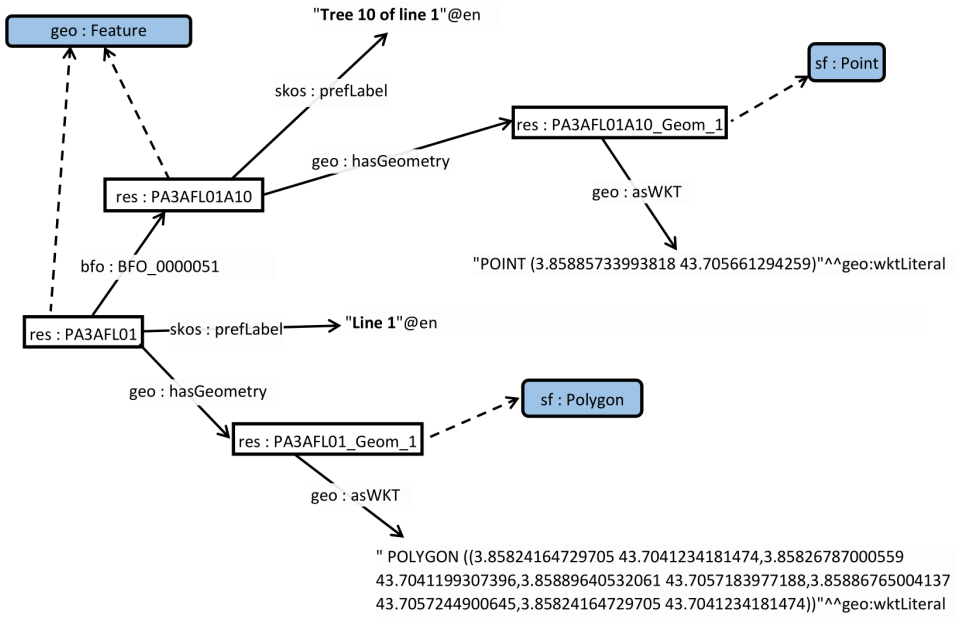


FIGURE 9.19 – *Diagramme RDF : Expression de la géolocalisation d'un arbre et d'une ligne d'arbres avec l'ontologie GeoSPARQL.*

– Le diagramme RDF (dérivé du diagramme d'instances (Fig. 9.13)) rend compte de la temporalité. Ainsi, l'ontologie OWL-Time permet de préciser l'intervalle temporel de présence d'un arbre sur la parcelle, entre le moment où il a été planté, et le moment où il a été abattu (Fig.9.20).

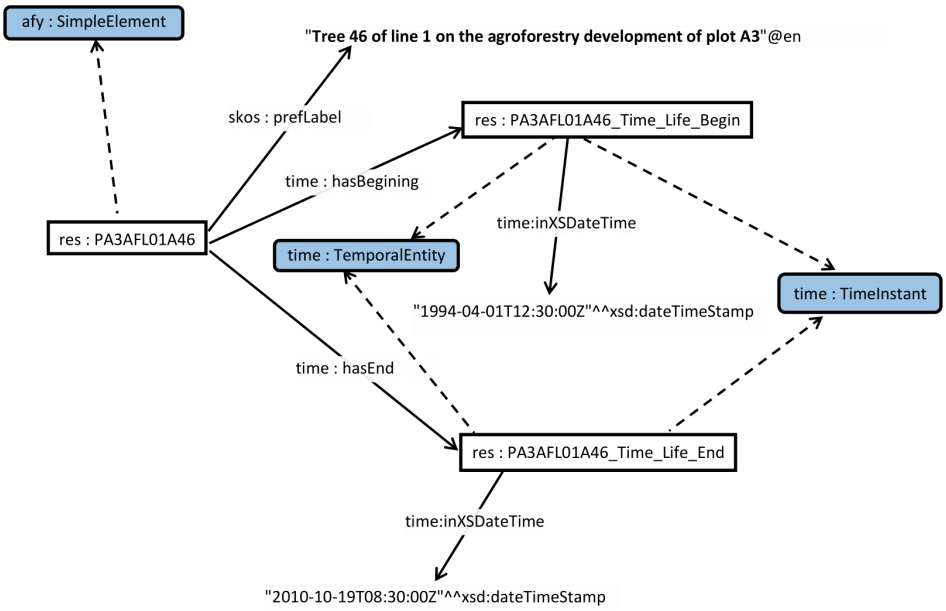


FIGURE 9.20 – *Diagramme RDF : Expression de la durée de vie d'un arbre avec l'ontologie Time.*

La base de connaissance Agroforestry

Nous pouvons donner quelques éléments de mesure précisant la volumétrie de la base de connaissance ainsi constituée. En nous aidant du logiciel Protégé [4], nous dénombrons :

- **Pour l'ontologie cadre Agroforestry :**

  Notre modèle, l'ontologie cadre Agroforestry qui sert de base à la description de notre cas d'étude est composé de 51 axiomes qui permettent de déclarer notamment 13 classes et 9 propriétés. Parmi ces axiomes on retrouve aussi 19 déclarations de subsomption de classes, 2 déclarations d'équivalences de classes, 2 déclarations d'inclusion de classes et 4 déclarations de disjonction de classes.

- **Au niveau de l'extension de notre modèle :**

  Lors de l'extension de notre modèle Agroforestry pour la description complète des parcelles A4 et B17 du domaine de Restinclières ainsi que de celles de leurs témoins agricole et forestier, nous totalisons 67 331 axiomes pour la définition de 14 661 individus représentant en grande majorité les 561 arbres et les observations faites sur leurs mesures dendrométriques annuelles durant trois années.
  D'autres axiomes supplémentaires pour la gestion de notre cas d'étude on été ajoutés comme par exemple cet axiome d'inclusion général de concept :
  $StructuralElement\ and\ (dcterms : type\ value\ afv : caf_27)\ SubClassOf$
  $StructuralElement\ and\ (dcterms : type\ value\ afv : caf_30)$
  qui énonce que tout élément du modèle clarifié par le terme "Témoin forestier" (concept caf_27 de notre ontologie terminologique AgroforestryVoc) est aussi par généralisation un élément de l'ensemble des éléments clarifié par le terme "Parcelle forestière" (concept caf_30 d'AgroforestryVoc).
  En incluant de plus les concepts importés, des ontologies cadres (SOSA, GeoS-

4. Protégé : Ontology Metrics. http://protegeproject.github.io/protege/views/ontology-metrics/.

PARQL et OWL-time) et terminologiques exploitées, la taille de notre base de connaissance est de 85 574 triplets

Un extrait très succinct de la base de connaissance, est proposé ci dessous, (Fig.9.21)) dans la syntaxe Turtle. Nous reprenons l'exemple de l'arbre 10 de la ligne 1 de la parcelle A3, qui a servi de fil d'Ariane tout au long de la construction des diagrammes d'instance et des graphes RDF illustratifs.

```
@prefix afy:          <http://www.afy.fr/Agroforestry/> .
@prefix res:          <http://www.afy.fr/Restinclieres/> .
@prefix skos:         <http://www.w3.org/2004/02/skos/core#> .
@prefix bfo:          <http://purl.obolibrary.org/obo/> .
@prefix geo:          <http://www.opengis.net/ont/geosparql#> .
@prefix sf:           <http://www.opengis.net/ont/sf#> .
@prefix om:           <http://www.ontology-of-units-of-measure.org/resource/om-2/> .
@prefix time:         <http://www.w3.org/2006/time#> .
@prefix sosa:         <http://www.w3.org/ns/sosa/> .
@prefix xsd:          <http://www.w3.org/2001/XMLSchema#> .
@prefix dcterms:      <http://purl.org/dc/terms/> .
@prefix agrovoc:      <http://aims.fao.org/aos/agrovoc/> .

res:PA3AFL01A10                                   Définition individu PA3AFL01A10
a                     afy:StructuralElement ;
skos:prefLabel        "Tree 10 of line 1 on the agroforestry development of plot A3"@en ;
bfo:BFO_0000050       res:PA3AFL01 , res:PA3AF ;
dcterms:type          agrovoc:c_7887 ;
afy:hasProperty       res:hasTreeTrunkDiameter ;
geo:hasDefaultGeometry  res:PA3AFL01A10_Geom_1 ;
time:hasTime          res:PA3AFL01A10_LifeTime ;
sosa:isFeatureOfInterestOf  res:PA3AFL01A10_Obs_2015_TreeTrunkDiameter .

res:PA3AFL01A10_Geom_1                            Géolocalisation
a                     sf:Point ;
geo:asWKT             "POINT (3.85885733993818 43.705661294259)" ^^geo:wktLiteral .

res:PA3AFL01A10_LifeTime                          Durée de vie
a                     time:Interval ;
time:hasBeginning     res:date_of_tree_planting .

                                  Ajout d'une observation sur le diamètre
res:PA3AFL01A10_Obs_2015_TreeTrunkDiameter
a                     sosa:Observation ;
sosa:hasFeatureOfInterest  res:PA3AFL01A10 ;
sosa:observedProperty  res:hasTreeTrunkDiameter ;
sosa:hasResult        res:PA3AFL01A10_Obs_2015_TreeTrunkDiameter_Measure ;
sosa:phenomenonTime   res:PA3AFL01A10_Obs_2015_TreeTrunkDiameter_Time ;
sosa:usedProcedure    res:Forest_mensuration .
sosa:madeBySensor     res:Lydie_Dufour ;

res:PA3AFL01A10_Obs_2015_TreeTrunkDiameter_Measure
a                     om:Measure ;
om:hasNumericalValue  3.25E1 ;
om:hasUnit            om:centimetre .

res:PA3AFL01A10_Obs_2015_TreeTrunkDiameter_Time
a                     time:Instant ;
time:inXSDgYear       "2015"^^xsd:gYear .
```

FIGURE 9.21 – *Extrait de la base de connaissance Agroforestry dans la syntaxe Turtle*

A ce stade, la base de connaissance Agroforestry peut être manipulée par des outils d'édition spécialisés comme le logiciel Protégé (voir Fig.9.22) [MALVIYA et al., 2011; JAIN et SINGH, 2013]. Néanmoins pour poursuivre notre démarche visant à mettre en avant l'utilité d'une base de connaissance pour la gestion des aménagements agroforestiers, nous avons conçu une application qui place la base de connaissance en son centre .

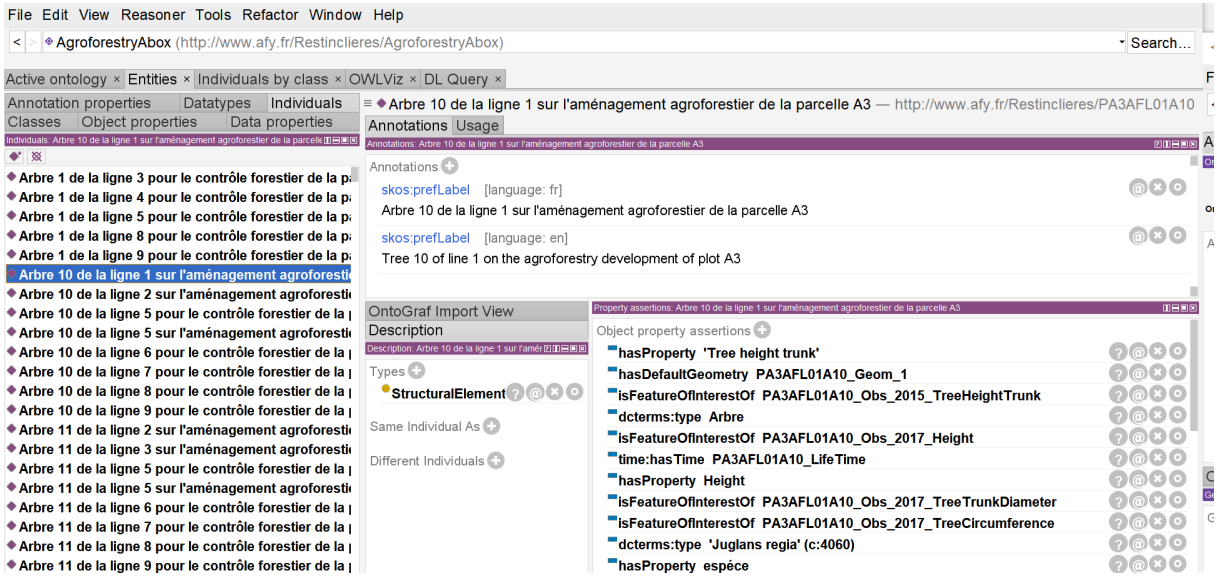


FIGURE 9.22 – *Détails de l'individu PA3AFL01A10 (qui est le dixième arbre de la première ligne de la parcelle A3) de notre base de connaissance visualisés avec le logiciel Protégé.*

9.1.6 L'application AOBRA

Vue générale de l'application

Nous avons conçu l'application avec une visée purement opérationnelle. Dans l'idée, l'agroforestier doit pouvoir tirer parti de ses jeux de données de manière quasi automatique. À cet effet, l'application est exploitable au travers de 3 phases qui vont se dérouler successivement dans le temps (Figure 9.23) :

- **La modélisation :**
  Les données fournies par les agriculteurs, les forestiers et les chercheurs, se présentent souvent sous la forme de classeurs Excel ou de bases de données le plus souvent relationnelles. À ce stade, nous allons transformer ces données et leurs modèles sous-jacents, en descriptions formelles qui pourront ensuite être utilisées sous une forme sémantisée. Ces données sont stockées dans un des formats de sérialisation OWL. L'adoption des standards du web sémantique permet de lier facilement les données collectées à d'autres bases de connaissance disponibles sur le web dans les mêmes formats.
- **Le raisonnement :**
  Des mécanismes d'inférence sont appliqués aux données afin d'en déduire de nouvelles connaissances. Au cours de cette phase, les données sont agrégées dans l'espace et le temps, et plusieurs calculs sont effectués, ce qui permet de consolider les données quantitatives pour la création d'indicateurs qualitatifs. Les différences mécanismes d'inférence seront déclenchés en séquentiel, et la base de connaissance va être appelée à grossir pour in fine contenir tous les individus d'intérêt pour l'agroforestier.

– **Les résultats :**
Les données déduites des activités de raisonnement sont de nouvelles données sémantiques qui peuvent être réutilisées pour participer à de nouvelles déductions, amplifiant ainsi la création de nouvelles connaissances. Les données natives et déduites sont ensuite extraites à l'aide du langage d'interrogation SPARQL, pour répondre aux requêtes des utilisateurs via l'interface de l'application ou pour alimenter d'autres applications (par exemple un site web via des API).

Au cours de la phase de modélisation, nous utilisons l'ontologie cadre Agroforestry décrite au dessus. Le modèle Agroforestry fournit non seulement la description physique des éléments biotiques et abiotiques présents sur une parcelle d'agroforesterie, mais aussi leurs interactions et leurs impacts sur la parcelle.

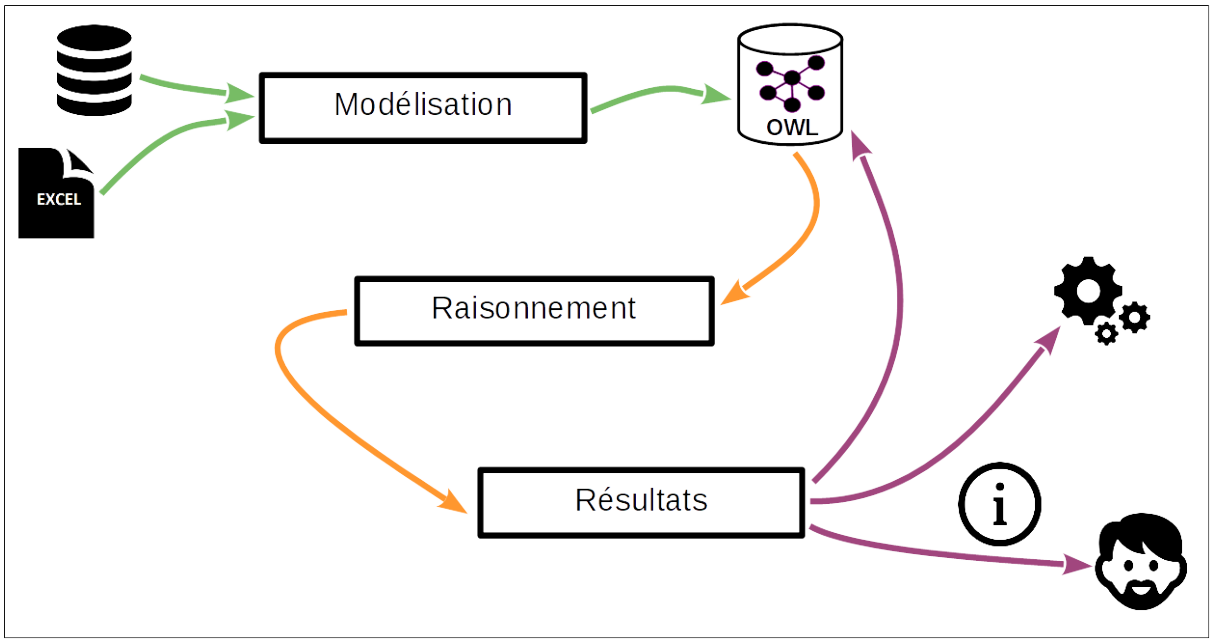


FIGURE 9.23 – *Vue d'ensemble de l'architecture de l'application AOBRA.*

Les mécanismes de raisonnement de l'application

Nous avons réalisé l'implémentation d'une couche applicative en Java qui s'appuie sur le framework Jena.[MCBRIDE, 2002]. Pour les activités de raisonnement, le processus général est présenté figure 9.24.
Le mécanisme d'inférence se déroule en plusieurs étapes successives :

– **Première étape : inférences avec l'outil PELLET et les axiomes OWL**
Le modèle initial comprenant l'ontologie Agroforestry (Tbox) et les individus de la Abox est soumis au moteur d'inférence PELLET. Cette activité aura pour effet de dégager certains faits (sous la forme de nouveaux triplets RDF) jusqu'alors implicites. Par exemple, la nature transitive et réflexive de la relation *BFO_0000051 "has part"* et de son inverse *BFO_0000050 (part of)*, permettront de rendre explicites les relations contenu/conteneur non exprimées dans le modèle original. Ce mécanisme d'inférence est rendu possible par l'utilisation du langage OWL qui fournit une formalisation pour la description des caractéristiques algébriques d'un prédicat utilisé dans un triplet RDF.

– **Deuxième étape : inférences basées sur les « Règles Jena »**
Les possibilités d'inférence basées sur le langage OWL sont limitées. Dans le but de dégager de nouveaux faits, nous utilisons les mécanismes basés sur des règles RDF définies au sein de l'environnement Jena (Jena Rules). Des fonctions créées par les utilisateurs en langage Java ou dans d'autres langages (tels que le langage R orienté vers les analyses statistiques) peuvent être appelées au sein des règles Jena, ce qui permet d'effectuer des calculs complexes, dont les résultats peuvent être utilisés lors de l'application des règles. L'intérêt pour nous est de pouvoir appuyer les règles sur un éventail de fonctions statistiques que nous construisons dans le langage R et que nous appelons dans le corps des règles et dans l'environnement Jena. Il est à noter que nous disposons d'un jeu de données essentiellement quantitatives et qui sont habituellement analysées à l'aide d'outils statistiques.

– **Troisième étape : l'extraction des données**
Le langage SPARQL nous permet d'interroger et de gérer les données, qu'elles soient natives ou inférées. Nous pouvons également mobiliser les fonctions statistiques au sein de requêtes SPARQL.

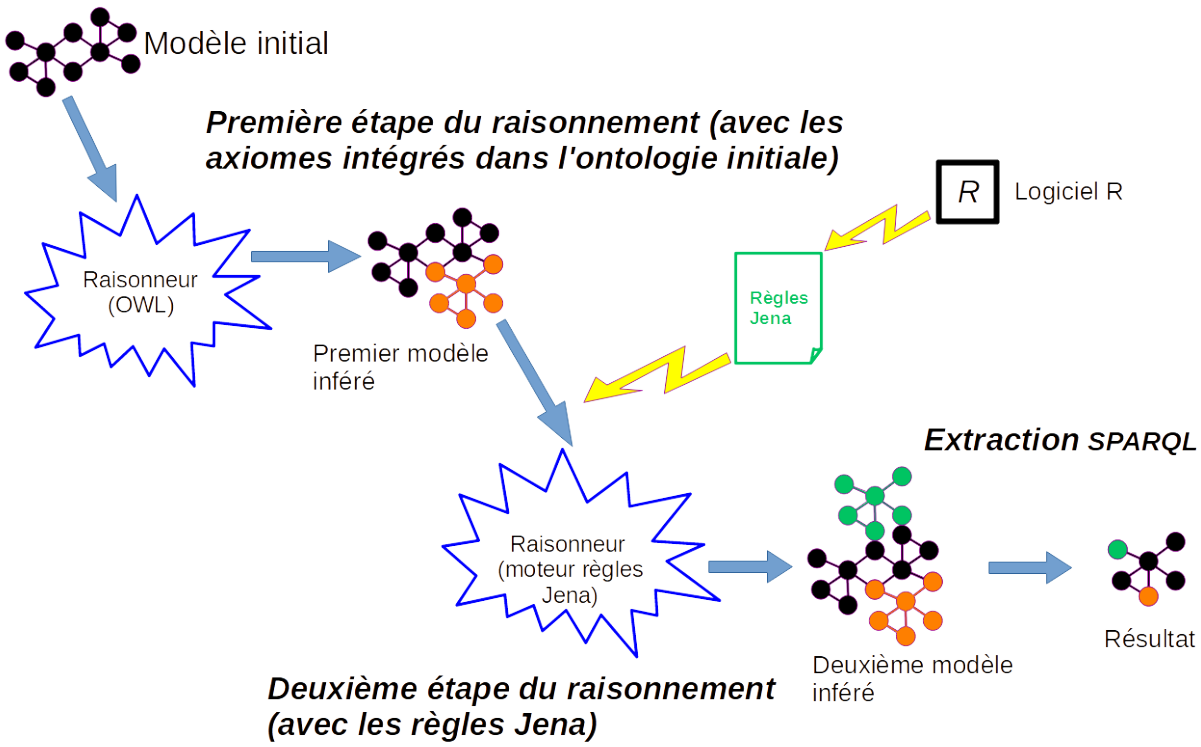


FIGURE 9.24 – *Schéma du mécanisme d'inférence de l'application AOBRA.*

9.1.7 Mode opératoire avec les axiomes OWL

La base de connaissance Agroforestry (Tbox + Abox) est soumise dans un premier temps au classifieur OWL PELLET (d'autres classifieurs OWL auraient pu se substituer à Pellet à l'exemple d'HermiT ou Fact++). Ceci aura pour effet de rendre explicites certains faits implicites en générant des triplets RDF supplémentaires sur la base des propriétés algébriques (réflexivité, transitivité, ...) des prédicats ou d'équivalence ou d'inclusion de concept présents initialement dans le modèle.

Utilisation des propriétés d'un prédicat

Pour exemple la transitivité et la réflexivité de la relation *BFO_0000051* et de son inverse *BFO_0000050* vont être exploitées pour expliciter les relations contenu/contenant non explicitées dans le modèle à l'origine. Ce mécanisme d'inférence est rendu possible par l'expressivité du langage OWL qui fournit les primitives de modélisation pour la description des caractéristiques algébriques d'un prédicat utilisé dans un triplet RDF comme le montre l'extrait en syntaxe Turtle (Fig. 9.25) qui définit la nature transitive et réflexive de la relation *BFO_0000051* et de son inverse *BFO_0000050*.

```
@prefix bfo:        <http://purl.obolibrary.org/obo/> .
@prefix owl:        <http://www.w3.org/2002/07/owl#> .
@prefix rdf:        <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs:       <http://www.w3.org/2000/01/rdf-schema#> .
[ rdf:type owl:Ontology] .
#################################################################
# Object Properties
#################################################################
### http://purl.obolibrary.org/obo/BFO_0000051
bfo:BFO_0000051   rdf:type        owl:ObjectProperty , owl:TransitiveProperty, owl:ReflexiveProperty ;
                  rdfs:label      "has part"@en .
### http://purl.obolibrary.org/obo/BFO_0000050
bfo:BFO_0000050   rdf:type        owl:ObjectProperty , owl:TransitiveProperty, owl:ReflexiveProperty ;
                  owl:inverseOf   bfo:BFO_0000051 ;
                  rdfs:label      "part of"@en .
```

FIGURE 9.25 – *Extrait en syntaxe Turtle de l'expression de la transitivité et de la réflexivité de la relation* BFO_0000051 *et de son inverse* BFO_0000050.

Un exemple du résultat de ce mécanisme est montré en figure 9.26 ou l'on voit que la présence explicite de relations *BFO_0000051* liant une parcelle à une ligne d'arbres et cette ligne d'arbres à des arbres dégage de nouvelles relations *BFO_0000051* entre la parcelle et les arbres et de nouvelles relations *BFO_0000050* entre les différents éléments liés par des relations *BFO_0000051*.

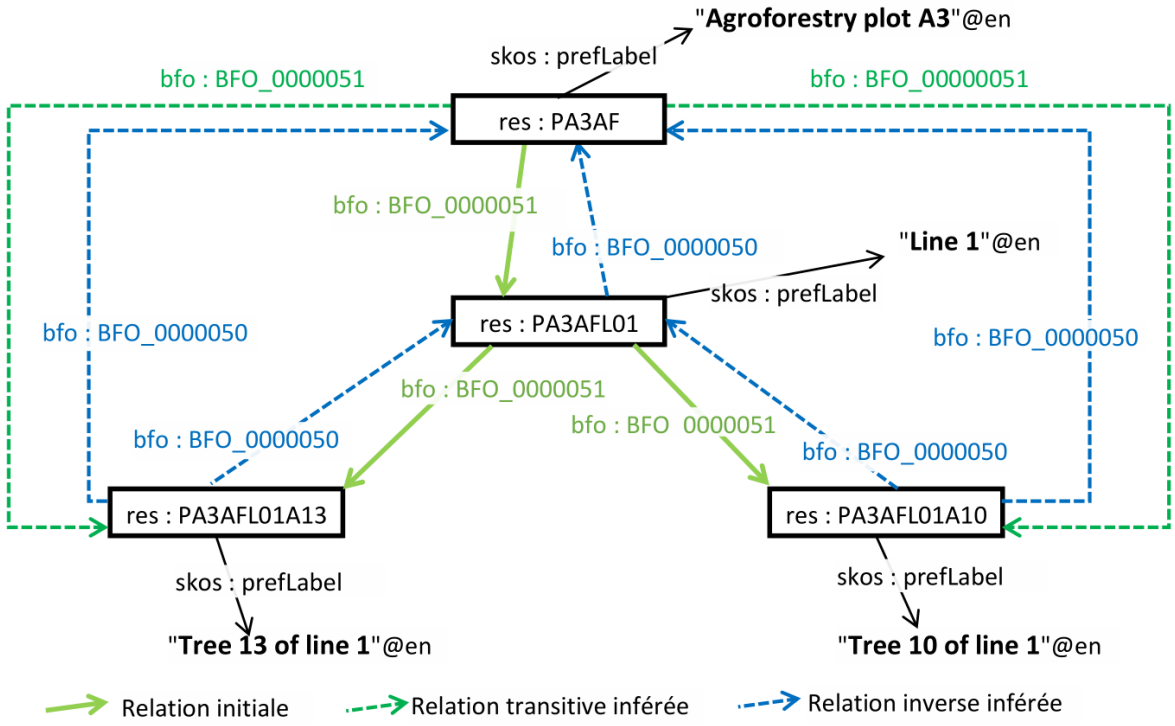


FIGURE 9.26 – *Exemple d'une inférence avec le moteur PELLET : Exploitation de la transitivité de la relation* BFO_0000051 *et de sa relation inverse* BFO_0000050.

Utilisation de la création d'un concept par extension

Ce nouvel exemple montre que le langage OWL permet aussi de matérialiser des concepts par extension comme avec cet extrait figure 9.27 qui définit les indidvidus des classes des arbres valides pour l'étude (**TreeForStudy**) et des arbres agroforestiers valides pour l'étude (**AgroforestryTreeForStudy**) par équivalence de concept. Les arbres valides pour l'étude sont définis comme des éléments de structure possédant les caractéristiques de diamètre à hauteur de poitrine (*breast_height_diameter*) et de hauteur de tronc (*height_trunk*) faisant chacune l'objet d'au moins une observation. Les arbres agroforestiers validés pour étude sont des arbres validés pour l'étude contenus dans une parcelle agroforestière.
Ainsi comme le montre la figure 9.28 l'individu "afy:PA3AFL01A0" qui satisfait les deux restrictions précédentes sera considéré après application des mécanismes d'inférence comme une instance des classes **TreeForStudy** et **AgroforestryTreeForStudy**.
On notera au passage qu'une relation issue d'un premier axiome comme ici l'appartenance de l'élément "afy:PA3AFL01A0" à la classe **TreeForStudy** peut être utilisée dans un second axiome qui dans notre cas infère son appartenance à la classe **AgroforestryTreeForStudy** à partir de son appartenance préalable à la classe **TreeForStudy**.

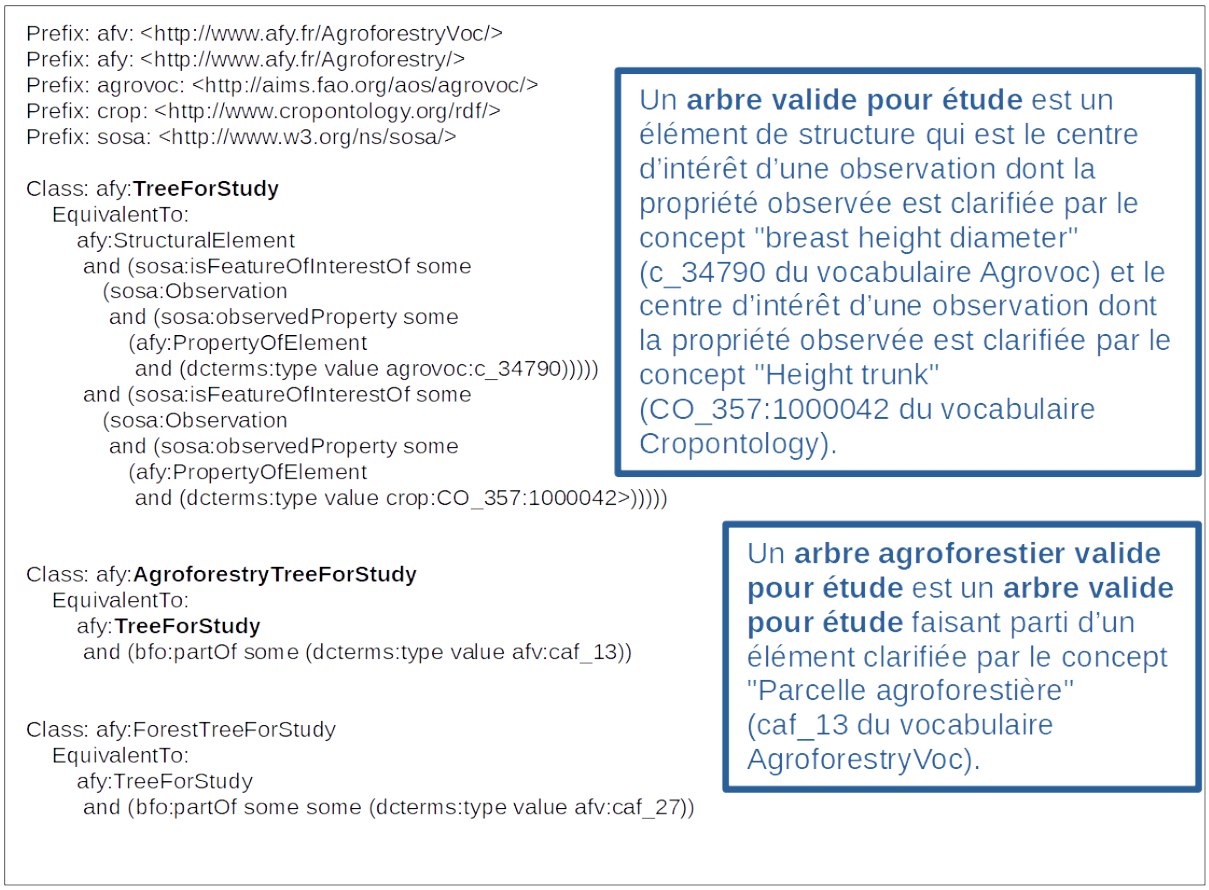


FIGURE 9.27 – *Exemple de la définition des classes* ***TreeForStudy*** *et* ***AgroforestryTreeForStudy*** *en syntaxe OWL Manchester.*

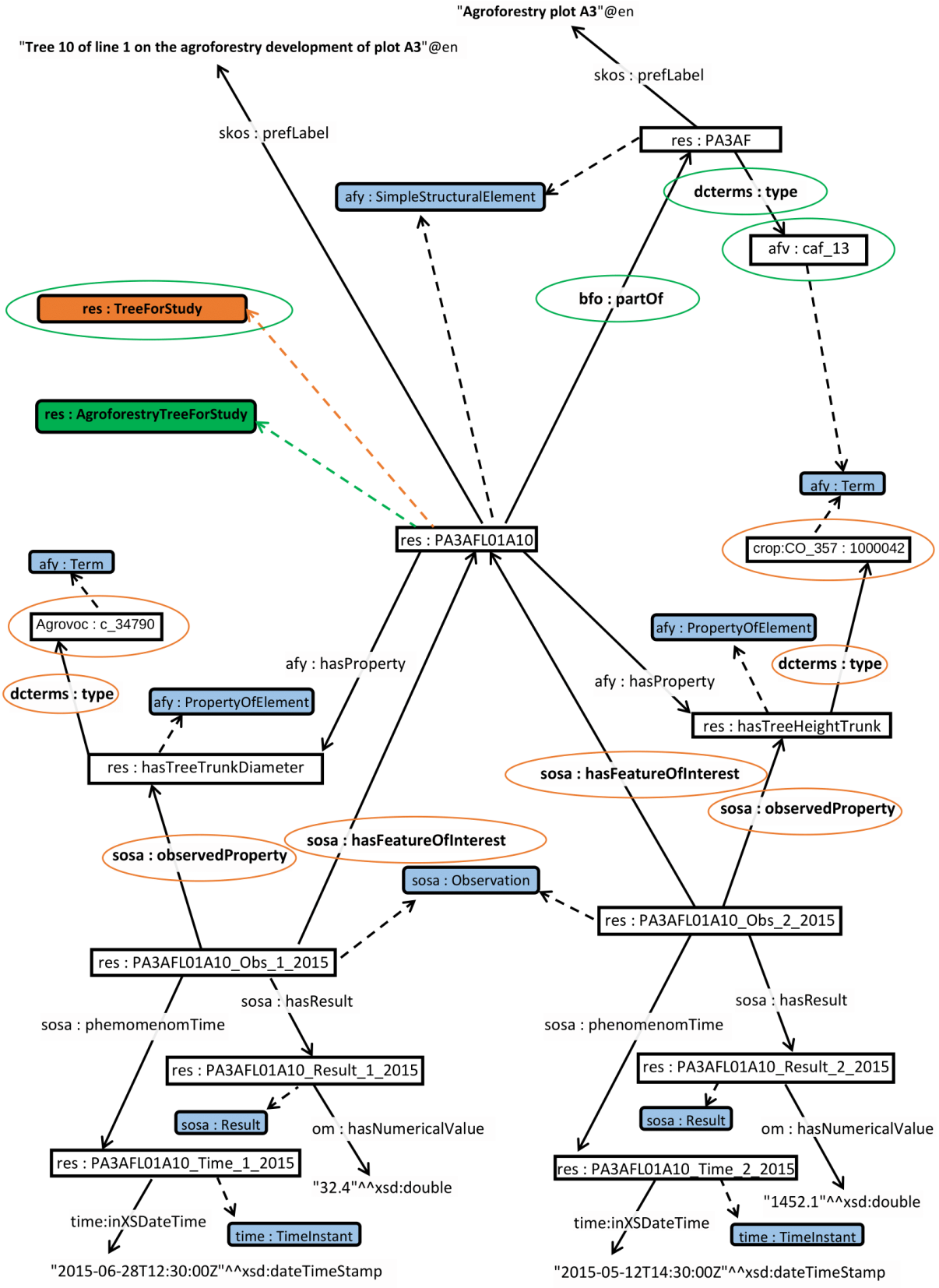


FIGURE 9.28 – *Exemple d'une inférence avec le moteur PELLET : équivalence de concept appliquée à l'individu PA3AFL01A10. Les ellipses orange indiquent les relations et valeurs prises en compte pour inférer l'appartenance de l'individu PA3AFL01A10 à la classe* ***TreeForStudy****. Les ellipses verte indiquent les relations et valeurs prises en compte pour inférer l'appartenance de l'individu PA3AFL01A10 à la classe* ***AgroforestryTreeForStudy****.*

Les relations nouvellement créées par des axiomes peuvent être alors exploitées et extraites au même titre que les relations natives du modèle.
Mais OWL ne permet pas d'exprimer toutes les relations possibles entre deux entités du monde réel. Il n'y a par exemple aucun moyen dans le vocabulaire d'OWL d'exprimer la relation entre des individus avec lesquels un individu aurait des relations, ce qui implique que l'on ne peut définir une classe dans laquelle une parcelle agroforestière aurait deux parcelles témoins, l'une témoin forestier et l'autre témoin agricole qui seraient toutes deux accolées. On peut toutefois étendre l'expressivité de OWL en ajoutant par exemple des règles SWRL (Semantic Web Rule Language) à une ontologie [HORROCKS, PATEL-SCHNEIDER, BOLEY et al., 2004]. Pour AOBRA, nous avons plutôt mobilisé le système de règles fourni par Jena.

9.1.8 Mode opératoire avec l'application de règles Jena

Les énoncés sous la forme "SI ... ALORS ... " qui vont être créés pour ce cas d'étude seront ensuite transposés dans les règles Jena. La transition vers les règles Jena est facilitée par le fait que les caractéristiques d'un individu sont représentées par des propriétés qui sont généralement décrites de manière standardisée dans des référentiels terminologiques. Ainsi, les instances de la classe **afy:StructuralElement** sont associées à un ensemble de propriétés via la relation *afy:hasProperty*. De même, les instances de la classe **sosa:Observation** sont liées à une propriété via la relation *sosa:observedProperty*. Cependant, les instances de la classe **afy:StructuralElement** sont liées par la relation *sosa:isFeatureOfInterestOf* aux instances de la classe **sosa:Observation** pour traduire le lien entre les observations et l'entité observée.

Pour illustrer le mode opératoire retenu pour la mise en place de règles Jena au sein d'AOBRA, nous en présentons quelque exemples dans les paragraphes suivants .

Traitement systématique des valeurs observées des caractéristiques d'éléments

Nous avons construit un ensemble de règles Jena complémentaires, s'appuyant sur des mécanismes de chaînage avant, qui sont avant tout des règles visant à enrichir la base de connaissance afin de traiter toutes les mesures collectées sur les parcelles de manière exhaustive. L'objectif est en effet de généraliser l'ajout de triplets dans la base de connaissance pour toutes les années, toutes les parcelles, toutes les observations, toutes les caractéristiques et leurs valeurs quantitatives mesurées. Nous serons ainsi en capacité de disposer de collections de valeurs par caractéristique, parcelle et année. La capacité de matérialisation des règles en chaînage avant est donc exploitée de façon systématique. Les règles construites se fondent sur le modèle cadre Agroforestry et tirent parti en particulier de son adossement au modèle SOSA.
Pour ensuite disposer d'une structure support pour les collections/agrégats de valeurs, nous avons défini de nouvelles règles dont certaines exploitent une stratégie hybride (chaînage avant/chaînage arrière). La définition des agrégats passe par l'explicitation de nœuds anonymes RDF (blank node) mobilisant la fonction prédéfinie "makeSkolem". Ainsi les nœuds anonymes seront des conteneurs de valeurs pour une caractéristique, une année et une parcelle donnée. Le nœud (et différentes valeurs associées) est manipulé sous la forme d'un vecteur noté (c (x,y,z,...)) et est lié à la parcelle concernée. Les données correspondantes au vecteur sont conservées au sein de la base de connaissance. Par suite, les différents vecteurs pourront alors être utilisés pour divers traitements statistiques (moyenne, quantiles, tests statistiques de Kolmogorof -Smirnov ou de Shapiro-Wilk ,...), ...).

Nous présentons certaines de ces règles. Pour illustrer la démarche, nous traitons un calcul élémentaire de moyenne à l'aide de la dernière règle présentée qui est nommée de ce fait "CalculationMean".

Nous listons les espaces de noms exploités dans les règles (voir figure 9.29), de manière à ne pas les rappeler pour chaque règle et ainsi faciliter la lecture. Les règles font donc appel à l'ontologie cadre Agroforestry, aux principales ontologies de support d'Agroforestry, à l'exemple de SOSA, GeoSPARQL, OWL-Time, SKOS, Dublin Core et OM. L'ontologie de domaine qui s'appuie sur le modèle cadre Agroforestry et qui a été définie pour le domaine de Restinclières est un composant central dans la construction des règles.

```
// STATISTICAL STUDIES OF THE MEASURES
@prefix  res:    <http://www.afy.fr/Restinclieres/> .
@prefix  afy:    <http://www.afy.fr/Agroforestry/> .
@prefix  ssn:    <http://www.w3.org/ns/ssn/> .
@prefix  sosa:   <http://www.w3.org/ns/sosa/> .
@prefix  rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix  om:     <http://www.ontology-of-units-of-measure.org/resource/om-2/> .
@prefix  time:   <http://www.w3.org/2006/time#> .
@prefix  bfo:    <http://purl.obolibrary.org/obo/> .
@prefix  geo:    <http://www.opengis.net/ont/geosparql#> .
@prefix  sf:     <http://www.opengis.net/ont/sf#> .
@prefix  dc:     <http://purl.org/dc/elements/1.1/> .
@prefix  dcterms:<http://purl.org/dc/terms/> .
@prefix  skos:   <http://www.w3.org/2004/02/skos/core#> .
@prefix  afv:    <http://www.afy.fr/AgroforestryVoc/> .
@prefix  sio:    <http://semanticscience.org/resource/> .
@prefix  agrovoc:<http://aims.fao.org/aos/agrovoc/> .
```

FIGURE 9.29 – *Liste des espaces de noms exploités dans nos règles.*

– **Première règle : identification de toutes les propriétés**

Il s'agit de déterminer l'ensemble des propriétés des éléments du modèle, qui ont pu faire l'objet d'une observation, et qui sont donc associées à au moins une valeur. Ces propriétés sont alors reliées à la valeur booléenne 'true' au moyen de la relation *isPropertyToStudy*. Un nom « local » leur est attribué (au regard de leur URI, il s'agit juste de l'alias, et l'espace de noms n'est pas conservé). Nous exploitons une règle en chaînage avant qui va permettre d'explorer l'ensemble des propriétés qui ont fait l'objet d'au moins une observation. Le modèle SOSA est central dans cette recherche. L'objectif est ici de préciser les propriétés et leurs mesures qui seront ensuite traitées. La figure 9.30 montre le déroulé de l'identification des ces propriétés susceptibles d'être étudiées avec cet énoncé sous la forme "SI ... ALORS ... ".

La règle Jena qui découle de cet énoncé est présentée en figure 9.31. Une règle Jena est définie entre crochets [...] et le premier littéral suivi du symbole : représente sa dénomination. Le corps et la tête de la règle (à l'image des clauses de Horn) sont séparés par le symbole ->. On retrouve au sein du corps et de la tête la structure d'un triplet RDF (sujet, prédicat, objet) pour chaque ligne délimitée entre parenthèses (...). Lorsqu'un élément d'un triplet est précédé d'un point d'interrogation ?, il s'agit alors d'une variable dont la portée s'étend sur l'ensemble de la règle (corps et tête). Lorsqu'une variable est introduite dans un triplet, elle prend tour à tour

Pour tous les éléments, propriétés et observations, **SI** il y a un élément qui a une propriété **ET** qu'il existe une observation par laquelle cette propriété est observée **ET** que cette observation a comme centre d'intérêt cet élément **ALORS** cette propriété est une propriété à étudier.

FIGURE 9.30 – *Algorithme d'identification des propriétés à étudier des différents éléments de la base de connaissance Agroforestry.*

```
 // Retrieving all properties related to the structural elements and observations
[RetrievingProperties:
        (?structuralElement afy:hasProperty ?property)
        (?observation sosa:observedProperty ?property)
        (?observation sosa:hasFeatureOfInterest ?structuralElement)
        ->
        (?property res:isPropertyToStudy 'true'^^xsd:Boolean)
]
```

FIGURE 9.31 – *Règle Jena pour identifier les propriétés à étudier.*

toutes les valeurs des éléments présents dans les triplets du modèle traité qui correspondent au schéma fourni par le triplet issu de la règle. Ainsi le premier triplet de la règle indique que l'on traitera tous les couples sujets objets liés par la relation *afy:hasProperty* dont l'objet (?property) sera lié à un autre sujet (?observation) par la relation *sosa:observedProperty*, lui-même lié par la relation *sosa:hasFeatureOfInterest* au sujet initial (?structuralElement). A l'exécution de cette règle, tous les éléments du modèle correspondant aux valeurs extraites de la variable ?property se verront liés par la relation *res:isPropertyToStudy* à la valeur booléenne 'true'^^xsd:Boolean.

Après application de la règle, les propriétés caractéristiques à étudier pourront donc être extraites du modèle inféré grâce à cette nouvelle relation *isPropertyToStudy*. La figure 9.32 présente la requête SPARQL de type consultation, qui va permettre de renvoyer les relations inférées. On retrouve là encore une organisation sous forme de triplets contenant des variables. On extrait ici tous les sujets du modèle liés par la relation *rdf:type* à l'objet afy:PropertyOfElement qui n'est autre que la classe des propriétés de notre modèle Agroforestry et liés par la relation *res:isPropertyToStudy* à la valeur booléenne 'true'.
La figure 9.33 présente les résultats en sortie de traitement.

```
 SELECT ?p
WHERE {
      ?p rdf:type afy:PropertyOfElement .
      ?p res:isPropertyToStudy 'true'^^xsd:Boolean .
} ORDER BY ?p ;
```

FIGURE 9.32 – *Requête SPARQL qui retourne les propriétés à étudier.*

Propriétés à etudier
Propriété
<http://www.afy.fr/Restinclieres/hasHeight>
<http://www.afy.fr/Restinclieres/hasTreeHeightTrunk>
<http://www.afy.fr/Restinclieres/hasTreeTrunkDiameter>
<http://www.afy.fr/Restinclieres/hasSpecies>

FIGURE 9.33 – *Résultat de l'extraction avec SPARQL des propriétés susceptibles d'être étudiées*

– **Deuxième règle : détermination des années d'étude**

  Il s'agit maintenant de matérialiser les différentes années à étudier comme autant d'instances de la classe **YearForStudy** de l'ontologie de Restinclières(il est à noter que les mesures dendrométriques sont annuelles, ce qui explique le pas de temps choisi). Les individus années sont liés par la propriété *hasYear* à une valeur numérique représentant l'année concernée (voir figure 9.34). Le corps de la règle définie en chaînage avant est vide, et son exécution va être systématique.

```
// Years of study
[YearsOfStudy:
        ->
        (:yearForStudy2015 rdf:type :YearForStudy)
        (:yearForStudy2016 rdf:type :YearForStudy)
        (:yearForStudy2015 :hasYear '2015'^^xsd:gYear)
        (:yearForStudy2016 :hasYear '2016'^^xsd:gYear)
        (:yearForStudy2017 :hasYear '2017'^^xsd:gYear)
        (:yearForStudy2015 :hasRank '1'^^xsd:int)
        (:yearForStudy2016 :hasRank '2'^^xsd:int)
        (:yearForStudy2017 :hasRank '3'^^xsd:int)
]
```

FIGURE 9.34 – *Règle Jena pour la détermination des années à étudier.*

Après avoir identifié les propriétés observées et déterminé les années à étudier, nous nous intéressons maintenant aux valeurs recueillies lors de ces observations. Au sein de nos unités de recherche, il est fréquent, d'étudier de manière statistique les résultats de mesures faites sur le terrain . Ces études se font en grande partie avec l'aide du "Langage R" [5] spécialisé en traitement statistique et qui offre entre autres avantages le fait d'être libre et gratuit et de bénéficier d'une communauté de support très active. L'utilisation de R va nous permettre de ne pas implémenter, dans le cadre de notre système, de fonctions statistiques en utilisant les nombreuses librairies de ce langage. Au travers des règles suivantes nous présentons la manière dont nous préparons les données d'observation et la façon dont nous exploitons R au sein des règles Jena.

– **Troisième règle : création de nœuds anonymes, conteneurs des valeurs à agréger**

Pour agréger les valeurs observées de chacune des propriétés par année et par parcelle, nous avons exploité la relation *bfo:hasPart* BFO_0000050 (has part) et sa relation inverse *bfo:partOf* BFO_0000050 (part of) entre les arbres et les parcelles dont les assertions manquantes ont été générées en raison du mécanisme d'inférence vu précédemment (voir figure 9.26). Le diagramme RDF de la figure 9.35 montre pour les arbres 10 et 13 de la parcelle A3 les relations qui les lient à une propriété observée (ici la propriété "res:hasTreeTrunkDiameter"), aux observations faites en 2015 et 2016 sur les valeurs de cette propriété et les résultats et estampilles temporelles de ces observations.
La règle suivante renvoie donc pour chaque parcelle, toutes les valeurs observées pour chaque propriété et année étudiées de chaque élément en reprenant le cheminement des relations de la figure 9.35.
La règle permet de dégager des nœuds anonymes pour chaque parcelle, chaque propriété à étudier, et chaque année d'étude (voir figure 9.36). Ces nœuds anonymes seront donc liés à l'ensemble des valeurs d'une propriété pour une parcelle donnée sur un intervalle de temps précisé. Cette propriété aura fait l'objet d'au moins une observation pour un élément au sein de la parcelle, et au cours d'une des années étudiées.

Nous commenterons cette règle ainsi :

a) : Pour toutes les parcelles, toutes les propriétés à étudier et toutes les années à étudier.

b) : La valeur numérique observée (variable ?Value) d'une propriété pour un élément d'une parcelle pour une année donnée est dans le corps de la règle.

c) : Il en va de même pour le nœud anonyme (désigné par ?blankNode) qui est défini à l'aide de la fonction prédéfinie Jena makeSkolem qui admet un nombre d'arguments arbitraire en fonction des besoins. L'URI du nœud anonyme (?blankNode) est construite par concaténation de l'espace de noms de l'ontologie ciblée et de l'identifiant de la parcelle, de celui de la propriété et de l'année considérée (par exemple : res :PA3AFhasHeightTrunk2015).

d) : Nous construisons aussi une URI pour une propriété (?uRIhasValue) composée de l'identifiant de l'élément concerné par la mesure (exemple : hasValuePA3AFL01A10).

e) : Dans la tête de la règle, nous retrouvons le nœud anonyme, qui est lié à la valeur numérique de la mesure concernée.

Les nœuds anonymes sont liés aux différentes valeurs concernées, recueillies à chaque itération, lors de l'exécution de cette règle. Le traitement pour un nœud ne sera donc complet qu'à l'issue du traitement.

5. Langage R : The R Project for Statistical Computing. https://www.r-project.org/.

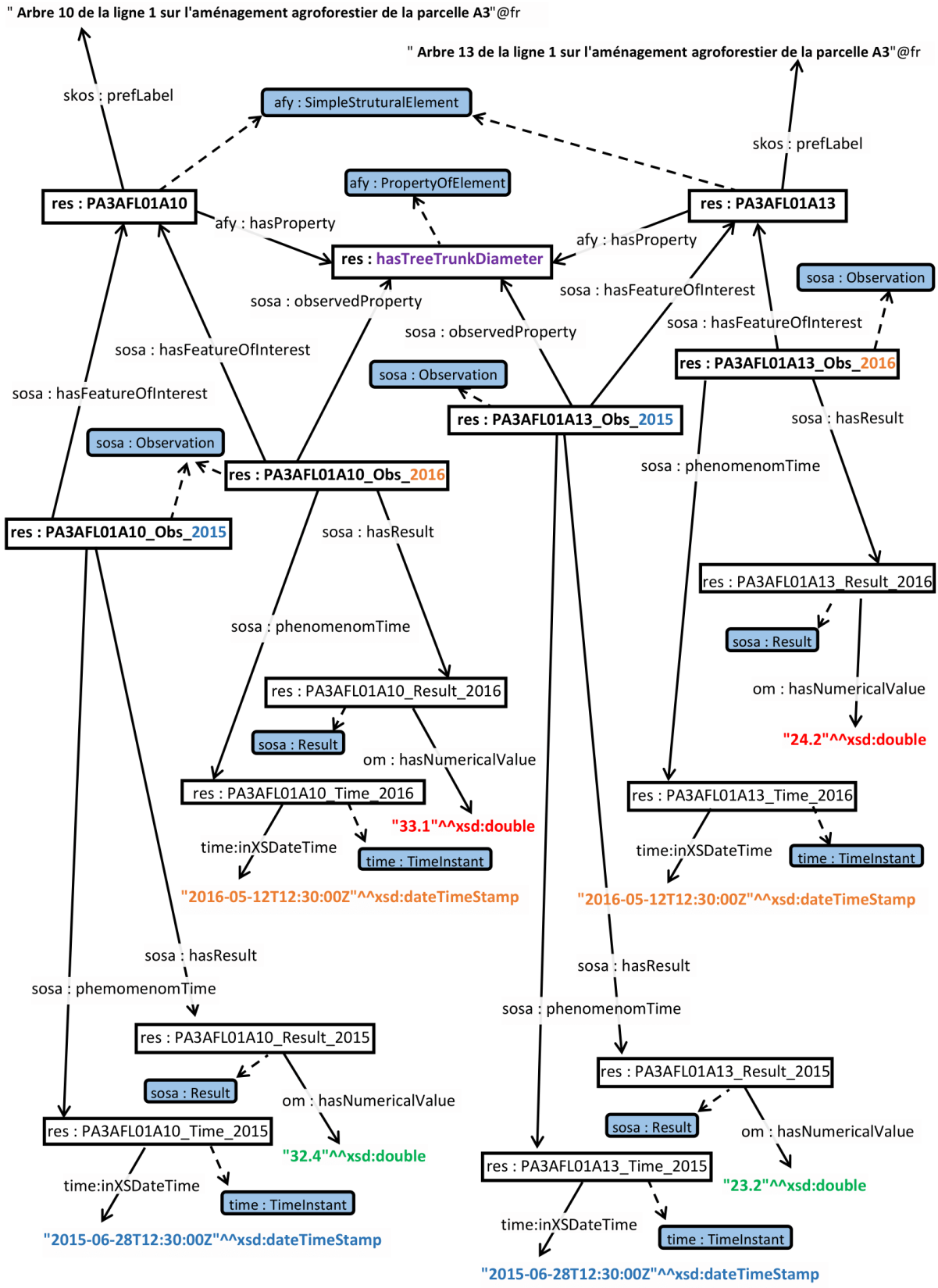


FIGURE 9.35 – *Diagramme RDF pour identifier les valeurs de la caractéristique* afy : Tree_Trunk_Diameter *par année.*

```
 // Recovery at the level of each plot of all observed values
// for each property and year studied with a blank node
[MakeBlankNodeObservedValuesForAYear:
        (?plot :isPlot 'true'^^xsd:Boolean)
        (?propertyStudied :isPropertyToStudy 'true'^^xsd:Boolean)        a)
        (?yearForStudy rdf:type :YearForStudy)
        (?yearForStudy :hasYear ?yearStudied)

        (?structuralElement bfo:partOf ?plot)
        (?structuralElement afy:hasProperty ?propertyStudied)
        (?structuralElement sosa:isFeatureOfInterestOf ?observation)
        (?observation sosa:observedProperty ?propertyStudied)
        (?observation sosa:hasResult ?measure)
        (?measure om:hasNumericalValue ?value)        b)
        (?observation sosa:phenomenonTime ?Time)
        (?Time time:inXSDgYear ?yearStudied)

        makeSkolem(?blankNode, ?plot, ?propertyStudied, ?yearStudied)        c)

        (:thisOntology :hasNameSpace ?ns)
        uriConcat(?ns, "hasValue", ?structuralElement, ?uRIhasValue)        d)
        ->
        (?blankNode ?uRIhasValue ?value)        e)
]
```

FIGURE 9.36 – *Règle Jena pour la création de nœuds anonymes pour concaténer les valeurs observées pour chaque arbre par parcelle, propriété étudiée et année.*

– **Quatrième règle : transformation du nœud anonyme**

La règle suivante (voir figure 9.37) exploite les nœuds anonymes créés précédemment afin de les proposer sous la forme d'un littéral de type vecteur de données dans le langage R 'c(x, y, z, ...)'.
La règle se décompose ainsi :

a) : Pour toutes les parcelles, toutes les propriétés à étudier et toutes les années à étudier.

b) : Si un nœud anonyme avec le même identifiant a déjà été créé (ici au travers de la règle précédente, il conserve ses liens avec les valeurs numériques déjà extraites .

c) : Nous faisons en sorte de construire l'identifiant d'une nouvelle propriété (?uRIProperty) par concaténation de l'espace de noms, du nom de la propriété, du mot 'Vector' et de l'année. (exemple 'res :hasTreeHeightTrunkVector2015')

d) : La tête de cette règle se compose également d'une sous-règle en chaînage arrière. Le moteur Jena en fonctionnement hybride (chaînage avant et arrière) traite en premier lieu toutes les règles de chaînage avant, et matérialisent tous les triplets pouvant en être explicités (les nœuds sont donc associés à toutes les valeurs possibles), avant de chercher à effacer les objectifs des règles en chaînage arrière. Nous nous servons de ce mode de fonctionnement, pour faire en sorte que la règle précédente 'MakeBlankNodeObservedValues' ait été exploitée entièrement.

```
 // Fabrication of property value vectors with blank nodes
[FabricationPropertyValuesVectorsForAYear:
      (?plot :isPlot 'true'^^xsd:Boolean)
      (?propertyStudied :isPropertyToStudy 'true'^^xsd:Boolean)       a)
      (?propertyStudied :hasLocalName ?propertyName)
      (?yearForStudy rdf:type :YearForStudy)
      (?yearForStudy :hasYear ?yearStudied)

      makeSkolem(?blankNode, ?plot, ?propertyStudied, ?yearStudied)       b)

      (:thisOntology :hasNameSpace ?ns)
      uriConcat(?ns,?propertyName,"Vector",?yearStudied,?uRIProperty)       c)
      ->
      [       d)
      (?plot ?uRIProperty ?vector)       e)
      <-
      calcMakeVector(?blankNode, ?vector)       f)
      ]
]
```

FIGURE 9.37 – *Règle Jena pour la création de vecteurs de valeurs à partir de nœuds anonymes pour concaténer les valeurs observées.*

e) : Le vecteur R (sous sa forme littérale) est lié à la parcelle par la relation identifiée au préalable.

f) : La fonction utilisateur calcMakeVector se charge de structurer les valeurs numériques liées au noeud sous la forme d'un vecteur R (ici ?vector).

– **Cinquième règle : utilisation des vecteurs en entrée de divers traitements**

La dernière des 5 règles présentée (voir figure 9.38) est une règle qui illustre les traitements statistiques pouvant être appliqués sur les vecteurs construits. Nous pouvons donc définir diverses « cinquième règle » en fonction des besoins. Dans l'exemple donné, les valeurs contenues dans un vecteur sont agrégées et une moyenne est calculée. Cette moyenne sera conservée et liée à la parcelle par une nouvelle relation.

Nous commenterons cette règle ainsi :

a) : pour toutes les parcelles, toutes les propriétés à étudier et toutes les années à étudier.

b) : les identifiants des propriétés qui lient chacune, une parcelle à un vecteur sont retrouvés en exploitant le même modèle de concaténation que la règle précédente.

c) : la fonction utilisateur calcRMeanVector calcule la moyenne (?mean) de l'ensemble des valeurs contenues dans le vecteur. Cette fonction utilise une interface avec le langage R qui fournit la méthode 'mean' qui calcule la moyenne sur les valeurs contenues dans le vecteur (?vector).

d) : l'identifiant d'une nouvelle propriété (?uRIPropertyMean) est construite sur la base d'une concaténation de l'espace de noms, du nom de la propriété, du mot 'Mean' et de l'année. (par exemple 'res :hasTreeHeightTrunkMean2015')

e) : la moyenne issue du vecteur de données est liée à la parcelle

```
// Calculation of the mean with R
[CalculationMean:
      (?yearForStudy rdf:type :YearForStudy)
      (?yearForStudy :hasYear ?yearStudied)
      (?plot :isPlot 'true'^^xsd:Boolean)        a)
      (?propertyStudied :isPropertyToStudy 'true'^^xsd:Boolean)
      ->
      [
      (?plot ?uRIPropertyMean ?mean)        e)
      <-
      (:thisOntology :hasNameSpace ?ns)
      (?propertyStudied :hasLocalName ?propertyName)
      uriConcat(?ns,?propertyName,"Vector",?yearStudied,?uRIProperty)        b)
      (?plot ?uRIProperty ?vector)
      calcRMeanVector(?vector, ?mean)        c)
      uriConcat(?ns,?propertyName,"Mean",?yearStudied,?uRIPropertyMean)        d)
      ]
]
```

FIGURE 9.38 – *Règle Jena pour le calcul de la moyenne des vecteurs de valeurs.*

Le diagramme RDF de la figure 9.39 reprend de manière schématique ce processus d'agrégation avec l'exemple des propriétés nouvellement créées *res:hasTreeTrunkDiameterMean2015* et *res:hasTreeTrunkDiameterMean2016* liant respectivement la parcelle 'res:PA3AF' aux résultats des moyennes de mesures du diamètre des arbres de cette parcelle pour les années 2015 et 2016 obtenus à partir des vecteurs correspondants *res:hasTreeTrunkDiameterVector2015* et *res:hasTreeTrunkDiameterVector2016*.

Cet ensemble de règles va donc faciliter le traitement des valeurs numériques issues d'observations de caractéristiques diverses des arbres et permettre l'accès aux résultats de manière simplifié comme le présente la figure 9.40.
D'autre règles à l'image de 'CalculationMean' peuvent utiliser les mêmes vecteurs et faire appel à d'autres fonctions statistiques de R. Le principe est donc généralisable.

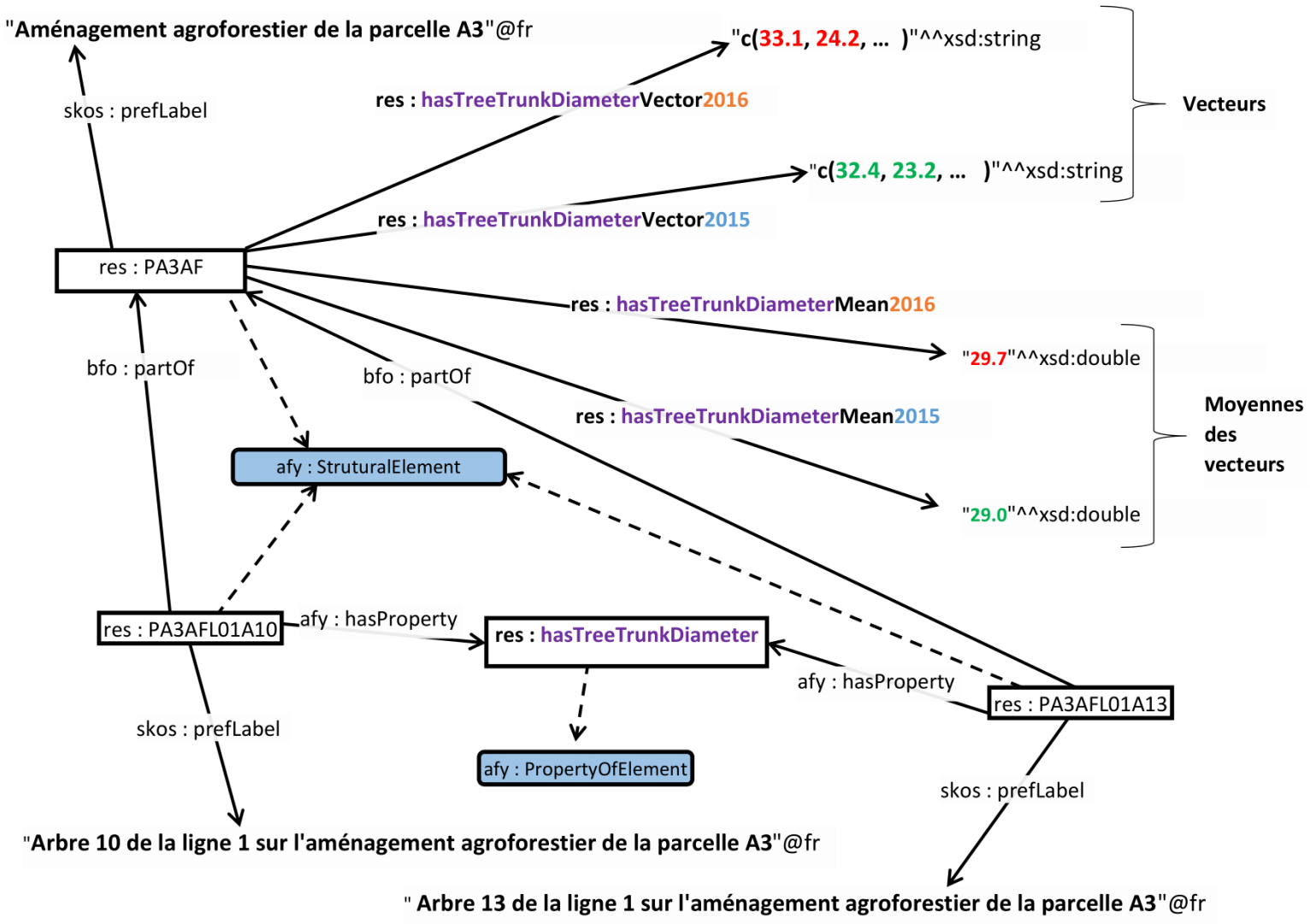


FIGURE 9.39 – *Diagramme RDF pour la création d'une moyenne des valeurs de la caractéristique* res : TreeTrunkDiameter *par année.*

Propriétés à étudier
Propriété
<http://www.afy.fr/Restinclieres/hasTreeHeightTrunk> <http://www.afy.fr/Restinclieres/hasTreeTrunkDiameter> ...

Années étudiées
Année
"2015" "2016" ...

```
SELECT (?plot AS ?Parcelle)
            (str(?valueMean) AS ?Moyenne_Diametres_Des_Troncs)
WHERE {
      ?plot res:hasTreeTrunkDiameterMean2015 ?valueMean .
} ORDER BY ?plot ;
```

FIGURE 9.40 – *Extraction des résultats statistiques avec une requête SPARQL. La connaissance des propriétés et des années étudiées facilite l'identification des relations liant une parcelle au résultat d'une fonction statistique appliquée sur les valeurs de ces propriétés.*

Création de nouvelles observations sur la base d'observations initiales

Des informations supplémentaires peuvent être déduites à partir d'observations antérieures. Dans l'exemple suivant, il s'agit de mettre en place une nouvelle observation sur la biomasse de chaque arbre à l'aide des informations obtenues sur leur hauteur d'élagage et leur diamètre.
La biomasse des arbres sera donc calculée par le système pour chaque année étudiée et sera retenue sous la forme d'une nouvelle observation.La règle utilisée pour accomplir cette tâche est visible en annexe B sur la figure B.1. Le calcul de la biomasse est effectué par une fonction utilisateur "calcTreeBiomass(?diameter, ?heightTrunk, ?biomass)", incorporée à la règle, qui à partir du diamètre et de la hauteur de l'arbre observés retourne la biomasse selon la formule approchée $Biomasse = ((\Pi * (Diamètre/2)^2) * Hauteur)/1000000$ en vigueur dans le milieu des négociants en bois.
Il en découle, pour chaque arbre d'une parcelle ayant fait l'objet d'observations simultanées sur sa hauteur et son diamètre, la création d'une observation sur la propriété *hasBiomass* ayant pour résultat numérique la biomasse en mètres cube valide pour la période concernée.
La figure 9.41 schématise le résultat de cette fonctionnalité du système, les arbres 1 et 2 sont enrichis d'un nouvelle observation sur la propriété *res:hasBiomass* dont les estampilles temporelles sont héritées de celles de leur observation sur leur hauteur et leur diamètre.

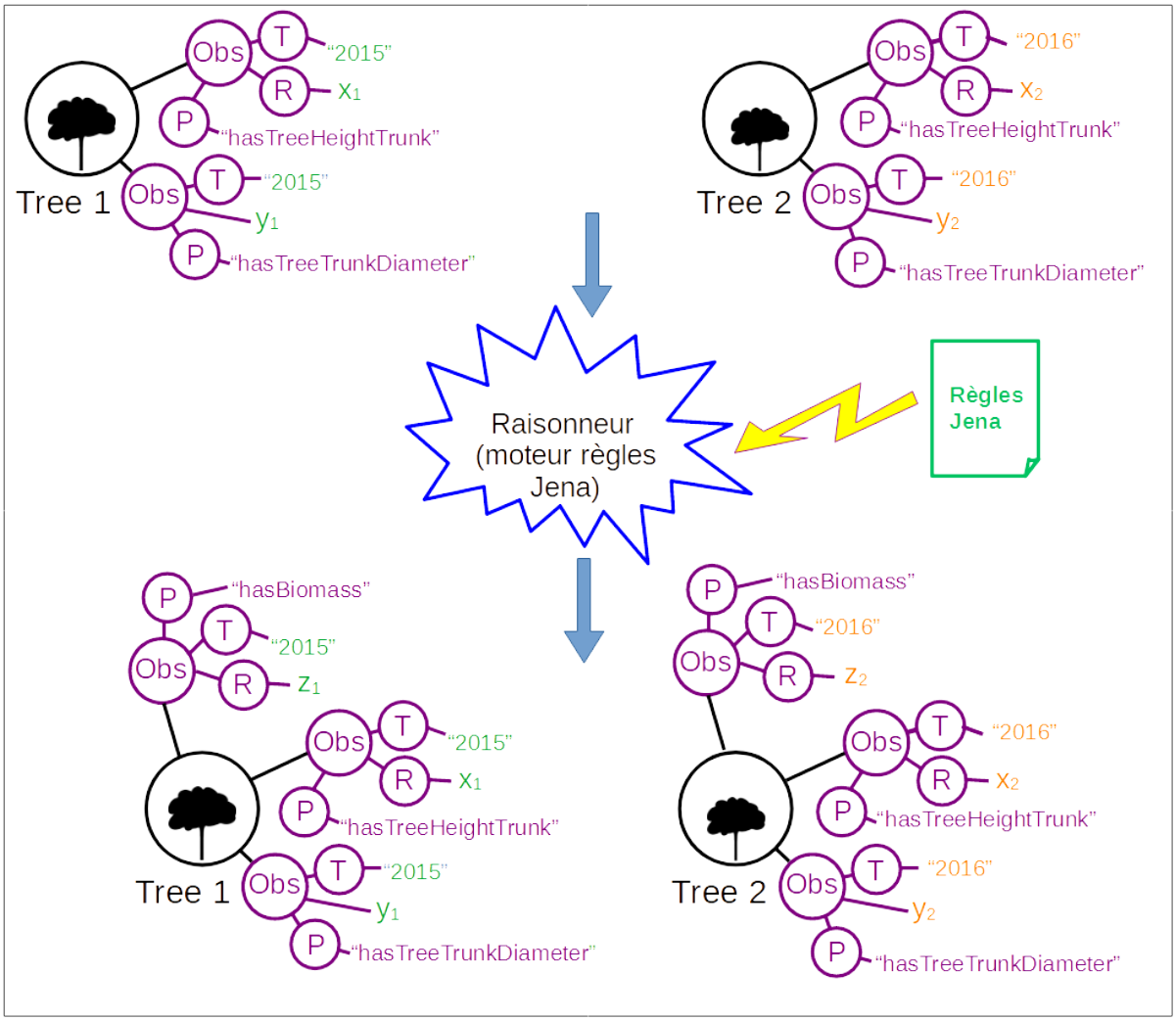


FIGURE 9.41 – *Mise en place d'une nouvelle observation sur la biomasse pour les arbres ayant été l'objet d'observations simultanées sur leur hauteur et leur diamètre pour une ou plusieurs des années étudiées.*

On note au passage que la création de cette nouvelle observation sur cette nouvelle propriété *hasBiomass* déclenche à nouveau l'exécution des règles vues précédemment ce qui conduit au niveau de la parcelle à la création d'un nouveau vecteur de valeurs de biomasse et le traitement de ce vecteur par des fonctions statistiques. Il est donc possible d'accéder par exemple à la médiane des valeurs de biomasse de chaque arbre de chaque parcelle en 2015 (voir figure 9.42).

```
SELECT (?plot AS ?Parcelle)
           (str(?value) AS ?Médiane_Biomasse_Des_Arbres)
WHERE {
     ?plot res:hasBiomassMedian2015 ?value .
} ORDER BY ?plot ;
```

Statistiques 2015 biomasse des arbres d'une parcelle

Parcelle	Médiane_Biomasse_Des_Arbres
<http://www.afy.fr/Restinclieres/PA3AF>	0.277
<http://www.afy.fr/Restinclieres/PA4TF>	0.100
<http://www.afy.fr/Restinclieres/PB17AF>	0.061
<http://www.afy.fr/Restinclieres/PB17TF>	0.056

FIGURE 9.42 – *Requête SPARQL pour l'extraction de la médiane des valeurs de biomasse des arbres d'une parcelle et son résultat.*

Création de nouvelles observations sur la base de critères imposés

Des éléments du modèle peuvent être catégorisés via une nouvelle relation sémantique en fonction de la valeur de l'une de leurs caractéristiques et de critères imposés par l'utilisateur.
L'exemple présenté en annexe B , section B.2 montre comment les arbres d'une parcelle sont catégorisés en fonction de leur diamètre selon des critères utilisés dans le négoce du bois.

Création de nouvelles observations sur la base d'une agrégation en fonction de la valeur observée d'une propriété particulière

Des agrégations sur la valeur d'une caractéristique des éléments d'une population peuvent être réalisées en fonction de la valeur d'une autre caractéristique afin d'étudier la corrélation entre celles ci.
On peut par exemple vouloir connaître la biomasse totale des arbres d'une parcelle en fonction de leur catégorie pour le négoce du bois car les revenus issus de la vente des billots de bois dépendra de la catégorie à laquelle il appartiennent.

Utilisation des capacités de traitement des données géospatiales de geoSPARQL

L'utilisation de geoSPARQL permet de représenter et d'interroger des données spatialisées au sein de notre application AOBRA. L'exemple suivant montre la capacité du système à détecter des lignes d'arbres sur la parcelle et à en donner l'orientation.
Voici les étapes nécessaires au déroulement de cette fonctionnalité :

- **Détermination du plus proche voisin**
  Dans cette étape , on détermine pour chaque arbre d'une parcelle agroforestière si il existe un arbre voisin (voir figure 9.43). Une première règle (voir figure 9.44) va donc déterminer pour chaque arbre d'une parcelle son plus proche voisin dans une limite de 6,5 mètres. Cette distance maximale a été choisie en raison de la littérature (la distance entre les lignes d'un aménagement d'agrisylviculture moderne descend rarement en deçà de cette limite pour permettre le passage des machines agricoles) afin d'éviter de prendre en compte les arbres d'une autre ligne dans le cas ou l'arbre de départ serait très isolé sur sa propre ligne. La fonction "calcNearestNeighbor" utilise le script SPARQL visible en figure 9.45. On obtient donc des couples d'arbres reliés ensemble par la relation sémantique nouvellement créée : *hasCloseNeighbor*.

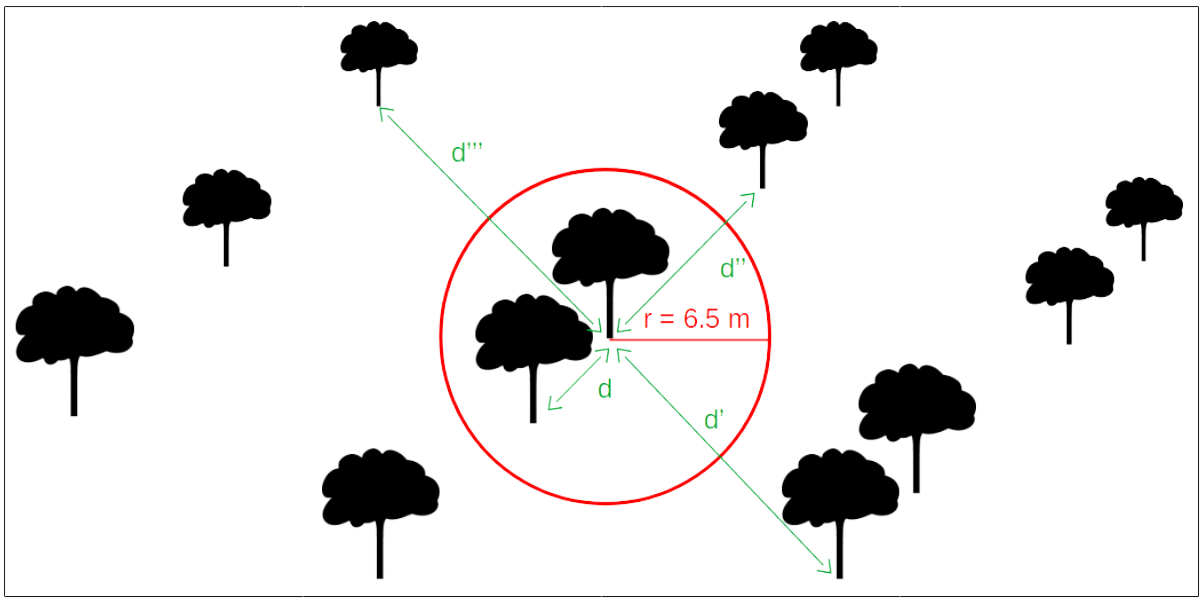


FIGURE 9.43 – *Identification du plus proche voisin d'un arbre dans un rayon de 6,5 mètres.*

```
[TheClosestNeighbor:
      (?plot res:isAFPlot 'true'^^xsd:Boolean)
      (?tree bfo:partOf ?plot)
      (?tree dcterms:type :term_Tree)
      calcNearestNeighbor(?tree, ?plot, "6.5", ?firstNeighbor)
      ->
      (?tree res:hasCloseNeighbor ?firstNeighbor)
      (?firstNeighbor res:hasCloseNeighbor ?tree)
]
```

FIGURE 9.44 – *Règle Jena pour déterminer au sein d'une parcelle le plus proche voisin d'un arbre dans une limite de 6,5 mètres.*

```
SELECT ?treeN ?d
WHERE {
      ?treeGeom geo:hasSerialization ?treeGeomLit .
      ?treeN bfo:partOf <node2.getURI()> .
      ?treeN dcterms:type :term_Tree .
      ?treeN geo:hasDefaultGeometry ?treeNGeom .
      ?treeNGeom geo:hasSerialization ?treeNGeomLit .
      FILTER (<node1.getURI()> != ?treeN)
      BIND(geof:distance(?treeGeomLit, ?treeNGeomLit, uom:metre) AS ?d)
      FILTER (?d < strDistanceMax)
} ORDER BY ASC(?d) LIMIT 1 ;
```

FIGURE 9.45 – *Script SPARQL de la fonction utilisateur "calcNearestNeighbor" pour renvoyer l'arbre le plus proche.*

– **Recherche d'arbres dans le prolongement de la ligne formée par un couple d'arbres voisins**
Pour chaque couple d'arbres voisins (identifiés par la relation sémantique *hasCloseNeighbor* qui les relie créée précédemment), on recherche alors un arbre dans le prolongement de la ligne formée par ce couple sur une distance maximale de 30 mètres et un écart angulaire maximal de 3 degrés (voir figure 9.43). La règle "TheNeighborOfNeighbors" visible en figure 9.47 utilise la fonction "calcNeighborOfNeighbors" dont le script SPARQL est montré en figure 9.48. Là encore l'arbre identifié sera relié à un arbre du couple par la relation *hasCloseNeighbor* créant ainsi un nouveau couple d'arbres pouvant être traité par la règle.

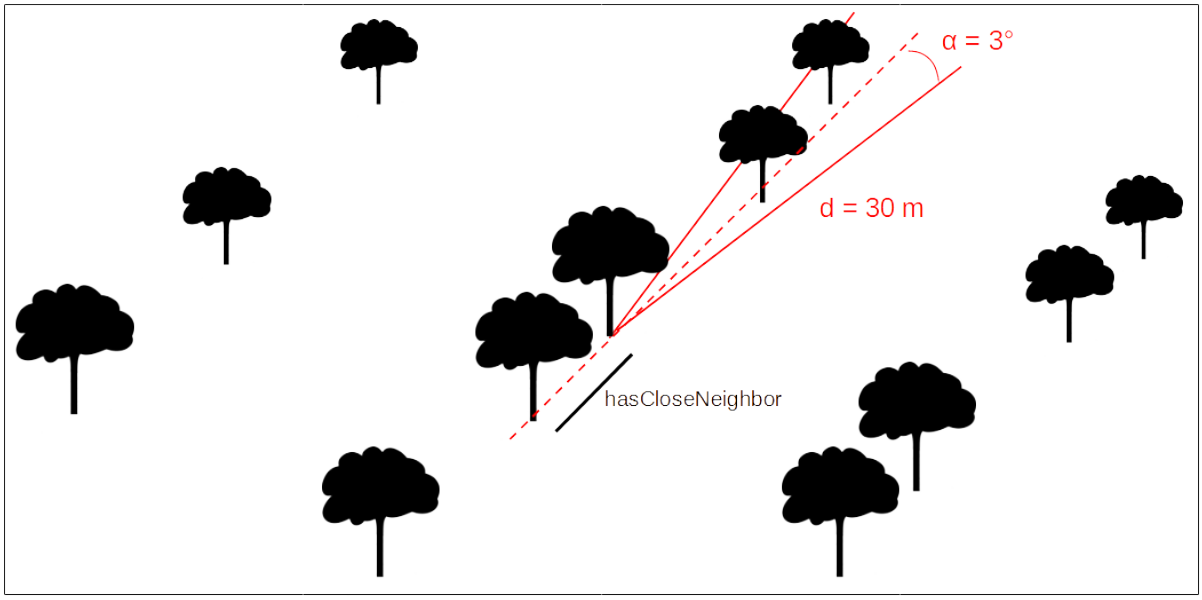


FIGURE 9.46 – *Détermination de la présence d'autres arbres sur un axe créé à partir de deux arbres liés par la relation sémantique* hasCloseNeighbor. *La recherche est limitée sur une distance de 30 mètres avec une tolérance angulaire de 3 degrés*

```
[TheNeighborOfNeighbors:
    (?plot res:isAFPlot 'true'^^xsd:Boolean)
    (?tree bfo:partOf ?plot)
    (?treeX res:hasCloseNeighbor ?treeY)
    (?treeY res:hasCloseNeighbor ?treeX)
    calcNeighborOfNeighbors(?treeX, ?treeY, ?plot, "30.0", "3.0", ?firstNeighbor)
    ->
    (?treeY res:hasCloseNeighbor ?firstNeighbor)
    (?firstNeighbor res:hasCloseNeighbor ?treeY)
]
```

FIGURE 9.47 – *Règle Jena pour déterminer au sein d'une parcelle le plus proche voisin sur la même ligne d'un couple d'arbres voisins avec une distance maximale de 30 mètres et un écart angulaire de 3 degrés.*

```
SELECT ?treeN ?d
WHERE {
  <node1.getURI()> geo:hasDefaultGeometry/geo:hasSerialization ?tree1GeomLit .
  BIND(xsd:decimal(replace( str(?tree1GeomLit),"^[^0-9.-]*([-]?[0-9.]+) .*$","$1")) as ?long1)
  BIND(xsd:decimal( replace( str(?tree1GeomLit),"^.* ([-]?[0-9.]+)[ ^0-9.]*$","$1")) as ?lat1)
  <node2.getURI()> geo:hasDefaultGeometry/geo:hasSerialization ?tree2GeomLit .
  BIND(xsd:decimal(replace( str(?tree2GeomLit),"^[ ^0-9.-]*([-]?[0-9.]+) .*$","$1")) as ?long2)
  BIND(xsd:decimal( replace( str(?tree2GeomLit),"^.* ([-]?[0-9.]+)[ ^0-9.]*$" ,"$1")) as ?lat2)
  BIND(spatialF:azimuthDeg(?lat1, ?long1, ?lat2, ?long2) AS ?azi)
  ?treeN bfo:partOf <node3.getURI()> .
  ?treeN dcterms:type :term_Tree .
  FILTER NOT EXISTS {?treeN res:hasCloseNeighbor* <node1.getURI()>}
  FILTER NOT EXISTS {?treeN res:hasCloseNeighbor* <node2.getURI()>}
  ?treeN geo:hasDefaultGeometry/geo:hasSerialization ?treeNGeomLit .
  BIND(xsd:decimal(replace(str(?treeNGeomLit),"^[^0-9.-]*([-]?[0-9.]+) .*$","$1")) as ?longN)
  BIND(xsd:decimal( replace( str(?treeNGeomLit),"^.* ([-]?[0-9.]+)[ ^0-9.]*$","$1")) as ?latN)
  BIND(spatialF:azimuthDeg(?lat1, ?long1, ?latN, ?longN) AS ?aziN)
  BIND(geof:distance(?tree2GeomLit, ?treeNGeomLit, uom:metre) AS ?d)
  BIND(xsd:decimal(ABS(?aziN - ?azi)) AS ?diffAzi)
  FILTER (?d <= strDistanceMax )
  FILTER ( ?diffAzi <= strAngleMax )
} ORDER BY ASC(?diffAzi) ASC(?d) LIMIT 1 ;
```

FIGURE 9.48 – *Script SPARQL de la fonction utilisateur "calcNeighborOfNeighbors" pour retourner l'arbre le plus proche d'un couple d'arbres selon l'axe sur lequel ce couple se place.*

- **Affectation d'arbres voisins à une ligne**
  La règle "InLine" présentée en figure 9.49 sera en charge d'affecter une ligne à chaque groupe d'arbres reliés par la relation *hasCloseNeighbor*

```
[InLine:
    (?plot res:isAFPlot 'true'^^xsd:Boolean)
    (?tree bfo:partOf ?plot)
    (?tree dcterms:type :term_Tree)
    calcIsInLine(?plot ?tree ?line)
    ->
    (?tree res:isInLine ?line)
    (?line bfo:partOf ?plot)
    (?line dcterms:type :term_Tree_Line)
]
```

FIGURE 9.49 – *Règle Jena pour la création de lignes avec des groupes d'arbres voisins.*

- **Création d'une géométrie pour les lignes**
  A partir des coordonnées des différents arbres d'une ligne, on met en place une géométrie pour chaque ligne grâce à la règle "GeomLine" (voir figure 9.50).

```
[GeomLine:
    (?plot res:isAFPlot 'true'^^xsd:Boolean)
    (?line bfo:partOf ?plot)
    (?line dcterms:type :term_Tree_Line)
    calcGeomLine(?plot ?line ?geom)
    ->
    (?line geo:hasDefaultGeometry ?geom )
]
```

FIGURE 9.50 – *Règle Jena pour affecter une géométrie à toutes les lignes d'une parcelle.*

– **Calcule de la distance entre les lignes**
La géométrie des lignes permet de mesurer la distance qui les sépare, permettant ainsi au besoin la requalification des parcelles concernées en fonction de la distance entre les lignes d'arbres (aménagement intensif ou extensif).

– **Calcul de l'azimut des lignes d'arbres**
Pour chaque ligne d'arbres, l'azimut est calculé en degrés grâce à la règle "Azimut-Line" (voir figure 9.51).

```
[RuleAzimutLine:
    (?plot res:isAFPlot 'true'^^xsd:Boolean)
    (?line bfo:partOf ?plot)
    (?line dcterms:type :term_Tree_Line)
    calcAzimutLine(?line ?azi)
    ->
    (?line res:hasAzimut ?azi)
]
```

FIGURE 9.51 – *Règle Jena pour affecter une valeur d'azimut à toutes les lignes d'une parcelle via la relation* res:hasAzimut.

– **Création de critères d'orientations cardinales**
Afin de passer d'une valeur d'azimut à une direction cardinale on met en place des directions cardinales (Nord Sud, Nord Nord Est, Nord Est, . . .) liées par les relations *hasMinAziValue* et *hasMaxAziValue* aux valeurs minimun et maximun d'azimut co-respondants (voir figure 9.52).

```
[CreateOrientation:
    ->
    (:north dcterms:type :term_cardinal_direction)
    (:north skos:prefLabel 'Nord Sud'^^@fr)
    (:north skos:prefLabel 'north south'^^@en)
    (:north :hasMinAziValue 0.0)
    (:north :hasMaxAziValue 11.25)
    (:north_north_east dcterms:type :term_cardinal_direction)
    (:north_north_east skos:prefLabel 'Nord Nord-Est'^^@fr)
    (:north_north_east skos:prefLabel 'north north-east'^^@en)
    (:north_north_east :hasMinAziValue 11.25)
    (:north_north_east :hasMaxAziValue 33.75)
    ...
    ...
]
```

FIGURE 9.52 – *Règle Jena pour la création d'orientations cardinales.*

– **Identification des directions cardinales des lignes d'une parcelle**
On identifiera ainsi la direction cardinale de chaque ligne (voir règle de la figure 9.53).

```
[IdentificationCardinalDirection:
    (?plot res:isAFPlot 'true'^^xsd:Boolean)
    (?line bfo:partOf ?plot)
    (?line dcterms:type :term_Tree_Line)
    (?line res:hasAzimut ?azi)
    (?cardinalDirection dcterms:type res:term_cardinal_direction)
    (?cardinalDirection res:hasMinAziValue ?AziCriterionMin)
    (?cardinalDirection res:hasMaxAziValue ?AziCriterionMax)
    greaterThan(?azi ?AziCriterionMin)
    le(?azi ?AziCriterionMax)
    ->
    (?line res:hasCardinalDirection ?cardinalDirection)
]
```

FIGURE 9.53 – *Règle Jena pour affecter à toute les lignes d'une parcelle possédant une valeur d'azimut, un direction cardinale via la relation* res:hasCardinalDirection.

– **Identification de la direction cardinale principale des lignes d'une parcelle**
On déduit la direction cardinale principale des lignes d'une parcelle en sélectionnant la direction cardinale la plus représentée aux niveaux des lignes de la parcelle. On pourra ainsi sélectionner différentes parcelles d'agrisylviculture en fonction de l'orientation générale de leurs lignes d'arbres, critère des plus important si l'on mène des études sur la compétition entre arbres et cultures pour la lumière.

Requalification d'un élément

Dans le paragraphe précédent nous avons identifié la présence de lignes d'arbres au sein d'une parcelle. Cette nouvelle information peut permettre de requalifier la parcelle

concernée comme par exemple avec la règle suivante visible en figure 9.54 qui stipule que toute parcelle contenant des lignes d'arbres et ayant une direction principale pour ces lignes sera clarifiée par le concept "caf_4" de notre ontologie terminologique AgroforestryVoc dont le libellé est "Disposition en alignement régulier discontinu".

```
[PlotIsRegularDiscontinuousLayout:
    (?line bfo:partOf ?plot)
    (?line dcterms:type :term_Tree_Line)
    (?plot res:hasMainDirectionOfLines ?MainDirectionOfLines)
    ->
    (?plot dcterms:type afv:caf_4)
]
```

FIGURE 9.54 – *Règle Jena pour la requalification des parcelles.*

Autres possibilités de traitements

D'autres traitements peuvent être réalisés par l'application AOBRA pour faciliter le travail des utilisateurs, comme par exemple la constitution de listes RDF rassemblant les concepts d'ontologies terminologiques à partir d'un terme et de critères de sélection (concepts plus étroits ou généralisant le concept dont le terme est fourni). Ces listes peuvent être utilisées par la suite dans l'écriture des règles Jena ou des requêtes SPARQL.

9.1.9 Études réalisées avec AOBRA et leurs résultats

Étude statistique des mesures

Une étude statistique est conduite à grande échelle et à partir de l'ensemble des mesures collectées sur les deux parcelles PA3AF (la parcelle agroforestière) et PA4TF (son témoin forestier). Le mode opératoire présenté en section 9.1.8 est retenu pour produire les différents vecteurs de valeurs fournis en entrée de fonctions statistiques d'intérêt empruntées à divers logiciels à l'exemple de R. L'objectif est de produire une analyse statistique exhaustive (moyenne, médiane, écart type, variance) de l'ensemble des valeurs observées pour chaque propriété d'élément par année et par parcelle. Ceci permet de vérifier la validité des données et également de contrôler que les données issues des observations de terrain sont exploitables et comparables entre elles. L'ensemble des résultats statistiques sont obtenus par le biais de requêtes de consultation SPARQL dont l'écriture est facilitée par la normalisation de l'identification des relations crées par le mécanisme d'application des règles comme le montrait la figure 9.40 vue précédemment.
Nous obtenons ainsi, différentes moyennes de mesures comme la hauteur moyenne des arbres de la parcelle agroforestière A3AF pour les années 2015, 2016 et 2017 ou le diamètre moyen des arbres de la parcelle A4TF pour ces mêmes années.
Les figures suivantes exposent une partie de ces résultats :
La figure 9.55 montre les moyennes des diamètres et des hauteurs des arbres de la parcelle agroforestière A3 et de son témoin forestier A4 pour l'année 2015. Il en va de même pour l'année 2016, et l'année 2017. On note d'emblée, une croissance des arbres en moyenne plus élevée en diamètre et en taille, en faveur de la parcelle agroforestière vis à vis de son témoin forestier.

Statistiques 2015 diamètres des troncs	
Parcelle	Moyenne des diamètres des troncs
<http://www.afy.fr/Restinclieres/PA3AF>	28.2
<http://www.afy.fr/Restinclieres/PA4TF>	18.6
Statistiques 2015 hauteurs des troncs	
Parcelle	Moyenne des hauteurs des troncs
<http://www.afy.fr/Restinclieres/PA3AF>	423.6
<http://www.afy.fr/Restinclieres/PA4TF>	342.6

FIGURE 9.55 – *Résultat des moyennes des diamètres et hauteurs des troncs par parcelle pour l'année 2015.*

Cette croissance plus importante des arbres en condition d'agroforesterie est confirmée par l'étude des quartiles de ces mêmes mesures illustrées dans les figures 9.56, 9.57 et 9.58. En effet on observe via ces résultats que durant les trois années consécutives le premier quart de la population des arbres en condition d'agroforesterie a un diamètre maximal plus élevé (25.0 contre 14.4 en 2015, 25.6 contre 14.9 en 2016 et 26.4 contre 15 en 2017) et une hauteur d'arbre maximale elle aussi plus élevée (408.7 contre 297.5 en 2015, 410.0 contre 298.7 en 2016 et 420 contre 293.7 en 2017). Cette différence se retrouve lorsqu'on considère la moitié et les trois quarts de la population avec des valeurs jusqu'à 30% plus élevées en faveur des arbres agroforestiers pour le diamètre et la hauteur des troncs.

Statistiques 2015 diamètres des troncs			
Plot	Quartile 25	Quartile 50	Quartile 75
<http://www.afy.fr/Restinclieres/PA3AF>	25.0	29.0	31.7
<http://www.afy.fr/Restinclieres/PA4TF>	14.4	19.5	23.2
Statistiques 2015 hauteurs des troncs			
Parcelle	Quartile 25	Quartile 50	Quartile 75
<http://www.afy.fr/Restinclieres/PA3AF>	408.7	420.0	440.0
<http://www.afy.fr/Restinclieres/PA4TF>	297.5	385.0	416.2

FIGURE 9.56 – *Résultat des quartiles pour les diamètres et hauteurs des troncs par parcelle pour l'année 2015.*

Statistiques 2016 diamètres des troncs			
Parcelle	Quartile 25	Quartile 50	Quartile 75
<http://www.afy.fr/Restinclieres/PA3AF>	25.6	29.7	32.5
<http://www.afy.fr/Restinclieres/PA4TF>	14.9	19.9	23.9
Statistiques 2016 hauteurs des troncs			
Parcelle	Quartile 25	Quartile 50	Quartile 75
<http://www.afy.fr/Restinclieres/PA3AF>	410.0	420.0	440.0
<http://www.afy.fr/Restinclieres/PA4TF>	298.7	380.0	420.0

FIGURE 9.57 – *Résultat des quartiles pour les diamètres et hauteurs des troncs par parcelle pour l'année 2016.*

Statistiques 2017 diamètres des troncs			
Parcelle	Quartile 25	Quartile 50	Quartile 75
<http://www.afy.fr/Restinclieres/PA3AF>	26.4	30.4	33.3
<http://www.afy.fr/Restinclieres/PA4TF>	15.0	20.1	24.4
Statistiques 2017 hauteurs des troncs			
Parcelle	Quartile 25	Quartile 50	Quartile 75
<http://www.afy.fr/Restinclieres/PA3AF>	420.0	435.0	450.0
<http://www.afy.fr/Restinclieres/PA4TF>	293.7	382.5	420.0

FIGURE 9.58 – *Résultat des quartiles pour les diamètres et hauteurs des troncs par parcelle pour l'année 2017.*

Ce résultat est conforme à la littérature qui relate le fait qu'un arbre isolé subit moins de concurrence vis à vis de ses voisins que dans des conditions de culture classique en foresterie et que dans des conditions d'agroforesterie la concurrence pour la ressource en eau avec la culture implique que les arbres développent un système racinaire plus profond, ceux qui leur donnent un avantage en été lorsque la disponibilité de l'eau vient à manquer.
Pour étudier l'avantage que cette augmentation du rythme de croissance pourrait procurer aux exploitants d'une parcelle agroforestière, nous avons étudié par la suite la répercussion de ce phénomène sur la biomasse par catégories d'arbres

Étude de la biomasse par catégorie d'arbres

Pour comparer le taux de profitabilité d'une parcelle agroforestière à celui de son témoin forestier, nous avons procédé en plusieurs étapes en accord avec le diagramme d'utilisation présenté figure 9.8 à la section 9.1.3 :

- **Identification des arbres d'une parcelle**
  Là encore, la relation *BFO_0000051* qui lie la parcelle à ses constituants est mobilisée. Pour affiner cette identification et s'assurer que le constituant d'une parcelle est bien un arbre nous pouvons nous référer à son appartenance à la classe **TreeForStudy** comme vu à la section 9.1.7 avec la figure 9.28.

- **Identification des arbres d'une parcelle éligibles au commerce du bois**
  Comme vu à la section 9.1.8 nous avons différencié les arbres au sein des parcelles en plusieurs catégories qui correspondent à celles utilisées pour le négoce du bois. Ces catégories sont axées sur le diamètre du tronc (Catégories I : 20-25 cm, Catégories II : 26-35 cm, Catégories III : 36-45 cm, . . .). Les arbres que nous allons traiter ici sont donc liés a une catégorie par la relation *hasCategory*. .

- **Calcul de la biomasse des arbres**
  Selon le mode opératoire vu à la section 9.1.8 la biomasse des arbres est calculée et placée sous forme d'observations SOSA (voir figure 9.59) pour tous les arbres de la parcelle quelle que soit leur catégorie.

- **Calcul du ratio la biomasse des arbres négociables sur la biomasse totale des arbres de la parcelle par catégorie**
  Pour étudier la biomasse des arbres d'une parcelle par catégorie, il nous suffit de déclarer la propriété *hasBiomass* comme étant une propriété à étudier par catégorie (voir figure 9.59), ce qui déclenche son traitement par la règle "PropertyValueByCategory" comme nous l'avons vu à la section 9.1.8. On aboutit ainsi à la création de propriétés comme *hasBiomassByCategoryIISum2015* liant la parcelle à la somme

des biomasses de chaque arbre de catégorie II en 2015. Les valeurs de ces nouvelles propriétés d'agrégation seront utilisées pour le calcul des différents ratios (voir figure 9.60).

```
[MakeTheHasBiomassPropertyAPropertyToStudyByCategory:
    (:thisOntology :hasNameSpace ?ns)
    uriConcat(?ns, "hasBiomass",?uRIProperty)
    ->
    (?uRIProperty :hasLocalName "hasBiomass")
    (?uRIProperty :isPropertyToStudyByCategory 'true'^^xsd:Boolean)
]
```

FIGURE 9.59 – *Règle Jena pour le calcul de la biomasse par catégorie d'arbres.*

```
[CalculationBiomassForWoodTradeRatio:
    (?plot :isPlot 'true'^^xsd:Boolean)
    (?yearForStudy rdf:type :YearForStudy)
    (?yearForStudy :hasYear ?yearStudied)
    (?category dcterms:type :term_Category)
    (?category :hasLocalName ?categoryName)
    ->
    [
    (?plot ?uRIPropertyRatio ?ratio)
    <-
    (:thisOntology :hasNameSpace ?ns)
    uriConcat(?ns, "hasBiomass","Sum",?yearStudied,?uRIPropertyTotalBiomass)
    (?plot ?uRIPropertyTotalBiomass ?totalBiomass)
    uriConcat(?ns, "hasBiomass","By",?categoryName,"Sum",?yearStudied,
     ?uRIPropertyCategoryTotalBiomass)
    (?plot ?uRIPropertyCategoryTotalBiomass ?categoryTotalBiomass)
    uriConcat(?ns, "hasBiomassBy", ?categoryName, "Ratio", ?yearStudied,
      ?uRIPropertyRatio)
    calcRatio(?categoryTotalBiomass, ?totalBiomass, ?ratio)
    ]
]
```

FIGURE 9.60 – *Règle Jena pour le calcul des ratios entre la biomasse des arbres d'une catégorie particulière et la biomasse totale des arbres d'une parcelle.*

Les résultats obtenus sont présentés sur les trois figures suivantes (voir figure 9.61, 9.62 et 9.63) et confirme le fait que les arbres se développent plus rapidement dans des conditions de culture agroforestières. Les résultats montrent en effet que le ratio de la biomasse des arbres de catégorie II est nettement plus élevé sur la parcelle agroforestière PA3AF que sur son témoin forestier PA4TF allant de plus du double en 2015 (76% contre 34%) à 30% d'écart en 2017. De plus on constate l'absence d'arbres de catégorie III pour la parcelle forestière alors que le ratio de biomasse de cette catégorie augmente de 9 à 16% durant les trois années étudiées pour la parcelle agroforestière.

Ces résultats sont des indicateurs économiques importants pour les agroforestiers car ils signifient que la production de biomasse individuelle sera plus importante, et que le produit final, l'arbre commercialisable sera obtenu plus rapidement.

Ratio de la biomasse des arbres de catégorie I/ biomasse totale			
Parcelle	Ratio 2015 (%)	Ratio 2016 (%)	Ratio 2017 (%)
<http://www.afy.fr/Restinclieres/PA3AF>	11.9	9.7	9.4
<http://www.afy.fr/Restinclieres/PA4TF>	40.1	36.9	36.7

FIGURE 9.61 – *Ratio en pourcentage de la biomasse des arbres de catégorie I sur la biomasse totale des arbres par parcelle et années étudiées.*

Ratio de la biomasse des arbres de catégorie II/ biomasse totale			
Parcelle	Ratio 2015 (%)	Ratio 2016 (%)	Ratio 2017 (%)
<http://www.afy.fr/Restinclieres/PA3AF>	75.9	74.6	71.7
<http://www.afy.fr/Restinclieres/PA4TF>	33.9	39.8	41.5

FIGURE 9.62 – *Ratio en pourcentage de la biomasse des arbres de catégorie II sur la biomasse totale des arbres par parcelle et années étudiées.*

Ratio de la biomasse des arbres de catégorie III/ biomasse totale			
Parcelle	Ratio 2015 (%)	Ratio 2016 (%)	Ratio 2017 (%)
<http://www.afy.fr/Restinclieres/PA3AF>	9.9	13.3	16.7

FIGURE 9.63 – *Ratio en pourcentage de la biomasse des arbres de catégorie III sur la biomasse totale des arbres par parcelle et années étudiées.*

Étude de l'influence de la proximité d'un cours d'eau sur la croissance des arbres

Afin de s'assurer des raisons d'un tel écart de croissance des arbres entre la parcelle agroforestière et la parcelle de contrôle forestier, alors même que les conditions météorologiques et pédologiques étaient identiques, il nous fallait étudier l'influence possible du fleuve Le Lez, qui jouxte les deux parcelles. Nous avons alors étudié la distance moyenne des arbres à ce cours d'eau par parcelle, année et catégorie d'arbres.
Le tableau suivant (voir figure 9.64) montre les résultats obtenus pour l'année 2015. Pour la parcelle de contrôle forestier, on note que la distance moyenne au Lez est légèrement inférieure pour la catégories d'arbres la plus élevée ($37m$ pour la catégorie II contre $45m$ pour la catégorie I). Au contraire, pour la parcelle agroforestière, le rapport est inversé, les arbres de la catégorie la plus élevée sont en moyenne les plus éloignés du Lez ($202m$ pour la catégorie III contre environ $172m$ en moyenne pour les catégories II et I). De plus on observe le fait que toutes catégorie confondues les arbres du contrôle forestier sont en moyenne plus proches du Lez que leurs homologues agroforestiers ($41m$ contre $181m$). Nous avons donc pu rejeter l'hypothèse selon laquelle la proximité de fleuve le Lez aurait un effet sur l'avantage en terme de croissance des arbres de la parcelle agroforestière A3AF.

Moyennes des distances au Lez par catégories en 2015

Parcelle	Catégorie I (m)	Catégorie II (m)	Catégorie III (m)
<http://www.afy.fr/Restinclieres/PA3AF>	170.6	172.8	201.6
<http://www.afy.fr/Restinclieres/PA4TF>	44.7	36.9	

FIGURE 9.64 – *Moyennes des distances en mètres des arbres au Lez pour les catégories I,II et III en 2015.*

Étude du rendement des cultures

Comme vu à la section 9.1.5 avec le diagramme RDF de la figure 9.17 l'utilisation du modèle SOSA nous permet de décrire des échantillonnages. Nous avons donc intégré à notre base de connaissance les données de différentes caractéristiques d'échantillons de cultures, telles que celles vues en figure 9.5, pour les différentes cultures successives de pois(2016), de blé dur (2017) et d'orge (2018) de la parcelle agroforestière PA3AF et de son témoin forestier PA3TA.
Là encore, la mise en place d'un ensemble de règles permet à l'application AOBRA d'agréger les observations faites sur les différent échantillons pour obtenir des résultats au niveau des parcelles. On obtient les résultats suivants (voir figures 9.65, 9.66 et 9.67).

Rendements des cultures en 2016

Échantillonnage	Nom culture	Rendement (qx/ha)
<http://www.afy.fr/Restinclieres/PA3AF_Echantilonnage_pois_2016>	"pois"	21.0
<http://www.afy.fr/Restinclieres/PA3TA_Echantilonnage_pois_2016>	"pois"	23.7

FIGURE 9.65 – *Rendements de la culture de pois en 2016 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA.*

Rendements des cultures en 2017		
Échantillonnage	Nom culture	Rendement (qx/ha)
<http://www.afy.fr/Restinclieres/PA3AF_Echantilonnage_ble_2017>	"blé dur"	23.3
<http://www.afy.fr/Restinclieres/PA3TA_Echantilonnage_ble_2017>	"blé dur"	35.3

FIGURE 9.66 – *Rendements de la culture de blé dur en 2017 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA.*

Rendements des cultures en 2018		
Échantillonnage	Nom culture	Rendement (qx/ha)
<http://www.afy.fr/Restinclieres/PA3AF_Echantilonnage_orge_2018>	"orge"	26.3
<http://www.afy.fr/Restinclieres/PA3TA_Echantilonnage_orge_2018>	"orge"	42.9

FIGURE 9.67 – *Rendements de la culture d'orge en 2018 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA.*

Le lien entre une parcelle agroforestière, et son témoin agricole nous est donné par la relation *hasAgriculturalWitness* qui lie la parcelle PA3AF à son témoin agricole PA3TA. Cela nous permet de comparer les rendements agricoles dans les conditions de culture en agroforesterie et dans les conditions d'agriculture classique.
On obtient ainsi la performance de la parcelle agroforestière pour les différentes cultures de pois, de blé dur et d'orge (voir figures 9.68, 9.69 et 9.70). On note donc une diminution conséquente de la performance du rendement des cultures dans les conditions d'agroforesterie avec une baisse de 11 à 39%. L'impact semble plus important dans le cas de la culture de céréales (34% de baisse pour le blé dur et 39% pour l'orge) que pour une culture de légumineuse (11% de baisse pour le pois).

Performance du rendement de culture du pois en 2016	
Parcelle	Ratio (%)
<http://www.afy.fr/Restinclieres/PA3AF>	88.8

FIGURE 9.68 – *Rendements de la culture de pois en 2016 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA.*

Performance du rendement de culture de blé dur en 2017	
Parcelle	Ratio (%)
<http://www.afy.fr/Restinclieres/PA3AF>	66.0

FIGURE 9.69 – *Performance en pourcentage du rendement de culture de blé dur en 2017 de la parcelle agroforestière PA3AF par rapport à son témoin agricole.*

Performance du rendement de culture de l'orge en 2018	
Parcelle	Ratio (%)
<http://www.afy.fr/Restinclieres/PA3AF>	61.0

FIGURE 9.70 – *Performance en pourcentage du rendement de culture d'orge en 2018 de la parcelle agroforestière PA3AF par rapport à son témoin agricole.*

Détermination de la performance d'un aménagement agroforestier

Sur la base des expérimentations faites en agroforesterie nous définissons une "bonne" parcelle agroforestière comme un aménagement dans lequel les arbres ont une croissance plus rapide que dans le une exploitation forestière, et dans lequel la baisse de rendement des cultures sera la plus faible possible. La baisse de rendement des cultures est en grande partie due à l'ombrage apporté par les houppiers des arbres et la densité des espèces associées sera déterminante sur la compétition qui se nouera entre arbres et cultures. Pour mesurer cette compétition au niveau des parcelles agroforestières d'agrisylviculture avec lignes d'arbres, il est d'usage d'avoir recours aux indices d'encombrement [DUPRAZ et LIAGRE, 2008].
Ces indices sont définis ainsi :

Soit H la hauteur des arbres,
$H\acute{e}$ la hauteur d'élagage des arbres,
L la distance entre les lignes d'arbres,
l la distance entre les arbres sur une ligne et
d la densité d'arbre par hectare $d = 10000/L/l$

- **Indice d'ouverture des allées cultivées**

$$IE_1 = H/L$$

- **Indice de densité des arbres**

$$IE_2 = H^2/(L * l)$$

- **Indice de longueur cumulée des houppiers**

$$IE_3 = (H - H\acute{e}) * d/500$$

La figure 9.71 schématise différents choix concernant la distance entre les lignes d'arbres, pouvant être faits au moment de la plantation des arbres.
L'indice IE_1 traduit ici l'encombrement qui sera obtenu au moment ou les arbres auront atteint leur hauteur maximale d'exploitation. Cet indice IE_1 étant lié à la hauteur des arbres, variera de la plantation à l'abattage des arbres comme le montre la figure 9.72.
L'indice IE_2 prend de plus en compte la distance entre les arbres sur la ligne. L'agroforestier peut influer sur cet indice en décidant par exemple d'abattre un arbre sur deux sur chaque ligne de sa parcelle si la culture intercalaire venait à trop manquer d'accès à l'ensoleillement.

L'indice IE_3 intègre quant à lui la hauteur du houppier calculée en retranchant à la hauteur de l'arbre la hauteur du tronc élagué $(H - H\acute{e})$ responsable en grande partie de l'ombrage. Là encore cet indice peut être modifié par l'agroforestier en augmentant la hauteur d'élagage mais le respect de l'équilibre biodynamique de l'arbre impose cer-

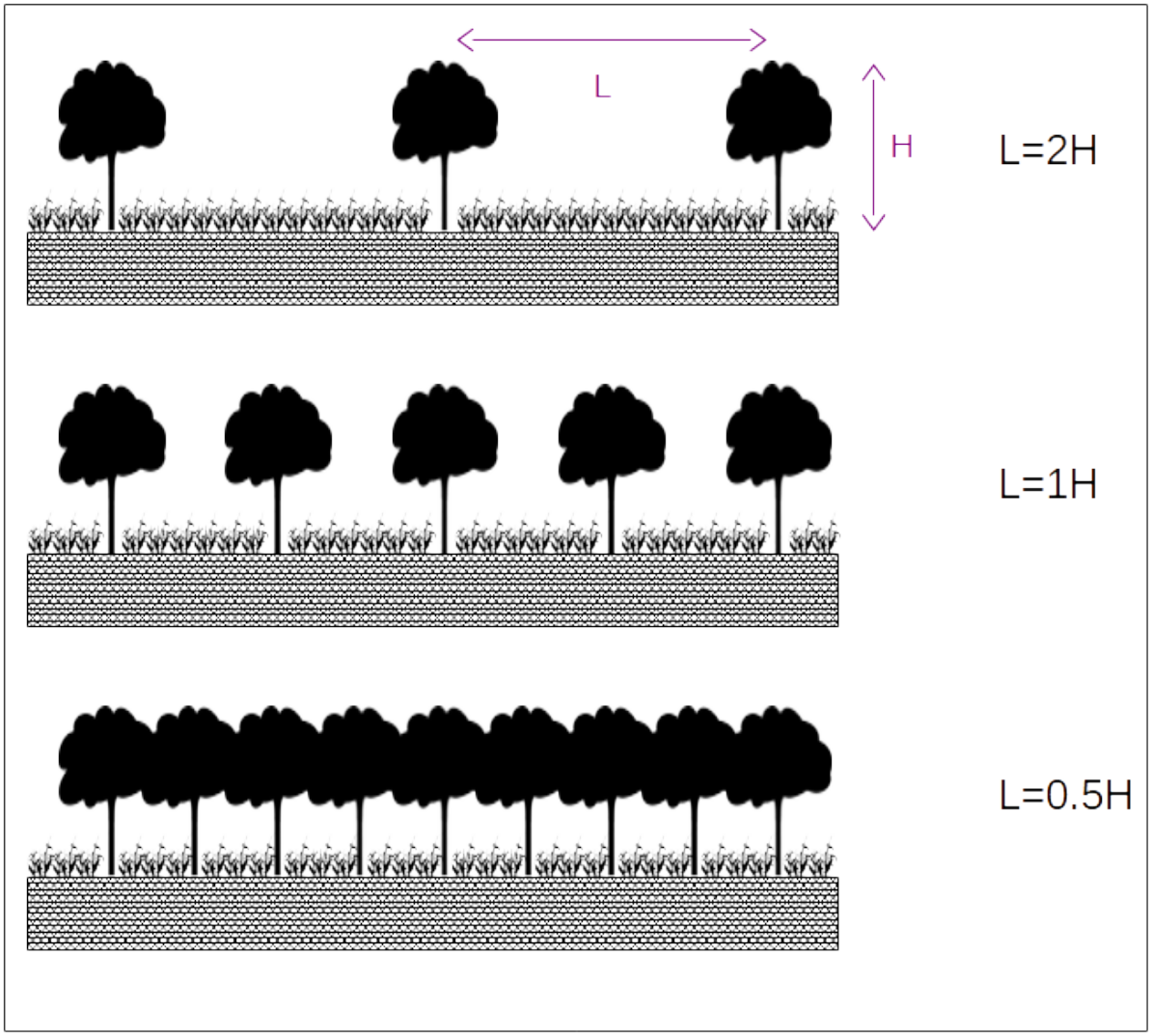


FIGURE 9.71 – *Encombrement sur différents schémas d'aménagement parcellaire estimé avec des arbres ayant atteint la hauteur finale nécessaire à leur exploitation.*

taines restrictions, ainsi que l'aspect pratique de l'opération (difficulté d'élaguer à plus de 4 mètres de hauteur).
Par expérience, les valeurs d'indices optimales pour un aménagement permettant de maintenir un rendement maximal des cultures tout au long du cycle de vie des arbres de la parcelle doivent être proches de 1.
La mise en place de règles adéquates permet à l'application AOBRA de calculer ces indices pour les différentes années étudiées comme le montre le tableau de la figure 9.73 pour les indices calculés de la parcelle agroforestière PA3AF pour l'année 2017. Les résultats indiquent pour les trois années étudiées des indices élevées de densité des arbres et de longueur cumulée des houppiers ce qui va dans le sens de la baisse de rendement constaté sur les cultures successives.

En fixant des critères de performance observés sur des expérimentations ayant déjà fait la preuve de leur rentabilité, nous pouvons par comparaison juger de l'efficacité d'un aménagement plus récent. Considérant par exemple, qu'une croissance des arbres supérieure de 20% et une perte de rendement de culture de 40% à la moitié de vie des arbres d'un aménagement agroforestier se compenseront financièrement sur un cycle complet de l'exploitation des arbres de l'aménagement.

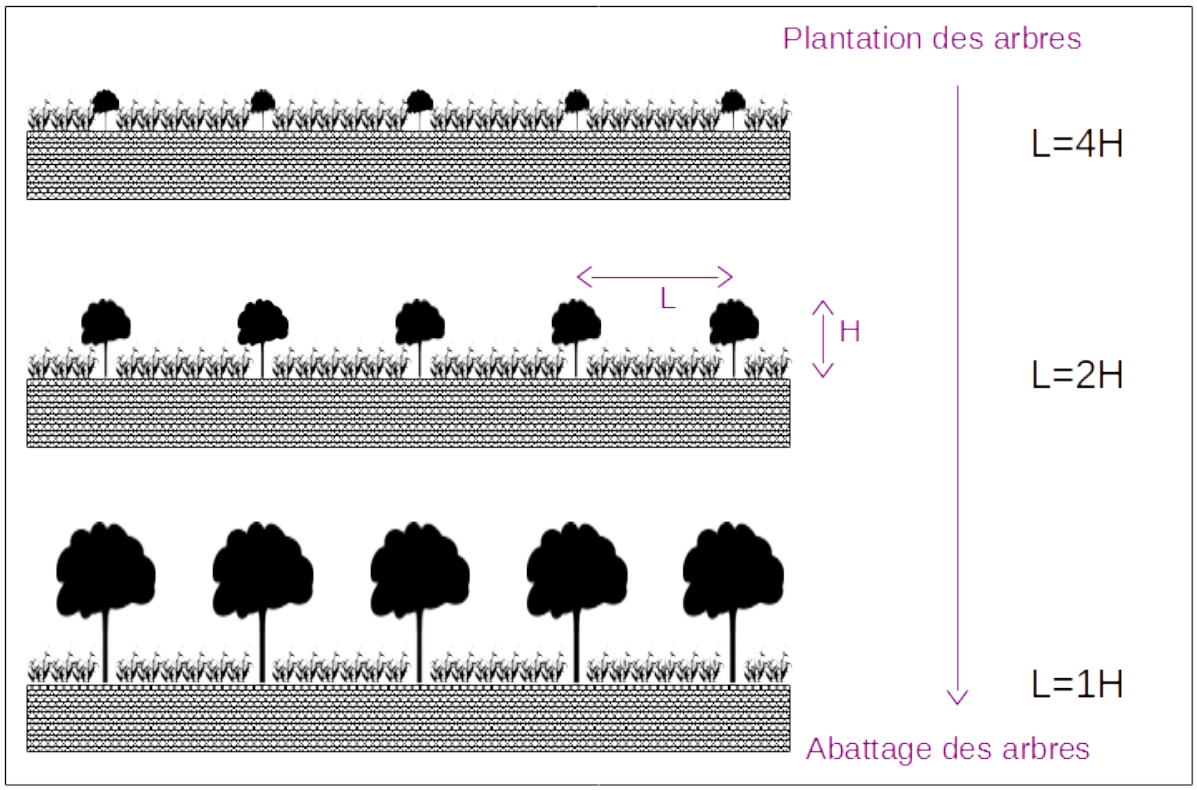


FIGURE 9.72 – *Encombrement sur une même parcelle estimé à différents niveaux de croissance des arbres*

Indices d'encombrement de la parcelle PA3AF en 2017			
Parcelle	IE_1	IE_2	IE_3
<http://www.afy.fr/Restinclieres/PA3AF>	1.09	2.66	2.77

FIGURE 9.73 – *Indices d'encombrement IE_1, IE_2 et IE_3 de la parcelle agroforestière PA3AF en 2017.*

En mettant en place des échelles de notation pour les rendements relatifs de culture et de production de bois (voir figures 9.74 et 9.75), on peut obtenir un indicateur pour l'efficience de l'aménagement d'une parcelle. L'échelle de notation utilisée est composée de six niveaux d'appréciations (mauvais, très moyen, moyen, bon, très bon et excellent) noté de 0 à 5.
Pour le rendement des cultures la valeur prise en compte est le pourcentage relatif R calculé en comparant le rendement de production des cultures en quintaux par hectare d'une parcelle agroforestière vis à vis de son témoin agricole.
Pour mesurer la performance de la croissance des arbres nous avons choisi la différence DR des ratios en pourcentage de la biomasse des arbres de la catégorie comparable la plus élevée par rapport à la biomasse totale des parcelles pour la parcelle agroforestière et son témoin forestier. Le choix de cette valeur DR repose sur le fait quelle ne prend pas en compte la superficie dédiée aux plantations d'arbres, donnée à laquelle nous n'avons pas accès de façon précise dans notre jeu de données.

Échelle de performance pour le rendement des cultures						
Rendement des cultures	R<=40%	40<R<=50	50<R<=60	60<R<=70	70<R<=80	80<R
Échelle	mauvais (0)	très moyen (1)	moyen (2)	bon (3)	très bon (4)	excellent (5)

FIGURE 9.74 – *Établissement d'une échelle de performance pour les cultures.*

Échelle de performance pour la croissance des arbres						
Différence de ratio	DR<=0%	0<DR<=10	10<DR<=20	20<DR<=30	30<DR<=40	40<DR
Échelle	mauvais (0)	très moyen (1)	moyen (2)	bon (3)	très bon (4)	excellent (5)

FIGURE 9.75 – *Établissement d'une échelle de performance pour la croissance des arbres. La mesure prise en compte est la différence des ratios de la biomasse des arbres de la catégorie la plus élevée par rapport à la biomasse totale pour une parcelle agroforestière et son témoin forestier*

On obtient ainsi un score annuel pour le rendement des cultures (voir figure 9.76) et la croissance des arbres (voir figure 9.77). A partir de ces scores, il est possible de calculer un score final pour apprécier l'efficience globale de l'aménagement étudié. La figure 9.78 montre en exemple que notre parcelle PA3AF a un score final de 3.5 (entre bon et très bon) en 2016 qui est la moyenne entre le score de la performance des arbres en 2016 (4) et le score de la culture récoltée en 2017 (3).

Score de la performance de la culture en 2017	
Parcelle	Score
<http://www.afy.fr/Restinclieres/PA3AF>	3

FIGURE 9.76 – *Score de la performance de la culture de blé dur en 2017 de la parcelle agroforestière PA3AF par rapport à son témoin agricole.*

Score de la performance pour la croissance des arbres en 2016	
Parcelle	Score
<http://www.afy.fr/Restinclieres/PA3AF>	4

FIGURE 9.77 – *Score de la performance de la croissance des arbres de catégorie II en 2016 de la parcelle agroforestière PA3AF par rapport à son témoin forestier.*

Score final en 2016	
Parcelle	Score
<http://www.afy.fr/Restinclieres/PA3AF>	3.5

FIGURE 9.78 – *Score final de la performance en 2016 de la parcelle agroforestière PA3AF par rapport à son témoin agricole.*

9.2 En ce qui concerne la validation et la confrontation de notre modèle avec la réalité

Par comparaison avec les travaux habituellement retrouvés dans le périmètre de de la mise en œuvre de bases de connaissance, notre démarche peut sembler inhabituelle. En effet, nous avons pris le parti de concentrer notre effort sur la définition et l'exécution de règles métier, statistiques pour leur grande majorité, et qui exploitent des données quantitatives. L'usage qui est fait des bases de connaissance est d'ordinaire plutôt symbolique, et cherche à dégager de nouveaux savoirs sur la base de données qualitatives. La raison est principalement à rechercher du côté de la nature du cas d'étude exploré qui comporte quasi exclusivement des données quantitatives et qui fait traditionnellement l'objet d'analyses statistiques.
Nous avons exploité les deux moteurs de raisonnement de Jena désignés sous le vocable "generic rules reasoner" qui permettent de mettre en œuvre des mécanismes de chaïnage avant, de chaînage arrière et des stratégies hybrides. Nous avons à ce sujet défini des règles tirant parti du chaînage avant et qui vont permettre de matérialiser tous les triplets qui peuvent en être déduits. Ces règles font un large usage de diverses fonctions utilisateurs construites à partir de fonctions R, et des fonctionnalités offertes par Jena pour définir de nouvelles fonctions utilisateurs visant à étendre les capacités du moteur de requêtage ARQ.
Nous avons aussi défini des règles hybrides, qui permettent de construire les vecteurs qui contiennent les valeurs agrégées obtenues à partir de l'union du graphe explicite et du graphe inféré, au moment où l'on applique la règle. Au niveau de l'architecture matérielle, nous avons utilisé une machine dotée d'un processeur intel(R) core(TM) i7-7920HQ cadencé à 3.10 Ghz, et de 32,0 Go de RAM. L'application AOBRA écrite en langage JAVA est outillée par un serveur web Tomcat v8.5. La base de connnaissance est rendue persistante au travers d'un triplestore TDB [6], autre fonctionnalité de la librairie Jena.
A titre indicatif, le chargement des modèles en mémoire (transfert et union des modèles de la base de données TDB en mémoire vive) est de l'ordre de moins de 3 secondes. Par la suite, la mise en place des moteurs d'inférence n'ont consommé que 0,3 seconde. A l'exécution les différents traitements, vus en section 9.1.9, comprenant les statistiques sur les relevés dendrométriques, l'étude des rendements agricoles, l'exploitation des données spatiales pour les distances séparant les arbres d'un cours d'eau, ...se sont déroulés en moins de 38 secondes. Le graphe explicite de départ contient 85 590 triplets, et l'ajout du graphe inféré conduit à un graphe total de 126 585 triplets, soit la création de plus de 40 000 triplets (47%) supplémentaires.

6. TDB : High performance RDF store. https://jena.apache.org/documentation/tdb/.

Par comparaison avec les travaux retrouvés dans la littérature, ayant déjà exploité le contenu informatif des données, l'application est à même de produire les mêmes conclusions. Un avantage indéniable est que l'ensemble de ces données est structuré, enrichi sémantiquement et peut être facilement partagé et réutilisé.

Pour poursuivre ce travail d'évaluation, il nous faudrait maintenant nous orienter vers un cas d'étude plus axé sur des données qualitatives.

X

Conclusion et perspectives

10.1 Difficultés rencontrées 215
10.1.1 le paradigme du monde ouvert a un prix 215
10.1.2 Mais où sont les données de l'agroforesterie ? 216
10.2 Perspectives 218
10.2.1 Rapprochement des domaines 218
10.2.2 Unification des sources de données complexes voire contradictoires 218
10.2.3 Développement d'une application pour la gestion des données liées 219
10.2.4 Développement d'une plateforme modulaire pour la gestion des données en agroforesterie 220

Ce dernier chapitre fait à la fois une synthèse du travail réalisé, revient sur les difficultés rencontrées, et expose quelques perspectives d'avenir dans le contexte de la gestion des données et connaissances en agroforesterie.
Nous sommes partis du constat, que de plus en plus de données sur des expérimentations agroforestières étaient actuellement produites ou devraient être prochainement produites, et que notre méconnaissance de ces données ne devait pas les rendre moins exploitables pour autant, pour montrer la pertinence d'aménagements réalisés ou futurs.
Les principales difficultés retenues à l'encontre d'une gestion efficace de ces données ont été en effet identifiées comme provenant de la pluralité des données elles-mêmes, en conséquence relevant du caractère systémique de l'agroforesterie et des multiples supports utilisés pour structurer, pérenniser, et valoriser les données produites.
Nous avons donc cherché à proposer une nouvelle démarche de capitalisation des données en agroforesterie à l'intention des agroforestiers et de leurs partenaires.
Cette démarche est basée sur l'utilisation des ontologies informatiques, et des langages et outils du web sémantique. Des approches similaires ont déjà été expérimentées avec succès dans d'autre domaines, en particulier dans les domaines de la génomique fonctionnelle et de la médecine.
Dans ses premiers chapitres, le manuscrit de thèse dresse un état de l'art sur les vocabulaires contrôlés (voir chapitre 2), soulignant ainsi leur rôle essentiel dans l'organisation de nos connaissances et de fait, de la recherche d'information, et de l'explicitation de nouvelles connaissances.

Cette présentation nous a mené jusqu'à la forme la plus aboutie des vocabulaire contrôlés, à savoir les ontologies informatiques. Ces ontologies sont sous-tendues par des formalismes rigoureux qui permettent de rendre les données interopérables, et de dégager de nouveaux savoirs en faisant jouer des mécanismes d'inférence. Une intégration de composants ontologiques accrue au cœur d'un système informatique, lui confère de nouvelles capacités telle qu'un socle de données de plus en plus important et la possibilité de réaliser de nouvelles inférences pour générer de nouvelles connaissances (effet boule de neige).
Les systèmes de gestion de données intégrant des ontologies sont appelés des bases de connaissance, et se différencient des autres systèmes par une « sémantisation » accentuée de la donnée. La donnée devient le sujet de ce que l'on cherche à transmettre, on lui donne un sens, on la conceptualise.
Dans le chapitre (voir chapitre 3), nous avons présenté l'agroforesterie pour en révéler toute sa diversité, et en faire ressortir ses intérêts multiples bien souvent sous la forme de services rendus aussi bien à la nature qu'à la société. Elle peut ainsi contribuer à la construction de systèmes de production agricole durables. Un avantage indéniable de l'agroforesterie est qu'elle peut être mise en pratique sans limitations relatives à la zone géographique, au climat ou à des considérations financières. Les exemples les plus réussis de systèmes agroforestiers ne sont souvent pas à rechercher dans les régions les plus fertiles, ni les plus mécanisées. Il est cependant à noter que les systèmes agroforestiers qui perdurent parfois depuis plusieurs siècles, sont des systèmes qui nécessitent un large savoir-faire parfois ancestral, et qui demandent aussi de croiser différentes disciplines à l'exemple de l'économie, la pédologie, la climatologie ou encore de l'agronomie. L'enseignement pour nous, est qu'il fallait trouver un dénominateur commun à tous ces systèmes pour en dégager une représentation partagée. Il nous a semblé que l'arbre était très souvent et très majoritairement l'élément pilier dans ces systèmes, et que le rapport entre d'une part les arbres entre eux, et d'autre part, les arbres avec le reste du système était également un socle fondateur. Nous avons donc décidé d'adosser nos travaux de modélisation conceptuelle sur une vision structurelle des éléments présents dans une parcelle, et de leurs relations topologiques.
L'objectif n'était pas de tout modéliser, mais plutôt de réutiliser l'existant, notamment en matière d'ontologies cadres (voir chapitre 4). L'agroforesterie se nourrit de diverses disciplines, nous avons exploré en conséquence différents composants ontologiques qui modélisent le temps, l'espace ou encore le paradigme d'observation. Nous avons également retenu différentes ontologies terminologiques pour ce qui touche à l'agronomie, à l'agriculture, aux espèces ou encore à l'alimentation. L'idée est de s'approprier les termes et les définitions standards de ces termes, tels que préconisés et partagés par chaque communauté d'expertise. Nous avons vu que l'agroforesterie était sous-tendue par le savoir-faire, nous allons donc faire l'acquisition de ce savoir-faire auprès des communautés. Nous nous sommes ensuite concentrés sur les formalismes logiques (chapitre 5), et en particulier sur les logiques de descriptions qui permettent de construire une ontologie sur la base d'une définition non ambiguë des concepts d'intérêt et de leurs relations. Une grande majorité des ontologies sont actuellement rendues disponibles au travers du langage ontologique OWL (voir chapitre 6), qui s'adosse aux logiques de description pour ce concerne une de ses deux sémantiques, nommée sémantique directe (direct semantics). Nous cherchons donc à exploiter différentes ontologies existantes sous la forme de composants, chacune s'attachant à proposer le meilleur modèle possible pour un univers d'intérêt donné. Notre double choix, à savoir une représentation générale d'éléments d'intérêt et de leurs relations (en particulier topologiques), et une mobilisation de composants ontologiques, nous a conduit ensuite à proposer trois grandes contributions :

1. Construction d'un référentiel terminologique
   Nous n'avions pas décidé de prime abord de construire un référentiel dédié. Il nous semblait en effet que nous pouvions trouver la totalité des termes de l'agroforesterie dans les référentiels des domaines connexes, à l'exemple d'Agrovoc ou d'AFO. Cette première vision était aussi supportée par le fait qu'il existe un thésaurus pour l'agroforesterie. Nous avons cependant modifié un peu notre jugement, et construit un référentiel terminologique « AgroforestryVoc » plus adapté à notre vision de l'aménagement agroforestier. Nous avons ainsi identifié et organisé les concepts clés au travers de plusieurs approches : analyses de discours et de documents de travail, fouille de texte, et alignement avec les référentiels existants. Ce travail est complémentaire aux efforts de modélisation autour « AgroforestryVoc », et nous conduit aux grandes notions traitées, à savoir, les observations, les aspect spatiaux et temporels, et la taxonomie.

2. Élaboration de l'ontologie cadre "Agroforestry"
   En accord avec l'ontologie terminologique « AgroforestryVoc », nous avons construit une ontologie cadre du nom de « Agroforestry ».
   Les notions d'engagement ontologique minimal et d'extensibilité dans la construction d'une ontologie [Thomas R. GRUBER, 1995] ont été au centre de nos préoccupations. « Agroforestry » se veut donc particulièrement générique et modulaire. Dans cet esprit, les ontologies cadres SOSA, GeoSPARQL et OWL-Time ont été mobilisées pour traiter respectivement les problématiques de l'observation, de l'espace et du temps en étendant notre modèle. Les éléments de structure de notre modèle ainsi que leur propriétés sont quand eux enrichis sémantiquement par des termes provenant de l'ontologie terminologique définie « AgroforestryVoc », et d'autres référentiels existants, en particulier, les référentiels retrouvés sur le portail AgroPortal. Pour rendre le composant sémantique « Agroforestry » opérationnel, les langages RDF, RDFS et OWL2, qui font partie intégrante des outils du web sémantique, ont été utilisés.

3. Élaboration de l'application "AOBRA"
   L'étude de cas, réalisée à partir de de données d'expérimentations agroforestières acquises sur le domaine départemental de Restinclières, a montré la pertinence de la démarche , et a fait l'objet du développement d'une application Java/Jena nommée "AOBRA" qui s'adosse au modèle de connaissances "Agroforestry". En effet "AOBRA" nous a permis d'analyser plusieurs indicateurs (à l'exemple de la taille et du diamètre des arbres, ou du rendement des cultures), à l'échelle d'une parcelle agroforestière, et par comparaison avec des parcelles dites de contrôle, ou bien encore d'autres parcelles (agro)forestières ayant des conditions similaires de pédologie, de climatologie, de culture et de période de plantation. Les ontologies tierces retenues pour ce cas d'étude, se révèlent être celles qui ont été reconnues et adoptées par les différentes communautés d'utilisateurs. Nous avons pu ainsi démontrer qu'une application centrée sur une connaissance modélisée peut fournir des fonctionnalités d'analyse pour la gestion de l'agroforesterie.

10.1 Difficultés rencontrées

10.1.1 le paradigme du monde ouvert a un prix

L'approche ontologique est, de part sa nature, indissociable du paradigme d'un monde ouvert. La finalité d'une base de connaissance ancrée dans le web sémantique

est, d'une part de permettre de dégager le plus de connaissances possibles à partir d'une base de faits, et d'autre part d'intégrer ou d'échanger des données et connaissances avec d'autres systèmes de connaissances eux aussi construits dans la sphère du web sémantique. Nous considérons que ces finalités doivent être prises en compte dès l'étape de conception de la base de connaissance. Ceci est d'autant plus vrai, si le langage de modélisation UML est mobilisé, comme cela a été notre cas. En effet, UML est pensé pour le génie logiciel et est aussi largement exploité pour la conception de schémas de bases de données relationnelles. Il est plutôt orienté vers le paradigme d'un monde clos et contraint. Ici nous définissons le monde clos comme un monde dans lequel tout ce qui n'est pas connu et modélisé, est considéré comme faux.

Le langage OWL, et les logiques de description sur lesquelles il base sa sémantique formelle, s'inscrivent au contraire dans l'hypothèse du monde ouvert, impliquant que l'on doit considérer qu'une chose qui n'est pas connue par le système peut néanmoins exister, et donc être vraie.
Une ontologie OWL peut donc être considérée comme une vue partielle du monde pouvant être étendue à l'infini par d'autres ontologies. Cette capacité d'extension entraîne de fait la possibilité de rencontrer des contradictions logiques entre les différents modules ontologiques lorsque leurs règles rentrent en conflit.
Pour exemple, prenons une première ontologie dans laquelle est définie une propriété objet *aPourBranche* qui a pour domaine la classe **Arbre** et pour cible la classe **Branche**. Prenons une seconde ontologie qui étend la première et qui contient un individu de la classe **Buisson** qui est rattachée à un individu de la classe **Branche** au travers de la même propriété *aPourBranche*. Si la classe **Buisson** est déclarée comme disjointe de la classe **Arbre** alors le système détectera une incohérence.

Avec le monde ouvert, une attention particulière doit aussi être portée aux modes opératoires comme le montre l'exemple suivant :

Si nous recherchions dans notre base de connaissance :

> *Tous les arbres ayant au moins un tronc et un système racinaire*

le système retournerait alors tous les arbres, pour lesquels il trouverait au moins un tronc et au moins un système racinaire. Mais sachant que tout les arbres ne possèdent qu'un tronc et qu'un système racinaire nous serions tentés de demander au système de nous extraire :

> *Tous les arbres ayant au moins un tronc et un système racinaire et au plus un tronc et un système racinaire.*

Cette dernière requête ne retournerait aucun résultat parce que le système envisagerait qu'il est toujours possible qu'un arbre puisse être équipé d'un deuxième tronc ou d'un deuxième système racinaire après exploitation de cet axiome.

10.1.2 Mais où sont les données de l'agroforesterie ?

Différentes bases de données sont actuellement disponibles via des interfaces utilisateurs sur le web. Pour exemple, nous nous intéresserons à trois d'entre elles, qui sont AFT [1], PLANTS [2] et GBIF [3].
AFT est maintenue par l'ICRAF, et organise des informations relatives à plus de 600 espèces d'arbres, qui peuvent être d'intérêt en agroforesterie. Ces informations relèvent

1. AFT : The Agroforestree (AFT) Database. http://apps.worldagroforestry.org/treedb2/index.php.
2. PLANTS : The USDA Plants Database. https://plants.usda.gov/home/plantProfile.
3. GBIF : Free and open access to biodiversity data. https://www.gbif.org/.

de la description botanique (caractéristiques de l'arbre, adaptation climatique, aire de distribution et rôle écologique). On peut au travers de l'interface proposée (voir Figure 10.1) rechercher par exemple un arbre correspondant à un besoin de fertilisation du sol (Soil Fertility) pour un pays en particulier [ORWA et al., 2009].
Bien que plus généraliste, et donc moins spécialisée en agroforesterie, les bases de

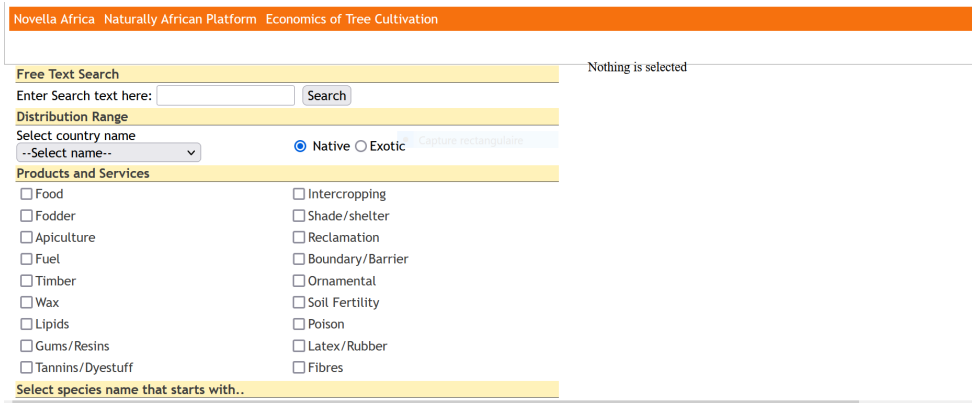


FIGURE 10.1 – *Interface web de la base de données AFT de l'ICRAF.*

données PLANTS de l'USDA [4] [USDA, NRCS, PLANTS, 2012] et GBIF [TELENIUS, 2011] offrent elles aussi des renseignements similaires pouvant être utiles aux agroforestiers sur les utilisations possibles de certaines espèces en terme de production de bois de chauffage, de fourniture de produits consommables pour l'homme, ou pour la production de médicaments ou d'arômes. Toutes ces données, qui pourraient faciliter le choix d'un agroforestier pour la plantation d'une nouvelle espèce en fonction de ses besoins de production, ne sont malheureusement pas ou partiellement accessibles via des web services qui pourraient être mis en place via des API REST ou SOAP.
En effet, il n'a pas été développé d'API pour AFT. L'API d'accès à PLANTS a été rendue indisponible lors de la dernière refonte du site web. L'API de GBIF ne permet pas d'accéder à toutes les informations qui pourraient être exploitables.
En règle générale dans le cadre de l'agroforesterie, nous avons donc souvent rencontré des difficultés à obtenir des données facilement accessibles au travers d'applications informatiques telles qu'AOBRA. Il en va de même pour les données d'expérimentation sur le terrain, dont l'absence ne nous a pas permis de montrer que notre application était généralisable.
Ce manque de données de et pour l'agroforesterie peut s'expliquer par le fait que le renouveau de l'agroforesterie est, comme nous l'avons vu au chapitre 3, assez récent (une vingtaine d'années) et que cette problématique ne s'est pas encore posée. Une autre raison peut aussi être invoquée pour les données déjà acquises mais non encore disponibles, qui est celle du coût. En effet rendre disponible des données implique du temps et des moyens financiers dont ne disposent pas toujours les agroforestiers et leurs partenaires.

Ce constat de manque de données accessibles nous conforte dans notre volonté de mettre en place les moyens nécessaires à la capitalisation et au partage des données agroforestières.

Mais nous pensons aussi que pour obtenir des simulations efficaces d'aménagements agroforestiers, il est impérativement nécessaire que les agroforestiers et leurs partenaires fournissent publiquement leurs données et en particulier les données brutes issues de leurs expérimentations. La disponibilité de ces données permettrait en effet le développement et la mise au point de nouveaux logiciels aptes à fournir une aide

4. USDA : Département américain de l'agriculture. https://www.usda.gov/.

efficace aux agroforestiers et d'encourager l'utilisation de cette technique agricole aux yeux de tous.

10.2 Perspectives

Nous présentons dans cette section plusieurs perspectives de travaux futurs qui ont été envisagées durant cette thèse.

10.2.1 Rapprochement des domaines

Afin de permettre une meilleure intégration des données systémiques et d'ouvrir des perspectives en termes de modélisation des systèmes complexes en agroforesterie, nous envisageons d'élargir le périmètre ontologique à d'autres domaines de connaissances autour de l'agroforesterie.

Parmi ces domaines connexes, on peut par exemple citer la génétique des arbres. Le challenge est de prévoir les réponses évolutives des espèces et des communautés au changement climatique.
On constate un impact anthropomorphique sur les forêts qui modifie le génome des arbres [LEFÈVRE, 2004], et les formes modernes de l'agroforesterie (utilisation de clones de plantes hybrides en agrisylviculture) doivent certainement contribuer dans ce sens à un appauvrissement des ressources génétiques des arbres. D'un autre coté, les techniques agroforestières traditionnelles comme le bocage contribue à protéger ce patrimoine génétique en maintenant en place certaines espèces là où la forêt a fait place à des cultures.
Dans cette voie, l'INRAe pilote le projet Evoltree « (»EVOLution of TREEs as drivers of terrestrial biodiversity) rassemblant les domaines de l'écologie, de l'évolution, de la génomique et de la génétique forestière. Les travaux au sein de ce réseau ont, d'ores et déjà, montré que la génomique peut désormais être intégrée à la recherche écologique pour étudier la réponse des espèces aux changements climatiques [KREMER et al., 2011].
L'ontologie médicale HPO (Human Phenotype Ontology) offre un bon exemple de ce rapprochement avec la génétique en permettant de lier des données génétiques provenant du domaine de la biologie moléculaire (expérimentation animale) avec des maladies humaines au travers de descriptions phénotypiques collectées par des médecins à l'hôpital (compte -rendu d'examen clinique fait sur les patients) [KÖHLER et al., 2014].
Nous pensons qu'un tel rapprochement serait nécessaire en agroforesterie d'autant que le choix des espèces à planter dépasse la problématique de la production de matières agricoles, si l'on prend en compte la nécessité d'obtenir un mélange d'espèces générateur d'effet de facilitation entre arbre et culture.

10.2.2 Unification des sources de données complexes voire contradictoires

Sur la demande faite par la Convention des Nations-Unies sur les Changements Climatiques et du Protocole de Kyoto, l'état français doit dresser un bilan annuel de l'occupation des sols et de ces répercussions sur les flux de carbone. Cet exercice qui se

nomme l'inventaire UTCATF [5] (Utilisation des Terres, Changements d'Affectation des Terres et Foresterie) nécessite l'analyse de nombreuses données complexes sur l'occupation des terres agricoles et leurs rôles dans le stockage du carbone. Fournir de telles données dans le cadre de l'agroforesterie via des applications comme AOBRA pourrait contribuer à cet effort dans le cadre de la lutte contre le changement climatique.

10.2.3 Développement d'une application pour la gestion des données liées

Comme nous l'avons souligné tout au long de ce manuscrit, les outils du web sémantique (RDF,OWL,SPARQL,...) qui sont parties intégrantes de la pile technologique du web, rendent interopérables des données (gérées dans des bases de données) et leurs signifiants (définis dans les ontologies), permettant ainsi à des agents logiciels d'acquérir des informations, et de les traiter via la sémantique des données. Il nous faut réfléchir à des agents qui puissent être exploitables par des agroforestiers non versés en informatique. Ces agents pourraient alors être à même de créer de nouveaux savoirs puis de les rendre disponibles sous la forme de documents, et notamment de page web "sémantisées". Ces pages web en respectant les recommandations du web sémantique fournies par le W3C offriraient sur le web des documents structurant des données directement exploitables pat les agroforestiers. L'idée est aussi de favoriser les échanges avec d'autre agents logiciels basés sur ces mêmes technologies du web sémantique et destinés à des services d'analyse.
Le schéma ci-dessous (Fig.10.2) illustre l'architecture d'une telle application exploitant des données liées.

Mais l'efficience d'une telle application repose sur la disponibilité de pages web déjà existantes traitant de sujets portés sur l'agroforesterie au format RDF et sur l'accès à des bases de données spécialisées en agroforesterie. Or comme nous l'avons déjà souligné les informations accessibles sous cette forme sont anecdotiques à l'heure d'aujourd'hui. Ceci nous a orienté vers une autre perspective présentée à la section suivante.

5. UTCAF : Utilisation des Terres, Changements d'Affectation des Terres et Foresterie. https://ec.europa.eu/clima/policies/forests/lulucf_fr.

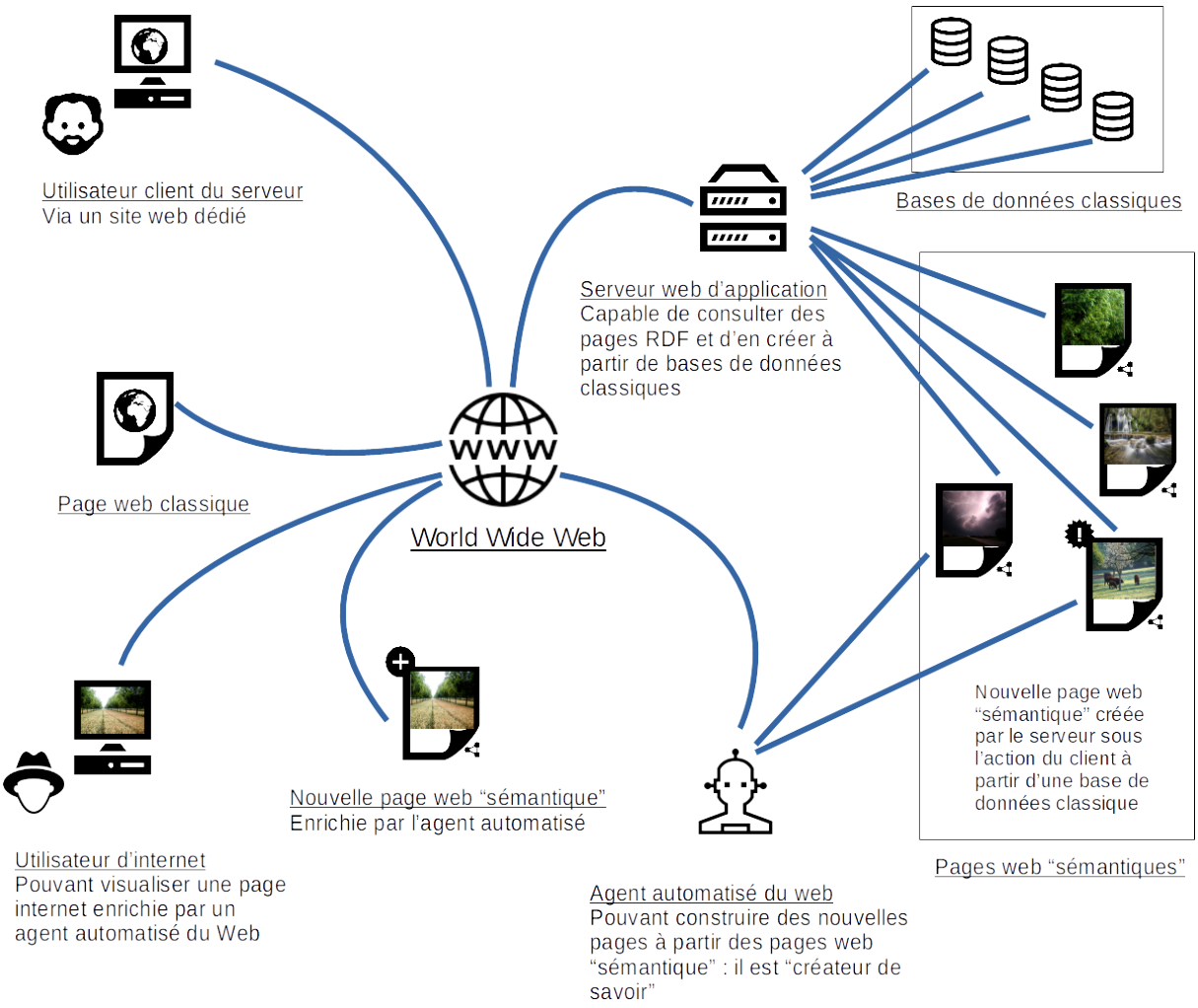


FIGURE 10.2 – *Illustration d'une application pour l'exploitation des données liées.*

10.2.4 Développement d'une plateforme modulaire pour la gestion des données en agroforesterie

Si l'on s'accorde sur la nécessité de mutualiser des efforts de recherche, qui rassemblent à la fois des concepts et des outils agronomiques et écologiques pour améliorer les prévisions de performance des systèmes multi-espèces comme dans le cas de l'agroforesterie [MALÉZIEUX et al., 2009], la tâche reste ardue.
En effet, si l'on considère que les avantages des techniques agroforestières reposent sur les réponses spécifiques d'une parcelle par les arbres et autres composants du système (culture animaux, ...) avec de fortes variations entre localisations des parcelles et politiques de gestion (irrigation, élagage, ...) [COE et al., 2014] et que de plus ces effets varient dans le temps du fait de la lente influence des arbres sur le sol [BARRIOS et al., 2012] les outils nécessaires à une étude efficace de l'efficience des pratiques agroforestières semblent difficiles à construire.
Nous pensons comme d'autres qu'un cadre de modélisation qui intégrerait entre eux des modèles déjà éprouvés serait une solution [MUETZELFELDT et F. L. SINCLAIR, 1993].
La figure 10.3 schématise l'organisation générale de cette plateforme.
Voici les fonctionnalités attendues de cette plateforme :

– Au travers d'une interface web, la plateforme devra permettre aux agroforestiers et à leurs partenaires de constituer un tableau (feuille tabulaire de classeur de type

excel) pour saisir leurs données. Les éléments à décrire, ainsi que les propriétés rattachées, sont proposés sous forme de listes de concepts pré-établies et tirant parti de concepts provenant d'ontologies terminologiques. Le tableau sera ainsi 'sémantisé' : le sujet, les éléments traités ainsi que leurs caractéristiques seront enrichis par des termes désignant des concepts. Pour exemple si un agroforestier veut saisir des mesures dendrométriques de sa parcelle, une en-tête de colonne proposée comme caractéristique de son élément arbre sera le concept "Diameter at breast height" http://opendata.inra.fr/anaeeThes/c3_2371 de l'ontologie AnaEE Thesaurus dont le synonyme et "DBH".

- Par la suite, une fois les données saisies, l'utilisateur peut transmettre ses données à la plateforme afin quelles puissent être transformées au format RDF. Cette intégration sera facilitée par la "sémantisation" préalable du tableau.

- Une fois les données transformées au format RDF, elles peuvent être traitées par le système via des mécanismes d'inférence. Elles peuvent par exemple être réinjectées en tant que paramètres d'entrées dans un modèle de croissance des arbres ou un modèle de bilan radiatif. Ce sont les règles exécutées par le moteur d'inférence de la plateforme qui vérifieront si les données sont valides pour être traitées par le modèle en utilisant leur description sémantique. Par exemple une règle pourra vérifier que l'espèce des arbres de la parcelle est précisée, si le simulateur de croissance requiert obligatoirement cette donnée. Ce sont ces mêmes règles qui seront à l'origine de l'exécution de la simulation. Les données en sortie seront réinjectées au sein de la base de connaissance.
  La modularité du système devra permettre d'interfacer facilement d'autres modèles. Cette capacité sera grandement facilitée par le fait qu'il suffira d'injecter de nouvelles règles (qui seront transmises à la plateforme sous de simples fichiers texte).

- Les données natives et inférées pourront par la suite être mises à disposition des utilisateurs et d'autres applications sur le web.

Nous pensons qu'une plateforme de ce type en facilitant la 'sémantisation' des données et pouvant de façon modulaire intégrer des modèles déjà éprouvés peut aider à une modélisation des plus efficaces d'un aménagement agroforestier dans sa globalité.

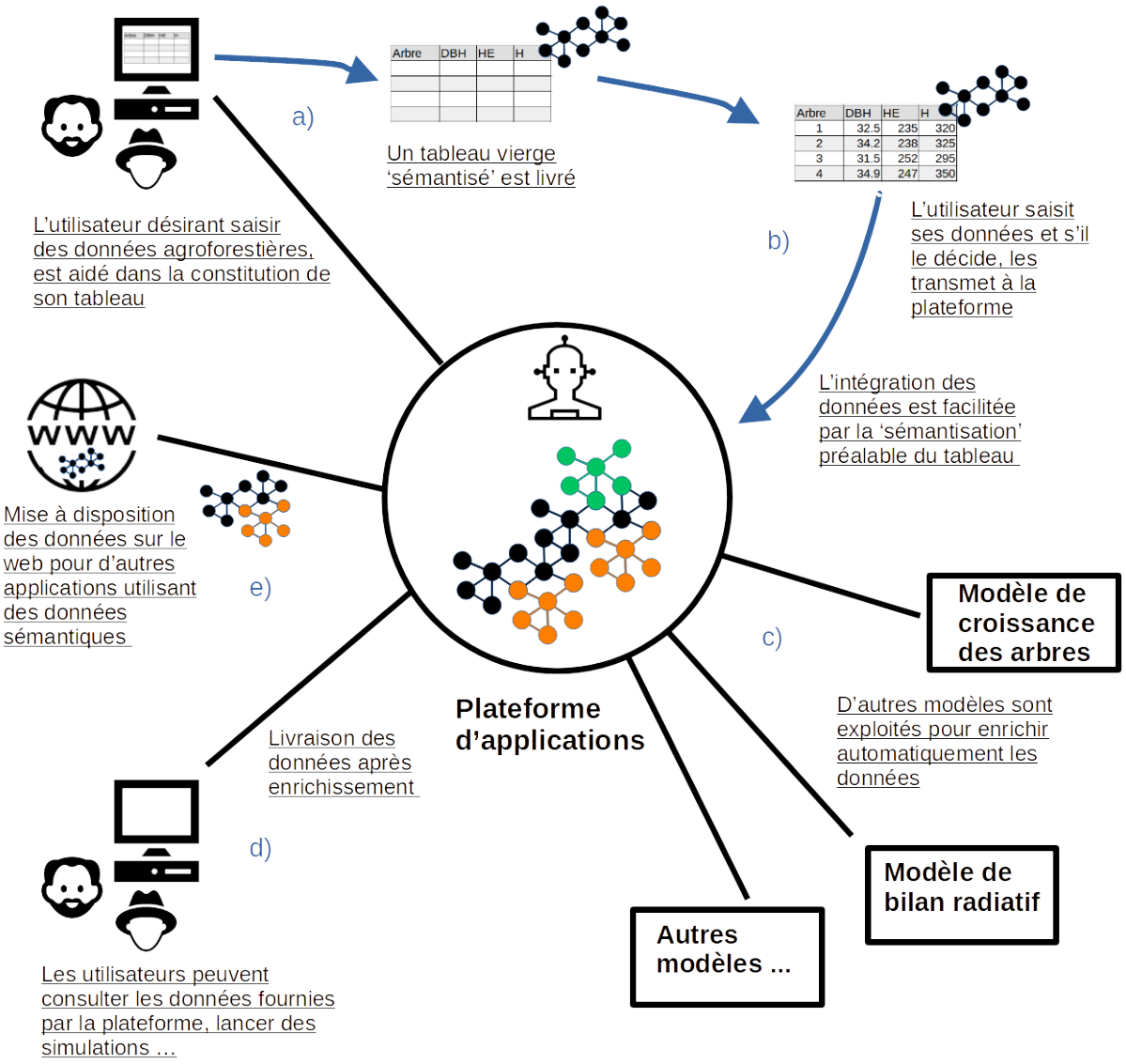


FIGURE 10.3 – *Illustration d'une plateforme d'applications modulaires pour l'exploitation des données agroforestières.* ***a)*** *Constitution d'un tableau 'sémantisé' adapté aux données que souhaite saisir l'utilisateur.* ***b)*** *Une fois les données saisies, elles peuvent être transmises à la plateforme et intégrées sous format RDF.* ***c)*** *Lors des mécanismes d'inférences, des modèles externes peuvent être utilisés.* ***d)*** *Les données natives et inférées de la plateforme sont consultables via une interface web.* ***e)*** *Les données sont aussi accessibles à d'autres applications sur le web.*

Remerciements

Réaliser une thèse "sur le tard" était pour moi un véritable challenge que je ne pensais pas pourvoir surmonter à l'époque. Forte heureusement j'ai rencontré bon nombre de personnes qui m'ont non seulement encouragé à me lancer dans cette formidable aventure, mais qui m'ont de plus aidé tout au long de celle çi. C'est toutes ces personnes que je voudrais remercier ici du fond du cœur.

Je commencerais par mes directrices de thèse Alexia Stokes et Isabelle Mougenot. Alexia, qui après m'avoir accueilli une première fois dans son laboratoire pour un stage de master, m'a incité à poursuivre ma voie dans la recherche scientifique. Isabelle pour avoir accepté de superviser mon travail de thèse en me fournissant une formation basée sur l'éthique et la rigueur.

Je remercie aussi Pierre Pompidor et Anne-Muriel Arignon Chifolleau qui ont suivi mon travail de thèse.

Je remercie l'équipe d'AGROOF, Fabien Liagre, Camille Beral, Nicolas Girardin, Hélène Le Gallic, Daniélé Ori et Ambroise Martin-Chave pour m'avoir transmis leur connaissances et leur passion pour l'agroforesterie.

Je remercie les rapporteurs, François Pinet et Nathalie Hernandez pour la relecture de ce manuscrit. Ainsi que les examinateurs Christian Pichot et Sylvie Ranwez.

Merci à Jérôme Perez pour avoir co-encadrer mon travail de thèse.

Merci à Thierry Fourcaud, Raphaël Pelissier et Pierre Couteron en tant que que directeurs et responsables du laboratoire UMR AMAP pour m'avoir fourni un cadre de travail des plus agréables.

Merci à Yannick Brohard pour ses talents de bibliothécaire et à Nathalie Hodebert, Miléna Dordevic-Giroud et Sylvie Beauclair pour leur soutien administratif et logistique.

Je remercie particulièrement Marilyne Laurans, pour m'avoir orienté dans ma quête de données concernant l'agroforesterie et mis en relation avec Lydie Dufour, que je remercie pour m'avoir fourni des données expérimentales sur les parcelle A3 et B17 du domaine de Restinclières.

Je remercie ausi Anne-Laure Decombeix, Brigitte Meyer-Berthaud, Laurence Gaume, Cédric Gaucherel, Gaëlle Viennois, Zhun Mao, Gilles LE Moghedec, François De Coligny, Yves Caraglio, Marc Jaeger et Gilles Dauby qui ont toujours répondu présents pour répondre à mes questionnements et m'indiquer le Nord.

Merci à mes collègue de bureau, Fabien Anthelme, Claire Fortunel et Jean-Luc Maeght pour leur constante humeur et leur disponibilité de tous les jours.

Merci à "ceux d'en face" Alain Ibrahim, Antoine Affouard, Julien Engel, Hervé Goeau, Mathias Chouet, Jeanne Clement, Pierre Bonnet, Emilie Peynaud et Colette Barboudine car au final tout les informaticiens finissent par se retrouver autour de la machine à café !

Merci à mes collèges Jing Ma, Nidhi Jha, Nereyda-Nathalie Cruz Maldonado, Camille Girard-Tercieux, Claudia Huertas Garcia, Asli Begum Kacamak, Daniela Krebber, Ca-

mille Salmon, Gislain Mofack, Beda Innocent Adji pour leur soutien et les bonnes soirées passées ensemble.

Merci à Merlin Ramel et Jean-François Barczi pour toutes ses conversations autour de tout et surtout de n'importe quoi ...

Merci enfin à Ghislain Vieilledent, Jérôme Munziger, Jean-François Molino, Isabelle Marechaux, Tom Hattermann, Philippe Verley, Rémi Vezy, Maxime Rejou-Mechain, Jean Dauzat, Nicolas Barbier, Patrick Heuret, Stéphane Fourtier avec qui j'ai partagé mon quotidien pendant ces trois années.

Bien sûr tout cela n'aurait pas pu être possible sans le concours de ma femme et de mes trois enfants qui m'ont soutenu tout au long de ma thèse.

Bibliographie

AERTSENS, Joris et al. (2013). « Valuing the carbon sequestration potential for European agriculture ». In : *Land Use Policy* 31, p. 584-594. ISSN : 0264-8377.

Agrof-MM (2016). URL : http://agrofmm.eu/ (visité le 02/09/2020).

AgroFE : projet Leonardo transfert d'innovation (2020). URL : https://eduter.fr/agrofe/ (visité le 25/08/2020).

AL-AJLAN, Ajlan (2015). « The comparison between forward and backward chaining ». In : *International Journal of Machine Learning and Computing* 5.2, p. 106. ISSN : 2010-3700.

ALAVI, Maryam et Dorothy E. LEIDNER (2001). « Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues ». In : *MIS Quarterly* 25.1, p. 107. ISSN : 02767783. DOI : 10.2307/3250961.

ALLIANCE OF AGRICULTURAL INFORMATION AND DOCUMENTATION SERVICES OF THE AMERICAS et al. (2019). *NAL Agricultural Thesaurus and Glossary*. DOI : 10.15482/USDA.ADC/1503889.

ALONSO, Y. (2004). « The National Agricultural Library–Providing Tools For Education And Development In Agriculture. Best Practices in Government Information: A Global Perspective. 2007, 205 ». In : *35981176*. ISSN : 35981176.

ASHBURNER, M. et al. (2000). « Gene ontology: tool for the unification of biology. The Gene Ontology Consortium ». In : *Nature genetics* 25.1, p. 25-29. ISSN : 1061-4036. DOI : 10.1038/75556.

BAADER, Franz, Sebastian BRANDT et al., éd. (2005). *Pushing the EL envelope*.

BAADER, Franz, Diego CALVANESE et al. (2003). *The description logic handbook: Theory, implementation and applications*. Cambridge university press. ISBN : 0521781760.

BAADER, Franz et Philipp HANSCHKE (1991). *A scheme for integrating concrete domains into concept languages*. Deutsches Forschungszentrum für Künstliche Intelligenz.

BACHIMONT, Bruno (2000). « Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances ». In : *Ingénierie des connaissances: évolutions récentes et nouveaux défis*, p. 305-323.

BAKER, Thomas, Sean BECHHOFER et al. (2013). « Key choices in the design of Simple Knowledge Organization System (SKOS) ». In : *Journal of web semantics* 20, p. 35-49. ISSN : 1570-8268.

BAKER, Thomas, Caterina CARACCIOLO et al. (2016). « GACS Core: Creation of a global agricultural concept scheme ». In : *Research Conference on Metadata and Semantics Research*. Springer, p. 311-316.

BAKER, Thomas et O. SUOMINEN (2014). « Global agricultural concept scheme (gacs): A multilingual thesaurus hub for linked data ». In : *Website: http://aims. fao. org/sites/default/files/GACS_ Integration_Proposal_1* 200.

BARAL, Chitta et Michael GELFOND (1994). « Logic programming and knowledge representation ». In : *The Journal of Logic Programming* 19, p. 73-148. ISSN : 0743-1066.

BARD, Jonathan B. L. et Seung Y. RHEE (2004). « Ontologies in biology: design, applications and future challenges ». In : *Nature reviews. Genetics* 5.3, p. 213-222. ISSN : 1471-0056. DOI : 10.1038/nrg1295.

BARRIOS, Edmundo (2007). « Soil biota, ecosystem services and land productivity ». In : *Ecological economics* 64.2, p. 269-285. ISSN : 0921-8009.

BARRIOS, Edmundo et al. (2012). « Agroforestry and soil health: linking trees, soil biota and ecosystem services ». In : *Soil ecology and ecosystem services* 14, p. 315-330.

BATTLE, Robert et Dave KOLAS (2012). « Enabling the geospatial Semantic Web with Parliament and GeoSPARQL ». In : *Semantic Web* 3.4, p. 355-370. ISSN : 1570-0844. DOI : 10.3233/SW-2012-0065. URL : https://content.iospress.com/articles/semantic-web/sw065.

BÉRAL, Camille et al. (2018). « PARASOL: AGROFORESTERIE EN SYSTÈME D'ÉLEVAGE OVIN-Étude de son potentiel dans le cadre de l'adaptation au changement climatique ». Thèse de doct. Agroof Inra Idele UniLaSalle.

BERGERON, Mélanie et al. (2011). « Reduced soil nutrient leaching following the establishment of tree-based intercropping systems in eastern Canada ». In : *Agroforestry Systems* 83.3, p. 321-330. ISSN : 0167-4366.

BERNERS-LEE, T. (2000). *Semantic Web - XML2000 - slide "Semantic Web on XML"*. URL : https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide1-0.html (visité le 19/10/2020).

BERNERS-LEE, T. et al. (2002). « A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities ». In : *Scientific American*.

BILAN, GIEC (2007). « des changements climatiques. Contribution des Groupes de travail I, II et III au quatrième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat ». In : *Équipe de rédaction principale. Geneva, Switzerland: GIEC*, p. 103.

BOBROW, Daniel G. et Terry WINOGRAD (1977). « An Overview of KRL, a Knowledge Representation Language ». In : *Cognitive Science* 1.1, p. 3-46. ISSN : 0364-0213. DOI : 10.1207/s15516709cog0101_2.

BOLLACKER, Kurt et al. (2008). « Freebase: a collaboratively created graph database for structuring human knowledge ». In : *ACM SIGMOD international conference on Management of data*, p. 1247-1250.

BORST, Willem Nico (1997). « Construction of engineering ontologies for knowledge sharing and reuse ». In.

BRACHMAN, Ronald J. et Hector J. LEVESQUE (1984). « The Tractability of Subsumption in Frame-Based Description Languages ». In : *AAAI*, p. 34-37.

BRACHMAN, Ronald J. et James G. SCHMOLZE (1985). « An Overview of the KL-ONE Knowledge Representation System ». In : *Cognitive Science* 9, p. 171-216. ISSN : 0364-0213.

— (1989). *Readings in artificial intelligence and databases*. Sous la dir. de John. MYLOPOULOS et Michael L.. BRODIE. San Mateo (Calif.) DOI : 10.1016/B978-0-934613-53-8.50019-4.

BREU, Ruth et al. (1997). « Towards a formalization of the Unified Modeling Language ». In : *ECOOP'97 - object-oriented programming*. Sous la dir. de Mehmet. AKSIT et Satoshi MATSUOKA. T. 1241. Lecture Notes in Computer Science. Berlin et New York : Springer, p. 344-366. ISBN : 978-3-540-63089-0. DOI : 10.1007/BFb0053386.

BURGESS, P. J. et al. (2019). « Modelling agroforestry systems ». In : *Agroforestry for Sustainable Agriculture*, p. 209-238.

BURGESS, Paul J. (1999). « Effects of agroforestry on farm biodiversity in the UK ». In : *0036-9217*. ISSN : 0036-9217.

BURRIEL, Charles et al. (2017). « Knowledge databank and repository service for agroforestry ». In : *Journal of Agricultural Informatics* 8.3. DOI : 10.17700/jai.2017.8.3.393.

BUTTIGIEG, Pier Luigi, Norman MORRISON et al. (2013). « The environment ontology: contextualising biological and biomedical entities ». In : *Journal of Biomedical Semantics* 4.1, p. 43. ISSN : 2041-1480. DOI : 10.1186/2041-1480-4-43.

BUTTIGIEG, Pier Luigi, Evangelos PAFILIS et al. (2016). « The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation ». In : *Journal of Biomedical Semantics* 7.1, p. 57. ISSN : 2041-1480. DOI : 10.1186/s13326-016-0097-6.

CADISCH, G. et al. (2004). « Safety-nets and filter functions of tropical agroforestry systems ». In : *Controlling nitrogen flows and losses*, p. 406-414.

CALAME, Matthieu (2016). *Comprendre l'agroécologie: origines, principes et politiques*. ECLM. ISBN : 2843772028.

CARACCIOLO, Caterina et al. (2013). « The AGROVOC Linked Dataset ». In : *Semantic Web* 4.3, p. 341-348. ISSN : 1570-0844. DOI : 10.3233/SW-130106. URL : https://content.iospress.com/articles/semantic-web/sw106.

CARDINAEL, Rémi et al. (2015). « Competition with winter crops induces deeper rooting of walnut trees in a Mediterranean alley cropping agroforestry system ». In : *Plant and Soil* 391.1-2, p. 219-235. ISSN : 0032-079X.

CHARLET, Jean (2004). *L'ingénierie des connaissances, entre science de l'information et science de gestion*.

CHENU, Karine et al. (2017). « Contribution of crop models to adaptation in wheat ». In : *Trends in plant science* 22.6, p. 472-490. ISSN : 1360-1385.

CIMIANO, Philipp et al. (2006). « Ontologies on demand ?-a description of the state-of-the-art, applications, challenges and trends for ontology learning from text ». In : *Information, Wissenschaft und Praxis* 57.6-7.

CLASTRE, Philippe et al. (2018). *Vocabulary for analysis and experimentation on terrestrial and aquatic ecosystems*. DOI : 10.15454/1.4894016754286177E12.

CLOBERT, Jean et al. (2018). « How to Integrate Experimental Research Approaches in Ecological and Environmental Studies: AnaEE France as an Example ». In : *Frontiers in Ecology and Evolution* 6. DOI : 10.3389/fevo.2018.00043.

COE, Richard et al. (2014). « Scaling up agroforestry requires research 'in'rather than 'for'development ». In : *Current Opinion in Environmental Sustainability* 6, p. 73-77. ISSN : 1877-3435.

COLLIGNON, Alain et Pascal CUXAC (2017). « ISTEX: des enrichissements au web de données ». In : *I2D Information, donnees documents* 54.4, p. 8-15. ISSN : 2428-2111.

COLLINS, Allan M. et M. Ross QUILLIAN (1969). « Retrieval time from semantic memory ». In : *Journal of verbal learning and verbal behavior* 8.2, p. 240-247.

CÔTÉ, Richard G. et al. (2006). « The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries ». In : *BMC bioinformatics* 7.1, p. 1-7. ISSN : 1471-2105.

COX, Simon J.D. (2016). « Ontology for observations and sampling features, with alignments to existing models ». In : *Semantic Web* 8.3, p. 453-470. ISSN : 1570-0844. DOI : 10.3233/SW-160214.

DALBERG, Ingetraut (2017). « Why a new universal classification system is needed ». In : *KO KNOWLEDGE ORGANIZATION* 44.1, p. 65-71. ISSN : 0943-7444.

DAML.org (2006). URL : http://www.daml.org/ (visité le 04/02/2021).

DAVIES, John et al. (2003). « Towards the semantic web ». In : *C hichester, John W iley & Sons*.

DBpedia (2021). URL : https://wiki.dbpedia.org/ (visité le 05/02/2021).

DCMI USAGE BOARD (2006). *DCMI: DCMI Metadata Terms*. URL : https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ (visité le 30/08/2021).

DELAHAYE, Jean-Paul (1995). *Logique, informatique et paradoxes*. Bibliothèque Pour la science. Paris : Pour la science et diff. Belin. ISBN : 2902918941.

DIENG, Rose et al. (2001). *Méthodes et outils pour la gestion des connaissances: une approche pluridisciplinaire du knowledge management*. Dunod. ISBN : 2100063006.

DIX, Mary Ellen et al. (1995). « Influences of trees on abundance of natural enemies of insect pests: a review ». In : *Agroforestry Systems* 29.3, p. 303-311. ISSN : 0167-4366.

DUBOIS, J. P. et al. (2008). « Agroforesterie: comportement des oies sous un couvert de noyers et effets sur les performances du verger ». In : *Proc. 8èmes Journées de la Recherche sur les Palmipèdes à Foie Gras*, p. 30-31.

DUCROS, D. et al. (2005). « Systèmes de production diversifiés ». In : *Les prés-vergers: une alternative à l'arboriculture intensive face à l'évolution de la PAC. Toulouse: ENFA Toulouse, Laboratoire Dynamiques rurales (UMR UTM-ENFA-INP ENSAT)*.

DUFY, L. et al. (2006). « GIEA : gestion des informations de l'exploitation agricole. Un projet au service de l'interopérabilité sémantique de la profession agricole ». In : *Ingénieries eau-agriculture-territoires* 48, p. 27-36. ISSN : 1264-9147. URL : https://hal.archives-ouvertes.fr/hal-00601326.

DUPRAZ, Christian (1994). « Les associations d'arbres et de cultures intercalaires annuelles sous climat tempéré ». In : *Revue Forestière Française*. ISSN : 1951-6827.

DUPRAZ, Christian, Daniel AUCLAIR et al. (1992). « Vingt ans de recherche agroforestière en Nouvelle-Zélande: quels enseignements pour l'Europe ? 1re partie: l'agroforesterie néo-zélandaise ». In : *Revue Forestière Française*. ISSN : 1951-6827.

DUPRAZ, Christian, Isabelle LECOMTE et al. (2002). « Technical plan for implementing the HySAFE biophysical model ». In.

DUPRAZ, Christian et Fabien LIAGRE (2008). *Agroforesterie: des arbres et des cultures*. France Agricole Editions. ISBN : 2855571502.

DUPRAZ, Christian, Kevin J. WOLZ et al. (2019). « Hi-sAFe: a 3D agroforestry model for integrating dynamic tree–crop interactions ». In : *Sustainability* 11.8, p. 2293.

FABIEN LIAGRE et al. (2004). « Will French farmers adopt agroforestry technology in the near future ? » In : URL : https://hal.archives-ouvertes.fr/hal-02759823/ (visité le 04/06/2020).

FANATICO, A. C. et al. (2016). « Effect of outdoor structural enrichments on the performance, use of range area, and behavior of organic meat chickens ». In : *Poultry science* 95.9, p. 1980-1988. ISSN : 0032-5791.

FELLBAUM, Christiane (2012). *WordNet*. The Encyclopedia of Applied Linguistics. American Cancer Society. DOI : 10.1002/9781405198431.wbeal1285.

FIORE, Nicola et al., éd. (2017). *EcoPortal: A Proposition for a Semantic Repository Dedicated to Ecology and Biodiversity*.

FORGY, Charles L. (1989). « Rete: A fast algorithm for the many pattern/many object pattern match problem ». In : *Readings in Artificial Intelligence and Databases*. Elsevier, p. 547-559.

FREGE, Gottlob (1948). « Sense and reference ». In : *The philosophical review* 57.3, p. 209-230. ISSN : 0031-8108.

GANDON, Fabien et Alexandre PASSANT (2015). *Folksonomies, Thésaurus et Ontologies: trois artefacts combinés dans la structuration des données du Web*.

GARGOMINY, O. et al. (2016). « TAXREF v8. 0, référentiel taxonomique pour la France: méthodologie, mise en œuvre et diffusion ». In : *Muséum national d'Histoire naturelle, Paris. Rapport SPN* 101, p. 144.

GARNIER, Eric et al. (2017). « Towards a thesaurus of plant characteristics: an ecological contribution ». In : *Journal of Ecology* 105.2, p. 298-309. ISSN : 0022-0477.

GAVALAND, André et Laurent BURNEL (2005). « Croissance et biomasse aérienne de noyers noirs en parcelle agroforestière à Les Eduts (Charente-Maritime). Dossiers" agroforesterie: produire autrement" ». In : *Chambres d'agriculture* 945, p. 20-22.

GERMAIN, Karine (2014). « Conduite de productions animales dans des couverts complexes. Production de volailles biologiques en parcours prairiaux et arborés ». In : *Innovations Agronomiques* 40, p. 125-132.

GINANE, C. et al. (2018). « Pâturage sur prairies agroforestières: quels impacts des arbres sur le comportement, le bien-être et les performances des ovins ? » In.

GKOUTOS, Georgios V. et al. (s. d.). « Entity/quality-based logical definitions for the human skeletal phenome using PATO ». In : *2009 annual international conference of the IEEE engineering in medicine and biology society*. IEEE, p. 7069-7072. ISBN : 1424432960.

GORIA, Stéphane (2006). « Knowledge Management et Intelligence Economique: deux notions aux passés proches et aux futurs complémentaires ». In : *Informations, Savoirs, Décisions et Médiations [Informations, Sciences for Decisions Making]* 27, p. 1-16.

GRAU, Bernardo Cuenca et al. (2008). « OWL 2: The next step for OWL ». In : *Journal of Web Semantics* 6.4, p. 309-322. ISSN : 1570-8268. DOI : 10.1016/j.websem.2008.05.001. URL : http://www.sciencedirect.com/science/article/pii/S1570826808000413.

GRAVES, Mike et al. (2007). « Foaf: Connecting people on the semantic web ». In : *Cataloging & classification quarterly* 43.3-4, p. 191-202. ISSN : 0163-9374.

GRUBER, Thomas R. (1995). « Toward principles for the design of ontologies used for knowledge sharing? » In : *International Journal of Human-Computer Studies* 43.5-6, p. 907-928. ISSN : 1071-5819. DOI : 10.1006/ijhc.1995.1081. URL : http://www.sciencedirect.com/science/article/pii/S1071581985710816.

GRUBER, Tom et al. (2009). « Encyclopedia of database systems ». In : *Ontology*.

GRUBER, Tom R. (1993). « Knowledge acquisition ». In : *A translation approach to portable ontology specifications* 5.199-220, p. 10.1006.

GUARINO, Nicola (1998). *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. IOS press. ISBN : 9051993994.

GUARINO, Nicola et al. (1994). « Formalizing ontological commitment ». In : *AAAI*, p. 560-567.

HAASE, Peter et al. (2008). « The neon ontology engineering toolkit ». In : *WWW*.

HALLER, Armin et al. (2018). « The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation ». In : *Semantic Web* 10.1, p. 9-32. ISSN : 1570-0844. DOI : 10.3233/SW-180320.

HAMON, Xavier et al. (2009). « L'agroforesterie, outil de séquestration du carbone en agriculture ». In.

HASSANPOUR, Saeed et al. (2009). « Exploration of SWRL rule bases through visualization, paraphrasing, and categorization of rules ». In : *International Workshop on Rules and Rule Markup Languages for the Semantic Web*. Springer, p. 246-261.

HEATH, Tom et Christian BIZER (2011). « Linked data: Evolving the web into a global data space ». In : *Synthesis lectures on the semantic web: theory and technology* 1.1, p. 1-136.

HOBBS, Jerry R. et Feng PAN (2004). « An ontology of time for the semantic web ». In : *ACM Transactions on Asian Language Information Processing (TALIP)* 3.1, p. 66-85. ISSN : 1530-0226. DOI : 10.1145/1017068.1017073.

HOCHMAN, Z. et al. (2009). « Re-inventing model-based decision support with Australian dryland farmers. 4. Yield Prophet® helps farmers monitor and manage crops in a variable climate ». In : *Crop and Pasture Science* 60.11, p. 1057-1070. ISSN : 1836-5795.

HODGE, Gail (2000). « Systems of knowledge organization for digital libraries ». In : *Digital Library Federation, USA*.

HOLZWORTH, Dean P. et al. (2014). « APSIM–evolution towards a new generation of agricultural systems simulation ». In : *Environmental Modelling & Software* 62, p. 327-350. ISSN : 1364-8152.

HORN, Alfred (1951). « On sentences which are true of direct unions of algebras ». In : *Journal of Symbolic Logic* 16.1, p. 14-21. DOI : 10.2307/2268661.

HORROCKS, Ian, Peter F. PATEL-SCHNEIDER, Harold BOLEY et al. (2004). « SWRL: A semantic web rule language combining OWL and RuleML ». In : *W3C Member submission* 21.79, p. 1-31.

HORROCKS, Ian, Peter F. PATEL-SCHNEIDER et Frank VAN HARMELEN (2003). « From SHIQ and RDF to OWL: The making of a web ontology language ». In : *Journal of web semantics* 1.1, p. 7-26. ISSN : 1570-8268.

https://gargantext.org (2021). URL : https://gargantext.org/ (visité le 24/02/2021).

HYVÖNEN, Eero (2010). « Preventing ontology interoperability problems instead of solving them ». In : *Semantic Web* 1.1, 2, p. 33-37. ISSN : 1570-0844.

HYVÖNEN, Eero et al. (2008). « Building a national semantic web ontology and ontology service infrastructure–the FinnONTO approach ». In : *European Semantic Web Conference*. Springer, p. 95-109.

IBAÑEZ, J. J. et al. (1981). « Propuesta de ordenamiento de los agrosistemas de dehesa en la península ibérica ». In : *Revista de Estudios Agrosociales* 114, p. 39-79. ISSN : 0034-8155.

IETF (2020a). *RFC 3986 - Uniform Resource Identifier (URI): Generic Syntax*. URL : https://tools.ietf.org/html/rfc3986 (visité le 03/10/2020).

— (2020b). *RFC 3987 - Internationalized Resource Identifiers (IRIs)*. URL : https://tools.ietf.org/html/rfc3987 (visité le 03/10/2020).

— (2020c). *RFC 8141 - Uniform Resource Names (URNs)*. URL : https://tools.ietf.org/html/rfc8141 (visité le 03/10/2020).

ISAAC, Antoine (2013). *Correspondence between ISO 25964 and SKOS/SKOS-XL models*. URL : https://groups.niso.org/apps/group_public/download.php/12351/Correspondence%20ISO25964-SKOSXL-MADS-2013-12-11.pdf.

ISO (2020a). *ISO 25964-1:2011*. URL : https://www.iso.org/standard/53657.html (visité le 12/10/2020).

— (2020b). *ISO 8879:1986*. URL : https://www.iso.org/standard/16387.html (visité le 15/10/2020).

ISO/TC et Simon J. D. COX, éd. (2014). *ISO/DIS 19150-2:2014 - Geographic information – Ontology – Part 2: Rules for developing ontologies in the Web Ontology Language (OWL)*.

JAIN, Vishal et Mayank SINGH (2013). « Ontology development and query retrieval using protégé tool ». In : *International Journal of Intelligent Systems and Applications* 9.9, p. 67-75.

JAISWAL, Pankaj et al. (2005). « Plant Ontology (PO): a Controlled Vocabulary of Plant Structures and Growth Stages ». In : *Comparative and functional genomics* 6.7-8, p. 388-397. ISSN : 1531-6912. DOI : 10.1002/cfg.496.

JANOWICZ, Krzysztof, Armin HALLER et al. (2019). « SOSA: A lightweight ontology for sensors, observations, samples, and actuators ». In : *Journal of Web Semantics* 56, p. 1-10. ISSN : 1570-8268. DOI : 10.1016/j.websem.2018.06.003.

JANOWICZ, Krzysztof, Pascal HITZLER et al. (2014). « Five stars of Linked Data vocabulary use ». In : *Semantic Web* 5.3, p. 173-176. ISSN : 1570-0844. DOI : 10.3233/SW-140135.

JONQUET, Clement et al. (2018). « Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal ». In : *Journal on Data Semantics* 7.4, p. 191-221. ISSN : 1861-2040.

JONQUET, Clément et al. (2018). « AgroPortal: A vocabulary and ontology repository for agronomy ». In : *Computers and Electronics in Agriculture* 144, p. 126-143. ISSN : 01681699. DOI : 10.1016/j.compag.2017.10.012.

JOSE, S. et al. (2004). « Interspecific interactions in temperate agroforestry ». In : *Agroforestry Systems* 61.1, p. 237-255. ISSN : 0167-4366.

JOSHI, Amit et al. (2021). « A Knowledge Organization System for the United Nations Sustainable Development Goals ». In : *European Semantic Web Conference*. Springer, p. 548-564.

KARKI, Uma et Mary S. GOODMAN (2010). « Cattle distribution and behavior in southern-pine silvopasture versus open-pasture ». In : *Agroforestry Systems* 78.2, p. 159-168. ISSN : 0167-4366.

KEET, C. Maria et Pablo Rubén FILLOTTRANI (2015). « An ontology-driven unifying metamodel of UML Class Diagrams, EER, and ORM2 ». In : *Data & knowledge engineering* 98, p. 30-53. ISSN : 0169-023X.

KEIL, Jan Martin et Sirko SCHINDLER (2019). « Comparison and evaluation of ontologies for units of measurement ». In : *Semantic Web* 10.1, p. 33-51. ISSN : 1570-0844.

Knowledge Interchange Format (KIF) (2021). URL : https://web.archive.org/web/20070212094221/http://www.ksl.stanford.edu/knowledge-sharing/kif/ (visité le 04/02/2021).

KÖHLER, Sebastian et al. (2014). « The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data ». In : *Nucleic acids research* 42.D1, p. D966-D974. ISSN : 1362-4962.

KOWALSKI, Robert A. (1988). « The early years of logic programming ». In : *Communications of the ACM* 31.1, p. 38-43. ISSN : 0001-0782. DOI : 10.1145/35043.35046.

KREMER, Antoine et al. (2011). « Forest ecosystem genomics and adaptation: EVOLTREE conference report ». In : *Tree Genetics & Genomes* 7.4, p. 869-875. ISSN : 1614-2950.

KRÖTZSCH, M. et al. (2014). « Description Logics ». In : *IEEE Intelligent Systems* 29.1, p. 12-19. ISSN : 1541-1672. DOI : 10.1109/MIS.2013.123.

KURT, Bollacker et al. (2007). « A platform for scalable, collaborative, structured information integration ». In : *Workshop on Information Integration on the Web (IIWeb'07)*, p. 22-27.

L'agriculture biologique (AB) (2020). URL : https://agriculture.gouv.fr/lagriculture-biologique-ab (visité le 15/11/2020).

LABELLE, Richard (1987). « Ten years of work in agroforestry information and documentation ». In : *Agroforestry systems* 5.3, p. 339-352. ISSN : 0167-4366.

LAPORTE, Marie-Angélique et al. (s. d.). « A semantic web faceted search system for facilitating building of biodiversity and ecosystems services ». In : *International Conference on Data Integration in the Life Sciences*. Springer, p. 50-57.

LASSILA, Ora et Ralph R. SWICK (1998). « Resource description framework (RDF) model and syntax specification ». In.

LAUBLET, Philippe et al. (2002). « Sur quelques aspects du Web sémantique ». In : *Assises du GDR I* 3, p. 59-78.

LE DEUFF, Olivier (2006). « Folksonomies ». In.

LEAKEY, Roger (2017). *Multifunctional agriculture: Achieving sustainable development in Africa*. Academic Press. ISBN : 0128093064.

LEFÈVRE, François (2004). « Human impacts on forest genetic resources in the temperate zone: an updated review ». In : *Forest Ecology and Management* 197.1-3, p. 257-271. ISSN : 0378-1127.

LEHMANN, Jens et al. (2015). « Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia ». In : *Semantic web* 6.2, p. 167-195. ISSN : 1570-0844.

LICKLIDER, J. C. R. et Welden E. CLARK (1962). « On-line man-computer communication ». In : *Proceedings of the May 1-3, 1962, spring joint computer conference on - AIEE-IRE '62 (Spring)*. Sous la dir. de G. A. BARNARD. New York, New York, USA : ACM Press, p. 113. DOI : 10.1145/1460833.1460847.

LINIGER, H. P. et al. (1998). « Pressure on land: the search for sustainable use in a highly diverse environment. Resources, actors and policies-towards sustainable regional development in the highland-lowland system of Mount Kenyae ». In.

Loom (2007). URL : https://www.isi.edu/isd/LOOM/ (visité le 04/02/2021).

LUEDELING, Eike et al. (2016). « Field-scale modeling of tree–crop interactions: Challenges and development needs ». In : *Agricultural Systems* 142, p. 51-69. ISSN : 0308-521X.

LUNDGREN, B. O. et J. B. RAINTREE (1982). « Agricultural Research for Development: Potential and Challenges in Asia ». In : *Sustained Agroforestry*, p. 37-49.

MACGREGOR, Robert et Mark H. BURSTEIN (1991). « Using a description classifier to enhance knowledge representation ». In : *IEEE Expert* 6.3, p. 41-46. ISSN : 0885-9000.

MADIN, Joshua et al. (2007). « An ontology for describing and synthesizing ecological observation data ». In : *Ecological Informatics* 2.3, p. 279-296. ISSN : 1574-9541. DOI : 10.1016/j.ecoinf.2007.05.004. URL : http://www.sciencedirect.com/science/article/pii/S1574954107000362.

MADS RASMUSSEN et al. (2017). « Recent changes in the Building Topology Ontology ». In : URL : https://hal-emse.ccsd.cnrs.fr/emse-01638305/ (visité le 20/11/2017).

MALÉZIEUX, Eric et al. (2009). « Mixing plant species in cropping systems: concepts, tools and models: a review ». In : *Sustainable agriculture*, p. 329-353.

MALVIYA, Naveen et al. (2011). « Developing university ontology using protégé owl tool: Process and reasoning ». In : *International Journal of Scientific & Engineering Research* 2.9, p. 1-8.

MANIEZ, Jacques (1999). « Des classifications aux thésaurus: du bon usage des facettes ». In : *Documentaliste-Sciences de l'Information* 36.4-5, p. 249-260. ISSN : 0012-4508.

MARTÍNEZ-GONZÁLEZ, M. Mercedes et María-Luisa ALVITE-DÍEZ (2019). « The support of constructs in thesaurus tools from a Semantic Web perspective: Framework to assess standard conformance ». In : *Computer Standards & Interfaces* 65, p. 79-91. ISSN : 0920-5489. DOI : 10.1016/j.csi.2019.02.003. URL : https://www.sciencedirect.com/science/article/pii/S0920548918302447.

MCBRIDE, Brian (2002). « Jena: A semantic web toolkit ». In : *IEEE Internet computing* 6.6, p. 55-59. ISSN : 1089-7801.

MENON, Bruno (2016). « Comprendre les standards du web de données ». In : *I2D Information, donnees documents* 53.2, p. 32-34. ISSN : 2428-2111.

MICHEL, Franck et al. (2017). « A model to represent nomenclatural and taxonomic information as linked data. Application to the French Taxonomic Register, TAXREF ». In : *ISWC 2017 Workshop on Semantics for Biodiversity (S4Biodiv 2017)*, p. 1-12.

MICHOTEY, Célia et al. (2019). *Woody Plant Ontology*. DOI : 10.15454/JB2WCE.

MINSKY, Marvin (1975). *A framework for representing knowledge. Pages 211–277in P. Winston (ed.), The psychology of computer vision*.

MISHRA, Ravi Bhushan et Sandeep KUMAR (2011). « Semantic web reasoners and languages ». In : *Artificial Intelligence Review* 35.4, p. 339-368. ISSN : 0269-2821.

MUETZELFELDT, R. I. et F. L. SINCLAIR (1993). « Ecological modelling of agroforestry systems ». In.

MULIA, R. et Christian DUPRAZ (2006). « Unusual fine root distributions of two deciduous tree species in southern France: what consequences for modelling of tree root dynamics ? » In : *Plant and Soil* 281.1, p. 71-85. ISSN : 1573-5036.

NAIR, P. RamachandranK (1985). « Classification of agroforestry systems ». In : *Agroforestry systems* 3.2, p. 97-128. ISSN : 0167-4366.

— (2011). « Agroforestry systems and environmental quality: introduction ». In : *Journal of environmental quality* 40.3, p. 784-790. ISSN : 0047-2425.

NARDI, Daniele et Ronald J. BRACHMAN (2003). « An introduction to description logics ». In : *Description logic handbook* 1, p. 40.

NOY, Natalya F. et al. (2009). « BioPortal: ontologies and integrated data resources at the click of a mouse ». In : *Nucleic acids research* 37.suppl_2, W170-W173. ISSN : 1362-4962.

O'CONNOR, Martin et al. (2005). « Supporting rule system interoperability on the semantic web with SWRL ». In : *International semantic web conference*. Springer, p. 974-986.

O'NEILL, Brian C. et al. (2017). « IPCC reasons for concern regarding climate change risks ». In : *Nature Climate Change* 7.1, p. 28-37. ISSN : 1758-678X. DOI : 10.1038/nclimate3179.

O'REILLY, Tim (2009). *What is web 2.0*. " O'Reilly Media, Inc." ISBN : 1449391079.

OMG (2012). *About the Object Constraint Language Specification Version 2.3.1*. URL : https://www.omg.org/spec/OCL/2.3.1/ (visité le 30/08/2021).

OPENTHESO (2016). *Maison de l'Orient et de la Méditerranée Jean Pouilloux*. URL : https://www.mom.fr/ressources-numeriques/opentheso (visité le 30/10/2020).

ORWA, C. et al. (2009). « Agroforestree Database: a tree reference and selection guide. Version 4 ». In : *Agroforestree Database: a tree reference and selection guide. Version 4*.

OWL 2 Web Ontology Language Direct Semantics (Second Edition) (2017). URL : https://www.w3.org/TR/2012/REC-owl2-direct-semantics-20121211/ (visité le 03/02/2021).

OWL 2 Web Ontology Language Document Overview (Second Edition) (2017). URL : https://www.w3.org/TR/2012/REC-owl2-overview-20121211/ (visité le 03/02/2021).

OWL 2 Web Ontology Language RDF-Based Semantics (Second Edition) (2017). URL : https://www.w3.org/TR/2012/REC-owl2-rdf-based-semantics-20121211/ (visité le 03/02/2021).

P. UDAWATTA, Ranjith et al. (2019). « Agroforestry and Biodiversity ». In : *Sustainability* 11.10, p. 2879. DOI : 10.3390/su11102879.

PALMA, Raúl et al. (2016). « An INSPIRE-Based Vocabulary for the Publication of Agricultural Linked Data ». In : *Ontology Engineering*. Sous la dir. de Valentina. TAMMA et al. T. 9557. Information Systems and Applications, incl. Internet/Web, and HCI. Cham : Springer International Publishing et Imprint: Springer, p. 124-133. ISBN : 978-3-319-33245-1. DOI : 10.1007/978-3-319-33245-1_13.

PEIRCE, Charles Sanders (1958). *Collected Papers: Science and philosophy and Reviews, correspondence, and bibliography*. Belknap Press of Harvard University Press.

POLI, Roberto (1996). « Ontology for knowledge organization ». In : *Advances in Knowledge Organization* 5, p. 313-319. ISSN : 0938-5495.

PORPHYRE (1998). *Isagoge*. Sic et non. Paris : J. Vrin. ISBN : 2711613445.

POSADA, Raúl Hernando et al. (2018). « Effects of soil physical and chemical parameters, and farm management practices on arbuscular mycorrhizal fungi communities and diversities in coffee plantations in Colombia and Mexico ». In : *Agroforestry Systems* 92.2, p. 555-574. ISSN : 0167-4366.

Qu'est-ce que l'agriculture de conservation ? (2020). URL : http://www.fao.org/conservation-agriculture/overview/what-is-conservation-agriculture/fr/ (visité le 15/11/2020).

Qu'est-ce que l'agroécologie ? (2020). URL : https://agriculture.gouv.fr/quest-ce-que-lagroecologie (visité le 15/11/2020).

QUILLIAN, M. Ross (1967). « Word concepts: A theory and simulation of some basic semantic capabilities ». In : *Behavioral science* 12.5, p. 410-430. ISSN : 0005-7940.

— (1969). « The teachable language comprehender: A simulation program and theory of language ». In : *Communications of the ACM* 12.8, p. 459-476. ISSN : 0001-0782.

QUINKENSTEIN, Ansgar et al. (2009). « Ecological benefits of the alley cropping agroforestry system in sensitive regions of Europe ». In : *Environmental Science & Policy* 12.8, p. 1112-1121. ISSN : 14629011. DOI : 10.1016/j.envsci.2009.08.008.

RAJBHANDARI, Sachit et Johannes KEIZER (2012). « The AGROVOC concept scheme–a walkthrough ». In : *Journal of Integrative Agriculture* 11.5, p. 694-699. ISSN : 2095-3119.

RAJMAN, Martin et al. (1997). *Format de description lexicale pour le français. Partie 2: Description morpho-syntaxique*.

RANDELL, David A. et al. (1992). « A spatial logic based on regions and connection ». In : *KR* 92, p. 165-176.

RANDS, Michael R. W. et al. (2010). « Biodiversity conservation: challenges beyond 2010 ». In : *science* 329.5997, p. 1298-1303. ISSN : 0036-8075.

RDF Schema 1.1 (2017). URL : https://www.w3.org/TR/rdf-schema/#ch_reificationvocab (visité le 07/10/2020).

REMACLE, Philippe (2010). *Traité des parties des Animaux*. URL : http://remacle.org/bloodwolf/philosophes/Aristote/tableparties.htm (visité le 08/09/2020).

RESEARCH, STANFORD CENTER FOR BIOMEDICAL INFORMATICS (2021). *protégé*. URL : https://protege.stanford.edu/ (visité le 04/03/2021).

RIJGERSBERG, Hajo et al. (2013). « Ontology of units of measure and related concepts ». In : *Semantic Web* 4.1, p. 3-13. ISSN : 1570-0844. DOI : 10.3233/SW-2012-0069. URL : https://content.iospress.com/articles/semantic-web/sw069.

ROBERT, Mélanie et Claude. CHEVERRY (2009). *Le sol: une contrainte pour la sécurité alimentaire mondiale*. ÉDITIONS QUAE.

ROBERTS, Norman (1984). « The pre–history of the information retrieval thesaurus ». In : *Journal of Documentation*. ISSN : 0022-0418.

ROUSSEY, Catherine et al. (2010). « Ontologies in agriculture ». In : *AgEng 2010, International Conference on Agricultural Engineering*. Cemagref, p-p.

ROWE, E. C. et al. (1999). « Testing the safety-net role of hedgerow tree roots by 15 N placement at different soil depths ». In : *Agroforestry for Sustainable Land-Use Fundamental Research and Modelling with Emphasis on Temperate and Mediterranean Applications*. Springer, p. 81-93.

RUMBAUGH, James. et al. (2004). *The unified modeling language reference manual*. 2nd edition. The Addison-Wesley object technology series. Boston (Mass.) : Addison-Wesley. ISBN : 032171895X.

RUSSELL, Stuart J. et al. (2010). *Artificial intelligence: A modern approach*. 3rd ed. Prentice Hall series in artificial intelligence. Upper Saddle River : Prentice Hall. ISBN : 9780136042594.

SAJOUS, Franck et al. (2020). « ENGLAWI: From Human-to Machine-Readable Wiktionary ». In : *12th Conference on Language Resources and Evaluation (LREC 2020)*, p. 3016-3026.

SAUSSURE, Ferdinand de (2011). *Course in general linguistics*. Columbia University Press. ISBN : 0231157274.

SCHMIDT-SCHAUSS, Manfred et Gert SMOLKA (1991). « Attributive concept descriptions with complements ». In : *Artificial intelligence* 48.1, p. 1-26. ISSN : 0004-3702.

SHAH, Nigam H. et al. (2009). « Comparison of concept recognizers for building the Open Biomedical Annotator ». In : *BMC bioinformatics*. T. 10. Springer, p. 1-9.

SILANIKOVE, Nissim (2000). « Effects of heat stress on the welfare of extensively managed domestic ruminants ». In : *Livestock Production Science* 67.1-2, p. 1-18. ISSN : 03016226. DOI : 10.1016/S0301-6226(00)00162-7.

SINCLAIR, Fergus L. (1999). « A general classification of agroforestry practice ». In : *Agroforestry systems* 46.2, p. 161-180. ISSN : 0167-4366. DOI : 10.1023/A:1006278928088.

SIRIN, Evren, Bernardo Cuenca GRAU et al. (2006). « From Wine to Water: Optimizing Description Logic Reasoning for Nominals ». In : *KR*, p. 90-99.

SIRIN, Evren et Bijan PARSIA (2004). « Pellet: An owl dl reasoner ». In : *Proc. of the 2004 Description Logic Workshop (DL 2004)*, p. 212-213.

SIRIN, Evren, Bijan PARSIA et al. (2007). « Pellet: A practical owl-dl reasoner ». In : *Journal of web semantics* 5.2, p. 51-53. ISSN : 1570-8268.

SKOS Simple Knowledge Organization System Reference (2018). URL : https://www.w3.org/TR/skos-reference/#semantic-relations (visité le 25/02/2021).

SMITH, Barry (2000). « Objects and their environments: from Aristotle to ecological ontology ». In : *Life and motion of socio-economic units*. CRC Press, p. 84-102.

SMITH, Barry et al. (2007). « The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration ». In : *Nature biotechnology* 25.11, p. 1251-1255. ISSN : 1546-1696.

SOERGEL, Dagobert et al. (2004). « Reengineering thesauri for new applications: the AGROVOC example ». In : *Journal of digital information* 4, p. 1-23.

SOWA, John F. (1976). « Conceptual graphs for a data base interface ». In : *IBM Journal of Research and Development* 20.4, p. 336-357. ISSN : 0018-8646.
— (1984). *Conceptual structures: information processing in mind and machine*. Addison-Wesley.
— (1987). « Semantic networks ». In : *Citeseer*.
— (2008). « Conceptual graphs ». In : *Foundations of Artificial Intelligence* 3, p. 213-237. ISSN : 1574-6526.
— (2011). « Peirce's tutorial on existential graphs ». In : *Semiotica* 2011.186. ISSN : 1613-3692. DOI : 10.1515/semi.2011.060.
— (2000). *Knowledge representation: Logical, philosophical, and computational foundations*. Pacific Grove, CA : Brooks/Cole. ISBN : 0534949657.
SPARQL 1.1 Overview (2018). URL : https://www.w3.org/TR/sparql11-overview/ (visité le 04/02/2021).
STAMPS, W. T. et M. J. LINIT (1997). « Plant diversity and arthropod communities: Implications for temperate agroforestry ». In : *Agroforestry Systems* 39.1, p. 73-89. ISSN : 0167-4366.
STEFANO, Andrea de et Michael G. JACOBSON (2018). « Soil carbon sequestration in agroforestry systems: a meta-analysis ». In : *Agroforestry Systems* 92.2, p. 285-299. ISSN : 0167-4366.
TELENIUS, Anders (2011). « Biodiversity information goes public: GBIF at your service ». In : *Nordic Journal of Botany* 29.3, p. 378-381. ISSN : 0107-055X.
TESNIERE, Lucien (1959). *Eléments de syntaxe structurale, Klincksieck*.
TORQUEBIAU, Emmanuel F. (2000). « A renewed perspective on agroforestry concepts and classification ». In : *Comptes Rendus de l'Académie des Sciences - Series III - Sciences de la Vie* 323.11, p. 1009-1017. ISSN : 0764-4469. DOI : 10.1016/S0764-4469(00)01239-7. URL : http://www.sciencedirect.com/science/article/pii/S0764446900012397.
Total number of Websites - Internet Live Stats (2020). URL : https://www.internetlivestats.com/total-number-of-websites/ (visité le 01/10/2020).
TSARKOV, Dmitry et Ian HORROCKS (2003). « DL Reasoner vs. First-Order Prover ». In : *Description Logics* 81, p. 152-159.
TULLY, Katherine L. et al. (2012). « More trees less loss: nitrogen leaching losses decrease with increasing biomass in coffee agroforests ». In : *Agriculture, Ecosystems & Environment* 161, p. 137-144. ISSN : 0167-8809.
USDA, NRCS, PLANTS (2012). « The PLANTS database ». In : *National Plant Data Team, Greensboro, NC 27401–4901 USA*.
VALENCIA GARCÍA, Rafael (2008). « Un entorno para la extracción incremental de conocimiento desde texto en lenguaje natural ». Thèse de doct. URL : https://digitum.um.es/digitum/handle/10201/209.
VAN ASSEM, Mark et al. (2006). « A method to convert thesauri to SKOS ». In : *European Semantic Web Conference*. Springer, p. 95-109. (Visité le 2006).
VAN HEIJST, Gertjan et al. (1997). « Using explicit ontologies in KBS development ». In : *International Journal of Human-Computer Studies* 46.2-3, p. 183-292. ISSN : 1071-5819.
VARAH, Alexa et al. (2020). « Temperate agroforestry systems provide greater pollination service than monoculture ». In : *Agriculture, Ecosystems & Environment* 301, p. 107031. ISSN : 0167-8809.
VO, Minh Hoang Lien et Quang HOANG (2020). « Transformation of UML class diagram into OWL Ontology ». In : *Journal of Information and Telecommunication* 4.1, p. 1-16. ISSN : 2475-1839. DOI : 10.1080/24751839.2019.1686681.

W3C (2017). *Extensible Markup Language (XML) 1.0*. URL : https://www.w3.org/TR/1998/REC-xml-19980210 (visité le 15/10/2020).

W3C Mission (2017). URL : https://www.w3.org/Consortium/mission.html (visité le 01/10/2020).

WHITE, Jeffrey W. et al. (2013). « Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards ». In : *Computers and Electronics in Agriculture* 96, p. 1-12. ISSN : 01681699. DOI : 10.1016/j.compag.2013.04.003.

Wikipédia, l'encyclopédie libre (2021). URL : https://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Accueil_principal (visité le 05/02/2021).

YOUNG, Anthony (1989). « Agroforestry for soil conservation ». In.

ZACKLAD, Manuel (2007). « Classification, thésaurus, ontologies, folksonomies: comparaisons du point de vue de la recherche ouverte d'information (ROI) ». In : *35e Congrès annuel de l'Association Canadienne des Sciences de l'Information*.

— (2010). « Évaluation des systèmes d'organisation des connaissances ». In : *Les Cahiers du numérique* 6.3, p. 133-166. ISSN : 1622-1494.

ZHANG, Qiang et al. (2018). « Impact of tree species on barley rhizosphere-associated fungi in an agroforestry ecosystem as revealed by 18S rDNA PCR-DGGE ». In : *Agroforestry Systems* 92.2, p. 541-554. ISSN : 0167-4366.

ZOU, Youyong et al. (2004). « F-owl: An inference engine for semantic web ». In.

Liste des figures

2.1 Notice d'autorité de l'agroforesteriea de la liste d'autorité RAMEAU . . . 12

2.2 Un exemple de taxonomie. 13

2.3 Diagramme général pour la modélisation d'un thésaurus. 14

2.4 Exemple d'une photographie étiquetée avec Flickr. 15

2.5 Les différents modes de classification. 17

2.6 Modèle UML d'un thésaurus selon la norme ISO 25964. 19

2.7 Classification de P.K.R Nair: Catégorisation des systèmes agroforestiers en fonction de la nature des composants (avec des exemples de sous-systèmes/pratiques communs à chaque système). 21

2.8 AFAF: Extrait du thésaurus Agrof-MM. 22

2.9 Description du concept c_207 ("Agroforesterie"), sur le site web d'AGROVOC. 24

2.10 Description du concept c_26823 ("Arbuste"), visualisé sous forme de page HTML par Loddy. 26

2.11 Description du label ("Arbuste"), visualisée sous forme de page HTML par Loddy. 26

2.12 Définition des termes "Reduced bone mineral density" et "Spinal muscular atrophy" dans une approche de description « entité/qualité » avec PATO. 27

2.13 Définition des termes "Augmentation de la biodiversité" et "Bande enherbée" dans une approche de description « entité/qualité » avec PATO. 28

2.14 Freebase: Description du physicien Richard Feynman Les propriétés ("Gender", "Date of Birth",...) a) sont celles rattachées au type "Person" b) de la ressouce. Les propriétés non renseignées comme "Parents" et "Sibling" c) peuvent complétées par un contributeur. Lors de la saisie du frère de Richard, la ressource Joan Freynman d) qui est déjà saisie dans freebase est proposé au contributeur. 30

2.15 Résultat de la recherche Google sur Tom Gruber: a) Page HTML issue de la recherche par indexation ; b) Le "Knowledge Panel" affiche les données issues de la recherche sémantique: images, définition de la page Wikipédia, propriétés et valeurs ; c) La propriété "Date de naissance" et sa valeur rattachée à l'entité nommée <Tom Gruber>. 31

3.1 Quelques uns des services écosystémiques apportés par une parcelle agroforestière. 39

3.2 Distribution racinaire des arbres dans des conditions d'agroforesterie. Les arbres s'enracinent plus profondément pour échapper à la concurrence des cultures. 40

3.3 Parcelle agroforestière. Site des Eduts (André Gavaland – INRA Toulouse). 44

3.4 Schéma d'une parcelle agroforestière avec des cultures en bande. 45

4.1 Les trois signes de C. S. Pierce : Le symbole, l'icône et l'indice 51

4.2 le triangle sémiotique proposé par les linguistes britanniques Charles Kay Ogden et Ivor Armstrong Richards. 51

4.3 Relations de subsomption entre différents concepts. 53

5.1 Arbre de Porphyre : a) Gravure du livre "Institutiones philosophicae I" de Purchotius, 1730 b) Interprétation de l'arbre de Porphyre 58

5.2 Schéma d'un réseau Sémantique pour le concept d'arbre. 58

5.3 Schéma des trois niveaux emboités d'un cadre. 59

5.4 Exemples de cadres définissant différentes représentations d'arbres. . . . 60

5.5 Graphe conceptuel de la phrase: "Nicolas est allé à Restinclières en tracteur". 62

6.1 Relation entre IRI, URI, URL et URN : Les URI sont des cas particuliers des IRI. Les URI peuvent être des URL des URN ou bien les deux 73

6.2 Exemple d'un code XML. 74

6.3 structure d'un triplet RDF. 75

6.4 Triplet RDF sous la forme d'un graphe étiqueté et orienté. 75

6.5 Pseudo code décrivant Robert et l'agroforesterie. 76

6.6 Graphe informel décrivant Robert et l'agroforesterie. 76

6.7 Graphe informel décrivant l'adresse de Robert sous forme de relation N-aire. 76

6.8 Pseudo code décrivant Robert et l'agroforesterie avec l'utilisation d'URI. 77

6.9 Document RDF sérialisé en RDF/XML. 78

6.10 Document RDF sérialisé en Turttle. 79

6.11 Vision ensembliste des classes OWL. 81

6.12 Un exemple de graphe RDF. 85

6.13 La pyramide du web sémantique. 86

6.14 Blocs fonctionnels de la pyramide du web sémantique. 87

7.1 Extrait d'un fichier tabulé fourni par le bureau d'études Agroof. 94

7.2 Relations d'inclusion topologique et exemples d'aménagement 96

7.3 Gargantext : interface listant les résultats de la recherche des mots les plus fréquents extraits d'un corpus dédié à l'agroforesterie 98

7.4 Gargantext : interface illustrant la recherche du terme sol ("soil") et des locutions comprenant ce terme dans le corpus dédié à l'agroforesterie . . 99

7.5 Nuage de mots relatifs à l'agroforesterie extraits par l'application Gargantext. 99

7.6 Classification des systèmes agroforestiers selon P.K.R. Nair [NAIR, 1985]. 101

7.7 Classification des systèmes agroforestiers selon E. Torquebiau [TORQUEBIAU, 2000]. 102

7.8 The levels of the thesaurus system for the Agrof-MM KDB – Les cinq critères et les trois différents niveaux du thésaurus de l'AFAF [BURRIEL et al., 2017]. 103

7.9 Ensemble des activités de la chaîne de traitement 106

7.10 Site web AmapVoc présentant le thésaurus sur l'agroforesterie de L'AFAF.107

7.11 Interface web de l'annotateur du site AgroPortal. 108

7.12 Exemples des différents concepts retrouvés pour les termes de la page de description de l'agroforesterie sur le site Wikipédia via le service Annotator 109

7.13 Diagramme de séquence de la chaîne de traitement pour l'annotation . . 109

7.14 Structure/résultats en sortie de chaîne 111

7.15 Détails pour les termes annotés. 112

7.16 Termes annotés triés par nombre de concepts retrouvés 113

7.17 Exemple de termes non annotés. 114

7.18 Détails pour les ontologies connexes 115

7.19 Les différentes ontologies contenant les concepts retrouvés avec les termes issus de l'extraction avec GarganText (classées par nombre de concepts retrouvés). 116

7.20 Les différentes ontologies contenant les concepts retrouvés avec les termes issus du thésaurus Agrof-MM (classées par nombre de concepts retrouvés). 117

7.21 Termes du vocabulaire Agrof-MM sans correspondance 118

7.22 AgroforestryVoc: Schéma des différents types d'aménagements agroforestiers. 119

7.23 AgroforestryVoc: Extrait en langage Turtle. 121

7.24 AgroforestryVoc: Exemple d'une ontologie terminologique pour l'agroforesterie visualisé avec le logiciel Protégé: a) Le concept "Alignement régulier" est sélectionné ; b) Les termes utilisés pour le nommer sont fournis en français et en anglais ; c) Une définition est donnée pour ce concept ; d) Ce concept est subsumé par le concept "Aménagement intra-parcellaire" ; e) Ce concept subsume entre autres les concepts "Alignement continu" et "Alignement discontinu" 122

8.1 Démarche retenue pour la construction et la maintenance de l'ontologie "Agroforestry". 127

8.2 Conceptualisation de l'ontologie "Agroforestry". 130

8.3 Diagramme de classes autour du module ontologique SOSA. 132

8.4 Diagramme de classes autour du module ontologique cœur de GeoSPARQL. 133

8.5 Diagramme de classes autour du module ontologique OWL-Time. . . . 134

8.6 Diagramme de classes intégré. 135

8.7 Diagramme de paquetages autour de l'ontologique cadre " Agroforestry ".136

8.8 Diagramme d'instances illustrant le modèle cœur de l'ontologie cadre. . 138

8.9 Diagramme d'instances illustrant l'utilisation du modèle SOSA. 139

8.10 Diagramme d'instances illustrant l'utilisation du modèle GeoSPARQL. . 140

8.11 Concepts d'Agroforestry. 141

8.12 Propriétés d'Agroforestry. 142

8.13 Adossement aux ontologies de plus haut niveau. 142

8.14 Exemple de peuplement de la ABox 142

8.15 Extrait pour la constitution de l'ontologie cadre Agroforestry avec les langages RDF et OWL. les formules logiques transcrites sont présentées en encadré 144

8.16 Concept défini pour classer les parcelles agroforestières 145

8.17 Concept défini pour classer les parcelles agroforestières en disposition linéaire 145

8.18 Equivalence de classe pour le concept *LinearArrangementAgroforestryPlot* en syntaxe Turtle 145

8.19 **Moteur hybride Jena** : a) Le moteur de chaînage avant s'exécute et maintient un pool de données déduites . b) Si les règles de chaînage avant affirment des règles en chaînage arrière, celles ci seront instanciées préalablement par le moteur avant, puis transmises au moteur arrière. 147

8.20 Règle Jena apparentée à l'axiome OWL 2 AgroforestryPlot 147

9.1 Plan de la parcelle agroforestière A3 du domaine de Restinclières. 151

9.2 Plan de la parcelle témoin forestier A4 du domaine de Restinclières. . . . 152

9.3 Extrait du tableau des caractéristiques des arbres de la parcelle agroforestière A3. 153

9.4 Plan des zone d'échantillonnage de la parcelle agroforestière A3 du domaine de Restinclières. 154

9.5 Tableau des caractéristiques des échantillons issus de la récolte de blé dur de la parcelle A3 en 2017. 155

9.6 Diagramme de cas d'utilisation UML pour les aménagements agroforestiers. 158

9.7 Diagramme de cas d'utilisation UML pour les aménagements agroforestiers : Étude des cultures. 159

9.8 Diagramme de cas d'utilisation UML pour les aménagements agroforestiers : Étude de la croissance des arbres. 160

9.9 Diagramme d'instances : Description de la composition d'éléments du domaine de Restinclières. 161

9.10 Diagramme d'instances : Description des caractéristiques des éléments et typage de ces éléments et caractéristiques. 162

9.11 Diagramme d'instances : Expression d'une observation de terrain avec l'ontologie SOSA. Ici le diamètre d'un arbre 162

9.12 Diagramme d'instances : Description d'une relation sémantique «Facilitive interaction between trees and crops» entre une ligne d'arbres et une culture intercalaire de la parcelle A3. 163

9.13 Diagramme d'instances : Expression de la géolocalisation et de la date de plantation d'un arbre de la parcelle A3. 164

9.14 Diagramme RDF : Description et composition d'éléments d'une exploitation agroforestière. 166

9.15 Diagramme RDF : étiquetage d'un élément et d'une propriété par des termes d'une ontologie terminologique. 167

9.16 Diagramme RDF : Observation du diamètre d'un arbre avec SOSA. . . . 167

9.17 Diagramme RDF : Exemple d'un échantillonnage avec SOSA. 168

9.18 Diagramme RDF : Relation entre la parcelle et le site. 169

9.19 Diagramme RDF : Expression de la géolocalisation d'un arbre et d'une ligne d'arbres avec l'ontologie GeoSPARQL. 169

9.20 Diagramme RDF : Expression de la durée de vie d'un arbre avec l'ontologie Time. 170

9.21 Extrait de la base de connaissance Agroforestry dans la syntaxe Turtle . 172

9.22 Détails de l'individu PA3AFL01A10 (qui est le dixième arbre de la première ligne de la parcelle A3) de notre base de connaissance visualisés avec le logiciel Protégé. 173

9.23 Vue d'ensemble de l'architecture de l'application AOBRA. 174

9.24 Schéma du mécanisme d'inférence de l'application AOBRA. 175

9.25 Extrait en syntaxe Turtle de l'expression de la transitivité et de la réflexivité de la relation *BFO_0000051* et de son inverse *BFO_0000050*. 176

9.26 Exemple d'une inférence avec le moteur PELLET : Exploitation de la transitivité de la relation *BFO_0000051* et de sa relation inverse *BFO_0000050*.177

9.27 Exemple de la définition des classes **TreeForStudy** et **AgroforestryTreeForStudy** en syntaxe OWL Manchester. 178

9.28 Exemple d'une inférence avec le moteur PELLET : équivalence de concept appliquée à l'individu PA3AFL01A10. Les ellipses orange indiquent les relations et valeurs prises en compte pour inférer l'appartenance de l'individu PA3AFL01A10 à la classe **TreeForStudy**. Les ellipses verte indiquent les relations et valeurs prises en compte pour inférer l'appartenance de l'individu PA3AFL01A10 à la classe **AgroforestryTreeForStudy**. 179

9.29 Liste des espaces de noms exploités dans nos règles. 181

9.30 Algorithme d'identification des propriétés à étudier des différents éléments de la base de connaissance Agroforestry. 182

9.31 Règle Jena pour identifier les propriétés à étudier. 182

9.32 Requête SPARQL qui retourne les propriétés à étudier. 182

9.33 Résultat de l'extraction avec SPARQL des propriétés susceptibles d'être étudiées 183

9.34 Règle Jena pour la détermination des années à étudier. 183

9.35 Diagramme RDF pour identifier les valeurs de la caractéristique *afy : Tree_Trunk_Diameter* par année. 185

9.36 Règle Jena pour la création de nœuds anonymes pour concaténer les valeurs observées pour chaque arbre par parcelle, propriété étudiée et année. 186

9.37 Règle Jena pour la création de vecteurs de valeurs à partir de nœuds anonymes pour concaténer les valeurs observées. 187

9.38 Règle Jena pour le calcul de la moyenne des vecteurs de valeurs. 188

9.39 Diagramme RDF pour la création d'une moyenne des valeurs de la caractéristique *res : TreeTrunkDiameter* par année. 189

9.40 Extraction des résultats statistiques avec une requête SPARQL. La connaissance des propriétés et des années étudiées facilite l'identification des relations liant une parcelle au résultat d'une fonction statistique appliquée sur les valeurs de ces propriétés. 190

9.41 Mise en place d'une nouvelle observation sur la biomasse pour les arbres ayant été l'objet d'observations simultanées sur leur hauteur et leur diamètre pour une ou plusieurs des années étudiées. 192

9.42 Requête SPARQL pour l'extraction de la médiane des valeurs de biomasse des arbres d'une parcelle et son résultat. 193

9.43 Identification du plus proche voisin d'un arbre dans un rayon de 6,5 mètres. 194

9.44 Règle Jena pour déterminer au sein d'une parcelle le plus proche voisin d'un arbre dans une limite de 6,5 mètres. 194

9.45 Script SPARQL de la fonction utilisateur "calcNearestNeighbor" pour renvoyer l'arbre le plus proche. 194

9.46 Détermination de la présence d'autres arbres sur un axe créé à partir de deux arbres liés par la relation sémantique *hasCloseNeighbor*. La recherche est limitée sur une distance de 30 mètres avec une tolérance angulaire de 3 degrés 195

9.47 Règle Jena pour déterminer au sein d'une parcelle le plus proche voisin sur la même ligne d'un couple d'arbres voisins avec une distance maximale de 30 mètres et un écart angulaire de 3 degrés. 195

9.48 Script SPARQL de la fonction utilisateur "calcNeighborOfNeighbors" pour retourner l'arbre le plus proche d'un couple d'arbres selon l'axe sur lequel ce couple se place. 196

9.49 Règle Jena pour la création de lignes avec des groupes d'arbres voisins. . 196

9.50 Règle Jena pour affecter une géométrie à toutes les lignes d'une parcelle. 197

9.51 Règle Jena pour affecter une valeur d'azimut à toutes les lignes d'une parcelle via la relation *res:hasAzimut*. 197

9.52 Règle Jena pour la création d'orientations cardinales. 198

9.53 Règle Jena pour affecter à toute les lignes d'une parcelle possédant une valeur d'azimut, un direction cardinale via la relation *res:hasCardinalDirection*.198

9.54 Règle Jena pour la requalification des parcelles. 199

9.55 Résultat des moyennes des diamètres et hauteurs des troncs par parcelle pour l'année 2015. 200

9.56 Résultat des quartiles pour les diamètres et hauteurs des troncs par parcelle pour l'année 2015. 200

9.57 Résultat des quartiles pour les diamètres et hauteurs des troncs par parcelle pour l'année 2016. 200

9.58 Résultat des quartiles pour les diamètres et hauteurs des troncs par parcelle pour l'année 2017. 201

9.59 Règle Jena pour le calcul de la biomasse par catégorie d'arbres. 202

9.60 Règle Jena pour le calcul des ratios entre la biomasse des arbres d'une catégorie particulière et la biomasse totale des arbres d'une parcelle. . . 202

9.61 Ratio en pourcentage de la biomasse des arbres de catégorie I sur la biomasse totale des arbres par parcelle et années étudiées. 203

9.62 Ratio en pourcentage de la biomasse des arbres de catégorie II sur la biomasse totale des arbres par parcelle et années étudiées. 203

9.63 Ratio en pourcentage de la biomasse des arbres de catégorie III sur la biomasse totale des arbres par parcelle et années étudiées. 203

9.64 Moyennes des distances en mètres des arbres au Lez pour les catégories I,II et III en 2015. 204

9.65 Rendements de la culture de pois en 2016 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA. 204

9.66 Rendements de la culture de blé dur en 2017 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA. 205

9.67 Rendements de la culture d'orge en 2018 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA. 205

9.68 Rendements de la culture de pois en 2016 pour la parcelle agroforestière PA3AF et son témoin agricole PA3TA. 205

9.69 Performance en pourcentage du rendement de culture de blé dur en 2017 de la parcelle agroforestière PA3AF par rapport à son témoin agricole.206

9.70 Performance en pourcentage du rendement de culture d'orge en 2018 de la parcelle agroforestière PA3AF par rapport à son témoin agricole. . . . 206

9.71 Encombrement sur différents schémas d'aménagement parcellaire estimé avec des arbres ayant atteint la hauteur finale nécessaire à leur exploitation. 207

9.72 Encombrement sur une même parcelle estimé à différents niveaux de croissance des arbres 208

9.73 Indices d'encombrement IE_1, IE_2 et IE_3 de la parcelle agroforestière PA3AF en 2017. 208

9.74 Établissement d'une échelle de performance pour les cultures. 209

9.75 Établissement d'une échelle de performance pour la croissance des arbres. La mesure prise en compte est la différence des ratios de la biomasse des arbres de la catégorie la plus élevée par rapport à la biomasse totale pour une parcelle agroforestière et son témoin forestier 209

9.76 Score de la performance de la culture de blé dur en 2017 de la parcelle agroforestière PA3AF par rapport à son témoin agricole. 209

9.77 Score de la performance de la croissance des arbres de catégorie II en 2016 de la parcelle agroforestière PA3AF par rapport à son témoin forestier. 210

9.78 Score final de la performance en 2016 de la parcelle agroforestière PA3AF par rapport à son témoin agricole. 210

10.1 Interface web de la base de données AFT de l'ICRAF. 217

10.2 Illustration d'une application pour l'exploitation des données liées. . . . 220

10.3 Illustration d'une plateforme d'applications modulaires pour l'exploitation des données agroforestières. **a)** Constitution d'un tableau 'sémantisé' adapté aux données que souhaite saisir l'utilisateur. **b)** Une fois les données saisies, elles peuvent être transmises à la plateforme et intégrées sous format RDF. **c)** Lors des mécanismes d'inférences, des modèles externes peuvent être utilisés. **d)** Les données natives et inférées de la plateforme sont consultables via une interface web. **e)** Les données sont aussi accessibles à d'autres applications sur le web. 222

A.1 Plan de répartition des arbres de la parcelle agroforestière A3. 256

A.2 Plan de répartition des arbres du témoin forestier A4. 257

B.1 Règle Jena pour la création d'une observation sur la biomasse d'un arbre. Une observation est créée pour chaque arbre de chaque parcelle chaque fois que ceux-ci sont liés a une observation sur leur taille (concept CO_357:1000042 de Crops Ontology) et une autre sur leur diamètre (concept c_34790 d'Agrovoc) effectuées à la même période. 260

B.2 Règle Jena pour la création de catégories utilisées pour le négoce du bois. 261

B.3 Les différentes catégories pour le négoce du bois et leurs valeurs minimales et maximales pour leur critère de diamètre. 262

B.4 Règle Jena pour la création d'une observation sur la catégorie d'un arbre. **a)** Pour toutes les parcelles, toutes les années étudiées et toutes les catégories. **b)** Pour tous les éléments d'une parcelle ayant une propriété clarifiée par le terme agrovoc:c_34790 (diamètre à hauteur de poitrine). **c)** Retour de la valeur numérique issue de l'observation de cette propriété et des valeurs numériques pour les critères de diamètre minimum et maximum d'une catégorie. **d)** Comparaison du diamètre observé pour l'élément avec les critères de la catégorie. **e)** Comparaison de l'année de l'observation avec une année étudiée. **f)** Préparation d'URI pour une nouvelle observation concernant la nouvelle propriété *hasCategory*. **g)** Constitution d'une nouvelle observation sur la propriété *hasCategory* de l'élément avec la même estampille temporelle que l'observation pris en compte pour le diamètre. 263

B.5 Règle Jena pour la création d'une observation sur la base d'une observation précédente et sur la catégorie d'un arbre. **a)** Pour toutes les parcelles, tous les éléments ayant une propriété à traiter par catégorie d'arbre, toutes les années étudiées et toutes les catégories. **b)** Pour tous les éléments possédant une observation sur une propriété à traiter réalisée au cours d'une année étudiée. **c)** Si cet élément possède aussi une observation sur sa catégorie pour la même année étudiée. **d)** Préparation d'URI pour une nouvelle observation concernant la nouvelle propriété *"NomDeLaPropriété"ByCategory*. **e)** Constitution d'une nouvelle observation avec la même estampille temporelle que l'observation pris en compte pour la propriété traitée. 265

B.6 Agrégation des valeurs de la propriété *hasBiomasse* par catégorie d'arbre pour l'année 2015. 266

B.7 Règle Jena pour la création d'une liste de concepts subsumés par la notion d'un terme prédéfini. 267

B.8 Script SPARQL pour la création d'un terme qui sera utilisé pour la création d'une liste de concepts. 268

B.9 Liste des concepts subsumés par le concept ayant pour terme 'Juglandaceae'. 269

B.10 Schéma de l'exploitation par AOBRA de vocabulaires et de taxonomies accessible sur le web via des services web pour la constitution de listes de concepts 269

B.11 Script SPARQL pour la recherche parmi tous les éléments d'une parcelle de ceux qui s'apparentent au concept de 'Juglandaceae'. Les concepts (?concept) de la liste "listNarrowerConceptsOfJuglandaceae" sont comparés au concept clarifiant des éléments (individu de la classe afy:Term relié par la relation *dcterms:type*). Les différents éléments d'une parcelle sont identifiés via la relation *bfo:partOf* qui les lie à celle-ci. 270

B.12 Résultat de la recherche des parcelles contenant des éléments identifié par le concept 'Juglandaceae' ou l'un des concepts qu'il subsume. 270

Liste des tableaux

5.1 Tableau des constructeurs et de leur syntaxe pour la logique $\mathcal{AL}$. 65

5.2 Constructeurs de rôles et concepts et contraintes sur l'interprétation des rôles pour L'extension de la logique $\mathcal{AL}$ 65

5.3 Sémantique des constructeurs 66

Annexes

Amménagements du domaine de Restinclières

La figure A.1 montre le plan de répartition de la parcelle A3 du domaine de Restinclières.
Les arbres absents (morts) au moment de la mesure (année 2004) sont indiqués en grisé.

La figure A.2 montre la répartition des arbres de la parcelle A4, témoin forestier de la parcelle A3.

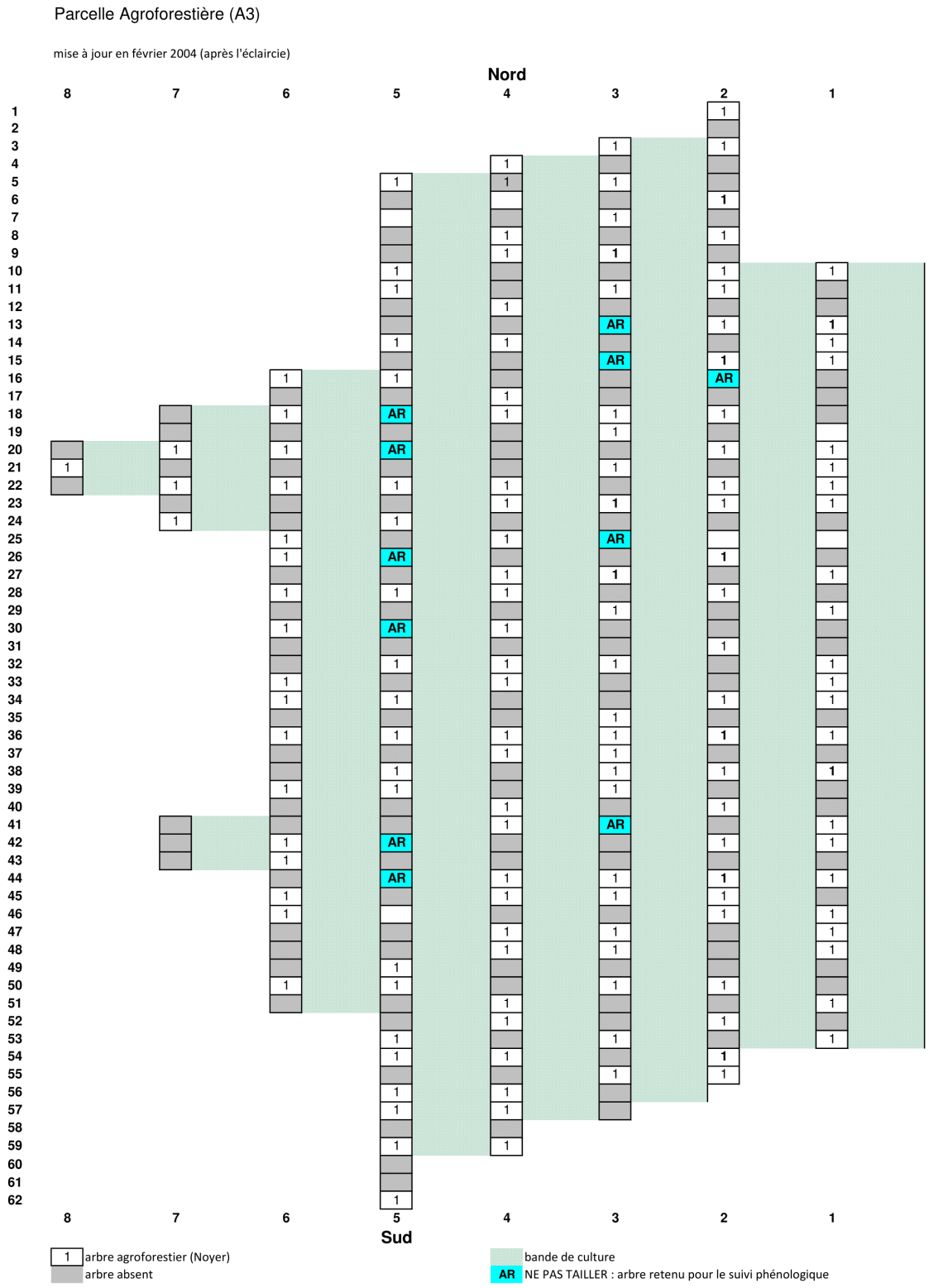


FIGURE A.1 – *Plan de répartition des arbres de la parcelle agroforestière A3.*

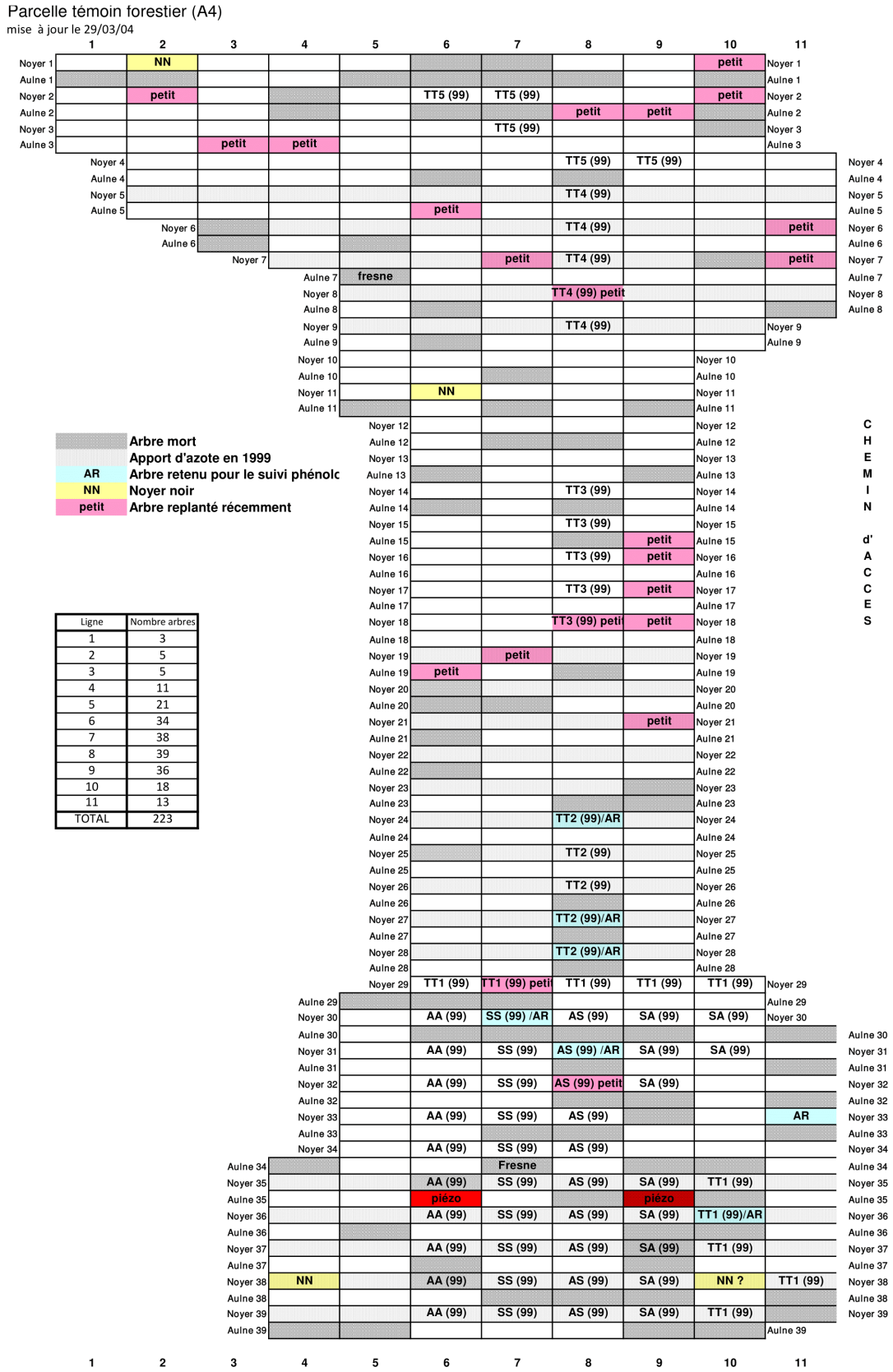


FIGURE A.2 – *Plan de répartition des arbres du témoin forestier A4.*

Échantillon de règles JENA utilisées avec l'application AOBRA

B.1 Création de nouvelles observations sur la base d'observations initiales

La règle Jena suivante permet de calculer pour tous les arbres d'une parcelle, ayant fait l'objet d'une observation sur leur taille et leur diamètre au cours des années étudiées, leur biomasse. Cette valeur de biomasse obtenue pour chaque arbre sera elle aussi consignée sous la forme d'une observation de la valeur de cette nouvelle propriété *res:hasBiomass* :

- **a)** Pour toutes les parcelles, tous les éléments contenus dans chaque parcelle et toutes les années étudiées.
- **b)** Chaque élément doit être l'élément d'intérêt d'une observation sur la valeur numérique d'une propriété clarifiée par le terme crop:CO_357:1000042 (Height trunk) au cours de l'année traitée.
- **c)** Cet élément doit aussi être l'élément d'intérêt d'une observation sur la valeur numérique d'une propriété clarifiée par le terme agrovoc:c_34790 (breast height diameter) au cours de la même année.
- **d)** Le calcul de la biomasse de cet élément se fait alors par une fonction utilisateur "calcTreeBiomass" utilisant les valeurs numériques recueillies du diamètre et de la hauteur du tronc pour restituer une valeur numérique pour la biomasse.
- **e)** Préparation des URI pour la constitution des triplets exprimant cette nouvelle observation.
- **f)** Écriture de cette nouvelle observation concernant cette propriété *res:hasBiomass*. Cette propriété sera typée par le terme agrovoc:c_926 (biomass).

```
[ CalculBiomass:
    (?plot :isPlot 'true'^^xsd:Boolean)
    (?structuralElement bfo:partOf ?plot)                    a)
    (?yearForStudy rdf:type :YearForStudy)
    (?yearForStudy :hasYear ?yearStudied)
    (?structuralElement afy:hasProperty ?propertyOfElementHeightTrunk)
    (?propertyOfElementHeightTrunk dcterms:type crop:CO_357:1000042)
    (?structuralElement sosa:isFeatureOfInterestOf ?observationHeight)      b)
    (?observationHeight sosa:observedProperty ?propertyOfElementHeightTrunk)
    (?observationHeight sosa:hasResult ?measureHeightTrunk)
    (?measureHeightTrunk om:hasNumericalValue ?heightTrunk)
    (?observationHeight sosa:phenomenonTime ?timeHeightTrunk)
    (?timeHeightTrunk time:inXSDgYear ?yearStudied)
    (?structuralElement afy:hasProperty ?propertyOfElementDiameter)
    (?propertyOfElementDiameter dcterms:type agrovoc:c_34790)               c)
    (?structuralElement sosa:isFeatureOfInterestOf ?observationDiameter)
    (?observationDiameter sosa:observedProperty ?propertyOfElementDiameter)
    (?observationDiameter sosa:hasResult ?measureDiameter)
    (?measureDiameter om:hasNumericalValue ?diameter)
    (?observationDiameter sosa:phenomenonTime ?timeDiameter)
    (?timeDiameter time:inXSDgYear ?yearStudied)
    calcTreeBiomass(?diameter, ?heightTrunk, ?biomass)       d)
    (:thisOntology :hasNameSpace ?ns)
    uriConcat(?structuralElement,"_Obs_",?yearStudied,"_Biomass",?uRIObservation)
    uriConcat(?ns, "hasBiomass",?UriBiomassProperty)                 e)
    uriConcat(?structuralElement,"_Obs_",?yearStudied,"_Biomass_Time",?uRItime)
    uriConcat(?structuralElement,"_Obs_",?yearStudied,"_Biomass_Measure",
      ?uRImeasure)
  ->
    (?structuralElement afy:hasProperty ?UriBiomassProperty)
    (?uRIObservation sosa:hasFeatureOfInterest ?structuralElement)
    (?structuralElement sosa:isFeatureOfInterestOf ?uRIObservation)
    (?uRIObservation sosa:hasResult ?uRImeasure)
    (?uRImeasure om:hasNumericalValue ?biomass)                          f)
    (?uRIObservation sosa:observedProperty ?UriBiomassProperty)
    (?UriBiomassProperty dcterms:type agrovoc:c_926)
    (?uRIObservation sosa:phenomenonTime ?uRItime)
    (?uRItime time:inXSDgYear ?yearStudied)
    (?uRIObservation sosa:madeBySensor :AOBRA)
]
```

FIGURE B.1 – *Règle Jena pour la création d'une observation sur la biomasse d'un arbre. Une observation est créée pour chaque arbre de chaque parcelle chaque fois que ceux-ci sont liés a une observation sur leur taille (concept CO_357:1000042 de Crops Ontology) et une autre sur leur diamètre (concept c_34790 d'Agrovoc) effectuées à la même période.*

B.2 Création de nouvelles observations sur la base de critères imposés

Dans l'exemple suivant des observations "SOSA" sont créées sur la base d'observations existantes et l'application de critères nouvellement imposés au système.

Pour le négoce du bois, les arbres vendus sont classés par catégories établies principalement en fonction d'un critère de diamètre. On commencera donc par intégrer ces catégories à la base de connaissance avec cette règle (voir figure B.2) dans laquelle chaque catégorie est liée par la relation *hasMinValue* à la valeur minimale de son critère de diamètre et par la relation *hasMaxValue* à la valeur maximale de son critère de diamètre. On obtient donc les quatre catégories visibles en figure B.3.

La catégorie à laquelle appartient un arbre étant dépendante de son diamètre, elle évoluera en fonction du temps et il a été choisi de la considérer comme la valeur d'une propriété observée nommée *hasCategory*. La règle montrée en figure B.4 a pour but de mettre en place pour chaque arbre de chaque parcelle une observation sur la catégorie chaque fois qu'une observation est faite sur le diamètre de ces arbres. La valeur de cette caractéristique *hasCategory* sera la catégorie ou la valeur observée du diamètre satisfera les critères de diamètre minimum et maximum et la date de cette nouvelle observation sera celle de l'observation faite sur le diamètre.
Chaque arbre qui possédera alors une observation sur sa catégorie pour une année donnée sera éligible au commerce du bois.

```
[CreateCategories:
        ->
        (:categoryI :hasLocalName "categoryI")
        (:categoryI dcterms:type :term_Category)
        (:categoryI :hasMinValue 20.0)
        (:categoryI :hasMaxValue 25.0)

        (:categoryII :hasLocalName "categoryII")
        (:categoryII dcterms:type :term_Category)
        (:categoryII :hasMinValue 25.0)
        (:categoryII :hasMaxValue 35.0)

        (:categoryIII :hasLocalName "categoryIII")
        ...
]
```

FIGURE B.2 – *Règle Jena pour la création de catégories utilisées pour le négoce du bois.*

Catégories pour le négoce du bois		
Catégorie	Valeur minimale incluse (cm)	Valeur maximale exclue (cm)
<http://www.afy.fr/Restinclieres/categoryI>	20	25
<http://www.afy.fr/Restinclieres/categoryII>	25	35
<http://www.afy.fr/Restinclieres/categoryIII>	35	45
<http://www.afy.fr/Restinclieres/categoryIV>	45	55

FIGURE B.3 – *Les différentes catégories pour le négoce du bois et leurs valeurs minimales et maximales pour leur critère de diamètre.*

```
[IdentificationOfTreeCategory:
  (?plot :isPlot 'true'^^xsd:Boolean)
  (?yearForStudy rdf:type :YearForStudy)                    a)
  (?yearForStudy :hasYear ?yearStudied)
  (?category dcterms:type :term_Category)

  (?structuralElement bfo:partOf ?plot)
  (?structuralElement afy:hasProperty ?propertyOfElementDiameter)          b)
  (?propertyOfElementDiameter dcterms:type agrovoc:c_34790)

  (?structuralElement sosa:isFeatureOfInterestOf ?observationDiameter)
  (?observationDiameter sosa:observedProperty ?propertyOfElementDiameter)
  (?observationDiameter sosa:hasResult ?measureDiameter)
  (?category :hasMinValue ?DiameterCriterionMin)                  c)
  (?category :hasMaxValue ?DiameterCriterionMax)
  (?measureDiameter om:hasNumericalValue ?diameter)

  greaterThan(?diameter ?DiameterCriterionMin)                d)
  le(?diameter ?DiameterCriterionMax)

  (?observationDiameter sosa:phenomenonTime ?timeDiameter)        e)
  (?timeDiameter time:inXSDgYear ?yearStudied)

  (:thisOntology :hasNameSpace ?ns)
  uriConcat(?structuralElement,"_Obs_",?yearStudied,"_Category",?uRIObservation)
  uriConcat(?ns,"hasCategory", uRIProperty)                   f)
  uriConcat(?structuralElement,_Obs_",?yearStudied,"_Category","_Time",?uRItime)
  (?structuralElement afy:hasProperty uRIProperty)
  ->
  (?structuralElement afy:hasProperty uRIProperty)
  (?uRIObservation sosa:hasFeatureOfInterest ?structuralElement)
  (?structuralElement sosa:isFeatureOfInterestOf ?uRIObservation)
  (?uRIObservation sosa:hasResult ?category)                                g)
  (?uRIObservation sosa:observedProperty uRIProperty)
  (uRIProperty dcterms:type sio:SIO_000137)
  (?uRIObservation sosa:phenomenonTime ?uRItime)
  (?uRItimeCategory time:inXSDgYear ?year)
  (?uRIObservation sosa:madeBySensor :AOBRA)
]
```

FIGURE B.4 – *Règle Jena pour la création d'une observation sur la catégorie d'un arbre.* ***a)*** *Pour toutes les parcelles, toutes les années étudiées et toutes les catégories.* ***b)*** *Pour tous les éléments d'une parcelle ayant une propriété clarifiée par le terme agrovoc:c_34790 (diamètre à hauteur de poitrine).* ***c)*** *Retour de la valeur numérique issue de l'observation de cette propriété et des valeurs numériques pour les critères de diamètre minimum et maximum d'une catégorie.* ***d)*** *Comparaison du diamètre observé pour l'élément avec les critères de la catégorie.* ***e)*** *Comparaison de l'année de l'observation avec une année étudiée.* ***f)*** *Préparation d'URI pour une nouvelle observation concernant la nouvelle propriété* hasCategory. ***g)*** *Constitution d'une nouvelle observation sur la propriété* hasCategory *de l'élément avec la même estampille temporelle que l'observation pris en compte pour le diamètre.*

B.3 Création de nouvelles observations sur la base d'une agrégation en fonction de la valeur observée d'une propriété particulière

Afin d'étudier la corrélation entre différentes caractéristiques d'éléments d'une population, il est nécessaire d'étudier l'ensemble des valeurs d'une propriété sur la base d'une autre caractéristique commune aux éléments étudiés.
La règle suivante montrée en figure B.5 traite toutes les propriétés (identifiées par leur relation *isPropertyToStudyByCategory*) de tous les éléments de toutes les parcelles en créant pour chacune une nouvelle propriété observée dont l'URI sera composé du nom de la propriété et de la catégorie de l'arbre observée (telle que nous l'avons défini dans l'exemple précédent) au moment de l'observation de cette propriété.
Dans le schéma visible en figure B.6 on visualise le résultat issu de l'application de cette règle avec la propriété *hasBiomass* qui donne naissance à une nouvelle observation sur une nouvelle propriété *hasBiomassByCategoryX*, "CategoryX" étant la catégorie observée de l'arbre pour la même année "2015".
Le choix de recréer une observation SOSA pour chaque élément permet le traitement automatisé de ces nouvelles observations par les règles que nous avons évoquées précédemment pour l'agrégation des valeurs. Le résultat final sera donc la mise à disposition au niveau de la parcelle de nouvelles propriétés comme dans cet exemple *hasBiomassByCategoryIISum2015* qui lie notre parcelle 'Plot1' à une valeur représentant la somme totale des biomasses observées pour les arbres de catégorie II de cette parcelle pour l'année 2015.

```
[PropertyValueByCategory:
  (?plot :isPlot 'true'^^xsd:Boolean)
  (?structuralElement bfo:partOf ?plot)
  (?propertyStudied :isPropertyToStudyByCategory 'true'^^xsd:Boolean)
  (?propertyStudied :hasLocalName ?propertyName)
  (?structuralElement afy:hasProperty ?propertyStudied)
  (?yearForStudy rdf:type :YearForStudy)                    a)
  (?yearForStudy :hasYear ?yearStudied)
  (?categoryStudied dcterms:type :term_Category)
  (?categoryStudied :hasLocalName ?categoryName)
   (:thisOntology :hasNameSpace ?ns)
  (?structuralElement sosa:isFeatureOfInterestOf ?observationProperty)
  (?observationProperty sosa:observedProperty ?propertyStudied)
  (?observationProperty sosa:hasResult ?measureProperty)
  (?measureProperty om:hasNumericalValue ?value)            b)
  (?observationProperty sosa:phenomenonTime ?timeProperty)
   (?timeProperty time:inXSDgYear ?yearStudied)
  uriConcat(?ns,"hasCategory", ?uRICategoryProperty)
  (?structuralElement afy:hasProperty ?UriCategoryProperty)
  (?structuralElement sosa:isFeatureOfInterestOf ?observationCategory)
  (?observationCategory sosa:observedProperty ?UriCategoryProperty)
  (?observationCategory sosa:hasResult ?categoryStudied)          c)
  (?observationCategory sosa:phenomenonTime ?timeCategory)
   (?timeCategory time:inXSDgYear ?yearStudied)
  uriConcat(?structuralElement,"_Obs_",?propertyName,"By",?categoryName,
      ?yearStudied,?uRIObservation)
  uriConcat(?ns,?propertyName,"By",?categoryName,?UriProperty)          d)
  uriConcat(?structuralElement,"_Obs_",?propertyName,"By",?categoryName,
      ?yearStudied,"_Time",?uRItime)
  uriConcat(?structuralElement,"_Obs_",?propertyName,"By",?categoryName,
      ?yearStudied,"_Measure",?uRImeasure)
  ->
  (?structuralElement afy:hasProperty ?UriProperty)
  (?uRIObservation sosa:hasFeatureOfInterest ?structuralElement)
  (?structuralElement sosa:isFeatureOfInterestOf ?uRIObservation)
  (?uRIObservation sosa:observedProperty ?UriProperty)
  (?uRIObservation sosa:hasResult ?uRImeasure)                  e)
  (?uRImeasure om:hasNumericalValue ?value)
  (?uRIObservation sosa:phenomenonTime ?uRItime)
  (?uRItime time:inXSDgYear ?yearStudied)
  (?uRIObservation sosa:madeBySensor :AOBRA)
]
```

FIGURE B.5 – *Règle Jena pour la création d'une observation sur la base d'une observation précédente et sur la catégorie d'un arbre.* ***a)*** *Pour toutes les parcelles, tous les éléments ayant une propriété à traiter par catégorie d'arbre, toutes les années étudiées et toutes les catégories.* ***b)*** *Pour tous les éléments possédant une observation sur une propriété à traiter réalisée au cours d'une année étudiée.* ***c)*** *Si cet élément possède aussi une observation sur sa catégorie pour la même année étudiée.* ***d)*** *Préparation d'URI pour une nouvelle observation concernant la nouvelle propriété* "NomDeLaPropriété"ByCategory. ***e)*** *Constitution d'une nouvelle observation avec la même estampille temporelle que l'observation pris en compte pour la propriété traitée.*

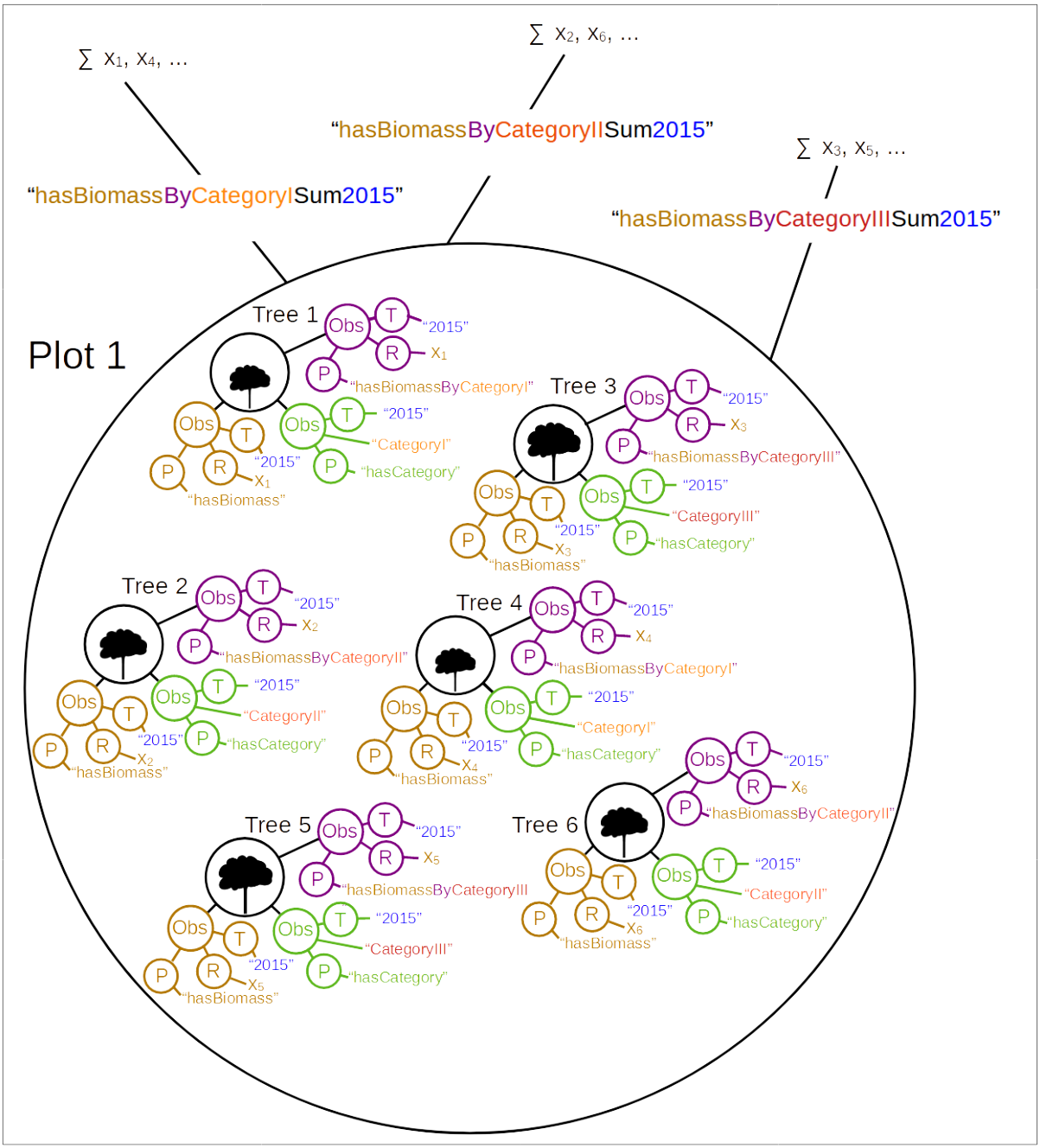


FIGURE B.6 – *Agrégation des valeurs de la propriété* hasBiomasse *par catégorie d'arbre pour l'année 2015.*

B.4 Autres possibilités de traitements

B.4.1 Utilisation de la taxonomie

Comme nous l'avons vu au chapitre 3, le choix des différentes espèces végétales (arbres et cultures) utilisées dans un aménagement agroforestier est essentiel à l'obtention d'un système performant. Les espèces choisies doivent non seulement être adaptées aux conditions pédoclimatiques de la parcelle mais aussi être compatibles entre elles afin de maximiser les synergies mises en place par leur rapprochement spatial tout en diminuant les effets de concurrence qu'elles pourraient entretenir.
L'exemple suivant montre la façon avec laquelle il nous est possible d'accéder à l'identification taxonomique des espèces présentes sur une parcelle en utilisant des vocabulaires spécialisés comme Agrovoc.
Agrovoc en utilisant le modèle SKOS permet de monter ou descendre un arbre hiérarchique en utilisant des relations "plus large que" et "plus étroit que" (voir chapitre 2, section 2.2.2) ce qui va nous permettre de rechercher une espèce présente sur une parcelle à partir d'un rang taxonomique plus élevé.
Dans une première étape on constitue une règle JENA (voir figure B.7) permettant de créer une liste de concepts "plus étroit que" à partir du label d'un élément clarifié par le terme "term_ForMakeNarrowerList". La fonction utilisateur nommée calcMakeListNarrowerConceptsWithTerm() est en charge de récupérer à partir d'un libellé les différents concepts se rapportant à ce terme ainsi que les concepts subsumés par ces premiers concepts en utilisant les web services mis à disposition par le vocabulaire Agrovoc et le portail Agroportal. Les concepts récupérés seront stockés sur forme d'une liste RDF identifiée par la concaténation des mots "listNarrowerConceptsOf" et du terme recherché.
Dans une deuxième étape (voir figure B.8) on procède à l'insertion dans le modèle

```
[MakeListNarrowerConceptsFromTerm :
    (?term dcterms:type :term_ForMakeNarrowerList)
    (?term :hasLabelForMakeList ?label)
    ->
    calcMakeListNarrowerConceptsWithTerm(?label)
]
```

FIGURE B.7 – *Règle Jena pour la création d'une liste de concepts subsumés par la notion d'un terme prédéfini.*

d'un terme qui sera traité par la règle vue précédemment. Le terme dans cet exemple est 'Juglandaceae', une famille taxonomique regroupant des plantes dicotylédones en englobant une cinquantaine d'espèces réparties en 8 à 11 genres. La figure B.9 montre

une partie des 175 concepts de la liste RDF "listNarrowerConceptsOfJuglandaceae" générée par l'introduction du terme dans le système. On y retrouve le concept du vocabulaire Agrovoc http://aims.fao.org/aos/agrovoc/c_4055 désigné par le label 'Juglandaceae' mais aussi les concepts http://id.nlm.nih.gov/mesh/D029525 de l'ontologie médicale MeSH (Librairie Nationale de Médecine aux États Unis) et http://lod.nal.usda.gov/nalt/22731 du thésaurus NAL (Département national de l'agriculture américaine) qui désigne le même concept. D'autres concepts comme http://aims.fao.org/aos/agrovoc/c_4060 désigné par le terme 'Juglans regia' proviennent de la hiérarchie descendante des premiers concepts. En effet le concept 'Juglans regia' réfère au nom de l'espèce du noyer commun du genre 'Juglans' (concept http://aims.fao.org/aos/agrovoc/c_4057 d'Agrovoc) de la famille des 'Juglandaceae'. la figure B.10 résume le processus de création de ces listes de concepts par AOBRA. L'accès aux différents vocabulaires sur le web se fait au moyen de services web fournis par des vocabulaires comme Agrovoc ou des portails de vocabulaires comme NCBO Bioportal regroupant plusieurs vocabulaires (NCBITaxon, MeSH, ...). AOBRA peut ainsi constituer des listes de concepts (concepts équivalents ; concepts proches ; concepts "plus étroit que" ...). Ces listes au format RDF qui regroupent les concepts désignés par un terme particulier pourront être utilisées ultérieurement par l'application pour par exemple identifier des éléments particuliers ou similaires au sein de la base de connaissance à partir de ce terme. Dans une troisième étape, on peut donc maintenant recher-

```
INSERT DATA {
    res:termToListJuglandaceae dcterms:type res:term_ForMakeNarrowerList .
    res:termToListJuglandaceae res:hasLabelForMakeList 'Juglandaceae' .
} ;
```

FIGURE B.8 – *Script SPARQL pour la création d'un terme qui sera utilisé pour la création d'une liste de concepts.*

cher dans notre base de connaissance les parcelles contenant des spécimens de la famille des 'Juglandaceae' en exploitant la liste de concepts créée. Le script SPARQL en figure B.11) montre la façon dont on exploite la relation *bfo:partOf* pour cibler les éléments de la parcelle et la relation *dcterms:type* qui lie ces mêmes éléments aux termes qui les clarifient.

Les résultats de cette requête SPARQL sont visibles en figure B.12 qui indique que l'on retrouve bien la parcelle agroforestière PA3AF et son témoin forestier PA4TF du fait de la présence en leur sein d'éléments clarifié par le concept http://aims.fao.org/aos/agrovoc/c_4060 ('Juglans regia').

Liste des concepts	
Concept	PrefLabel
<http://aims.fao.org/aos/agrovoc/c_1348>	Carya
<http://aims.fao.org/aos/agrovoc/c_1349>	Carya pecan
<http://aims.fao.org/aos/agrovoc/c_4055>	Juglandacee
<http://aims.fao.org/aos/agrovoc/c_4057>	Juglans
<http://aims.fao.org/aos/agrovoc/c_4058>	Juglans cinerea
<http://aims.fao.org/aos/agrovoc/c_4059>	Juglans nigra
<http://aims.fao.org/aos/agrovoc/c_4060>	Juglans regia
<http://aims.fao.org/aos/agrovoc/c_a7f88329>	Pterocarya fraxinifolia
<http://aims.fao.org/aos/agrovoc/c_b7471669>	Pterocarya
<http://d-nb.info/gnd/4189026-7>	Juglans regia
<http://d-nb.info/gnd/4230377-1>	Juglans nigra
<http://d-nb.info/gnd/4294466-1>	Carya illinœnsis
<http://id.nlm.nih.gov/mesh/D029525>	Juglandaceae
<http://lod.nal.usda.gov/nalt/17825>	Juglans nigra
<http://lod.nal.usda.gov/nalt/20642>	Juglans cinerea
<http://lod.nal.usda.gov/nalt/22730>	Carya
<http://lod.nal.usda.gov/nalt/22731>	Juglandaceae
<...>	

FIGURE B.9 – *Liste des concepts subsumés par le concept ayant pour terme 'Juglandaceae'.*

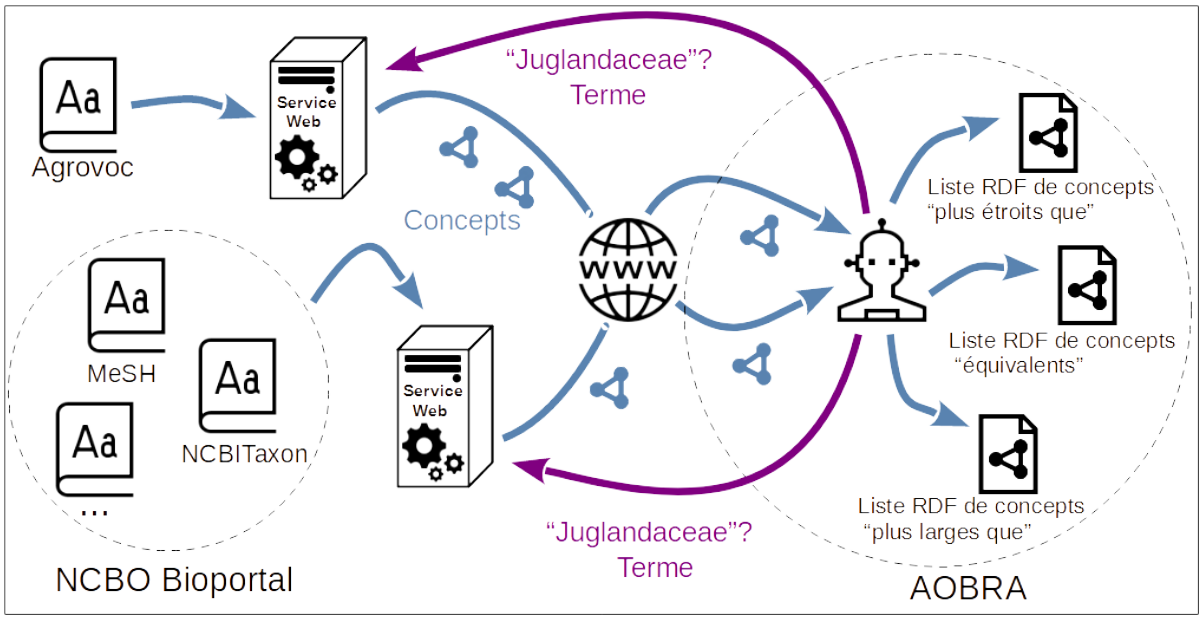


FIGURE B.10 – *Schéma de l'exploitation par AOBRA de vocabulaires et de taxonomies accessible sur le web via des services web pour la constitution de listes de concepts*

```
SELECT DISTINCT (?plot AS ?Parcelle) (?concept AS ?Concept)
WHERE {
      ?plot res:isAFPlot ?true .
      res:listNarrowerConceptsOfJuglandaceae rdf:rest*/rdf:first ?concept .
      ?structuralElement bfo:partOf ?plot. .
      ?structuralElement dcterms:type ?concept .
} ORDER BY ?plot ?concept ;
```

FIGURE B.11 – *Script SPARQL pour la recherche parmi tous les éléments d'une parcelle de ceux qui s'apparentent au concept de 'Juglandaceae'. Les concepts (?concept) de la liste "listNarrowerConceptsOfJuglandaceae" sont comparés au concept clarifiant des éléments (individu de la classe afy:Term relié par la relation* dcterms:type*). Les différents éléments d'une parcelle sont identifiés via la relation* bfo:partOf *qui les lie à celle-ci.*

Parcelle contenant des 'Juglandaceae'	
Parcelle	Concept
<http://www.afy.fr/Restinclieres/PA3AF>	<http://aims.fao.org/aos/agrovoc/c_4060>
<http://www.afy.fr/Restinclieres/PA4TF>	<http://aims.fao.org/aos/agrovoc/c_4060>

FIGURE B.12 – *Résultat de la recherche des parcelles contenant des éléments identifié par le concept 'Juglandaceae' ou l'un des concepts qu'il subsume.*