UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Le marché de la publication scientifique à l'ère du numérique : une analyse économique de l'*Open Access*

THÈSE

présentée et soutenue publiquement le 6 décembre 2021
pour l'obtenition du grade de

Docteur de l'Université de Lorraine
en Sciences Economiques
par
Jaele Bechis

Composition du jury

Rapporteurs :	Thierry Pénard	Professeur, Université de Rennes 1
	Fabrice Rochelandet	Professeur, Université Sorbonne Nouvelle
Examinateurs :	Julien Pénin	Professeur, Université de Strasbourg
	Dominique Torre	Professeur émérite, Université Côte d'Azur
Président du jury :	Myriam Doriat-Duban	Professeure, Université de Lorraine
Directrice de thèse :	Sophie Harnay	Professeure, Université Paris Nanterre

Bureau d'Économie Théorique et Appliquée -UMR CNRS 7522 UFR Droit, Sciences Économiques et Gestion - 13 place Carnot C.O. 70026 - 54035 Nancy cedex

L'Université de Lorraine n'entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Elles doivent être consiidérée comme étant propres à l'auteur.

Remerciements

Cette thèse aurait été différente si je n'avais pas partagé ce chemin avec de nombreuses personnes qui m'ont accompagnée pendant ces années de doctorat. C'est pour cette raison que je tiens à les remercier dans les lignes suivantes.

Avant tout, mes remerciements sont à l'intention de ma directrice de thèse Sophie Harnay. Je la remercie sincèrement d'avoir encadré mon travail de thèse, toujours avec enthousiasme et de m'avoir transmis sa passion de la recherche. Ses remarques et suggestions m'ont permis d'avancer et d'améliorer ce travail. Je la remercie pour son support, son écoute et pour le fait d'avoir été une figure fondamentale dans ce parcours.

Je remercie les membres du jury qui me font l'honneur d'avoir accepté d'évaluer cette thèse. Je remercie Thierry Pénard et Fabrice Rochelandet d'avoir accepté d'être rapporteurs et Myriam Doriat-Duban, Julien Pénin et Dominique Torre d'être examinateurs. Je mesure l'honneur qui m'est fait.

Le Bureau d'économie théorique et appliquée (BETA) m'a accueillie il y a quatre ans et je tiens à exprimer ma gratitude envers ses membres. Je remercie les directeurs successifs Bruno Jeandidier, Agnès Gramain et Serge Garcia, de m'avoir offert d'excellentes conditions de travail et de leur attention envers les doctorants. Je remercie également les membres de mon comité de suivi de thèse Julien Pénin et Bérangère Lauren Szostak pour leur encouragement constant et pour leurs suggestions. Mes remerciements vont également aux membres de l'axe économie du droit, et plus précisément à Yannick Gabuthy, Cécile Bourreau-Dubois, Eve-Angéline Lambert, Thomas Lanzi et Sébastien Massoni et ainsi qu'à Agnès Gramain pour leur aide crucial. Je remercie également les membres du service administratif du laboratoire qui nous offrent leur aide tous les jours. Je remercie Sylviane Utereiner et Julie Mansuy pour leurs sourires et leurs bonbons.

Je remercie les doctorants et jeunes docteurs du BETA, d'avoir toujours été à mes côtés. Je remercie Camille Aït-Youcef, Julien Grandjean et Dylan Martin-Lapoirie pour leurs relectures, suggestions et surtout pour leur support dans les moments les plus difficiles. Je remercie également Sarah Van Driessche, Jeanne Poulain et Julien Tavard pour leur soutien indéfectible. Je remercie Georges El Haddad pour nos discussions et ses pilules d'histoire française. Je cherchais des collègues, j'ai trouvé des amis précieux.

Mes remerciements les plus tendres se dirigent vers mes proches. Je remercie mes sœurs Carola et Gaia et mon frère Umberto d'avoir été présents lors des longues présentations auxquelles ils ont été obligés d'y assister, pour leur bonne humeur et leurs encouragements. Je remercie mon père Walter, pour tout ce qu'il a fait pour me permettre d'être ici aujourd'hui et pour avoir été toujours prêt à m'accueillir pendant les moments difficiles de ce parcours. Je remercie ma maman Anna et ma grand-mère Rosanna d'avoir toujours cru dans mes capacités. J'espère les rendre fiers. Je remercie Piero, Lorenza, Stefano et Edoardo, ma deuxième famille. Je remercie toutes les personnes qui, depuis toujours, ont été à mes côtés et les nouvelles, que cette aventure m'a permis de rencontrer. Je remercie Lorella, pour son encouragement et son soutien. Je remercie Guillaume et Julien pour leur douceur et leur capacité de me faire sentir chez-moi. Je remercie Valentina pour nos discussions toujours intéressantes et pour ses mots d'encouragement. Je remercie Giulia pour son écoute et sa sensibilité. Je remercie enfin Alexia, d'avoir continué à croire dans notre amitié malgré la distance et toutes les difficultés.

À mon père

Table des matières

Introduction **1**

0.1 Contexte et motivation 4

0.1.1 Numérisation, *Open Science* et *Open Access* 4

0.1.2 Nouvelles modalités de partage du savoir scientifique : du *Closed Access* à l'*Open Access* ? 8

0.1.3 Les évolutions paradoxales du cadre juridique : entre promotion de la science ouverte et renforcement de la protection intellectuelle . . . 10

0.2 Contribution à la littérature 12

0.2.1 L'économie du numérique et les communs de la connaissance 12

0.2.2 L'analyse économique du droit de la propriété intellectuelle 13

0.2.3 La nouvelle économie de la science 15

0.3 Présentation des chapitres, méthodologie et principaux résultats 16

Bibliographie de l'introduction générale 23

Annexe 29

1 La régulation de l'accès et de la circulation de la connaissance scientifique à l'ère du numérique : un état de l'art **33**

Introduction 34

1.1 La régulation de l'accès à la connaissance scientifique à travers le droit d'auteur 40

1.1.1 Le cadre juridique du droit d'auteur et du *copyright* 40

1.1.2 L'évolution de l'analyse économique du droit d'auteur 42

1.1.3 Le rôle du droit d'auteur sur le marché de la publication scientifique 47

1.1.4 Le droit d'auteur à l'âge du numérique 51

1.2 L'accès ouvert au savoir scientifique : le modèle *Open Access* 54
1.2.1 Le mouvement *Open Access* 55
1.2.2 Les différentes formes d'*Open Access* 56
1.2.3 L'impact de l'*Open Access* pour les maisons d'édition 64
1.2.4 L'utilisation de l'*Open Access* par la communauté scientifique 66
1.3 L'accès ouvert et la connaissance scientifique en tant que commun 70
1.3.1 La numérisation de l'information : de biens communs à communs informationnels 71
1.3.2 La gestion commune du savoir scientifique 74
Conclusion 81
Bibliographie du chapitre 1 83
Annexe 95

2 Rule vs. Communality Norm in the Scholarly Publishing Market : Institutional Overlap as a Barrier to Publishing in Open Access journals ? 97
Introduction 98
2.1 Intellectual property and Open Access in the scholarly publishing market : a complex interplay of legal rules and social norms 101
2.1.1 A tale of two coexisting systems 101
2.1.2 The complex interplay of norms-based and law-based IP in scholarly publishing 103
2.1.3 Open Access at the crossroads between norms-based and law-based IP 106
2.2 The evolving institutional dynamics in scholarly publishing : an historical overview 107
2.2.1 The long-lasting prevalence of social norms in the governance of scientific exchanges 107
2.2.2 The takeover of copyright on norms-based IP 110
2.3 The salience of copyright as a barrier for the migration to OA publishing . 114
2.3.1 The rise of copyright 114
2.3.2 The new copyright era and its stickiness 116
2.4 The complementarity between the two IP systems : further comments . . . 118

Conclusion . . . 119
Bibliography of Chapter 2 . . . 121

3 Who wants Open Access ? Copyright Reform and Scientific Knowledge Dissemination : Empirical Evidences Based on French Data 127

Introduction . . . 128
3.1 The economic analysis of barriers to the development of Green Open Access 131
3.1.1 The Green Open Access model . . . 131
3.1.2 Copyright infringement as a barrier to Green Open Access . . . 133
3.1.3 Scholars' attitudes towards Green Open Access . . . 134
3.2 Empirical strategy . . . 136
3.2.1 The French Law "*Pour une république numérique*" . . . 136
3.2.2 Data collection and descriptive statistics . . . 138
3.2.3 The theoretical model . . . 141
3.2.4 The empirical model . . . 144
3.3 Results . . . 145
3.3.1 The effect of the copyright reform on increasing free access to published papers . . . 145
3.3.2 Time lag . . . 149
3.3.3 Results discussion . . . 150
Conclusion . . . 152
Bibliography of Chapter 3 . . . 155
Annex . . . 158

4 The star effect of self-archiving practice : how journal rank affects green Open Access choices 193

Introduction . . . 194
4.1 Literature review . . . 198
4.1.1 Citation advantage and the Open Access choice . . . 198
4.1.2 Journal quality and the Open Access choice . . . 200
4.2 Theoretical model . . . 202

4.3 Data analysis . . . 205
4.4 The empirical model . . . 209
4.5 Results . . . 210
4.5.1 National journal ranking indices . . . 210
4.5.2 International journal ranking indices . . . 212
4.6 Discussion and policy implications . . . 214
Conclusion . . . 217
Bibliography of Chapter 4 . . . 219
Annexe . . . 223

5 La gouvernance du marché de la publication scientifique à l'ère du numérique : une analyse de la connaissance scientifique en tant que *commons* 227
Introduction . . . 228
5.1 La connaissance scientifique à l'aune de l'*Institutional Analysis Development framework* . . . 232
5.2 De la numérisation de l'information à la connaissance scientifique en tant que commun . . . 234
5.2.1 Les conditions de l'existence de la connaissance scientifique en tant que communs . . . 235
5.2.2 Quelle communauté pour la gestion de la connaissance scientifique ? 236
5.2.3 La régulation du marché de la publication scientifique : entre public order et private ordering . . . 237
5.3 L'arène de l'action : coordination ou compétition entre intérêts ? . . . 240
5.3.1 Le *Closed Access* à l'ère du numérique . . . 240
5.3.2 Différents modèles pour différents intérêts : Les modèles *Open Access* 241
5.4 Entre innovation et adaptation : l'évolution de l'*Open Access* . . . 247
5.4.1 La chaîne de valeur de la connaissance scientifique : complémentarité de modèles pour la *Grande Conversation scientifique* . . . 247
5.4.2 Adaptation, innovation et réaction : quelle gestion commune du savoir scientifique ? . . . 250
Conclusion . . . 252

Bibliographie du chapitre 5 254

Conclusion générale **261**

Résumé **265**

Abstract **267**

Introduction

La connaissance est un facteur de production essentiel dans les économies modernes de la connaissance (Foray, 2018), où l'innovation en tant que dissémination, réutilisation et valorisation du savoir devient fondamental pour la croissance économique (Solow, 1967). L'innovation est l'un des termes les plus utilisés dans notre quotidien : elle fait l'objet d'étude de différentes disciplines dont on dispose de presque une définition pour chaque domaine d'application (Damanpuor et Schneider, 2006). Bien qu'elle reste sans définition universelle, incapable d'englober toutes les facettes qui la caractérisent, l'innovation est devenue une référence pour mesurer la puissance d'un État, et donc la cible sensible des politiques publiques pour la croissance économique.

Comme la Commission européenne l'a souligné (Hilty et al., 2008), la science est à la base de ce processus de création et de valorisation de l'innovation. La circulation et la réutilisation de la connaissance sans barrières sont donc devenues l'une des priorités des politiques publiques de l'Union européenne (ci-aprèsUE). La libre diffusion de la recherche scientifique est l'étape préliminaire au développement de la science ouverte [1], définie comme « [. . .] an approach to research based on greater access to public research data, enabled by ICT tools and platforms, and broader collaboration in science, including the participation of non-scientists, and finally, the use of alternative *copyright* tools for diffusing research results », et qui conduira vers la prochaine révolution industrielle (OCDE, 2016).

La question de la science ouverte est alors centrale au sein des différents pays de l'Union européenne. Presque tous les pays membres ont adopté ou sont en train de mettre en place des politiques publiques, visant à promouvoir l'*Open Science* au sein de la communauté scientifique (Sveinsdottir et al., 2020). De même, la communauté scientifique s'est intéressée à la question de la science ouverte, qui croise de façon transversale plus de 150 domaines de la recherche scientifique [2].

L'intérêt à la fois de la communauté scientifique et des pouvoirs publics envers l'Open Science montre bien l'importance de rendre disponibles les résultats scientifiques, non seulement en permettant une libre circulation et réutilisation au sein de la communauté scientifique, mais aussi au sein de la société. Afin de rendre moins abstrait le caractère central de

1. ou indifféremment *Open Science*, terme issu de la littérature anglaise
2. Voir Figure 1 en annexe 1 donnant un aperçu de la transdisciplinarité de la science ouverte au sein de la communauté scientifique.

la question de la science ouverte, nous proposons à titre d'exemple deux cas, qui montrent bien son importance sociale. Comme souligné par l'OCDE (2020), dans le cadre de la pandémie COVID-19, la science ouverte a permis d'éliminer les obstacles à la libre circulation des données de la recherche et aux nouvelles idées développées, en accélérant ainsi l'avancée de la science. C'est véritablement dans ce contexte de crise sanitaire que de nombreuses initiatives ont vu le jour, telles que l'Open COVID Pledge, le CORD-19 et la Public Health Emergency COVID-19 Initiative, avec l'objectif de rendre accessibles les nouveaux résultats et données scientifiques, sans barrières tariffaires et légales qui en pourraient empêcher une utilisation dérivée. Ces initiatives ont permis l'élargissement des collaborations internationales et l'accélération de la production de nouveaux résultats scientifiques, dans l'effort commun de faire face à la crise sanitaire de la Covid-19. Sans aller jusqu'au cas extrême d'une crise sanitaire mondiale, Houghton et Sheehan (2009) ont montré les bénéfices économiques de l'*Open Access*, l'un des piliers de la science ouverte. À travers l'utilisation du modèle Solow-Swan, modifié en introduisant les variables d'accessibilité et d'efficience des résultats scientifiques, les auteurs montrent comment ce modèle de publication amène à des rendements sur les investissements en R&D beaucoup plus important par rapport à la dissémination traditionnelle du savoir scientifique, qui impose des barrières de prix à l'accès et de réutilisation des données de la recherche.

Conscients du potentiel économique et social de l'*Open Access*, les gouvernements des différents pays membres de l'UE ont posé les bases d'une réforme institutionnelle capable de valoriser et de rendre viable ce nouveau modèle de dissémination de la connaissance. Cependant, et malgré les efforts accomplis jusqu'aujourd'hui, ce modèle reste faiblement utilisé par rapport au modèle standard, appelé *Closed Access*[3]. Ce dernier prévoit la cession du droit d'auteur de la part de son titulaire à la maison d'édition, en contrepartie de la publication de l'article scientifique, dont elle devient le nouveau titulaire. Grâce à ce titre juridique, la maison d'édition est en mesure de rendre excluable l'accès à la ressource, en imposant un prix à l'accès.

Puisque l'*Open Access* permet une plus large diffusion des résultats scientifiques, en éliminant toute barrière légale et économique, pourquoi reste-t-il moins utilisé par rapport

3. Ou *Accès fermé* dans la littéarture française

à un système de circulation de la connaissance qui limite l'accès à la littérature scientifique et la réutilisation de ses résultats ? Cette thèse vise donc à comprendre et étudier les barrières à l'utilisation d'un modèle de publication qui assure un libre accès à la littérature scientifique à tout utilisateur. Pour cela, nous devons avant tout comprendre les différents modes d'organisation de la diffusion du savoir scientifique qui caractérisentn à l'ère du numérique, le marché de la publication scientifique et les institutions qui en permettent le fonctionnement.

0.1 Contexte et motivation

La numérisation de l'information a fait en sorte que l'idée d'une science ouverte puisse devenir l'une des priorités de la communauté scientifique (1.1). Sa mise en place a impacté le fonctionnement traditionnel du marché de la publication scientifique (1.2) et a demandé de concevoir de nouveaux modèles pour la circulation du savoir scientifique (1.3). Cela requiert l'imposition d'exceptions à l'application du droit d'auteur dans le domaine de la recherche scientifique, ayant pour conséquence la transformation du cadre juridique où un assouplissement du droit d'auteur coexiste avec un renforcement de la protection accordée par cet instrument de propriété intellectuelle (1.4).

0.1.1 Numérisation, *Open Science* et *Open Access*

La numérisation de l'information a permis de concevoir la production et la circulation du savoir scientifique différemment. Les barrières géographiques qui empêchaient ou limitaient la co-production du savoir entre chercheurs de différents pays ont été affaiblies. Le support papier, à travers lequel la connaissance était diffusée, est désormais devenu un instrument de partage du savoir obsolète. À l'ère de l'internet, les collaborations entre scientifiques et les échanges entre les membres de la communauté scientifique se sont accélérés et enrichis de nouveaux instruments pour le partage des données de la recherche. Les adresses électroniques, les réseaux sociaux (Ovadia, 2014), l'accès numérique aux articles scientifiques assuré par différents acteurs, tels que les bibliothèques universitaires et les archives en ligne (Björk et al., 2014), mais aussi les nouveaux instruments de veille

scientifique, comme *Google Scholar* (Cecchino, 2010), ont favorisé la « grande conversation scientifique » (Guédon, 2014) et un accès élargi au savoir scientifique.

Ces nouvelles opportunités ont été essentielles pour la naissance de la science ouverte, que l'EU soutient énergiquement. À travers la recommandation C(2012) 4890 de la Commission européenne, les États-membres sont encouragés à supporter la diffusion dans le domaine public de la recherche financée par des fonds publics. Plus précisément, l'Union européenne demande la mise en place de politiques publiques pour soutenir la dissémination et la préservation des résultats et des données de recherche. Ainsi, les pays membres ont adopté, ou sont en train de mettre en place, des mesures visant la libre circulation des résultats scientifique et leur réutilisation [4]. Ce phénomène d'ouverture de la science ne connaît pas de frontière. On peut prendre à titre d'exemple le case des États-Unis, où, à partir de 2005, le National Institute of Health (NIH) a obligé ses chercheurs à déposer une copie des produits de la recherche au sein du NIH, dans leur version finale, acceptée pour la publication dans une revue (Swan, 2013).

Le mouvement de l'Open Science a autant intéressé les pouvoirs publics que la communauté scientifique. Etant donné le grand intérêt pluridisciplinaire pour la science ouverte, la littérature se réfère à ce sujet de manière très hétérogène. Ainsi, on parle d'Open Science en termes de connaissance (David, 2004 ; Fry et al., 2009), de transparence de la recherche (Leonelli et al., 2015), de connaissance accessible (Partha et David, 1994 ; Grand et al., 2016), de connaissance partagée (Schmidt et al., 2016), ou encore de développement collaboratif du savoir (David, 1998 ; Fry et al., 2009). En réalité, chaqu'une de ces définitions ne correspond qu'à l'une des différentes sphères qui, dans leur ensemble, constituent la science ouverte, comme le montre la figure 1.

4. Voir Tableau 1 en annexe 1 pour les différentes politiques publiques adoptées par chaque pays membre de l'Union européenne.

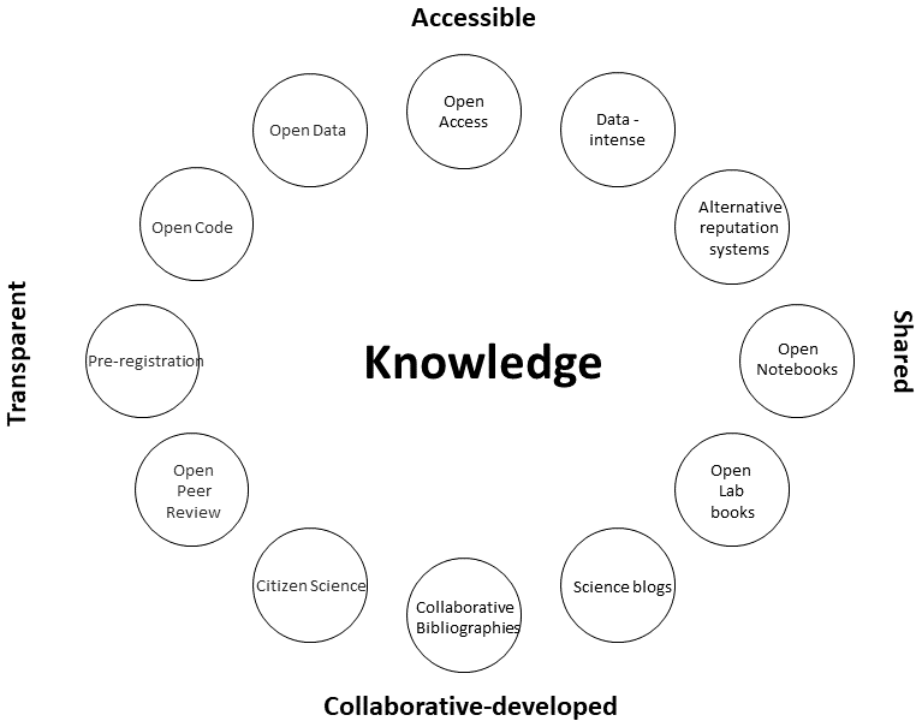

FIGURE 1 – Les différentes dimensions de l'*Open Science* - Source : Vicente-Saez et Martinez-Fuentes (2018)

On peut le voir, c'est dans ce contexte qui s'inscrit le mouvement *Open Access* (ou accès ouvert ou encore libre accès dans la littérature française), conçu en réaction à l'augmentation des prix des revues scientifiques et aux stratégies de discrimination par les prix opérées par les maisons d'édition (Suber, 2012 ; Willinsky, 2009). En tant que porteur d'un nouveau modèle de dissémination du savoir scientifique, il vise l'élimination de toute barrière de prix et des restrictions imposées par le droit d'auteur à la libre circulation et réutilisation de la connaissance (Suber, 2012).

1.2 Le marché de la publication scientifique Le marché de la publication scientifique occupe alors un rôle central dans la mise en place de la science ouverte. La valeur de ce marché, défini comme le lieu d'échange du savoir scientifique sous forme d'articles, considérés comme le standard de la communication scientifique (Origgi, 2010), a été estimée autour de 25.2 milliards de dollars en 2013 [5] (Ware et Mabe, 2015), avec plus de 2.5 millions d'articles scientifiques publiés chaque année (Plume et Van Weijen, 2014).

5. Ces données ne font référence qu'au marché STM (Science, Technique et Médicine). Pour le domaine des sciences humaines et sociales, à notre connaissance il n'existe pas d'estimation de la valeur de ce marché.

À partir de la seconde guerre mondiale, grâce aux investissements importants en R&D, on assiste à un profond changement dans l'organisation de la communauté scientifique (Ziman, 2000 ; Kellog, 2006). L'élargissement de cette communauté et l'augmentation conséquente de la production scientifique ne font que souligner le rôle central du marché de la publication scientifique qui, lui aussi, a subi des changements profonds. D'un marché caractérisé par une gestion de la connaissance assurée par les sociétés de savants, telles que la Royal Society de Londres ou l'Académie Royale de Paris (Vittu, 2002), les maisons d'édition commerciales commencent à cibler des opportunités à saisir sur ce marché. La mise en place de stratégies de marketing et l'adoption d'économies d'échelle ont poussé les maisons d'édition commerciales à devenir de plus en plus spécialisées dans cette nouvelle activité, en rendant la publication scientifique rentable (Fyfe, 2015). Sans une organisation commerciale capable d'entrer en compétition avec ces nouveaux acteurs, les sociétés savantes ont créé des alliances avec les maisons d'édition commerciales, qui ont adopté le système de l'évaluation par les paires, typique des sociétés savantes, afin de légitimer leur activité de certification de la science (Baldwin, 2015).

Le système à la base de la publication scientifique, appelé *Closed Access*, se fonde sur une cession du droit d'auteur de la part de l'auteur de l'output intellectuel à la maison d'édition, qui en devient le nouveau titulaire, en contrepartie de la publication des résultats scientifiques. Grâce à l'exploitation de ce pouvoir de monopole temporaire sur l'accès à la ressource intellectuelle (Foray, 1999 ; Menell, 2016), les maisons d'édition sont ainsi en mesure de rémunérer les investissements nécessaires pour assurer la dissémination de la connaissance scientifique. Cette activité ne se résume pas que dans la production de la copie sous format papier, mais intégre aussi sa distribution et l'organisation nécessaire pour la validation des résultats de la science.

La demande inélastique pour l'accès au savoir scientifique combiné avec un pouvoir de monopole sur la ressource intellectuelle ont permis l'instauration d'un marché en oligopole, où cinq maisons d'édition, et plus précisément Reed-Elsevier, Springer Nature, Wiley-Blackwell, Taylor & Francis et Sage, assurent encore aujourd'hui plus de 50% de la publication d'articles scientifiques (Johnson et al., 2017 ; Larivière et al., 2014). C'est véritablement à partir des années 1980 que cette concentration pose problème. Le pouvoir

assuré par le monopole sur la ressource intellectuelle a permis aux maisons d'édition commerciales d'augmenter les prix des revues, qui ont subi une augmentation plus importante que l'évolution des prix à la consommation. Différentes études montrent une augmentation des prix supérieur à 210% sur la période 1986-2001 (Edwards et Shulenburger, 2003[6]). Cela a donné lieu à ce que la littérature appelle la *Serial Crisis*, situation dans laquelle l'accès à la connaissance est devenu fortement limité, en raison de l'augmentation des prix des abonnements (Kyrillidou et Young, 2008 ; Edwards et Shulenburger, 2003).

Etant donné le caractère cumulatif de la production du savoir scientifique, un accès limité à la littérature devient un véritable obstacle à l'avancement de la science. Il suffit de penser aux problèmes qu'un accès limité peut entraîner dans le domaine médical. Comme l'ont soutenu un groupe de chercheurs du Liberia, la crise sanitaire provoquée par le virus Ebola en 2015 aurait pu être empêchée si la littérature scientifique n'était pas soumise à un prix à l'accès (Pells, 2020).

0.1.2 Nouvelles modalités de partage du savoir scientifique : du *Closed Access* à l'*Open Access* ?

Avec la numérisation de l'information et la baisse des prix marginaux de reproduction et de diffusion du savoir presqu'à zéro, on aurait pu s'attendre à une baisse des prix des revues scientifiques et, surtout, à une ouverture du marché de la publication scientifique à la concurrence. Dans les faits, la structure en oligopole du marché a été préservée (Larivière et al., 2014), et le paiement d'un prix à l'accès a créé des fortes tensions entre la communauté scientifique et les maisons d'édition. En France, en 2018, un nombre important de centres de recherche n'a pas eu accès aux articles produits par la maison d'édition Springer, à cause du manque d'accord sur le prix des abonnements à payer (Kwon, 2018).

Le maintien d'une telle structure de marché a été possible grâce à la présence de barrières à l'entrée qui protègent ce marché, telles que la réputation à établir des journaux scientifiques. Comme expliqué par Edlin et Rubinfeld (2004), un journal doté d'une réputation importante peut être caractérisé comme un équilibre de coordination. Les auteurs

6. Voir aussi Kyrillidou, M., & Young, M. (2008). ARL Statistics, 2005-06 : A Compilation of Statistics from the One Hundred and Twenty-Three Members of the Association of Research Libraries. In Association of Research Libraries. Association of Research Libraries ; Bergstrom, T. C. (2001). Free Labour for Costly Journals ? Journal of Economic Perspectives, 15(4), 183–198.

expliquent que la valeur d'une revue est déterminée à partir du consensus simultané d'une multitude d'acteurs (*e.g.* auteurs, lecteurs, éditeurs etc.), ce qui rend très difficile pour une revue qui vient d'entrer sur le marché le fait d'atteindre ce type d'équilibre.

C'est dans ce contexte que l'*Open Access* a vu le jour en tant que nouveau modèle de diffusion de la recherche scientifique. Dans sa définition la plus large, l'*Open Access* regroupe toute initiative visant la subversion du régime de propriété intellectuelle de l'intérieur (Dusollier, 2007), c'est-à-dire à partir de l'initiative d'une communauté spécifique, dans notre cas la communauté scientifique.

En tant que modèle de publication, l'*Open Access* vise alors l'élimination de toute barrière de prix à l'accès à la connaissance scientifique et des barrières légales, imposées par le droit d'auteur, qui empêchent la libre circulation et utilisation des résultats de la recherche (Suber, 2012). Contrairement au modèle de publication traditionnel, le *Closed Access*, ce modèle ne prévoit pas la cession du droit d'auteur à la maison d'édition par le chercheur, qui reste le seul titulaire. La maison d'édition n'a donc pas le pouvoir d'imposer un prix à l'accès.

Cependant, il faut remarquer que ce modèle n'est pas forcément synonyme de gratuité. Avant tout, il faut distinguer entre modèles d'*Open Access* et modèles d'affaire adoptés pour leur mise en place, concepts souvent mélangés dans la littérature (*e.g.* Feess et Scheufen, 2016 ; Shavell, 2010 ; McCabe et Snyder, 2018). L'*Open Access* peut se présenter sous deux formes différentes, à savoir le *Green Open Access* et le Gold *Open Access*. Le premier, consiste dans l'archivage d'articles scientifiques, publiés ou pas dans une revue scientifique, dans une archive numérique, souvent financée par des organisations publiques, dans le cadre de la mise en place de l'Open Science. Le *Green Open Access* constitue alors un instrument alternatif aux maisons scientifiques pour avoir accès à la littérature scientifique. Afin de protéger leur activité éditoriale, une série de mesures ont alors été adoptées par les maisons d'édition. D'un côté, un certain nombre de restrictions a été imposé à l'archivage d'articles scientifiques, telles que l'imposition d'une période d'embargo à respecter, avant de pouvoir déposer un article dans une archive (Björk et al., 2014). De l'autre côté, certains acteurs du marché de la publication scientifique ont été capables de faire évoluer leur activité de dissémination du savoir scientifique, en intégrant le modèle *Green Open Access* dans leurs

activités, comme le montre l'achat de l'archive SSRN (Social Science Research Network) par Elsevier en 2016 (Gordon, 2016). Le deuxième modèle, à savoir le Gold *Open Access*, est un modèle de publication d'articles scientifiques, comme dans le cas du *Closed Access*, mais qui, contrairement à lui, n'impose pas de prix à l'accès au savoir. S'il est vrai qu'un certain nombre de journaux ont été mis en place par des membres de la communauté scientifique même, en adoptant des modèles d'affaires à but non lucratif, comme la revue « economics e-journal », les maisons d'édition commerciales ont été capables de rendre rentable ce modèle. C'est ainsi qu'elles ont commencé à imposer un prix à la publication (*Article Processing Charges* ou APC) ou, mieux encore, à adopter des stratégies mixtes comme l'hybrid *Open Access*, qui prévoit la publication d'un article en *Open Access*, soumis à des frais à la publication, dans une revue en *Closed Access*.

Malgré les différentes formes d'*Open Access* qui sont à disposition des chercheurs, le modèle de l'*Open Access* reste très peu utilisé. Comme le montrent les études de Green (2018) et Fortson (2019), l'accès à 80% de la littérature scientifique reste encore aujourd'hui soumis à une barrière de prix. De plus, la large utilisation des sites pirates (Himmelstein et al., 2018) pour avoir accès à la littérature scientifique montre que l'accès à la littérature scientifique reste encore aujourd'hui entravé.

0.1.3 Les évolutions paradoxales du cadre juridique : entre promotion de la science ouverte et renforcement de la protection intellectuelle

S'il n'est pas synonyme de gratuité, l'OA n'est pas non plus synonyme d'absence de régulation à l'accès et à la réutilisation des résultats scientifiques. Comme on l'a vu, le chercheur reste le titulaire du droit d'auteur et donc le seul à pouvoir autoriser toute reproduction et réutilisation de son œuvre intellectuelle. C'est justement pour permettre une plus grande réutilisation et circulation des résultats de la recherche que les licences *copyleft*, surtout sous la forme de licences *Creative Commons*, ont été introduites dans le domaine scientifique. Elles posent une exception à la règle de tous droits réservés, imposée par le droit d'auteur, et permettent au titulaire de décider quelles utilisations dérivées sont autorisées et valables pour tout utilisateur (Carrol, 2006 ; Cogo, 2010).

La régulation *bottom-up* a été suivie par différentes politiques publiques, favorisant la

libre circulation et la réutilisation du savoir scientifique dans le monde entier. Cependant, l'UE reste l'un des acteurs les plus impliqués dans la promotion de l'accès ouvert à la connaissance scientifique [7]. S'il est vrai que plus de 700 politiques publiques ont été adoptées au niveau mondial, très souvent elles ne sont que le résultat d'initiatives menées dans le cadre d'instituts de recherche ou d'universités. À travers la recommandation C(2012) 4890 de la Commission européenne, l'UE marque la science ouverte comme l'une de ses priorités. Malgré cela, la France, avec la loi *Pour une république numérique* reste le seul État à avoir créé d'exceptions à l'application du droit d'auteur dans le milieu scientifique (Sveinsdottir et al., 2020). Désormais, les auteurs financés au moins pour la moitié par des fonds publics ont le droit de mettre en libre accès dans des archives institutionnelles ouvertes les résultats de leurs recherches, même publiés dans des revues scientifiques. Par ailleurs, plusieurs laboratoires ont adopté des politiques de dépôt obligatoire dans les archives ouvertes [8].

La tension entre l'exigence d'un accès élargi à la littérature scientifique pour les chercheurs et la volonté de protéger la titularité du droit d'auteur de la part des maisons d'édition est bien montrée par les nombreux arrêts qui ont caractérisé le marché de la publication scientifique à partir de sa numérisation. En 2011, aux États-Unis, le cas *United States of America vs. Aaron Swartz* (Criminal No. 11-10260-NMG) témoigne de façon manifeste la volonté des maisons d'édition de s'engager dans la lutte contre la piraterie scientifique, même si dans ce cas l'action illicite n'avait pas encore été commise. La même année, un projet de loi avait été présenté au Congrès américain, visant l'interdiction pour les agences scientifiques d'imposer à leur chercheurs la publication en accès libre (Benjamin, 2012). En 2017, Elsevier gagne son contentieux contre A. Elbakyan, créatrice des sites pirates Libgen et Sci-hub, la cour du district de New York lui imposant une amende de 15 millions de dollars (Schiermeir, 2017). En France, le Tribunal de grande instance de Paris a ordonné à trois fournisseurs d'accès à internet de bloquer l'accès aux sites internet Libgen et Sci-hub (TGI Paris, 3e ch. 4e sect. jugement en la forme des réf. le 7 mars 2019). Une telle tension juridique fait écho aux batailles juridiques qui ont eu lieu autour de la piraterie dans le domaine de la musique dans les années 2000, phénomène qui a totalement changé le visage du marché et du partage des contenus musicaux.

7. Voir Figure 3 en annexe 2 pour la distribution des politiques publique dans les différents continents
8. Voir l'archive RoarMap pour une liste exhaustive `http://roarmap.eprints.org/`

0.2 Contribution à la littérature

Cette thèse contribue principalement à trois domaines de la littérature. Avant tout, elle s'intéresse à l'impact de la numérisation sur le marché de la publication scientifique surtout du point de vue de la gestion de la ressource (2.1). Cette numérisation a amené à une réforme du droit d'auteur, non seulement à travers l'introduction de nouvelles lois mais aussi grâce à l'application de nouveaux instruments juridiques, étudiés par l'analyse économique de la propriété intellectuelle (2.2). Puisque le droit d'auteur ne représente que l'une des institutions régulant le marché de la publication scientifique, nous contribuons aussi à la littérature de la nouvelle économie de la science, qui s'intéresse avant tout au fonctionnement de la communauté scientifique en tant qu'organisation (2.3).

0.2.1 L'économie du numérique et les communs de la connaissance

De façon très large, l'économie du numérique s'intéresse à l'impact de la numérisation sur différents secteurs de l'activité économique. Comme souligné par Bourreau et Pénard (2016), elle ne peut pas être réduite à une sous-catégorie de l'économie industrielle, car son domaine de recherche se place bien au-delà de ces frontières. Elle se fonde surtout sur l'analyse des big data, collectées à partir de l'internet et ensuite appliquées à différents champs de l'économie.

Notre thèse contribue à cette littérature, à travers son application au marché de la publication scientifique et aux conséquences de la numérisation de l'information. Notre contribution est double.

D'une part, nous proposons d'analyser l'impact de la numérisation sur la chaîne de valeur de la diffusion de la connaissance scientifique. L'étude des effets de la numérisation dans ce marché n'est pas une littérature complètement nouvelle. Plusieurs auteurs ont largement contribué à cette littérature, en étudiant le phénomène de persistance d'une structure de marché en oligopole même après la numérisation de l'information (Larivière et al., 2015), et en expliquant, à travers l'analyse des cinq forces de Porter, les raisons qui ont permis le maintien d'une telle structure de marché (Björk, 2017). Nous montrons comment, en réalité, de nouveaux acteurs ont essayé de pénétrer ce marché, en offrant des instruments complémentaires à la publication des revues scientifiques pour la diffusion

de la connaissance, et comment les maisons d'édition ont été capables de différencier leurs activités, comme le montre le cas du rachat de l'archive SSRN par Elsevier en 2016 (Gordon, 2016).

D'autre part, l'impact de la numérisation a permis l'émergence de nouvelles formes de gestion du savoir scientifique. Les communs (ou commons dans la littérature anglophone) sont des biens non rivaux et non excluables, dont la gestion est soumise à une communauté spécifique (Boyle, 2003 ; Hess et Ostrom, 2007 ; Coriat, 2015). Dans l'objectif de contribuer à la littérature sur les communs numériques, nous utilisons l'IAD *framework* (*Institutional Analysis Development*) pour identifier et analyser les différences entre les diverses formes d'*Open Access* et les modèles d'affaires utilisés pour rendre viables les modèles de partage de la connaissance (Kiser, Ostrom, 1982 ; Ostrom, 2009) – ces différences n'étant pas clairement établies dans la littérature (*e.g.* Feess et Scheufen, 2016 ; Shavell, 2010 ; McCabe et Snyder, 2018). Cette littérature, dont l'application au marché de la publication scientifique ne compte que quelques rares études (Boyle, 2007 ; Waters, 2007 ; Suber, 2007), nous permet de comprendre l'importance de la complémentarité entre le modèle en *Open Access* et les licences *Creative Commons* pour assurer une véritable ouverture et la libre circulation des résultats de la recherche.

0.2.2 L'analyse économique du droit de la propriété intellectuelle

L'une des variables qui permettent la viabilité d'une économie de la connaissance est le cadre institutionnel, qui doit fournir des incitations pour une utilisation, dissémination et création efficientes de la connaissance (Chen et Dahlman, 2005). L'octroi d'un pouvoir de monopole, qui permet de rendre excluable l'accès à la ressource, ayant les caractéristiques d'un bien collectif, assure ainsi au producteur de rémunérer l'investissement soutenu pour la production et la diffusion de son œuvre, à travers la négociation du droit de propriété intellectuelle entre le titulaire et les autres parties prenantes (Foray, 2018 ; Scheufen, 2015). Afin d'étudier le cadre institutionnel du marché de la publication scientifique, nous nous appuyons sur la littérature de l'analyse économique du droit de la propriété intellectuelle. Dans le cas du marché de la publication scientifique, le droit d'auteur est l'une des institutions principales du marché de la publication scientifique. Pendant longtemps, il a assuré la

circulation de la connaissance, à travers l'échange de ce droit entre chercheurs et maisons d'édition.

Cependant, avec la numérisation de l'information, la nécessité d'un tel droit a fait débat dans la littérature. Puisque la numérisation permet une diffusion de la connaissance qui ne connaît pas de frontières, avec un coût marginal de reproduction de la première copie presqu'égal à zéro, pourquoi l'accès à la connaissance scientifique devrait être soumis au payement d'un prix à l'accès ? Plus précisément, à l'ère du numérique, le droit d'auteur assure-t-il encore sa fonction économique primaire, à savoir inciter la production d'un bien qui ne serait autrement pas produit en absence de ce cadre juridique ? Un certain nombre de solutions ont été proposées, allant de la suppression du droit d'auteur (Shavell, 2010) à l'introduction d'un éventail d'exception à son application dans le domaine de la recherche scientifique (Scheufen, 2015 ; Mueller-Langer et Watt, 2010).

Cette thèse vise à comprendre dans quelles mesures le droit d'auteur est devenu une véritable barrière au développement de l'*Open Access*. Notre hypothèse est que le droit d'auteur n'est que l'une des institutions régissant le marché de la publication scientifique et qu'une analyse ne prenant en considération que les normes légales n'est pas suffisante. D'autres communautés professionnelles ou épistémiques (Loshin, 2007 ; Fauchart et von Hippel, 2008) ont un fonctionnement proche de celui de la communauté scientifique, en se fondant sur un double système de normes de propriété intellectuelle, légales et sociales. Les normes légales sont promulguées et appliquées par une autorité centrale (McAdams, 2001), quand les normes sociales sont le résultat de la production endogène de l'interaction individuelle (Bicchieri, 2011 ; Lewis, 1969). C'est pour ces raisons que notre contribution à la littérature de l'analyse économique du droit est double. D'un côté, nous analysons les normes formelles et informelles de ce marché, afin d'en étudier le degré de substituabilité et de complémentarité et leur processus de coévolution éventuelle. De l'autre côté, nous testons empiriquement si la réforme du droit d'auteur pourrait à elle seule favoriser le développement de l'*Open Access*, comme proposé par la littérature, en changeant le comportement des chercheurs face à leurs choix de publication, ou si une réforme de tout le système institutionnel sur le marché de la publication devrait être envisagée, afin de rendre plus efficientes les incitations à la dissémination de la connaissance scientifique.

0.2.3 La nouvelle économie de la science

La nouvelle économie de la science est le résultat d'une synthèse entre l'approche classique de l'économie de la science, s'intéressant à l'efficience allocative des ressources dans l'activité de la recherche, et la sociologie de la science, dont Merton (1957, 1973) est l'un des fondateurs, notamment en ce qui concerne l'étude des normes sociales régissant la rémunération indirectement monétaire de l'activité de recherche (Partha et David, 1994 ; Stephan, 1996), où la publication d'articles scientifiques devient un actif valorisable sur le marché du travail des chercheurs (Heckman et Moktan, 2020). Traditionnellement, cette branche de l'économie a toujours accordé une grande importance à la relation entre normes sociales et légales, surtout en ce qui concerne l'analyse des brevets (Romer, 2010 ; Eisenberg, 1989). Cependant, une attention réduite a été accordée au droit d'auteur, à propos duquel on trouve relativement peu d'études dans ce sous-champ disciplinaire de l'économie (Rai, 1999).

Cette thèse contribue à combler cette lacune, en montrant l'interaction existante entre les deux systèmes de protection de la propriété intellectuelle – le droit et les normes sociales - et en soulignant l'importance du droit d'auteur qui perdure aujourd'hui encore à l'heure du numérique. En effet, les licences *Creative Commons*, permettant la libre circulation du savoir scientifique et des données de la recherche, ne peuvent se mettre en place qu'à partir du droit d'auteur, qui continue en ce sens à occuper un rôle important en matière de régulation de la diffusion de la connaissance sur le marché de la publication scientifique – même si ce rôle a pu évoluer.

Comme le montre la nouvelle économie de la science, les citations jouent un rôle fondamental dans la communauté scientifique, en tant qu'actifs intangibles, nécessaires pour l'avancement de carrière des chercheurs (Hamermesh et Pfann, 2009 ; Heckman et Moktan, 2020). Puisque l'*Open Access* augmente la probabilité pour un chercheur d'être cité, grâce à l'élimination de la barrière du prix à l'accès, on pourrait légtimement s'attendre à ce que ce modèle soit largement utilisé par les membres de la communauté scientifique. Cependant, de nombreuses études montrent que ce n'est pas le cas (voir par exemple Laasko et al., 2011 ; Laasko et Björk , 2012). En utilisant une base de données entièrement originale que nous créons à partir de l'archive ouverte HAL, nous étudions les stratégies sous-jacentes

au choix du *Green Open Access* des chercheurs, en prenant en considération le rôle joué par la qualité de la revue scientifique dans laquelle l'article a été préalablement publié. L'intérêt est de voir si le *Green Open Access* peut s'interpréter uniquement comme un instrument de marketing, permettant à un chercheur de mettre la communauté scientifique au courant des derniers résultats de la recherche, ou s'il est au contraire considéré comme un instrument offrant un accès différent à celui proposé par les maisons d'édition.

0.3 Présentation des chapitres, méthodologie et principaux résultats

Pour comprendre quelles sont les barrières à l'utilisation d'un modèle qui assure un libre accès à la connaissance scientifique à l'ère du numérique, la question de recherche est développée et décomposée en cinq chapitres.

Le **premier chapitre** propose une revue de la littérature qui s'interroge sur les différentes courants d'étude qui se sont intéressées à la régulation du marché de la publication scientifique. Il offre une vision d'ensemble du fonctionnement du système traditionnel de régulation de l'accès et de la circulation du savoir scientifique assuré par le droit d'auteur, et sa remise en cause par le mouvement de numérisation de l'information. Nous montrons que ces approches analysent de façon différente mais complémentaire le marché de la publication scientifique et que cette complémentarité est essentielle pour en comprendre le fonctionnement et l'impact du numérique.

En tant que bien collectif, la connaissance scientifique est soumise aux défaillances de marché. Son caractère non rival et non excluable empêche le marché en concurrence pure et parfaire d'atteindre un équilibre efficace, d'où la nécessité d'une intervention publique à travers l'octroi d'un droit de propriété intellectuelle. Nous allons reconstruire l'évolution de l'analyse économique du droit d'auteur et son application au marché de la publication scientifique. Plus précisément, nous allons montrer les spécificités du contrat d'édition dans ce marché et la structure en oligopole qui découle de la concentration au sein de quelques maisons d'édition des droits d'auteur.

La numérisation avait été accueillie par au moins une partie de la communauté scien-

tifique comme une possibilité de sortie concernant la crise engendrée par les stratégies de prix des maisons d'édition, qui avait conduit à une augmentation des prix des revues scientifiques bien au-delà de l'évolution des prix à la consommation (Bergstrom, 2001). Au contraire, la numérisation a permis aux maisons d'édition de garder des marges de profit très importantes (Larivière et al., 2014), à travers la mise en place de stratégies de discrimination de prix et de *bundling*. C'est dans ce contexte que le modèle traditionnel de la circulation du savoir, fondé sur le droit d'auteur, est remis en cause et que lors de la conférence de Budapest en 2002, l'*Open Access* devient une priorité dans le milieu scientifique.

De ce premier chapitre, nous tirons plusieurs enseignements importants. d'un côté, nous montrons que si les maisons d'édition ont su adapter leur modèle d'affaire aux nouvelles formes de circulation du savoir, tout en maintenant une activité éditoriale rentable, de l'autre côté, le modèle d'*Open Access* reste cependant encore, vingt ans après son idéation, globalement moins utilisé que le modèle de publication standard. Si les bénéfices d'un tel type de publication sont bien mis en évidence par la littérature, notamment en ce qui concerne l'augmentation des citations reçues par les auteurs, les causes de sa sous-utilisation restent nettement moins étudiées. Parmi ces causes, nous en identifions deux plus spécifiquement envisagées dans la littérature. D'une part, le droit d'auteur empêcherait le passage d'un mode de publication à l'autre (Shavell, 2010 ; Scheufen, 2015 ; Mueller-Langer et Watt, 2010). D'autre part, un niveau d'adhésion au mouvement *Open Access* caractériserait les sous-communautés scientifiques (Kling et McKim, 2000 ; Spezi et al., 2013 ; Eger et al., 2015).

Deuxièmement, la numérisation a aussi permis d'inscrire la connaissance scientifique dans la catégorie des nouveaux communs, biens ayant les caractéristiques des biens collectifs, mais dont la gestion est assurée par une communauté spécifique. Notre revue de la littérature met ainsi en évidence le rôle central joué par les normes sociales dans le fonctionnement de la communauté scientifique. Cells-ci semblent en effet jouer un rôle aussi important que le droit d'auteur dans la gestion de la circulation de la connaissance scientifique. Si une nouvelle littérature très prolifique sur les communs numériques s'est rapidement développée dans les années récentes, la connaissance scientifique n'est encore

que peu fréquemment étudiée sous cet angle, les normes légales et informelles régissant les choix de publication des chercheurs étant rarement analysées simultanément en tant que contraintes au comportement des chercheurs dans leurs choix de publication. C'est précisément le constat de cette aporie dans la littérature qui motive notre choix de proposer dans la thèse une analyse de l'impact de la numérisation du savoir scientifique sur la chaîne de valeur de la connaissance scientifique en tant que commun.

Le fonctionnement du marché de la publication scientifique a longtemps été analysé en suivant deux axes de recherche différents. D'un côté, l'analyse économique du droit, et plus précisément l'analyse économique de la propriété intellectuelle, a mis en évidence les barrières créées par le droit d'auteur à la libre circulation du savoir scientifique. De l'autre, la nouvelle économie de la science s'est focalisée sur les normes sociales qui contribuent au fonctionnement du marché de la publication scientifique. Le **deuxième chapitre** analyse les changements institutionnels ayant caractérisé l'évolution de ce marché dans le temps et la situation d'inertie qui le caractérise selon nous aujourd'hui, expliquant les obstacles actuels à une transition du régime de *Closed Access* vers le modèle d'*Open Access*. Pour ce faire, nous mobilisons simultanément la littérature de la nouvelle économie de la science et de l'analyse économique du droit d'auteur pour étudier l'interaction entre le système juridique du droit d'auteur et les normes sociales de la communauté scientifique. Nous mettons en évidence un processus de coévolution entre ces deux formes de régulation de la diffusion de la connaissance et expliquons par un phénomène d'influence réciproque les obstacles actuels à la formation d'un nouvel équilibre sur le marché de la publication scientifique ayant comme point focal le modèle en OA. Cette analyse nous amène à proposer une analyse nouvelle et originale de la persistance et de la dominance du régime de *Closed Access* sur le marché de la publication scientifique aujourd'hui.

La littérature a amplement interprété le droit d'auteur comme l'une des barrières principales au développement de l'*Open Access* (Shavell, 2010). Comme nous le verrons, la régulation du marché de la publication scientifique est assurée en même temps par des normes sociales (informelles) et juridiques (formelles), dont les interactions structurent et déterminent les choix effectifs de publication des individus. Dès lors, une réforme du système institutionnel régulant les échanges de connaissance sur le marché scientifique, à

travers par exemple une réforme du droit d'auteur, ne saurait être que partielle, puisqu'elle ne prendrait pas en compte le rôle des normes sociales dans le processus de régulation. Les tentatives des autorités publiques destinées à promouvoir l'utilisation de l'*Open Access* risquent dans ce contexte de ne produire que des effets limités, dès lors que d'autres barrières, peu prises en considération par la littérature, existent et influencent le comportement des chercheurs.

En France, la loi *Pour une république numérique* semble clairement inspirée par cette branche de l'analyse économique du droit qui pose le droit d'auteur au centre de la sous-utilisation de l'*Open Access* et recommande dans cette perspective une réforme du cadre juridique de ce droit de propriété intellectuelle. Comme nous l'avons déjà indiqué, cette loi accorde le pouvoir aux auteurs de mettre en libre accès leurs résultats scientifiques après la publication dans une revue scientifique. C'est cette vision que nous nous employons à questionner dans le **troisième chapitre**, à travers une analyse empirique visant à déterminer si une réforme du droit d'auteur est effectivement en mesure de promouvoir l'utilisation de l'*Open Access*. A cette fin, nous élaborons une base de données entièrement originale, construite à partir de l'archive ouverte HAL. Nous obtenons plusieurs résultats. Premièrement, dans le prolongement de la littérature existante (Kling et McKim, 2000 ; Spezi et al., 2013 ; Eger et al., 2015), nos résultats confirment l'existence de niveaux d'engagement différents dans l'utilisation des archives ouvertes en fonction des communautés scientifiques d'appartenance. Deuxièmement, nous montrons aussi qu'une réforme des normes juridiques n'est pas suffisante pour promouvoir une plus large utilisation de l'*Open Access* dans sa forme d'auto-archivage.

En outre, la réforme opérée par la loi *Pour une république numérique* a également un impact sur la période d'embargo, définie comme le délai à respecter entre la publication d'un article scientifique dans une revue et son téléchargement dans une archive ouverte. Le deuxième objectif du chapitre trois est par conséquent de tester l'effet de la loi sur cette période d'embargo. Nos résultats établissent une tendance à la diminution de ce délai suite à la mise en place de la loi. Ce résultat nous permet d'avancer certaines hypothèses pour expliquer pourquoi une réforme partielle du cadre institutionnel régissant le fonctionnement du marché de la publication scientifique n'a pas abouti aux résultats prévus théoriquement

par la littérature. En effet, la diminution du délai suggère que la loi a eu un impact positif sur ceux qui utilisaient les archives avant même la réforme du droit d'auteur. Cela nous amène à conclure que la réforme doit s'accompagner d'initiatives visant la promotion de l'*Open Access* au sein de la communauté scientifique, la sensibilisation à la question de la science ouverte et à l'explication des nouveaux droits qui découlent de la réforme du droit d'auteur.

Dans le prolongement de l'analyse précédente, le chapitre suivant de la thèse se focalise également sur le modèle du *Green Open Access* dont il approfondit l'analyse. Ce modèle, dont le coût d'adoption est très faible pour les chercheurs, permet l'élimination des barrières économiques à l'accès à la connaissance scientifique, tout en augmentant la probabilité pour les chercheurs d'être cités, grâce au plus large accès que cet instrument assure (Piwowar et al., 2018). Si le droit d'auteur n'est que l'une des barrières possibles à l'adoption de l'*Open Access*, quelles sont alors les autres possibles barrières qui en limitent l'utilisation ?

Le quatrième chapitre entend répondre à cette question en étudiant le comportement des chercheurs vis-à-vis de l'adoption du *Green Open Access* en fonction de la qualité de la revue dans laquelle l'article téléchargé sur HAL a été préalablement publié. L'objectif est de déterminer si la qualité de la revue peut influencer le choix d'utilisation du *Green Open Access* et si les comportements des chercheurs diffèrent en fonction de la qualité de leur article. Nous obtenons là aussi plusieurs résultats. Tout d'abord, de façon intéressante, nos résultats montrent que les téléchargements sont influencés par la qualité de la revue. Plus la qualité d'un article est élevée, plus il est susceptible d'être mis en libre accès sur HAL. De ce point de vue, il apparaît que le modèle *Green* est alors utilisé par les chercheurs comme un instrument de publicité pour augmenter l'attention de la communauté scientifique sur leurs travaux. On pourrait s'attendre à ce que le modèle du *Green* access soit surtout utilisé par les chercheurs qui ont publié des articles dans les revues dont la qualité est moins attestée, et dont l'accès à travers les bibliothèques universitaires est fortement limité à cause des contraintes budgétaires. Plus qu'un instrument complémentaire, l'*Open Access* est dans ce cas la seule voie dont disposent les auteurs pour assurer l'accès à leur production scientifique. A cet égard, notre second résultat montre que, quand on analyse les données en utilisant les classements internationaux (*Scimago Journal Rank et Clarivate's Journal*

Citation Report), la sous-utilisation du modèle d'*Open Access* par les chercheurs publiant des articles non-classés est plus marquée par rapport au scénario dans lequel on utilise les classements français (CNRS, HCERES et FNEGE). En définitive, le chapitre quatre nous permet donc de clarifier les différentes raisons poussant les chercheurs à adopter le modèle *Green* et d'établir que leur stratégie de dissémination des résultats scientifiques diffère en fonction de la qualité de leur output. Cette contribution à la littérature nous conduit à recommander la mise en place de politiques publiques destinées à corriger les comportements de sous-utilisation du *Green Open Access*.

Si le phénomène de numérisation de l'information a suscité un intérêt renouvelé pour la littérature sur les commons, son application au domaine de la connaissance scientifique reste une exception. C'est pour cette raison que nous proposons dans le cinquième chapitre une analyse de la connaissance scientifique en tant que commun, à travers l'application de l'IAD *framework* (Kiser et Ostrom, 1982 ; Ostrom, 2009), pour comprendre comment le changement technologique peut affecter la gestion d'une ressource. Ce chapitre analyse les enjeux de la transition d'une régulation fondée quasi-exclusivement sur le droit d'auteur vers une gestion de la ressource caractérisée par plusieurs exceptions à l'application du droit d'auteur.

À partir des différents modèles d'affaire adoptés par les maisons d'édition, nous développons une analyse des coûts et des bénéfices de chaque type de modèle d'*Open Access*, en soulignant leur impact sur les intérêts des différentes parties prenantes. Les interactions entre les agents économiques sont étudiées dans le cadre de la chaîne de valeur de la circulation de la connaissance scientifique, afin de montrer comme certaines stratégies d'adaptation et d'innovation survenues sur le marché de la publication scientifique en ont changé la structure.

Cette analyse globale de la chaîne de valeur nous permet de montrer comment des intérêts apparemment rivaux sur le marché de la publication scientifique ne le sont pas forcément et qu'il peut, au contraire, exister entre eux une forme de complémentarité, dès lors que l'on considère les différents nouveaux services proposés. Ce n'est donc pas une application du fonctionnement traditionnel du marché de la publication scientifique à l'ère du numérique, fondé sur le *Closed Access* qui doit orienter les prochaines étapes de la

régulation de ce marché, mais plutôt une collaboration existante mais renouvelée entre les principaux acteurs de ce marché et élargie aux nouveaux agents qui s'insèrent dans les parties prenantes du marché de la circulation du savoir scientifique.

Bibliographie de l'introduction générale

Baldwin, M. (2015). Credibility, peer review, and Nature, 1945–1990. *Notes and Records : the Royal Society Journal of the History of Science 69*(3), 337–352. Royal Society.

Benjamin, A. (2012). Harvard rejoint les universitaires pour un boycott des éditeurs. *Le Monde.fr*.

Bergstrom, T. C. (2001). Free Labour for Costly Journals ? *Journal of Economic Perspectives 15*(4), 183–198.

Bicchieri, C. (2014). Norms, Conventions, and the Power of Expectations. In *Philosophy of Social Science : A New Introduction* (Nancy Cartwright, Eleonora Montischi ed.)., pp. 208–229. Oxford University Press.

Björk, B.-C. (2017). Scholarly journal publishing in transition- from restricted to open access. *Electronic Markets 27*(2), 101–109.

Björk, B.-C., M. Laakso, P. Welling, and P. Paetau (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology 65*(2), 237–250.

Bourreau, M. and T. Pénard (2016). Introduction. L'économie numérique en question. *Revue d'economie industrielle n° 156*(4), 11–15. De Boeck Supérieur.

Boyle, J. (2003). The Second Enclosure Movement and the Construction of the Public Domain. *Law and Contemporary Problems 66*(1/2), 33–74. Duke University School of Law.

Boyle, J. (2007). Mertonianism Unbound ? : Imagining Free, Decentralized Access to Most Cultural and Scientific Material. In *Understanding Knowledge as a Commons : From Theory to Practice* (Charlotte Hess and Elinor Ostrom ed.)., pp. 123–144. The MIT Press.

Carroll, M. W. (2006). Creative Commons and the New Intermediaries W(h)ither the Middleman : The Role and Future of Intermediaries in the Information Age. *Michigan State Law Review 2006*(1), 45–66.

Cecchino, N. J. (2010). Google Scholar. *Journal of the Medical Library Association : JMLA 98*(4), 320–321.

Chen, D. H. C. and C. J. Dahlman (2005). The Knowledge Economy, the Kam Methodology and World Bank Operations. SSRN Scholarly Paper ID 841625, Social Science Research Network, Rochester, NY.

Cogo, A. (2010). *I contratti di diritto d'autore nell'era digitale*. Giappichelli.

Coriat, B. (2015). *Le retour des communs : & la crise de l'idéologie propriétaire.* Paris : Les Liens Qui Libèrent.

Damanpour, F. and M. Schneider (2006). Phases of the Adoption of Innovation in Organizations : Effects of Environment, Organization and Top Managers1. *British Journal of Management 17*(3), 215–236.

David, P. A. (1998). Common Agency Contracting and the Emergence of "Open Science" Institutions. *The American Economic Review 88*(2), 15–21. American Economic Association.

David, P. A. (2004). Can "Open Science" be Protected from the Evolving Regime of IPR Protections ? *Journal of Institutional and Theoretical Economics (JITE) 160*(1), 9–34. Tübingen Mohr Siebeck.

Dusollier, S. (2007). Sharing Access to Intellectual Property through Private Ordering. *Chicago-Kent Law Review 82*(3), 1391.

Edlin, A. S. and D. L. Rubinfeld (2004). Exclusion or Efficient Pricing - The Big Deal Bundling of Academic Journals Symposium : Integrating New Economic Learning with Antitrust Doctrine. *Antitrust Law Journal 72*(1), 119–158.

Edwards, R. and D. Shulenburger (2003). The High Cost of Scholarly Journals : (And What To Do About It). *Change : The Magazine of Higher Learning 35*(6), 10–19. Routledge.

Eger, T., M. Scheufen, and D. Meierrieks (2015, June). The determinants of open access publishing : survey evidence from Germany. *European Journal of Law and Economics 39*(3), 475–503.

Eisenberg, R. S. (1989). Patents and the Progress of Science : Exclusive Rights and Experimental Use. *The University of Chicago Law Review 56*(3), 1017–1086. University of Chicago Law Review.

Europe, S. (2019, August). An Analysis of Open Science Policies in Europe v4. Technical report.

Fauchart, E. and E. von Hippel (2008). Norms-Based Intellectual Property Systems : The Case of French Chefs. *Organization Science 19*(2), 187–201. INFORMS.

Feess, E. and M. Scheufen (2016). Academic copyright in the publishing game : a contest perspective. *European Journal of Law and Economics 42*(2), 263–294.

Foray, D. (2000). *L'économie de la connaissance* (La Découverte ed.). Repères.

Fortson, D. (2019). A Napster moment for ivory towers. *Sunday Times (Business Section)*.

Fry, J., R. Schroeder, and M. den Besten (2009, January). Open science in e-science : contingency or policy ? *Journal of Documentation 65*(1), 6–32. Emerald Group Publishing Limited.

Fyfe, A. (2015). Journals, learned societies and money : Philosophical Transactions, ca. 1750–1900. *Notes and Records : the Royal Society Journal of the History of Science 69*(3), 277–299. Royal Society.

Gordon, G. (2016). SSRN—the leading social science and humanities repository and online community—joins Elsevier.

Grand, A., C. Wilkinson, K. Bultitude, and A. F. T. Winfield (2016, January). Mapping the hinterland : Data issues in open science. *Public Understanding of Science 25*(1), 88–103. SAGE Publications Ltd.

Green, T. (2018, September). We're Still Failing to Deliver Open Access and Solve the Serials Crisis : To Succeed We Need a Digital Transformation of Scholarly Communication Using Internet-Era Principles. SSRN Scholarly Paper ID 3333907, Social Science Research Network, Rochester, NY.

Guédon, J. Le libre accès et la « Grande conversation » scientifique. In *Pratiques de l'éditionnumérique* (Michaël E. Sinatra & Marcello Vitali-Rosai ed.)., pp. 111–126. Montréal : Presses de l'Université de Montréal.

Hamermesh, D. S. and G. A. Pfann (2009). Markets for Reputation : Evidence on Quality and Quantity in Academe. Working Paper 15527, National Bureau of Economic Research. Series : Working Paper Series.

Heckman, J. J. and S. Moktan (2020). Publishing and Promotion in Economics : The Tyranny of the Top Five. *Journal of Economic Literature 58*(2), 419–470.

Hess, C. and E. Ostrom (2011). *Understanding Knowledge as a Commons : From Theory to Practice*. MIT Press.

Hilty, R., S. Krujatz, B. Bajon, A. Frueh, A. Kur, J. Drexl, C. Geiger, and N. Klass (2008). European Commission - Green Paper : Copyright in the Knowledge Economy - Comments by the Max Planck Institute for Intellectual Property, Competition and Tax Law. SSRN Scholarly Paper ID 1317730, Social Science Research Network, Rochester, NY.

Himmelstein, D. S., A. R. Romero, J. G. Levernier, T. A. Munro, S. R. McLaughlin, B. Greshake Tzovaras, and C. S. Greene (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife 7*, e32822. eLife Sciences Publications, Ltd.

Houghton, J. and P. Sheehan (2009, March). Estimating the Potential Impacts of Open Access to Research Findings. *Economic Analysis and Policy 39*(1), 127–142.

Johnson, R., M. Fosci, A. Chiarelli, S. Pinfield, and M. Jubb (2017). Towards a Competitive and Sustainable OA Market in Europe - A Study of the Open Access Market and Policy Environment. Technical report, Research Consulting, OpenAire.

Kellogg, D. (2006). Toward a Post-Academic Science Policy : Scientific Communication and the Collapse of the Mertonian Norms. SSRN Scholarly Paper ID 900042, Social Science Research Network, Rochester, NY.

Kiser, L. and E. Ostrom (1982). The three worlds of action : a metatheoretical synthesis of institutional approaches. *Strategies of Political Inquiry*, 179–222.

Kling, R. and G. McKim (2000). Not just a matter of time : Field differences and the shaping of electronic media in supporting scientific communication. *Journal of the American Society for Information Science 51*(14), 1306–1320.

Kwon, D. (2018). French Universities Cancel Subscriptions to Springer Journals. *The Scientists*.

Kyrillidou, M. and M. Young (2008). *ARL Statistics, 2005-06 : A Compilation of Statistics from the One Hundred and Twenty-Three Members of the Association of Research Libraries*. Association of Research Libraries. Publication Title : Association of Research Libraries.

Laakso, M. and B.-C. Björk (2012). Anatomy of open access publishing : a study of longitudinal development and internal structure. *BMC Medicine 10*(1), 124.

Laakso, M., P. Welling, H. Bukvova, L. Nyman, B.-C. Björk, and T. Hedlund (2011). The Development of Open Access Journal Publishing from 1993 to 2009. *PLOS ONE 6*(6), e20961. Public Library of Science.

Larivière, V., S. Haustein, and P. Mongeon (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE 10*(6), e0127502. Public Library of Science.

Leonelli, S., D. Spichtinger, and B. Prainsack (2015). Sticks and carrots : encouraging open science at its source. *Geo : Geography and Environment 2*(1), 12–16.

Lewis, D. (1969). *Convention : A Philosophical Study*. John Wiley & Sons.

Loshin, J. (2007). Secrets Revealed : How Magicians Protect Intellectual Property without Law. SSRN Scholarly Paper ID 1005564, Social Science Research Network, Rochester, NY.

McAdams, R. H. (2001). Signaling Discount Rates : Law, Norms, and Economic Methodology Book Review. *Yale Law Journal 110*(4), 625–690.

McCabe, M. J. and C. M. Snyder (2018). Open Access as a Crude Solution to a Hold-Up Problem in the Two-Sided Market for Academic Journals. *The Journal of Industrial Economics 66*(2).

Menell, P. S. (2016). Economic Analysis of Copyright Notice : Tracing and Scope in the Digital Age. *Boston University Law Review 96*, 967.

Merton, R. K. (1957). Priorities in Scientific Discovery : A Chapter in the Sociology of Science. *American Sociological Review 22*(6), 635–659. American Sociological Association, Sage Publications, Inc.

Mueller-Langer, F. and R. Watt (2010). Copyright and Open Access for Academic Works. SSRN Scholarly Paper ID 1647586, Social Science Research Network, Rochester, NY.

OCDE (2020). Why open science is critical to combatting COVID-19.

OECD (2016). *Open Government The Global Context and the Way Forward : The Global Context and the Way Forward*. OECD Publishing.

Origgi, G. (2010). Epistemic Vigilance and Epistemic Responsibility in the Liquid World of Scientific Publications. *Social Epistemology 24*(3), 149–159. Routledge.

Ostrom, E. (2009). The Institutional Analysis and Development Framework and the Commons Response. *Cornell Law Review 95*(4), 807–816.

Ovadia, S. (2014). ResearchGate and Academia.edu : Academic Social Networks. *Behavioral & Social Sciences Librarian 33*(3), 165–169. Routledge.

Partha, D. and P. A. David (1994, September). Toward a new economics of science. *Research Policy 23*(5), 487–521.

Pells, R. (2020). Coronavirus and Ebola : could open access medical research find a cure ?

Piwowar, H., J. Priem, V. Larivière, J. P. Alperin, L. Matthias, B. Norlander, A. Farley, J. West, and S. Haustein (2018). The state of OA : a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ 6*, e4375. PeerJ Inc.

Plume, A. and D. van Weijen (2014). Publish or perish ? The rise of the fractional author... - Research Trends. *Research Trends* (38).

Rai, A. K. (1999). Regulating Scientific Research : Intellectual Property Rights and the Norms of Science. *Northwestern University Law Review 94*(1), 77–152.

Romer, P. (2010). Technologies, Rules, and Progress : The Case for Charter Cities. Working Paper, eSocialSciences.

Scheufen, M. (2015). *Copyright Versus Open Access : On the Organisation and International Political Economy of Access to Scientific Knowledge*. International Law and Economics. Springer International Publishing.

Schiermeier, Q. (2017). US court grants Elsevier millions in damages from Sci-Hub. *Nature*. Nature Publishing Group.

Schmidt, B., A. Orth, G. Franck, I. Kuchma, P. Knoth, and J. Carvalho (2016, June). Stepping up Open Science Training for European Research. *Publications 4*(2), 16. Multidisciplinary Digital Publishing Institute.

Shavell, S. (2009, December). Should Copyright of Academic Works be Abolished ? SSRN Scholarly Paper ID 1525667, Social Science Research Network, Rochester, NY.

Solow, R. M. (1967). Some Recent Developments in the Theory of Production. In *NBER Chapters*, pp. 25–53. National Bureau of Economic Research, Inc.

Spezi, V., J. Fry, C. Creaser, S. Probets, and S. White (2013, January). Researchers' green open access practice : a cross-disciplinary analysis. *Journal of Documentation 69*(3), 334–359. Emerald Group Publishing Limited.

Stephan, P. E. (1996). The Economics of Science. *Journal of Economic Literature 34*(3), 1199–1235. American Economic Association.

Suber, P. Creating an Intellectual Commons through Open access. In *Understanding Knowledge as a Commons : From Theory to Practice* (The MIT Press ed.)., pp. 171–208. The MIT Press.

Suber, P. (2012). *Open Access*. MIT Press.

Sveinsdottir, T., V. Proudman, and J. Davidson (2020). An Analysis of Open Science Policies in Europe, v6.

Swan, A. (2013). *Principes directeurs pour le développement et la promotion du libre accès*. Paris : Organisation des Nations Unies pour l'éducation, la Science et la culture.

Vicente-Saez, R. and C. Martinez-Fuentes (2018, July). Open Science now : A systematic literature review for an integrated definition. *Journal of Business Research 88*, 428–436.

Vittu, J.-P. (2002). La formation d'une institution scientifique : le Journal des Savants de 1665 à 1714. *Journal des Savants 2*(1), 349–377. Persée - Portail des revues scientifiques en SHS.

Ware, M. and M. Mabe (2015). The STM Report : An overview of scientific and scholarly journal publishing. *Copyright, Fair Use, Scholarly Communication, etc.*.

Waters, D. J. Preserving the Knowledge Commons. In *Understanding Knowledge as a Commons : From Theory to Practice* (Charlotte Hess and Elinor Ostrom ed.)., pp. 145–168. The MIT Press.

Willinsky, J. (2009). The Stratified Economics of Open Access. *Economic Analysis and Policy 39*(1), 53–70.

Ziman, J. (2000). *Real Science : What it Is and What it Means.* Cambridge : Cambridge University Press.

Annexe

1. Domaines de la recherche scientifique analysant la science ouverte

La communauté scientifique s'est largement intéressée à l'analyse de l'*Open Science*. En utilisant les données de Web of Science, nous pouvons voir (Figure 2) les différentes disciplines qui, dans leur ensemble, produisent 70% des articles scientifiques portant sur la science ouverte. Le 30% restant regroupe plus de 100 domaines de la recherche scientifique, produisant au moins un article sur le sujet. Ce qui est intéressant est de voir la pénétration de l'*Open Science* dans un nombre surprenant de disciplines scientifiques, un sujet qui touche non seulement à la façon de faire de la science, mais qui devient aussi un sujet lui-même de recherche scientifique, qui ne connaît pas de frontières entre les différentes sous-disciplines de la communauté scientifique.

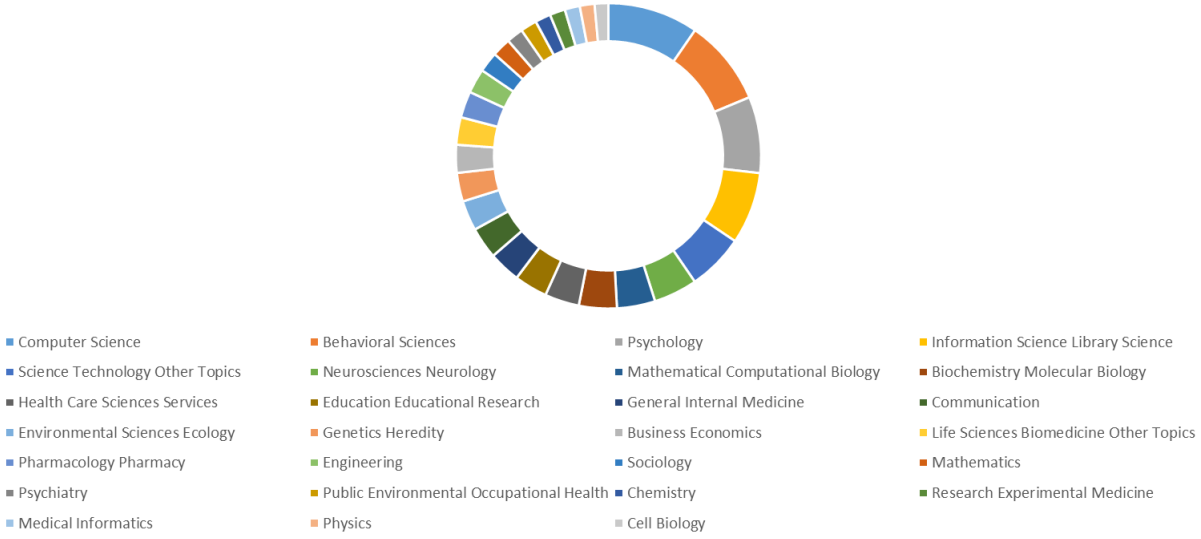

FIGURE 2 – Disciplines analysant l'*Open Science*, données Web of Science - all databases

2. Principales politiques publiques des pays de l'Union européenne pour la promotion de la science ouverte

Pays	Politiques publiques des pays membres de l'UE
AT	Open Access Policy: Open Access to Research Data (Policy of national funder FWF)
BE	Code of Ethics
CY	National Policy of the Republic of Cyprus for Open Access to Scientific Information
CZ	Action Plan for Implementation of the National Strategy on Open Access to Scientific Information of the Czech Republic for 2017–2020
DK	National Strategy for Open Access
DE	Principles for the Handling of Research Data (Policy of Alliance of German Science Organisations) and the DFG Guidelines on the Handling of Research Data
ES	State Plan for Research, Development and Innovation 2017-2020 in January 2018
FR	National Plan for Open Science
FI	Open Science and data - Action Programme for the Finnish Scholarly Community
LT	Guidelines on Open Access to Scientific Publications and Data (Policy of Research Council of Lithuania)
NL	National Plan Open Science
PT	Policy on management and sharing of data and other results arising from FCT-funded research (Policy of national funder FCT)
SI	National Strategy of Open Access to Scientific Publications and Research Data in Slovenia 2015- 2020
SK	The Open Government Partnership National Action Plan of the Slovak Republic 2017 – 2019 includes Open Science.
UK	UKRI Common Principles on Data Policy (Policy of UK national funder organisation UK Research and Innovation) Concordat on Open Research Data (Policy of a UK multi-stakeholder group, including research funders and higher education associations)
Pays	Politiques publiques des pays membres de l'UE - en cours de réalisation
BG	The National Science Strategy (2017-2030)
HR	Research and Innovation Infrastructures Roadmap
CZ	The Czech National Action Plan on OA
EE	Recommendations for the development of a national policy
EL	Re-formulation of a National Open Science High Level Task Force
HU	Joint committee on open science
IE	National Action Plan on the Transition to an Open Research Environment
IT	New Working group ICDI (Italian Computing and Data Infrastructure)
LU	National open access to science plan f
MT	Open Access/Open Science policy
PL	New OA policy
PT	National Open Science Policy
RO	Proposal for an Open Science national strategy
SE	National open research data initiative
Pays	Pays membres de l'UE sans politiques publiques
LV	No policy for the open science implementation

TABLE 1 – Politiques publiques pour l'*Open Science* des pays membres de l'Union européenne - source des données « An Analysis of Open Science Policies in Europe, v4 », SPARC (2019)

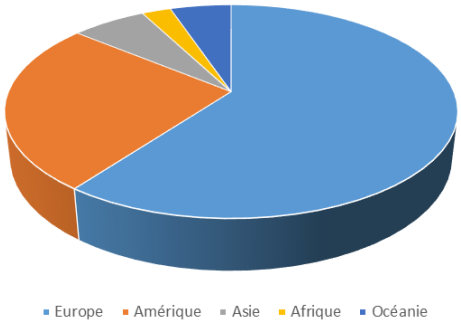

FIGURE 3 – Politiques publiques pour la promotion de l'*Open Access* - source des données ROARMAP

Chapitre 1

La régulation de l'accès et de la circulation de la connaissance scientifique à l'ère du numérique : un état de l'art

Résumé

Le régime d'ordre public a longtemps assuré, à travers le droit d'auteur, le fonctionnement du marché de la publication scientifique. Cette revue de la littérature offre une analyse du fonctionnement du système traditionnel de la régulation de l'accès au savoir scientifique, fondé sur le droit d'auteur, et sa remise en cause suite à la numérisation de l'information.

L'*Open Access*, en tant que système visant la subversion du régime propriétaire, permet alors d'inscrire la recherche scientifique dans le domaine des communs, où une forme renouvelée de gestion opérée par la communauté scientifique peut mieux assurer l'accès, la circulation et la réutilisation des résultats de la recherche.

Introduction

Le savoir, élaboration intellectuelle de l'information, est l'un des facteurs de la croissance économique et un pilier principal de l'économie de la connaissance, définie comme la branche de l'économie fondée sur des activités intensives en savoir (OCDE, 1996). Ces activités émergent à partir de l'interaction des agents, avec un effort commun pour coproduire de la nouvelle connaissance (David et Foray, 2002). Si la relation existant entre l'avancement scientifique et technologique et la croissance économique constitue depuis plusieurs décennies une interrogation centrale de l'analyse économique, elle n'est pas remise en cause et il est admis que la croissance ne peut être atteinte qu'à travers l'accès à de la nouvelle connaissance produite et son utilisation.

Comme pour tout bien collectif, la production de nouvelle connaissance est sujette aux défaillances de marché engendrées par la nature économique de ce bien. La connaissance se définit en effet comme un bien non-rival, dont la consommation par un individu n'empêche pas sa consommation par un autre agent, et non-excluable, dans la mesure où il est impossible ou très coûteux d'exclure les consommateurs qui n'ont pas payé de la jouissance du bien. Ces propriétés empêchent les producteurs de couvrir les coûts de production et de distribution, ce qui crée un effet dissuasif à la production du savoir (Arrow, 1962 ; Nelson, 1959). L'incapacité du marché en concurrence pure et parfaite à produire un niveau

optimal d'innovation et de créativité requiert ainsi une intervention publique, afin de créer des incitations capables d'assurer l'équilibre du marché.

En tant que forme particulière de connaissance, la connaissance scientifique est soumise à ces mêmes défaillances de marché, d'où la nécessité d'une intervention publique pour assurer un niveau optimal d'output de recherche par la communauté scientifique, une « *delicate piece of social and institutional machinery* » (Partha et David, 1994 p. 489). Une première solution consiste en un financement direct de la production de la recherche scientifique. À travers les impôts, l'État est en mesure de faire contribuer l'ensemble de la population au financement de la recherche scientifique. L'allocation de dépenses publiques au financement des laboratoires de recherche, les appels à projet et les salaires versés aux chercheurs et enseignant-chercheurs en sont des exemples, qui s'inscrivent dans la théorie du niveau optimal de dépenses publiques de Samuelson (1954). En 1959, Nelson souligne que l'exigence d'un tel financement public devient encore plus centrale lorsqu'on considère la recherche scientifique fondamentale, dont les résultats ne seront économiquement exploitables que sur la longue période et pour laquelle les risques liés à l'incertitude sont le plus souvent plus élevés que ceux associés à la recherche appliquée.

La deuxième solution aux défaillances de marché classiquement préconisée en économie publique consiste dans la production privée du bien collectif, à travers l'octroi d'un droit de propriété (Coase, 1960). L'allocation d'un tel droit recouvre un rôle clé dans le cas de la connaissance, si bien qu'il a conduit à la mise en place d'un nouveau système de normes juridiques. Les droits de propriété intellectuelle (DPI) peuvent ainsi être considérés dans leur ensemble comme une institution visant la coordination des agents, en réduisant l'incertitude liée aux interactions humaines et donc le coût de la coopération (North, 1990), dans le cadre de la production de nouvelle connaissance. Les brevets et le droit d'auteur permettent dans ce contexte d'internaliser les bénéfices dérivant de la production d'un bien collectif, à travers l'octroi d'un pouvoir de monopole sur le bien sous-jacent. Ces droits de propriété rendent excluables l'accès et l'utilisation du bien et permettent au producteur d'imposer un prix à la jouissance du bien, afin de couvrir les coûts de production. Ils rendent de ce fait rentable la production de nouvelle connaissance.

À titre d'exemple, le brevet permet d'internaliser les profits dérivant de la production

d'une invention. Grâce à l'octroi d'un monopole, qui s'entend sur un espace temporel et spatial déterminé, le titulaire peut interdire aux parties tierces toute réutilisation de son produit, sans son autorisation préalable. Ce pouvoir exclusif lui permet de couvrir les coûts fixes très élevés dérivant de la production de l'invention, à travers la vente exclusive du bien ou la cession d'un droit d'exploitation secondaire à une ou plusieurs contreparties. Si la négociation du brevet et la valeur économique qu'il engendre sont évidentes (on peut penser à titre d'exemple à la production d'un nouveau vaccin), le rôle du droit d'auteur appliqué à la connaissance scientifique mérite d'être détaillé. Le caractère collectif de l'avancement de la science, qui se fait « sur les épaules des géants » (Newton et Hooke, 1675), n'est possible qu'à travers la dissémination, et donc la publication, des derniers résultats scientifiques. S'il est vrai que le financement public de la recherche est essentiel pour la production de nouvelle connaissance, ce n'est qu'à travers le droit d'auteur qu'elle peut être disséminée au sein de la communauté scientifique. Pour disséminer les résultats de leurs recherches, les chercheurs recourent à des maisons d'édition spécialisées dans la production de revues scientifiques. En tant que titulaires du droit d'auteur, les chercheurs cèdent aux maisons d'édition les droits économiques incorporés dans ce droit de propriété, et notamment le droit exclusif de reproduction de l'œuvre, en contrepartie de la publication de leurs résultats. Les maisons d'édition, qui deviennent ainsi les nouvelles titulaires du droit d'auteur, sont alors en mesure d'appliquer un prix à la consommation du bien et de rémunérer leurs investissements. Du côté de l'auteur, la publication lui permet, le cas échéant, d'être reconnu comme un contributeur à l'avancement du savoir scientifique et de bénéficier de l'estime des pairs, à travers le système des citations.

Cette réallocation du droit de la propriété intellectuelle recouvre aussi d'autres fonctions. En effet, le caractère de bien d'expérience de la nouvelle connaissance produite ne permet de connaître sa qualité, et donc sa valeur dans le processus d'avancement de la science, qu'*a posteriori*, à savoir après avoir consommé le bien. Après vérification de la véracité scientifique de la nouvelle connaissance produite à travers des mécanismes institutionnalisés, tels que la relecture des articles par les comités de lecture avant leur publication (Merton, 1973 ; David et Foray, 2002), la publication permet de réduire les asymétries d'information existant entre producteur et consommateurs. Selon la qualité de la revue dans

laquelle l'article est publié, les chercheurs-consommateurs peuvent déterminer a priori la qualité de l'article et orienter leur consommation de connaissance. C'est ainsi que sur le marché de la publication scientifique, défini comme le lieu d'échange du savoir scientifique sous forme d'articles, considérés comme le standard de la communication scientifique (Origgi, 2010), le droit d'auteur accomplit les deux fonctions typiques des droits de propriété intellectuelle. D'un côté, il fournit une incitation économique à la circulation du savoir scientifique, considérée comme une étape faisant partie de la production du bien collectif de la connaissance, et de l'autre il réduit les asymétries d'informations à travers un signal de qualité (Demsetz, 1974 ; Benkler, 2002).

Néanmoins, l'octroi d'un monopole doit être considéré comme un optimum de second rang, puisque les droits de la propriété intellectuelle représentent un *trade-off* entre efficacité statique et dynamique. L'absence d'un droit exclusif d'exploitation sur le bien intellectuel amènerait certes à un niveau de production sous-optimal. Pour autant, l'instauration d'un monopole engendre une perte sèche pour la société. Cette perte de bien-être dans le cas de la connaissance est double. D'un côté, les prix imposés sont supérieurs à ceux de concurrence pure et parfaite. De l'autre, l'exploitation de la ressource par d'autres acteurs se trouve de ce fait limitée et cela pose une barrière importante au développement de nouvelle connaissance, étant donné le caractère cumulatif du savoir. La non-perpétuité du droit d'auteur, en ce qui concerne les intérêts économiques, sert justement à compenser les effets négatifs, en termes d'efficacité dynamique, de manière à maximiser le bien-être social sur la longue période (Nordhaus, 1969) et, une fois le monopole expiré, le bien intellectuel sous-jacent revient à sa nature de bien collectif.

La numérisation de l'information a profondément changé le mode de fonctionnement de la circulation du savoir scientifique, en affectant toute la chaîne de valeur. L'élimination du support papier a permis aux maisons d'édition, en tant qu'entreprises éditoriales, de diminuer leurs coûts de production et donc d'augmenter leurs profits. Quant à la communauté scientifique, elle a interprété l'âge du numérique comme une opportunité pour élargir l'accès aux résultats scientifiques et pour concevoir d'autres modalités de diffusion du savoir. C'est ainsi qu'en 2002, lors de la Conférence de Budapest, l'*Open Access* (OA), en tant que nouveau modèle pour la diffusion du savoir scientifique, voit le jour, avec pour objectif

premier l'élimination de toute barrière à l'accès à la littérature scientifique, qu'elle soit de nature financière ou légale (Suber, 2012). Ce modèle se différencie du modèle traditionnel, dit *Closed Access* (CA), en ce sens qu'il n'impose aucun prix à l'accès à la connaissance scientifique et pose, à travers des licences *copyleft*, des exceptions au droit d'auteur en ce qui concerne la redistribution et la réutilisation de l'œuvre. Il remet donc en cause la légitimité du régime propriétaire exploité par les maisons d'édition, eu égard notamment à la perte sèche et aux conséquences négatives en termes d'efficacité dynamique entraînées par le droit d'auteur. C'est ainsi que l'*Open Access*, en tant que mouvement, vise à renverser la logique propriétaire fondée sur les droits de la propriété intellectuelle (Dusollier, 2007), en proposant un nouveau schéma de partage et donc de gestion de la propriété.

C'est dans ce contexte institutionnel nouveau que la numérisation a permis d'inscrire la connaissance scientifique dans la catégorie des communs (ou *commons* dans la littérature anglophone), biens ayant les caractéristiques des biens collectifs, mais dont la gestion de l'accès et de l'utilisation de la ressource est assurée par une communauté spécifique (Hess, 2008 ; Hess et Ostrom, 2007). C'est ainsi que, grâce à la numérisation, des nombreuses initiatives ont été mises en place, telles que ArXiv et PLOS One, mais aussi la promotion des licences *copyleft* dans le domaine de la recherche scientifique, afin de promouvoir une gestion du savoir scientifique alternative à celle fondée sur le droit d'auteur et représentée par le modèle de *Closed Access*. D'un tel système de type ordre privé [1], on attend alors notamment une meilleure exploitation des possibilités offertes par la numérisation de la connaissance, en permettant une plus grande réutilisation et donc une valorisation socio-économique plus importante, en faveur d'une efficacité dynamique accrue, par rapport au *trade-off* imposé par le droit d'auteur. Plus particulièrement, en s'appuyant sur le droit d'auteur même, ce nouveau système de gestion prévoit un ensemble de licences standardisées, permettant de rendre légitimes certaines actions qui ne le seraient autrement pas dans le système traditionnel prévu par le droit d'auteur. Elles permettent, par exemple, de redistribuer l'œuvre sans demander une autorisation préalable au titulaire du droit d'auteur, étape requise dans le système légal traditionnel, et favorisent ainsi la circulation du savoir.

1. Les mécanismes d'ordre privé se basent sur des arrangements contractuels qui prévoient des exceptions au système légal. Tel est le cas, par exemple, des licences créatives, qui prévoient des exceptions à l'application du droit d'auteur, notamment en ce qui concerne la redistribution de l'œuvre protégée.

Dans ce chapitre, nous proposons une revue de la littérature des différents modes de régulation juridique et économique de l'accès à la connaissance scientifique à travers le marché de la publication sous forme d'articles scientifiques. Cette revue nous permet d'analyser les différents courants d'étude qui, de façon différente mais complémentaire, se sont intéressés à la régulation de ce marché spécifique, sur lequel interviennent notamment les chercheurs à la fois en tant que producteurs de connaissance et en tant que principaux consommateurs de savoir scientifique. Dans une première section, nous nous focalisons sur l'analyse économique du droit d'auteur, qui a constitué la branche dominante de la littérature consacrée au marché de la publication scientifique jusqu'à l'avènement de la numérisation de l'information. Parce que cette institution en tant que système principal d'incitations à la production et diffusion des œuvres de l'esprit rencontrait déjà avant l'ère du numérique certaines limites, nous en exposons également les faiblesses. La deuxième section est alors consacrée à l'*Open Access* en tant que nouvelle modalité d'accès au savoir scientifique, mais aussi de circulation de ce savoir, en suivant deux axes de présentation. D'un côté, nous examinons les contributions de l'économie industrielle à l'étude des différentes formes d'OA et des différents modèles d'affaires sous-jacents à sa mise en place. De l'autre, nous montrons en quoi l'économie de la science permet d'étudier d'une part les préférences des chercheurs en matière de choix de publication et, d'autre part, le système d'incitations nécessaire au basculement effectif du système de *Closed Access* vers le modèle ouvert d'OA. Dans un dernier temps, le marché de la publication scientifique est analysé à travers la littérature portant sur les communs, à partir des travaux pionniers de Hess et Ostrom (2007). Les communs sont dans ce cadre analysés comme une forme de gestion alternative du droit d'auteur et à l'ordre public qu'il représente, permettant de mieux répondre aux exigences des acteurs du marché de la publication scientifique – chercheurs, maisons d'édition et centres de recherche – à l'ère du numérique [2]

2. Pour un résumé de ces différents courants d'étude, voir en annexe le tableau récapitulatif 1.5.

1.1 La régulation de l'accès à la connaissance scientifique à travers le droit d'auteur

Le droit d'auteur a constitué, depuis la deuxième guerre mondiale, l'institution principale favorisant la dissémination des résultats scientifiques à travers la publication de revues spécialisées. Dans cette section, nous définissons cette modalité particulière de droit de propriété intellectuelle, en insistant sur ses différences et similitudes par rapport au *copyright*, forme de droit de propriété en vigueur dans les régimes juridiques de *Common Law*. Nous nous attachons également à décrire l'évolution de l'analyse économique du droit d'auteur et les spécificités de l'application du contrat d'édition sur le marché de la publication scientifique, en soulignant son rôle dans la structuration de ce marché autour d'un oligopole de quelques maisons d'édition.

La numérisation de l'information, souvent associée à une idée de partage et d'accès libre, s'est pourtant dans les faits accompagnée d'un durcissement de la protection garantie par les droits de propriété intellectuelle. Les maisons d'éditions ont en effet montré leur capacité à exploiter le pouvoir de monopole garanti par le droit d'auteur pour accroître leurs marges de profit et limiter le libre accès à la littérature scientifique. Cette situation a contribué à la remise en cause de ce régime juridique, qui est aujourd'hui au centre du débat de l'économie du droit.

1.1.1 Le cadre juridique du droit d'auteur et du *copyright*

Même si la régulation du droit des œuvres de l'esprit reste très fragmentée et spécifique à chaque pays (Scheufen, 2015), deux grandes approches peuvent être distinguées au niveau global, sur la base de justifications philosophiques opposées. Dans les pays de *Common Law*, le *copyright* trouve sa raison d'être dans l'école utilitariste de Bentham. Cette approche met en avant le bien-être de la société qui doit dans cette perspective accorder un droit exclusif sur les profits dérivant de la création de nouvelles idées et de travaux créatifs utiles à la société (David, 1993). Dans ce contexte, le *copyright* permet seulement à son titulaire de tirer des copies légitimes et donc d'exploiter économiquement son œuvre de façon légale, mais le produit final est considéré comme un produit de la société qui s'inscrit

dans un processus collectif de création. Si cette première approche se fonde donc sur la protection des intérêts économiques garantis à l'auteur, afin de lui donner des incitations à la création de nouvelles œuvres de l'esprit, dans les pays de *Civil Law* l'accent est mis sur la contribution de chaque individu à l'avancement de la science et de l'art (Rigamonti, 2007). En s'inscrivant dans le courant philosophique du droit naturel de Locke, le droit d'auteur est alors considéré comme un droit sur lequel l'auteur a une titularité par nature. Le droit d'auteur englobe alors non seulement les droits économiques assurés par le *copyright*, mais aussi une dimension morale absente dans les systèmes de *Common Law*. Le droit d'auteur recouvre ainsi un ensemble de droits qui n'ont pas été conçus pour protéger des intérêts économiques. C'est au contraire véritablement la sphère morale qui différencie et oppose le *copyright* au droit d'auteur.

En réalité, ces divergences se sont aujourd'hui estompées. Bien qu'il existe encore des différences entre les deux systèmes, celles-ci portent plus sur les justifications philosophiques à leur origine que sur le résultat final de la protection accordée (Goldstein, 2001 ; Ginsburg, 1990). L'opposition entre les deux systèmes devient dès lors de plus en plus artificielle, puisque le droit moral a été progressivement intégré dans les systèmes de *Common Law* (Benhamou et Farchy, 2009), et cela grâce à la prolifération des accords internationaux. L'émergence de plusieurs systèmes de protection des œuvres de l'esprit a en effet demandé la création d'une législation commune, de manière à limiter l'incertitude juridique à laquelle auteurs et éditeurs se trouvaient faire face (Scheufen, 2015). Initialement, on a ainsi assisté à la création d'accords bilatéraux entre pays. Mais c'est surtout grâce à la convention de Berne en 1886 qu'un niveau commun de protection juridique a été accordé aux œuvres de l'esprit. Modifiée et renégociée à plusieurs reprises, cette convention, qui compte aujourd'hui 166 pays signataires, établit en effet un niveau minimal de protection juridique auquel tous les pays signataires doivent se conformer. L'article 2 de la convention de Berne 1971[3] (modifié en 1979) accorde ainsi spécifiquement une protection juridique aux œuvres dérivant des domaines littéraires, scientifiques et artistiques. La seule condition de la protection est l'originalité de l'œuvre, qui doit être le résultat de la création intel-

3. La Convention de Berne a été stipulée en 1886 et modifiée à plusieurs reprises : 1895 (Paris), 1908 (Berlin), 1928 (Rome), 1948 (Brussels), 1967 (Stockholm), 1971 (Paris) et 1979 (Berne).

lectuelle et personnelle de l'auteur [4]. Comme spécifié à l'article 9 alinéa 2 des TRIPS [5], la protection est assurée seulement à l'expression de l'œuvre, et non pas aux idées contenues dans l'œuvre même.

Ce système de protection juridique accorde aux auteurs le droit exclusif de copier, reproduire, redistribuer, adapter, exécuter et montrer leurs œuvres. L'octroi de ce droit se traduit alors par un pouvoir de monopole à travers lequel l'auteur est rémunéré des investissements qu'il a consacrés à la production et distribution de l'œuvre intellectuelle (Maurer et Scotchmer, 2004). Tous les droits sont réservés aux titulaires et l'accès et la réutilisation de la ressource intellectuelle sont soumis à l'autorisation préalable du titulaire.

En définitive, malgré le fait que les systèmes juridiques nationaux restent les principaux responsables de la conception de la protection des œuvres de l'esprit et qu'un droit international n'a pas encore vu le jour (Scheufen, 2015), on a cependant assisté à un rapprochement des deux systèmes juridiques de *Common* et *Civil Law*, grâce aux traités mentionnés ci-dessus, tant en ce qui concerne les droits moraux que ceux économiques.

Notre thèse porte plus spécifiquement sur les aspects économiques du droit de propriété intellectuelle. Bien conscient des différences sémantiques entre les deux termes, on utilisera par conséquent les concepts de droit d'auteur et de *copyright* de façon interchangeable. Ce faisant, nous nous alignons sur la pratique de la Convention de Berne, qui utilise indifféremment les deux termes pour faire référence au même concept juridique dans ses versions en français et anglais en désignant ainsi, par les termes de *copyright* et droit d'auteur, à la fois la protection des œuvres littéraires et artistiques prévue par la Convention de Berne.

1.1.2 L'évolution de l'analyse économique du droit d'auteur

Une revue de la littérature sur l'analyse économique du *copyright* n'est pas une tâche simple à accomplir, à cause de la richesse de ce domaine de recherche et des différents axes de recherche dont se revendiquent les différents auteurs (Towse et al., 2008). Cependant, une distinction fondamentale s'est imposée comme un standard dans cette littérature, opérant

4. Pour les systèmes de Civil Law, voir la Directive 93/98/EEC du Conseil européen. Pour les Pays de Common law, ce principe est inclus dans §§ 102(a) 17 USC.

5. Les TRIPS (Trade-Related Aspect of Intellectual Property Rights) sont un accord créé par l'OMC (World Trade Organization), contraignants pour tous les États membres. Ils sont d'une importance particulière puisqu'ils obligent tous les États à se conformer à la Convention de Berne.

une différence entre l'analyse économique du *copyright* et celle du copiage (*economics of copying*) (Landes et Posner, 1989 ; Towse et al., 2008).

L'analyse économique du copiage

Les premiers travaux portant sur l'analyse économique du droit de la propriété intellectuelle, ayant pour objet l'étude le droit d'auteur, sont attribués à Plant (1934). Ce dernier s'intéresse au rôle de ce droit dans la production et la diffusion des livres et, plus précisément, aux effets des comportements déviants, où le copiage est vu comme une menace à la production optimale de ces biens. L'idée est que, sans la protection du droit d'auteur, d'autres acteurs pourraient s'approprier l'œuvre intellectuelle et éviter les coûts importants de production de la première copie. En effet, le marché est caractérisé par des coûts fixes de création de la première copie très élevés et par des coûts marginaux de reproduction du bien très faibles (O' Hare, 1985 ; Pollock, 2007). Une situation de concurrence amènerait dans ce contexte le prix d'équilibre du marché au-dessous des coûts supportés par le producteur de l'œuvre originale, ce qui empêcherait la rémunération de l'auteur et de la maison d'édition chargée de la production de la première copie (Breyer, 1970).

L'analyse économique du copiage s'intéresse alors aux effets du copiage dans un environnement affecté par le changement technologique. Plus précisément, cette branche de l'économie s'intéresse à la relation existant entre les coûts supportés par le titulaire du droit pour la création de la première copie et les coûts associés à la réalisation par autrui de copies non autorisées (Handke, 2010). Pour déterminer s'il est rentable de s'engager dans l'activité de production d'une œuvre de l'esprit, l'auteur doit calculer les bénéfices dérivant de la vente de son œuvre nets des coûts de production et de reproduction non autorisée. Le *copyright* s'analyse alors comme l'instrument juridique qui peut empêcher, pendant une période déterminée par la loi, le copiage de l'œuvre originelle (Hurt et Schuchman, 1966).

Cette littérature a mis en évidence que le risque de perte de profit dérivant du copiage et de la production de copies non autorisées est en réalité limité. Le droit d'auteur permet en effet à son titulaire d'adopter des stratégies de discrimination par les prix limitant les effets négatifs d'une reproduction non autorisée de l'œuvre originale. Le propriétaire, par la fixation d'un prix de monopole, anticipe la pratique de partage de la ressource et inclut

dans le prix la disponibilité à payer des agents entre lesquels l'œuvre sera partagée (Menell et Scotchmer, 2007). Le phénomène de copiage dépendant largement de l'environnement technologique, cette littérature a mis en avant les limitations naturelles au copiage avant la numérisation de l'information, en redimensionnant le coût des reproductions non autorisées de l'œuvre originale. En particulier, parce que la qualité des copies se dégrade avec l'augmentation du nombre de copies, la reproduction devient impossible au-delà d'un certain seuil (Liebowitz et Margolis, 1982).

Si cet argument ne semble plus adapté à l'ère du numérique, car la dématérialisation de l'information permet de créer un nombre illimité de copies sans subir de pertes en termes de qualité, certains auteurs soulignent cependant les bénéfices du partage non-autorisé pour le titulaire du droit de la ressource, même après la numérisation de l'information. Varian (2000) explique par exemple qu'il pourrait être moins coûteux pour le titulaire de faciliter la reproduction de son œuvre, en chargeant un prix qui tient compte des possibles copies dérivées, que de produire le bien pour chaque utilisateur. Le titulaire anticipe alors la consommation collective du bien et, à travers des stratégies de discrimination des prix, peut s'approprier les bénéfices dérivant de la consommation collective d'un bien unique. Tel est le cas des contrats entre maisons d'édition et bibliothèques, où le prix imposé reflète la multitude de consommateurs qui auront accès aux livres et revues qui font l'objet du contrat.

L'analyse économique du droit d'auteur

Si Breyer (1970) avait déjà étudié certaines différences entre l'analyse économique du *copyright* et l'analyse économique du copiage, la littérature a principalement attribué cette distinction à Landes et Posner (1989). Dans leur article « *An Economic Analysis of copyright Law* », ces auteurs s'intéressent, à travers une analyse positive, au *copyright* en tant qu'instrument pour promouvoir une allocation efficiente des ressources. Dans le prolongement de ces travaux, l'analyse économique du *copyright* s'oriente alors plus en général vers l'étude des problèmes associés à la production de l'information en tant que bien collectif et aux effets de l'intervention publique pour inciter à la production de ce type particulier de bien.

Le *copyright* est analysé principalement à travers deux approches. La première est celle des biens collectifs, où les autorités publiques accordent un pouvoir de monopole qui permet d'internaliser les externalités positives. Cette approche qui s'intéresse aux comportements déviants (Gordon, 1982), et notamment à l'appropriation illicite des œuvres d'autrui, se situe à l'intersection entre l'analyse économique du *copyright* et l'analyse économique du copiage, les frontières entre ces deux courants d'analyse étant d'ailleurs souvent difficiles à discerner. La deuxième approche est celle des droits de propriété, qui s'intéresse aux caractéristiques du *copyright* et à sa capacité à assurer un résultat économiquement efficace par rapport à la situation de concurrence pure et parfaite. Cette deuxième approche s'inscrit alors en complément de la première, en procédant à la comparaison entre les coûts et bénéfices dérivant de l'octroi d'un pouvoir de monopole (Leibowitz et Watt, 2006). Précisément, cette seconde approche étudie l'efficacité de l'allocation des droits de propriété intellectuelle, sur un marché caractérisé par la présence d'externalités, du point de vue de la production privée du bien collectif de la connaissance. L'octroi d'un droit de propriété est vu comme permettant la marchandisation du savoir (Ramello, 2005), passant par une négociation autour du droit entre son titulaire et les autres parties, dans le but de maximiser le bien-être collectif (Foray, 2000 ; Scheufen, 2015).

La comparaison entre les coûts et bénéfices dans un tel cadre juridique se situe au cœur de l'analyse économique du droit d'auteur. Avant tout, l'octroi d'un pouvoir de monopole conduit à l'apparition d'un coût social résultant de l'accès restreint à la nouvelle œuvre produite, qualifié par la littérature de coût statique (Landes et Posner, 1989 ; Benkler, 2002). Le deuxième type de coût, ou coût dynamique, dérive de l'augmentation des coûts associés au fait de générer des œuvres dérivées. En effet, la création d'une nouvelle œuvre, qui intègre une ouvre antécédente, nécessite la rémunération et l'autorisation préalable de l'auteur de l'œuvre originelle. Cette remarque est particulièrement importante dans le domaine de la science, où le *copyright*, et de manière plus générale, les droits de propriété intellectuelle, peuvent devenir des barrières à la circulation du savoir et à la capacité de chercheurs à contribuer à l'avancement de la science (David, 2003 ; Lessig, 2002 ; Heller et Eidenberg, 1998). Comme souligné par Posner (2005), outre les coûts de transaction liés à la protection et au contrôle du respect des normes, il existe un problème d'allocation

des ressources liée aux profits de monopole dérivant de l'octroi d'un droit de propriété. En effet, les profits découlant du droit de propriété intellectuelle fournissent une incitation à allouer les ressources à ces activités, plutôt qu'à des usages alternatifs qui donneraient lieu à des rentes inférieures, car allouées sur un marché en concurrence, mais à un produit social de valeur supérieure.

Les études qui s'inscrivent dans l'analyse économique du *copyright* s'interrogent par ailleurs également sur la durée optimale de la protection accordée par ce droit de propriété (Watt, 2004 ; Landes et Posner, 2003) et sur ce que Watt (2014) appelle l''*optimal copyright scope*', à savoir quels sont les actes légaux et quels sont les aspects précisément protégés par le droit d'auteur. Il s'agit d'un balancement entre les intérêts des auteurs d'un côté et de la société de l'autre, c'est-à-dire d'une comparaison entre les bénéfices générés par l'octroi d'un pouvoir de monopole et le coût ainsi généré. (Wu, 2006). La rétro-ingénierie en est un exemple. Elle demande de créer des copies d'un logiciel soumis à la protection du droit d'auteur, ce qui entraînerait une violation de la propriété intellectuelle. Cependant, les courts ont posé des exceptions, en limitant la protection assurée par le droit d'auteur et en rendant ainsi légale la pratique de la rétro-ingénierie (Sag, 2006).

Toujours dans la catégorie de l''*optimal copyright scope*', on peut également inclure les études portant sur les exceptions au *copyright* (Shavell, 2010 ; Scheufen, 2015) et notamment l'analyse économique du *fair use*. Le *fair use*[6] (17 U.S.C. §107) se définit comme un ensemble de règles qui déroge à la discipline générale du *copyright*, en rendant légitimes des actes qui ne le seraient pas autrement, sous la règle générale. La théorie de l'allocation efficiente des droits de propriété, sur laquelle se fonde l'analyse économique du *copyright*, requiert une hypothèse de coûts de transaction nuls ou très faibles (Coase, 1960). Or, outre l'augmentation des prix qu'il entraîne, le pouvoir de monopole octroyé à travers le *copyright* empêche toute réutilisation du bien intellectuel sans autorisation préalable du titulaire, ce qui engendre des coûts de transactions potentiellement très élevés, notamment dans le domaine de l'enseignement et de la recherche scientifique. Pour illustrer ces coûts,

6. Le *fair use* est un ensemble de règle de droits prévu par le système juridique des États-Unis. Cependant, des exceptions similaires sont prévues aussi dans plusieurs systèmes juridiques. L'Union Européenne (Directive 2001/29/CE) prévoit que « [L]es États membres devraient avoir la faculté de prévoir certaines exceptions et limitations dans certains cas tels que l'utilisation, à des fins d'enseignement ou de recherche scientifique [...] ». Dans le système français, ces exceptions sont prévues par l'article L.122-5 du Code de propriété intellectuelle.

on peut penser à l'exemple de l'utilisation d'un texte pour des finalités d'enseignement. Dans un système qui ne prévoit pas d'exceptions au droit d'auteur, un enseignant décidant d'utiliser ce texte devrait demander à chaque utilisation l'autorisation de l'auteur. Aussi, afin de limiter les effets négatifs entraînés par ces coûts de transactions, il a fallu mettre en place un instrument de régulation permettant de simplifier les négociations entre les parties prenantes. Dans cette perspective, Gordon (1982) propose de définir le *fair use* comme la réponse corrective aux limites du droit de propriété, en ce qu'il limite les coûts de transaction en permettant une reproduction de l'œuvre qui fait exception au droit d'auteur. En d'autres termes, l'autorisation préalable du titulaire n'est plus nécessaire pour toute réutilisation définie par la loi de l'œuvre sous-jacente. Si dans certains domaines le *fair use* n'est qu'une exception, il constitue presque la règle dans le domaine de l'enseignement et de la recherche scientifique.

1.1.3 Le rôle du droit d'auteur sur le marché de la publication scientifique

Dès 1962, Arrow considère les droits de propriété intellectuelle comme une solution de second rang, car incapable d'éliminer totalement les défaillances de marché. En effet, la situation de monopole induite par ces droits peut diminuer le bien-être collectif (Heller et Eisenberg, 1998). Dans cette section, nous examinons les spécificités du contrat d'édition sur le marché de la publication scientifique par rapport à la négociation des droits économiques incorporés dans le droit d'auteur des contrats d'édition standard. Nous montrons que ce sont ces spécificités qui ont amené à une concentration de la détention du *copyright* sur les différentes œuvres intellectuelles au sein de quelques maisons d'édition, structurant ainsi le marché de la publication scientifique autour d'un oligopole.

Le droit d'auteur et le contrat d'édition

L'attribution du droit d'auteur permet l'échange de ce droit entre différentes parties, afin qu'il soit alloué à l'agent le mieux capable de le valoriser. De manière générale, le contrat d'édition fournit une illustration de cette logique, en ce sens qu'il vise à répartir efficacement le droit entre la maison d'édition et le créateur de l'œuvre intellectuelle. Ce

type de contrat permet une division du travail entre création de l'œuvre et production et distribution de l'œuvre même, qui permet un gain d'efficacité, à partir du moment où l'exploitation est réalisée par un agent spécialisé à un moindre coût. Le contrat résulte alors de la répartition des profits dérivant de la vente de l'œuvre entre le créateur et la maison d'édition. Cette répartition peut se réaliser à travers des *royalties* au montant fixe ou au prorata des ventes (Watt, 2000).

Le contrat d'édition est aussi une forme de partage du risque portant sur la demande de travaux originaux (Liebowitz, 1987). Les solutions envisageables sont multiples. L'auteur peut verser une rémunération fixe à la maison d'édition, sans ainsi partager aucun risque. Au contraire, la solution opposée, où le risque n'est supporté que par la maison d'édition, consiste dans une cession totale des droits à un prix fixe à la maison d'édition. En réalité, très souvent c'est une solution intermédiaire qui se met en place, par laquelle le partage des droits entre créateur et maison d'édition est limité dans le temps.

Quelle que soit la forme du contrat, l'échange monétaire a toujours lieu. Cela n'est toutefois pas le cas concernant le contrat d'édition sur le marché de la publication scientifique. La relation contractuelle existant entre chercheurs et maison d'édition constitue en effet une catégorie *sui generis*[7] . En premier lieu, l'auteur transfère la composante économique du droit d'auteur à la maison d'édition, qui devient ainsi le nouveau titulaire. L'objet du contrat n'est pas la concession d'une licence pour la reproduction et la distribution de l'œuvre, mais un véritable transfert de propriété. En contrepartie, la maison d'édition assure deux fonctions primaires. Premièrement, elle sélectionne les articles méritant d'être publiés, et qui contribuent donc à l'avancement de la science, à travers le processus interne de révision des résultats, permettant ainsi la validation du contenu des articles scientifiques.

Deuxièmement, les maisons d'édition assurent la publication et donc la circulation de la connaissance scientifique. La publication des travaux de l'auteur permet de rendre visibles les outputs intellectuels produits par celui-ci, et de porter à la connaissance de la communauté scientifique sa contribution à l'avancement de la science. Cette rémunération, non pas directement monétaire, mais sous forme de services de publication, est essentielle,

7. Ce type de contrat est très peu analysé par la littérature. En effet, ces contrats sont presqu'entièrement standardisés et les auteurs ne sont pas en position de négocier les clauses, vis-à-vis du pouvoir contractuel très fort des maisons d'édition, d'où le manque d'intérêt pour la littérature d'étudier ce type de contrat.

si l'on considère que la carrière d'un chercheur dépend de la qualité et de la quantité de ses publications (Hamermesh et Pfann, 2009 ; Heckman et Moktan, 2020). La publication, rendue possible par la réallocation du droit d'auteur au sein de la maison d'édition, devient ainsi un actif incorporel pour l'*homo academicus* (Bourdieu,1984).

En ce qui concerne le risque économique lié à la publication, la maison d'édition devient le seul acteur concerné. Cependant, comme l'expliquent Edlin et Rubinfeld (2004), la structure des maisons d'édition leur permet de diminuer ce risque : les consommateurs sont souvent des abonnés à la revue dans laquelle l'article est publié, ce qui permet aux éditeurs de prévoir la demande. De plus, l'achat de la revue ne dépend pas du nouvel article, dont les consommateurs ne connaissent pas la qualité, mais de la qualité de la revue même, connue par la communauté scientifique, grâce à des indicateurs de qualité tels que l'*impact factor* (Borgman et Furner, 2002 ; Vinkler, 2004 ; Weingart, 2005).

Le droit d'auteur et la structure du marché de la publication scientifique

Si la négociation du droit d'auteur a représenté la seule solution possible pour la diffusion des résultats scientifiques avant la numérisation de la connaissance, la concentration au sein d'organisations spécialisées du droit d'auteur des outputs scientifiques s'est traduite par un marché en oligopole (Larivière et al., 2014). En outre, si jusqu'aux années 1960 la plupart des publications était contrôlée par des organisations à but non lucratif, la pénétration du marché de la publication scientifique par les maisons d'édition commerciales a considérablement réduit la partie du marché contrôlée par les organisations à but non lucratif. Les *scholarly societies* ont par exemple créé des partenariats avec les éditeurs commerciaux, afin de minimiser les coûts de gestion et de production de leurs revues scientifiques (McGuigan et Russell, 2008). En 2013, on a estimé que cinq maisons d'édition (Reed-Elsevier, Springer Nature, Wiley-Blackwell, Taylor & Francis and Sage) contrôlaient plus de 50% des publications d'articles scientifiques (Johnson et al., 2017 ; Larivière et al., 2014). Le fait que le droit d'auteur soit attaché à des biens non-substituables empêche la formation d'un marché en concurrence pure et parfaite. L'analyse de Porter proposée par Björk (2017) détaille les faiblesses d'une telle structure de marché. En raison de la réputation que les journaux ont su construire dans le temps, et grâce à la demande inélastique

à l'accès à la nouvelle connaissance produite (Nevo et al., 2005), la concurrence sur ce marché est presqu'inexistante. La construction de la réputation, en termes de contrôle de la qualité des nouveaux produits de la recherche, est très coûteuse et difficile à développer, d'où la difficulté à pénétrer ce marché. Comme expliqué par Björk (2017), les journaux qui sont entrés en premier sur le marché de la publication scientifique sont protégés par leur réputation. Cela est le résultat d'un équilibre de coordination entre le choix d'où publier pour les chercheurs-producteurs et le choix de quoi consommer, en termes d'achat de revues, pour les chercheurs-consommateurs (Bergstrom, 2001). La valeur d'une revue est alors déterminée à partir du consensus qui se forme entre les différents agents de la communauté scientifiques (*e.g.* éditeurs, comités de lecture, chercheurs en tant que producteurs et consommateurs de savoir scientifique, bibliothèques, etc.). C'est le nombre élevé d'acteurs concernés par la prise de décision sur la valeur de la nouvelle connaissance scientifique produite qui rend cet équilibre extrêmement stable et qui rend difficile l'entrée de nouveaux concurrents (Björk, 2017). Plus qu'à une compétition sur le marché, le contrôle des droits de propriété, et donc la situation de monopole d'un nombre restreint de maisons d'édition, a amené à une compétition pour le marché entre ces organisations spécialisées dans la distribution du savoir scientifique (Silva et Ramello, 2000). La structure oligopolistique du marché de la publication scientifique est à l'origine de ce qu'il est convenu d'appeler la *Serial Crisis*, une augmentation des prix des revues scientifiques bien au-delà de l'augmentation des prix à la consommation. L'Association for Research Libraries a estimé qu'entre 1986 et 2005 le prix de souscription aux revues scientifiques pour les bibliothèques a augmenté de 321%, contre une hausse des revues achetées de $1,9\%$ et une augmentation des prix à la consommation sur la même période de 73% (Kyrillidou et Young, 2008). Ces résultats ont été confirmés par d'autres études : une variation positive de 215% entre 1986 et 2001 (Edwards et Shulenburger, 2003) et de 273% entre 1986 et 2004 (Ramello, 2010). Cette augmentation des prix est le résultat de l'exploitation du pouvoir de monopole des maisons d'édition et de la demande inélastique de l'accès aux résultats scientifiques. Si on prend les dix premières revues les plus citées en 2001, on remarque qu'elles sont gérées par des maisons d'édition sans but lucratif et que leur prix a augmenté de 80% entre la période 1985-2001, surtout à cause de l'augmentation générale des prix à la consommation, contre

une augmentation de 279% pour les journaux commerciaux (Bergstrom, 2001).

1.1.4 Le droit d'auteur à l'âge du numérique

Avec la numérisation de l'information, la structure du marché de la publication scientifique a subi des modifications, surtout concernant la structure des coûts de production et de diffusion de la connaissance. Pour tenir compte de cette évolution, l'analyse économique du *copyright* appliquée aux revues scientifiques, notamment à travers les contrats entre bibliothèques et maisons d'édition (McCabe, 2004 ; Jeon et Menicucci, 2006), s'est orientée vers l'étude de l'exploitation du pouvoir de monopole des maisons d'édition à l'ère du numérique et de la pertinence d'un tel droit pour assurer le fonctionnement du marché de la publication scientifique.

L'évolution de l'exploitation du droit d'auteur par les maisons d'édition à l'ère du numérique

La numérisation de l'information a avant tout remis au centre de l'analyse les comportements déviants liés à la violation du droit d'auteur. Si le phénomène du copiage à l'ère de l'analogique avait en soi des barrières naturelles au nombre de copies possibles, il représente avec la numérisation une véritable menace à l'appropriabilité de la ressource par son auteur (Menell et Scotchmer, 2007 ; *United States of America vs. Aaron Swartz – Criminal No. 11-10260-NMG*). Une certaine pression est alors exercée sur le législateur de la part des maisons d'édition, afin de créer des limitations et des exceptions applicables dans l'environnement numérique de manière à assurer une protection accrue du droit d'auteur (Reichman et Okediji, 2012). L'application de mesures techniques, c'est à dire d'instruments technologiques qui limitent le partage de la ressource numérisée, tel que les TPM (*Technical Protection Measures*), et la mise en place d'instruments électroniques visant le contrôle de la ressource, comme le DRM (*Digital Right Management*) ont assuré aux maisons d'édition une protection automatique des données et de l'information dans ce nouvel environnement (Reichman et al., 2007)[8]. La numérisation a été suivie par un

8. Voir, à titre d'exemple, l'application dans les systèmes de Common Law contenue en 17 USC §§ 1201-1205 (2006). Le durcissement de la protection accordée par le droit d'auteur trouve une application dans les systèmes de *Civil Law* aussi, comme bien le montre la protection accordée aux bases de données par l'Union européenne à partir de 1996 (96/9/EC du Parlement Européen et du Conseil du 11/03/1996).

durcissement de la propriété des œuvres de l'esprit, non seulement à travers la création de barrières technologiques, mais aussi à travers une protection juridique accrue, dont témoigne par exemple la protection des bases de données assurée par l'Union européenne (Directive 96/9/CE du Parlement européen et du Conseil européen 11 mars 1996).

Cet élargissement de la protection a suscité un débat au sein de la communauté scientifique sur les dangers liés à l'extension de la protection de la propriété intellectuelle, notamment pour le partage et la réutilisation du savoir scientifique. En effet, non seulement ces droits de propriété intellectuelle renforcés sont susceptibles d'être contre-productifs en termes d'incitation au développement de nouvelle connaissance, mais ils comportent aussi le risque d'introduire des distorsions dans la distribution des profits entre les parties prenantes. L'augmentation des prix ayant caractérisé le marché de la publication scientifique à partir des années 1980 constitue un exemple avéré de ce risque. Même au moment de la numérisation de l'information, et donc dans un contexte de baisse des coûts de production et de distribution des revues scientifiques, les maisons d'édition ont continué à augmenter les prix, au détriment de la communauté scientifique. C'est justement la baisse des coûts de production et de distribution liée à l'élimination du support papier qui a permis aux maisons d'édition d'atteindre des profits marginaux autour de 35% (Larivière et al., 2015).

Cette augmentation des profits s'explique aussi par les nouvelles techniques de discrimination de prix que la numérisation a permis de mettre en place. Au lieu de créer une demande pour une seule revue, les maisons d'édition ont commencé à offrir un ensemble de revues scientifiques. Ce nouveau contrat, qualifié de *Big Deal*, est contraignant pendant plusieurs années : les bibliothèques, qui représentent la part la plus importante de la demande d'accès au savoir scientifique, achètent désormais un droit à l'accès à plusieurs revues scientifiques, à un prix inférieur à la somme des prix de chaque revue, pour une période de plusieurs années (Bergstrom et al., 2014). Le *Big Deal* constitue en ce sens une stratégie de *bundling* non seulement entre journaux, mais aussi entre les formats numériques et papier des revues scientifiques. Dans le premier cas, cette stratégie permet aux maisons d'édition de vendre un ensemble de revues, incluant même celles que les bibliothèques n'auraient pas achetées. Dans le deuxième cas, le Big Deal crée une forme de dépendance à long terme, car si les bibliothèques s'engagent sur un contrat sur plusieurs

années, elles auront accès à une large littérature par voie numérique, mais elles pourront aussi recevoir à des prix inférieurs le format papier des revues, sous condition de ne pas interrompre le contrat (Frazier, 2001 ; Edlin et Rubinfeld, 2004).

Cette technique permet aux bibliothèques d'avoir accès à des revues qu'elles n'auraient pas pu acheter avec un contrat standard, mais crée aussi des distorsions sur le marché, détaillées par Edlin et Rubinfeld (2004). Puisque les bibliothèques consacrent une partie importante de leurs ressources financières à ces contrats, leur contrainte budgétaire limite leur capacité à souscrire d'autres contrats avec d'autres maisons d'édition. C'est ainsi que le Big Deal renforce les barrières à l'entrée sur le marché de la publication scientifique. Si, avant la numérisation, la concurrence entre maisons d'édition se fondait sur une seule revue, objet du contrat entre le consommateur et le titulaire de la revue, avec cette nouvelle technique c'est un ensemble de revues qui fait désormais l'objet de la concurrence, ce qui rend encore plus difficile pour les nouveaux entrants de commencer leur activité avec une offre suffisamment importante de journaux scientifiques. Comme Edlin et Rubinfeld (2004) le soulignent, cette stratégie a permis une discrimination par les prix parfaite : le prix du bouquet de revues étant établi à partir des souscriptions aux revues que les bibliothèques achetaient avant la mise en place du *Big Deal*, les éditeurs ont en effet été en mesure de déterminer avec précision la disposition à payer des bibliothèques.

La remise en cause du droit d'auteur

La monopolisation de l'exploitation économique du droit d'auteur par les maisons d'édition (Hilty et al., 2008 ; Bernault, 2016) ne semble plus être justifiée à l'ère du numérique, étant donné la baisse importante des coûts de production et de distribution des revues scientifiques, d'où la proposition d'éliminer le droit d'auteur dans le domaine de la recherche scientifique (Litman, 2006 ; Hilty et al., 2008 ; Shavell 2010). Shavell (2010) considère ainsi que, même sans ce droit de propriété, la communauté scientifique serait en mesure d'organiser et d'assurer la production et la diffusion du savoir scientifique, à travers des normes autres que légales, et qui assuraient déjà auparavant le fonctionnement de la circulation du savoir scientifique, comme les normes sociales de la communauté scientifique, objet d'études de Merton (1957, 1973). L'élimination du droit d'auteur permettrait d'un côté la

gestion de la circulation des résultats de la recherche au sein de la communauté scientifique, mais elle créerait de l'autre un vide quant aux modalités de diffusion et de réutilisation de l'output intellectuel. Le caractère non nécessaire d'exercer sur les chercheurs une incitation économique via le droit d'auteur à produire de la connaissance a été aussi soulignée par d'autres auteurs (Müller-Langer et Scheufen, 2013), mais des solutions plus nuancées que l'élimination pure et simple du droit d'auteur ont également été proposées. À cet égard, Müller-Langer et Watt (2010) envisagent l'élimination du droit d'auteur comme *extrema ratio* seulement dans le cas où d'autres modèles de dissémination du savoir ne pourraient pas se mettre en place et devenir le standard de la communication scientifique à cause de la dominance du modèle de publication s'appuyant sue les monopoles garantis par le droit d'auteur. Scheufen (2015) propose dans cette perspective la création d'un éventail d'exceptions à l'application du droit d'auteur dans le contexte académique.

Dans un contexte d'augmentation de prix et de limitation à l'accès à la connaissance scientifique, le système d'ordre public pour la gestion et la régulation du marché de la publication scientifique imposé par le droit d'auteur ne semble plus être viable. Les barrières légales et économiques à l'accès et à la réutilisation ont amené certains auteurs à déclarer préférer subir une violation de leur propriété intellectuelle à travers le piratage plutôt que de voir ignorer leurs travaux (Stiglitz, 2007).

1.2 L'accès ouvert au savoir scientifique : le modèle *Open Access*

De manière générale, la numérisation de l'information a été accueillie par la communauté scientifique comme la solution pour sortir de la crise des prix des revues scientifiques, étant donné les coûts marginaux de reproduction presqu'égaux à zéro. Dans ce contexte caractérisé par la rapidité de l'échange de l'information et un accès immédiat et gratuit aux contenus, le modèle traditionnel de diffusion du savoir, fondé sur l'échange du *copyright*, a été remis en cause. C'est ainsi que lors de la conférence de Budapest en 2002, certains membres du milieu scientifique ont inscrit la viabilité de la diffusion du savoir dans les priorités de la communauté scientifique et proposé le modèle *Open Access* comme une al-

ternative au modèle d'ordre public qui avait jusqu'alors régulé le marché de la publication scientifique. Dans son sens le plus général, l'*Open Access* désigne toute pratique de partage et de gestion en commun d'une certaine ressource (Dusollier, 2007). Dans le cadre de la circulation du savoir, il désigne aussi plusieurs modèles de dissémination de la connaissance, visant à garantir un accès sans barrières à la littérature scientifique, que cette barrière prenne la forme d'un prix ou d'une barrière juridique (Suber, 2012). Après avoir défini ces nouveaux modèles de dissémination des résultats de la recherche scientifique, dans cette section nous allons détailler les différentes formes d'*Open Access* et leur impact au sein des maisons d'édition et de la communauté scientifique.

1.2.1 Le mouvement *Open Access*

Même si la naissance officielle de l'*Open Access* date de 2002, avec la conférence de Budapest, différentes initiatives promouvant la libre circulation du savoir scientifique existaient déjà auparavant. C'est surtout le cas des archives en ligne, des espaces numériques de collecte de travaux scientifiques, publiés ou non, qui ont caractérisé le marché de la publication scientifique à partir des années 1990. Tel est le cas de ArXive, fondée en 1991, en tant qu'espace de stockage de papiers scientifiques dans le domaine des sciences techniques, médicales, et économiques (Ginsparg, 2004), suivi en 1994 par SSRN (*Social Science Research Network*) et en 1997 par RePEc *Research Papers in Economics* (Bernault, 2016).

Cependant la conférence de Budapest (2002), suivie par celle de Berlin et de Bethesda (2003), fait de l'*Open Access* un mouvement international, et non plus uniquement une initiative prenant la forme de quelques niches. L'objectif de ces conférences est en effet précisément de faire converger des projets déjà existants pour atteindre un résultat plus large et plus rapide dans la diffusion du savoir scientifique[9]. Dans la déclaration de Budapest, l'*Open Access* est défini comme « la libre disposition de la littérature scientifique sur internet, qui permet à tout utilisateur de lire, télécharger, copier, distribuer, imprimer, rechercher ou créer un lien vers le texte intégral de ces articles, les explorer pour les indexer, les transmettre en tant que données à un logiciel ou les utiliser pour toute autre finalité

9. Pour une chronologie des différentes initiatives *Open Access*, voir P. Suber Timeline of the Open Access Movement (`https://webcache.googleusercontent.com/search?q=cache:hLcEvLUfmuMJ:https://legacy.earlham.edu/~peters/fos/timeline.htm+&cd=2&hl=fr&ct=clnk&gl=it`)

légale, sans barrières financières, juridiques ou techniques autres que celles nécessaires pour accéder à Internet lui-même » [10].

Les différentes formes d'*Open Access*, que nous allons examiner dans la section suivante, permettent alors un accès gratuit à la littérature scientifique, l'auteur gardant tout ou partie des droits de propriété sur son œuvre. L'auteur peut ainsi soit rester le titulaire du droit de propriété intellectuelle, en conservant les droits déterminant les conditions de réutilisation de son œuvre intellectuelle, soit il peut poser des exceptions au droit d'auteur détenu par les maisons d'édition, en mettant en libre accès un article, même après sa publication dans une revue scientifique en *Closed Access*.

1.2.2 Les différentes formes d'*Open Access*

En tant que modèle de diffusion de la connaissance scientifique, l'*Open Access* peut se présenter sous trois différentes formes : le *Green*, le *Gold* et le *Black Open Access*. Une littérature riche et hétérogène a pour objet l'étude des différentes formes d'*Open Access*. Cependant, les classifications proposées confondent le modèle d'affaires avec la forme d'*Open Access* adopté. C'est pourquoi nous nous focalisons ici avant tout sur la forme sous laquelle l'*Open Access* peut se présenter et, ensuite, sur les différents modèles d'affaires présidant à sa mise en place.

Le *Green Open Access*

En tant que modèle de diffusion des résultats de la recherche scientifique, le *Green Open Access* est un outil complémentaire au modèle de diffusion du savoir traditionnel. En effet, ce modèle ne prévoit pas la publication des résultats scientifiques stricto sensu, à savoir une publication des résultats entendue comme processus de certification du savoir opéré par les comités de lecture des revues scientifiques. Il constitue une pratique d'auto-archivage des œuvres de l'esprit, préalablement ou pas publiées dans une revue scientifique. En effet, la numérisation de l'information a permis aux auteurs de mettre immédiatement à la disposition des lecteurs tous leurs travaux scientifiques, appelés pre-print s'ils n'ont pas encore bénéficié d'une publication dans une revue scientifique, ou post-print s'ils sont déjà parus

10. `budapestopenaccessinitiative.org/read`, traduction de l'auteur.

dans un journal scientifique, dans des espaces numériques appelées archives, à travers lesquelles le consommateur bénéficie d'un libre accès aux outputs intellectuels (Björk, 2014). Puisque l'auteur ne doit pas payer pour archiver sa propre œuvre intellectuelle[11], le *Green Open Access* est la forme la plus utilisée. Ainsi, en 2008, 20% des articles publiés était librement accessibles et 12% l'étaient sous la forme du modèle *Green* (Björk et al., 2010).

Différentes classifications du *Green Open Access* ont été proposées dans la littérature. Une première classification distingue ainsi les archives spécifiques à des disciplines (*disciplinary repositories*), dont RePEc constitue un exemple pour le domaine de l'économie, ainsi que les archives institutionnelles, mises en place par un centre de recherche, afin de mettre à la libre disposition du public les outputs de ses chercheurs (Frosio, 2014 ; Xia, 2008 ; Darby et al., 2009). Au contraire, d'autres classifications définissent les *disciplinary repositories* comme une sous-catégorie des archives institutionnelles, en les opposant à celles qui mélangent toutes les disciplines scientifiques (Parker, 2007). Souvent, une troisième catégorie, à savoir la page web personnelle de l'auteur, est ajoutée dans la classification des archives ouvertes (Björk et al., 2014).

Au lieu de classer les archives en fonction de l'organisme qui s'occupe de leur gestion, d'autres classifications reposent sur le degré d'interopérabilité de l'archive. Si l'objectif final du *Green Open Access* est de rendre accessible une copie d'un article sur une page internet autre que celle de la maison d'édition, les utilisateurs doivent pouvoir trouver un article sans savoir que l'archive existe, où elle est et ce qu'elle contient. On parle alors d'archives interopérables, en ce sens qu'elles utilisent la même structure de collecte de données, en permettant ainsi une plus vaste circulation du savoir (Suber, 2007 p.173), en opposition à celles qui ne suivent pas ce protocole, géré par l'*Open Archives Initiative Protocol for Metadata Harvesting*, une initiative qui vise à promouvoir des standards communs pour la création des archives ouvertes et donc à coordonner les différentes initiatives locales.

La possibilité pour les auteurs de partager leurs résultats après la publication par d'autres canaux que les revues scientifiques, a créé des tensions entre l'intérêt des cher-

11. Il a été estimé que le coût d'archivage pour un auteur est de l'ordre de $33, compte tenu du temps qu'il consacre à la création de la notice bibliographique et du téléchargement du format numérique du papier (Houghton et al., 2009). Le coût de création de la structure numérique de l'archive est supporté généralement par un ou plusieurs instituts de recherche. Ces coûts de gestion et de création peuvent varier entre £26.000-210.000 (Swan, 2010), mais aller aussi jusqu'à $1.5 (Willinsky, 2009), en fonction de la taille de l'archive.

cheurs à maximiser leur visibilité et celui des maisons d'éditions à garder un contrôle exclusif sur l'œuvre dont ils sont les titulaires du droit de propriété intellectuelle. En effet, le *Green Open Access* permet aux bibliothèques d'être indépendantes des abonnements qu'elles sont obligées de souscrire pour garder l'accès à la connaissance antérieure (Merret, 2006). C'est ainsi que les maisons d'édition ont commencé à imposer plusieurs restrictions à l'archivage des articles publiés et même pour les versions avant publication. Ces restrictions peuvent concerner autant l'embargo, c'est-à-dire le délai à respecter imposé par la maison d'édition entre la publication et l'archivage de l'article sur une autre plateforme, que la version de l'article (version publiée ou dernière version avant la publication), ou encore le type d'archive dans laquelle l'article peut être déposé (institutionnelle, commerciale, personnelle) (Björk et al., 2014). Afin d'aider les auteurs à comprendre quelles sont les politiques d'auto-archivage adoptées par les différentes maisons d'édition, SHERPA/RoMEO a mis en place un système de couleur qui, de façon très rapide, explique quels sont les droits des auteurs accordés par les différents journaux pour la mise en libre accès de leur manuscrit, comme expliqué dans le tableau 1.1.

Couleur RoMEO	**Type de document susceptible d'archivage**
Green	Pre-print, post-print et version PDF de la revue
Blue	Post-print et version PDF de la revue
Yellow	Pre-print (avant soumission)
White	Aucun

TABLE 1.1 – Couleurs SHERPA/RoMEO pour le classement des politiques des maisons d'édition

Ce système de couleurs a permis de mettre en évidence l'augmentation des mesures de restrictions à l'auto-archivage opérée par les maisons d'édition. En analysant 107 maisons d'édition à partir de la base de données de SHERPA/RoMEO sur la période 2004-2015, Gadd et Covey (2016) trouvent que les restrictions portant sur quelle version de l'article, où et quand l'article peut être archivé ont augmenté respectivement de 119%, 190% et 1000%.

Le *Gold Open Access*

La principale alternative au modèle de publication traditionnelle est le *Gold Open Access*, qui garantit la publication dans une revue scientifique à comité de lecture, capable

ainsi de satisfaire les exigences de certification de la connaissance, en réduisant les asymétries d'information typiques de la connaissance. Ce qui change par rapport au modèle *Closed Access* est que sa viabilité économique n'est pas assurée par les consommateurs et que le chercheur reste le titulaire du droit d'auteur. Tout comme dans le *Green Open Access*, pour le *Gold Open Access* la littérature a proposé différentes classifications. Pour Laakso et al. (2011), le *Gold Open Access* peut se présenter sous trois différentes formes :

- l'*Open Access* direct, où la revue est entièrement accessible en *Open Access* ;
- le *delayed Open Access*, où l'accès aux articles les plus récents reste soumis au paiement d'un prix, mais où les articles deviennent librement accessibles au fur et à mesure que le temps passe ;
- l'*Open Access* hybride, où des articles on *Open Access* sont publiés dans une revue qui garde le *Closed Access* comme modèle d'affaires principal.

D'autres auteurs font référence à cette classification, mais ils y ajoutent aussi toute pratique d'auto-archivage (Willinsky, 2009), en éliminant donc la distinction entre le *Green* et le *Gold Open Access*. On voit bien qu'il existe donc une véritable confusion entre les différents modèles d'*Open Access* – c'est-à-dire la forme sous laquelle la connaissance scientifique peut être mise en ligne - et les différents modèles d'affaires adoptés par les maisons d'édition pour couvrir leurs coûts des productions. En effet, même si les coûts de production et de diffusion ont remarquablement baissé, les tâches accomplies par les maisons d'édition à travers la coordination de la certification et de la distribution du savoir doivent toujours être rémunérées, que les éditeurs soient des entreprises à but lucratif ou pas (May, 2020).

Cette confusion entre modèle d'*Open Access* et modèle d'affaires nous conduit à proposer ici notre propre classification du *Gold Open Access* en fonction du modèle d'affaires adopté par la maison d'édition. Le modèle le plus diffusé est celui de l'*Article Processing Charges* (APC), qui consiste à imposer un prix à la publication à la charge de l'auteur. Ce modèle, qui représente le modèle le plus répandu parmi les éditeurs commerciaux (Jubb et al., 2015 ; Pinfield et al., 2016) est souvent, et à tort, assimilé au *Gold Open Access* lui-même (*e.g.* Feess et Scheufen, 2016 ; Shavell, 2010 ; McCabe et Snyder, 2018), puisqu'il ne constitue que l'un des modèles d'affaires possibles parmi d'autres. Une autre forme sous laquelle le *Gold Open Access* peut se présenter est l'*Open Access* hybride, souvent

utilisé par les maisons d'édition commerciales, car il permet de publier un article en accès ouvert dans une revue en *Closed Access*. Dans ce cas, l'auteur a la possibilité de payer des frais à la publication pour garantir un accès immédiat et gratuit à son article aux chercheurs-consommateurs. Ce modèle s'avère être particulièrement rentable pour les maisons d'édition, car il permet d'ajouter aux profits dérivant de la vente en *Closed Access* de la revue scientifique les charges imposées à la publication dans la même revue d'un article en *Open Access*.

Le fait que, dans le modèle Gold APC, les recettes de la maison d'édition augmentent avec le nombre d'articles publiés, a mené à l'émergence de revues prédatrices, que l'on peut définir comme des revues caractérisées souvent par un manque de transparence concernant la composition du comité éditorial et la pratique de sollicitations massives auprès des chercheurs via leurs adresses électroniques (Fuchs et Sandoval, 2013), afin de les inciter à soumettre leurs manuscrits. Dans ce type de revues, aujourd'hui regroupées dans la liste Beall (Beall, 2012), le processus de lecture par les référés est souvent absent, ce qui implique une circulation du savoir sans une véritable certification des résultats obtenus par les pairs.

Comme on l'a déjà indiqué, l'imposition d'un prix à la publication n'est que l'un des choix possibles parmi d'autres. En effet, certaines revues n'imposent aucun prix à la publication. Ce modèle, dit Platinum ou encore Diamant (Haschak, 2007 ; Björk, 2017), est promu par des organisations non-commerciales, et il assure non seulement un accès libre aux résultats scientifiques, mais aussi leur libre réutilisation pour toute finalité non-commerciale (Fuchs et Sandoval, 2013).

Quel que soit le modèle adopté, le *Gold Open Access* combine donc un accès élargi, gratuit et immédiat à la littérature scientifique pour le consommateur final avec le principe de la certification du savoir scientifique, et donc une garantie de validité de nouveaux résultats contribuant à l'avancement de la science. Nous résumons dans le tableau 1.2 les différents modèles sous lesquels le *Gold Open Access* peut se présenter.

Modèle de Gold Open Access	**Type de financement**
APC	Frais à la publication
Hybride	Frais à la publication + prix de la revue
Diamant	Autofinancement de la revue

TABLE 1.2 – Les différents modèles de business des revues en Gold *Open Access*

Le *Black Open Access*

Très souvent, la littérature ajoute à cette classification dichotomique une troisième catégorie, à savoir le *Black Open Access*, ou modèle Robin Hood (Archambault et al., 2014). Ce modèle se fonde sur une appropriation illégitime et illégale de l'output intellectuel protégé par le droit d'auteur (Björk, 2017). Appelé aussi bibliogifts (Cabanac, 2016) ou encore *Guerilla Open Access* (Bodò, 2008), ce modèle permet de télécharger gratuitement l'article scientifique sans payer le prix imposé par la maison d'édition. Sci-hub en est l'exemple le plus connu dans la communauté scientifique. Né en 2011 de l'initiative de la chercheuse A. Elbakyan, ce site a téléchargé plus de 50 millions d'articles scientifiques à partir des plateformes de différentes maisons d'édition (Bohannon, 2016), articles qui sont ensuite stockés dans un espace numérique appelé Libgen. Dans une étude menée en 2014, Cabanac (2015) a croisé, à partir de l'identifiant unique des articles (DOI), les articles archivés dans Libgen avec ceux contenus dans les archives de différentes maisons d'édition. Le résultat est que 36% de la littérature scientifique était à cette date disponible sur cette plateforme pirate, qui est parvenue à couvrir 77% des publications d'Elsevier. L'une des explications possibles pour cette large utilisation de sites illégaux est qu'ils sont considérés comme moralement acceptables par la communauté scientifique (Travis, 2016).

Une partie de la littérature considère les ASNs (Academic Social Networks) tels que Research Gate ou Academia.edu comme faisant partie aussi de la catégorie *Black Open Access*, puisque très souvent ils permettent le téléchargement d'articles scientifiques protégés par le droit d'auteur, sans autorisation préalable de la revue scientifique titulaire du droit d'auteur (Björk, 2017 ; Travis, 2016). C'est justement au motif de la violation du droit d'auteur que la maison d'édition Elsevier a demandé que des milliers de papiers soient retirés du site internet de Academia.edu (Howard, 2013).

Un autre exemple de *Black Open Access* est le #icanhazPDF, tweet à travers lequel les chercheurs demandent à la communauté de Twitter de leur envoyer des articles scientifiques auxquels ils n'ont pas accès, à cause du prix imposé par la maison d'édition (Gardner et Gardner, 2017). Les 4/5 des articles scientifiques publiés en 2017 restent non-disponibles via les options *Open Access* légales (Green, 2017) et la large utilisation des sites pirates montre bien que le *Black Open Access* a pris le dessus sur le *Green* et le Gold *Open Access*

(Green, 2017 ; Björk, 2017). À ce titre Himmelsteir et al. (2018) montrent comme Sci-hub permet l'accès à 85% des articles publiés, quand l'utilisation du modèle *Green Open Access* pour garantir un accès libre et légal reste très limitée.

La piraterie dans la recherche scientifique (*piracy of scientific papers*[12]) est donc promue par au moins certains des producteurs du savoir piraté eux-mêmes, ce qui semble montrer la nécessité d'un changement dans les pratiques de circulation du savoir scientifique, dans le respect des lois et des intérêts de toutes les parties concernées dans le marché de la publication scientifique.

L'évolution de la diffusion de l'*Open Access*

Reconstruire l'évolution de l'utilisation de l'*Open Access* par la communauté scientifique en tant que modèle de publication à travers la littérature n'est pas une tâche simple, à cause des différences importantes qui existent entre les différentes analyses proposées par les auteurs. Souvent, la définition de l'*Open Access* qu'ils retiennent n'est pas la même, cela constituant ainsi la première source de différence. De plus, les résultats sur l'évolution de l'*Open Access* et sur sa pénétration dans le marché de la publication scientifique varient selon le type d'*Open Access* analysé (*Green* ou Gold), selon la base de données utilisée et en fonction des revues scientifiques analysées.

L'un des premiers auteurs à avoir analysé l'évolution des journaux en *Open Access* est Crawford (2002), qui étudie les données recueillies par l'*Association of Research Libraries* entre la période 1995 et 2001. Son analyse qualitative montre ainsi qu'au début de la période d'étude, il n'existait que 86 revues en *Open Access* et que six ans après, seulement 57% d'entre elles étaient encore actives. Pour lui, le problème est qu'après une première phase qualifiée d'enthousiaste, le journal n'arrive pas à augmenter le nombre d'articles publiés et finit donc par ne plus publier d'articles scientifiques.

Une analyse de l'évolution de l'*Open Access* a été proposée par Laakso et al. (2011), en identifiant trois périodes différentes à partir des données de DOAJ, ISI (Institute for

12. Voire à titre d'exemple Machin-Mastromatteo, J. D., Uribe-Tirado, A., & Romero-Ortiz, M. E. (2016). Piracy of scientific papers in Latin America : An analysis of Sci-Hub usage data. Information development, 32(5), 1806-1814 ; Woolston, C. (2016). Paper piracy sparks online debate. Nature News ; Tran, D. (2016). Elsevier V. Sci-hub : Piracy of knowledge and the jurisdictional reach of US copyright law. Wake Forest J. Bus. & Intell. Prop. L., 17, 198.

Scientific Information) et Scopus, comme montré sur la figure 1.1.

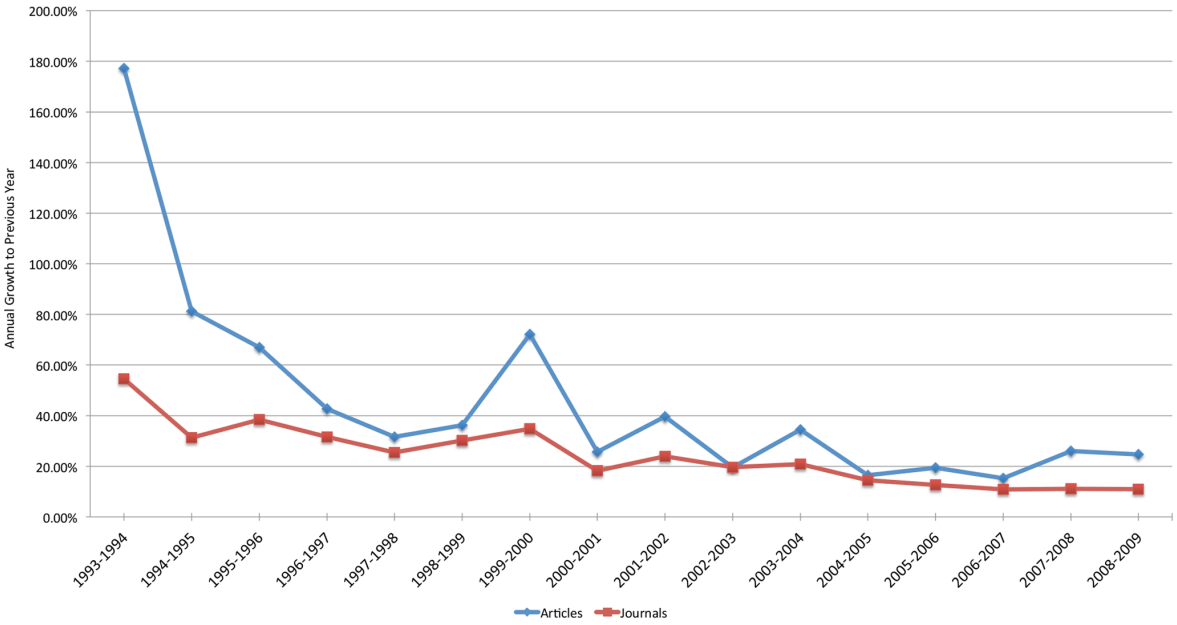

FIGURE 1.1 – Evolution des revues et articles en *Open Access* - Source : Laakso M, Welling P, Bukvova H, Nyman L, Björk BC, et al. (2011) The Development of Open Access Journal Publishing from 1993 to 2009.

Les années dites années pionnières (1993-1999) se caractérisent par un nombre total d'articles et de journaux *Open Access* modeste. Le financement à la publication est à la charge des auteurs eux-mêmes, et la plupart du travail de publication est accomplie gratuitement [13]. C'est pendant les années d'innovation (2000 – 2004) qu'on assiste à une véritable croissance du nombre des articles publiés en *Open Access* et à l'émergence de nouveaux journaux proposant la publication en *Open Access*. Dans cette deuxième phase, l'utilisation de licences créatives et l'émergence de nouveaux modèles d'affaires destinés à rendre rentable cette nouvelle méthode de publication ont permis la croissance et le développement de l'*Open Access*. C'est véritablement dans la dernière phase, appelée années de consolidation (2005 – 2009) qu'on assiste à l'émergence de nombreuses infrastructures soutenant l'*Open Access*, telle que la *Directory of Open Access Journals* (DOAJ), qui s'occupe d'indexer les nouveaux journaux en *Open Access* dans le marché de la publication scientifique, comme le montre la figure 1.2.

13. On peut penser à ce titre au travail non rémunéré des rapporteurs, des éditeurs, mais aussi des chercheurs même en ce qui concerne désormais la mise en page du manuscrit. Pour une analyse détaillée, voir Bergstrom, T. C. (2001). Free labour for costly journals?. Journal of Economic Perspectives, 15(4), 183-198.

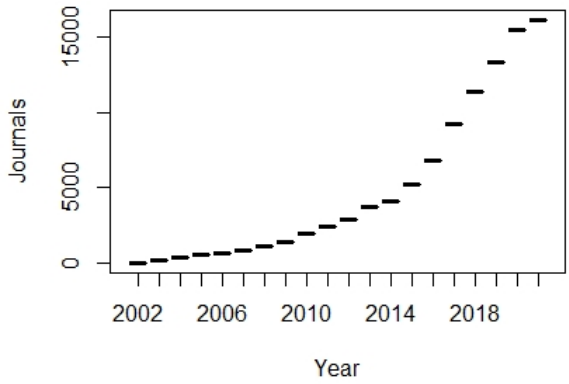

FIGURE 1.2 – Journaux *Open Access* dans DOAJ

On peut le voir, le nombre de journaux en *Open Access* augmente fortement sur l'ensemble de la période analysée. Néanmoins, la plupart des publications restent inaccessibles au public en *Open Access*, et le modèle de *Closed Access* conditionnant l'accès au paiement d'un prix demeure le modèle dominant. L'utilisation de l'*Open Access* reste donc faible (Johnson et al., 2017). À partir des données de Laakso et Björk (2012) et de Redhead (2015), Björk (2017) montre que la part des articles publiés en Gold *Open Access* reste faible, autour de 10% si on utilise la base de données ISI et de 15% quand on utilise celle de Scopus et WoS (Web of Science) sur la période 2000-2015. Des résultats similaires ont été mis en évidence par Green (2018) et Fortson (2019), qui montrent que 80% des articles scientifiques restent encore aujourd'hui soumis au régime de l'accès payant. De plus, si l'on regarde le modèle d'affaires adopté par les revues en *Gold Open Access*, le modèle APC est devenu de plus en plus commun, représentant à peu près la moitié de tous les journaux en *Open Access* (Laakso et Björk, 2012).

1.2.3 L'impact de l'*Open Access* pour les maisons d'édition

La littérature s'est intéressée à l'impact de l'*Open Access* sur les maisons d'édition en se focalisant sur les prix imposés par celles-ci à la publication d'articles scientifiques. Même si l'élimination du support papier a permis de faire baisser les coûts de production et de dissémination de la connaissance, les éditeurs soulignent que des coûts très élevés liés à l'activité éditoriale perdurent, allant jusqu'à $40.000 par article (Van Noorden, 2013). En revanche, quand on regarde le coût moyen de production des intermédiaires à but

non lucratif, le coût de production d'un article baisse jusqu'à $1.100 pour PLoS (Ware et Mabe, 2015) et $290 pour Hindawi (Van Noorden, 2013). La différence des coûts entre producteurs se reflète sur les APC qu'ils imposent aux auteurs. Pour la production d'un article, Odlyyzo (1995) avait estimé des prix à la publication autour de $300 – $1.000. Si cela semble être cohérent avec les APC imposées par des revues telles que PLoS, la moyenne des prix imposés par la plupart des revues commerciales est bien au-delà de ce seuil. Dans le modèle de Gold *Open Access* APC, les frais à la publication tournent autour de $1.500 (Johnson et al., 2017 ; Björk et Solomon, 2014) et elles passent à $2.500 quand le modèle utilisé est l'hybride (Björk et Solomon, 2014).

Le prix à la publication a été aussi analysé pour comprendre quel type de maison d'édition a le plus d'intérêt à changer son modèle d'affaires, en passant du modèle *Closed Access* au Gold APC. McCabe et Snyder (2005) suggèrent que, du point de vue de l'intérêt de la revue, ce sont les journaux avec une qualité plus faible qui pourraient être incités à adopter ce modèle avant ceux dont la réputation est établie et forte. En effet, puisqu'il y a des frais à la publication, cela pourrait entraîner une dégradation de la qualité des revues, étant donné leur intérêt à augmenter le nombre d'articles acceptés et publiés. Or, plus une revue publie d'articles, plus elle peut augmenter ses profits. La dégradation de la qualité des revues scientifiques, suite à cet effet de distorsion provoqué par le prix à la publication, se reflète alors sur le choix des consommateurs, pour lesquels la publication dans une revue ne refléterait plus forcément un processus de sélection et donc de certification de la nouvelle connaissance produite. Les auteurs suggèrent alors une possible solution, qui serait d'imposer des frais à la soumission et des frais à la publication, ce qui pourrait atténuer les intérêts opposés entre la maximisation du profit d'un côté et la certification de la qualité de l'autre.

Un autre obstacle au développement de l'*Open Access* pourrait résider dans le prix imposé à la publication, puisque les auteurs ne sont pas forcément en mesure de financer le coût pour publier leurs articles. Shavell (2010) explique que si ces coûts sont à la charge des auteurs, le danger d'une dissuasion à la publication en accès ouvert est réel. En revanche, si ces coûts sont supportés par les universités ou les institutions auxquelles les chercheurs sont affiliés, le problème disparaît.

Cependant, un tel modèle de publication constitue une dépense importante pour les centres de recherche et les universités qui publient le plus. Le modèle théorique de Müller-Langer et Watt (2010) montre que l'imposition d'un prix à la publication pourrait ainsi être un modèle non-viable pour les centres de recherche caractérisés par un nombre de publications élevé. Pour ces institutions, un tel modèle de publication se traduirait par une dépense majeure par rapport au coût d'accès aux revues scientifiques actuellement payé. Ainsi, en définitive, il s'agit de trouver une solution à un problème de contribution au développement d'un bien collectif, où les premiers joueurs, qui s'avèrent aussi être les plus grands producteurs de connaissance scientifique, doivent supporter des coûts importants au bénéfice de toute la communauté scientifique.

De plus, étant donné l'objectif de maximisation du profit des maisons d'édition commerciales, celles-ci pourraient imposer des prix à la publication qui deviendraient des véritables barrières à la publication, non seulement pour les centres de recherches aux ressources financières les plus modestes, mais aussi pour les pays où le financement de la recherche représente déjà un problème (Scheufen, 2015). C'est ainsi qu'un nouveau système de financement pour l'accès à la connaissance peut être envisagé, pour assurer la transition vers le modèle d'*Open Access* au savoir scientifique.

1.2.4 L'utilisation de l'*Open Access* par la communauté scientifique

Dans cette section, nous allons détailler les résultats de la littérature visant à expliquer quels sont les avantages d'une publication en *Open Access* par rapport à ceux d'une publication standard en *Closed Access* et quelles sont les barrières qui, encore aujourd'hui, limitent l'utilisation de ce nouveau modèle alors qu'il existe pourtant depuis vingt ans.

Les bénéfices de la publication en *Open Access*

En plus de permettre un accès illimité à la littérature scientifique, l'*Open Access* comporte aussi des bénéfices pour les chercheurs au niveau individuel. En effet, l'élimination des barrières de prix devrait permettre une plus grande visibilité des résultats des chercheurs et donc une augmentation de leur réputation au sein de la communauté scientifique. La réputation scientifique des chercheurs est un actif intangible, puisque leur carrière dé-

Source	Méthodologie	Résulats
Lawrence (2001)	Ils analysent 119 924 articles dans le domaine des computer sciences	Les articles en libre accès sont cités 157% plus que les articles en closed access
Antelman (2004)	Les auteurs ont choisi les articles publiés entre 2001 et 2002 dans les 10 meilleurs journaux scientifiques, à partir de la base de données de Web of Science, en se focalisant sur différentes disciplines.	Chaque discipline registre une augmentation du nombre de citations reçues, allant de 45% à 91%
Hajjem et al. (2006)	1 307 038 articles appartenant à 12 différentes disciplines ont été analysés, entre la période 1992 - 2003, à partir de la base de données ISI	Le nombre de citations reçues varie en fonction de la discipline et de la période considérée, mais les auteurs soulignent un effet positif général, qui varie entre 36% - 172 %
Eysenbach (2006)	Les auteurs comparent les articles publiés dans une même revue (Proceedings of the National Academy of Sciences) entre la période	Les articles en *Open Access* sont deux fois plus cités (odds ration = 2,1) dans les premiers 10 mois, avec un odds ratio qui augmente jusqu'à 2,9 après 16 mois de la publication
McCabe et Snyder (2014)	Ils analysent 100 journaux, considérés comme les meilleurs, de différentes disciplines, en utilisant les données fournies par ISI entre la période 1996 - 2005.	Dans ce cas, l'augmentation est plus faible, avec une augmentation du nombre de citations reçues de 8%.
Ottaviani (2016)	Ils analysent 3 850 articles scientifiques *Open Access*, publiés entre la période 1990 - 2013, en les comparant avec 89 895 articles en *Closed Access*, en utilisant la base de données Web of Science et JCR	Leur résultat montre que les citations reçues par les articles en *Open Access* augmentent de 19%.
Piwowar et al. (2018)	Ils utilisent une base de données de 100 000 articles à partir de la base de données Web of Science, publiés entre 2009 et 2015.	Les articles en accès libre bénéficient d'une augmentation du nombre de citations de 18%. Cet effet est dû surtout à la mise en ligne de l'article après la publication dans une archive ouverte (*Green Open Access*) ou à la publication avec le modèle Gold *Open Access* hybride.

TABLE 1.3 – L'*Open Access* et l'augmentation de la visbilité des travaux scientifiques

pend du nombre et de l'importance de leurs publications (Heckman et Moktan, 2020). La perception de l'importance d'un article scientifique est mesurée à partir des citations que l'article reçoit et la quantification de cette perception est devenue l'objet d'étude d'une branche de la science, à savoir la bibliométrie (Ball, 2017). Avec la mise en libre accès des articles scientifiques, la bibliométrie s'est intéressée à l'effet *Open Access*, c'est-à-dire à l'impact de l'élimination de la barrière de prix sur le nombre de citations reçues.

Bien que quelques études suggèrent qu'il pourrait ne pas exister d'effet *Open Access* (Kurtz et al., 2005), la littérature reste quasi-unanime dans l'affirmation que ce modèle augmente le nombre de citations reçues. Le tableau 1.3 offre un aperçu des travaux les plus importants de cette littérature.

En revanche, la discussion portant sur le lien entre qualité des articles mis en libre accès et l'augmentation du nombre de citations reste ouverte. Gargouri et al. (2010) créent

une base de données de 27 197 articles publiés entre 2002 et 2006 dans 1984 journaux. Ils trouvent que l'effet *Open Access*, c'est-à-dire l'augmentation des citations reçues par un article scientifique publié en *Open Access ceteris paribus*, est plus important pour les articles de meilleure qualité. Les chercheurs - consommateurs sélectionnent les articles qu'ils citent en fonction de leur qualité, et le fait d'avoir un libre accès à cette littérature permet d'augmenter le nombre de citations reçues pour ces articles de bonne qualité. Les auteurs expliquent qu'une fois éliminée la barrière de prix, l'augmentation des citations entre les articles de bonne qualité en open et *Closed Access* et de qualité plus faible est plus importante pour les articles du premier type. Pourtant, Xia et Nakamishi (2012) remettent en cause cet impact de l'*Open Access* en termes de citations en fonction de la qualité de l'article. En se focalisant sur la discipline de l'anthropologie, ils étudient 58 journaux à partir de la liste proposée par JCR (*Journal Citation Reports*) parmi lesquels ils choisissent les 10 meilleurs et les 10 dernières journaux, classés en termes d'*impact factor*. Pour chaque article dans chaque journal, ils utilisent *Google* et *Google Scholar* pour déterminer si une copie de l'article est librement accessible. Grâce au nombre de citations enregistrées par JCR, ils constatent que les articles publiés dans les revues ayant un *impact factor* plus faible bénéficient d'une augmentation du nombre de citations plus importante une fois mis en libre accès par rapport aux articles publiés dans les revues les mieux classées.

Les barrières au développement de l'*open acces*

Puisque le modèle de l'*Open Access* augmente la probabilité d'avoir accès à la ressource intellectuelle et donc d'être cité par ses pairs, il pourrait inciter les chercheurs à publier de préférence sous cette forme. En effet, les citations sont une forme d'actif incorporel, déterminant pour la carrière d'un chercheur. Plus un chercheur est connu par la communauté scientifique, notamment grâce à sa contribution au développement de la science, reconnue à travers les citations, plus sa probabilité d'avoir un avancement de carrière et de travailler pour des institutions prestigieuses augmente (Heckman et Moktan, 2020). Pourtant, un nombre important d'éléments a un effet dissuasif sur le choix individuel de l'*Open Access* comme modèle de publication.

Avant tout, la réputation des nouvelles revues adoptant ce modèle est très faible. Il

est en effet long et difficile pour une revue d'établir une réputation solide. Il en résulte une préférence des chercheurs pour la publication dans les revues traditionnelles et bien établies (Schroter et al., 2005 ; Coonin et Younce, 2009). Même si certains chercheurs déclarent souhaiter contribuer au développement de l'OA en choisissant ce modèle pour leurs publications (Uhl, 2009 ; Dallmeier-Tiessen et al., 2011), d'autres barrières en limitent le développement. Ainsi, l'absence d'un système de financement pour couvrir les frais de publication oblige les chercheurs à dédier une partie de leur temps à la recherche de fonds pour la publication, ce qui induit un effet dissuasif à l'adoption de ce modèle (Warlik et Vaughan, 2007 ; Solomon et Björk, 2012).

Même dans le cas du *Green Open Access*, qui s'avère être un choix moins risqué en termes de perte de réputation, car les articles peuvent bénéficier d'une publication préalable à l'archivage dans les répertoires en ligne, l'incertitude juridique liée à son utilisation est dissuasive à son adoption (Cullen et Chawner, 2011 ; Rowley et al., 2017). Très souvent, les auteurs ne savent pas s'ils ont le droit de mettre leur article en ligne après la publication dans une revue scientifique et les politiques très hétérogènes des maisons d'édition augmentent l'incertitude quant à une possible violation du droit d'auteur.

Il existe aussi des différences importantes entre les différentes disciplines scientifiques concernant l'adoption du modèle de publication en *Open Access*. En effet, certaines disciplines, surtout dans les domaines des STM (Science, Technologie et Médicine), tels que les mathématiques, l'informatique et la médecine, enregistrent un taux d'utilisation de l'*Open Access*, dans ses formes *Green* et Gold, supérieur à celui des sciences humaines et sociales (Kling et McKim, 2000 ; Spezi et al., 2013 ; Eger et al., 2015), où le taux de pénétration de l'*Open Access* reste faible. Cependant, ces différences pourraient s'être réduites dans le temps. Rowley et al. (2017), à partir d'une analyse qualitative se fondant sur l'enquête menée en 2014 par la maison d'édition Taylor & Francis, montrent ainsi que les comportements des chercheurs face à l'adoption de l'*Open Access* ne diffèrent pas considérablement entre disciplines et ne vérifient pas l'opposition traditionnelle entre science, technologie et médecine d'un côté et les sciences humaines sociales de l'autre dans le cadre de leur analyse qualitative.

Le désir de contribuer à la création d'une science ouverte, à travers le partage de la

connaissance en *Open Access* est confronté à une réalité, où l'incertitude juridique et un système de financement de la publication peu flexible freinent donc la transition vers le libre accès au savoir scientifique.

1.3 L'accès ouvert et la connaissance scientifique en tant que commun

L'un des apports majeurs de E. Ostrom est d'avoir enrichi la classification économique des biens, en ajoutant au degré d'excluabilité d'un bien (Samuelson, 1954 ; Musgrave, 1959) un niveau plus ou moins élevé de rivalité (Ostrom et Ostrom, 1977). Selon le degré d'excluabilité et de rivalité, la connaissance scientifique peut se présenter sous différentes formes, présentées dans le tableau 1.4.

		Rivalité	
		Faible	Élevée
Excluabilité	Difficile	Bien collectif *Idées, théories*	Bien commun *Bibliothèque*
	Simple	Bien de club *Abonnement à une revue*	Bien privé *Un livre acheté*

TABLE 1.4 – La classification des biens (à partir de Hess et Ostrom, 2007)

En tant que bien collectif, la connaissance se présente sous forme intangible et acquiert un caractère de non-rivalité : chaque individu peut y avoir accès sans empêcher les autres agents de bénéficier du même bien. Pour illustrer cette dimension de bien collectif, on peut penser à une théorie scientifique, dont la connaissance par un individu n'empêche pas les autres de pouvoir l'apprendre. Quand la connaissance est explicitée à travers un support papier, sa nature change en fonction de la gestion des droits qui en règlent l'accès et l'utilisation. Prenons à titre d'exemple une revue scientifique. Celle-ci peut être considérée comme un bien privé si l'achat a été effectué par un et un seul agent, qui dispose d'un pouvoir d'exclusion à la jouissance de sa propre copie par rapport à tout autre consommateur. En revanche, quand l'accès à cette revue est effectué à travers une bibliothèque, celle-ci devient un bien commun, puisque la jouissance de ce bien par un membre de la bibliothèque empêche la consultation de la revue par tous les autres membres. La même revue peut

aussi être considérée comme un bien de club, à partir du moment où sa consultation est soumise à un prix, mais cela n'empêche pas les autres utilisateurs abonnés d'avoir accès à la même revue, puisque chacun est en mesure de disposer de sa propre copie. Mis à part pour le cas d'un bien collectif, les autres catégories se différencient par la gestion du droit d'auteur auquel la revue est soumise et qui détermine les conditions d'accès et d'utilisation du savoir scientifique incorporé dans le journal.

Avec la numérisation de l'information, et la dématérialisation des supports à travers lesquels la connaissance est explicitée, le modèle propriétaire de gestion du droit d'auteur est remis en cause. L'*Open Access* peut alors être conçu comme une initiative privée qui vise à subvertir le régime de propriété intellectuelle et qui s'oppose à la marchandisation croissante de la publication scientifique (Dusollier, 2007). Néanmoins, si une partie de la littérature définit l'*Open Access* comme l'instrument permettant l'abolition du droit d'auteur dans le contexte académique (Feess et Schefen, 2016), cette institution reste centrale dans la gestion des droits liés à la circulation et l'exploitation des résultats de la recherche. L'objectif est plutôt de renouveler la gestion du faisceau de droits, à travers une négociation privée qui devient un instrument complémentaire au modèle d'ordre public. L'*Open Access* peut ainsi plutôt être défini comme l'instrument de gestion de type ordre privé de la propriété intellectuelle (Dusollier, 2007), permettant d'inscrire la connaissance scientifique dans la littérature des communs, où la gestion du savoir est assurée par une communauté spécifique, à travers des mécanismes autres que la propriété privée.

Dans cette section, nous allons étudier quelles sont les caractéristiques de ces communs et comment la littérature économique analyse la gestion commune du savoir scientifique à l'ère du numérique.

1.3.1 La numérisation de l'information : de biens communs à communs informationnels

La numérisation de l'information a donc permis la naissance des nouveaux communs, appelés aussi communs informationnels ou encore communs créatifs (Hess et Ostrom, 2007). Afin d'éclairer ce concept et d'éviter toute confusion, Hess (2008) propose une classification trichotomique : les biens communs, la propriété commune et les communs. Les biens com-

muns sont des biens analysés à partir de la classification standard des biens en économie, à savoir à travers leur degré de rivalité et d'excluabilité. Ils sont alors des biens rivaux, mais difficilement excluables. L'exemple typique est la pêcherie, utilisée par Hardin (1968)dans son article *The Tragedy of the Commons* pour expliquer les effets néfastes de l'exploitation des ressources naturelles en l'absence de droits de propriété. Le problème de l'analyse de Hardin est que cet auteur a confondu le manque d'un droit formel avec toute absence de régulation. Dans plusieurs travaux, Ostrom souligne qu'une régulation de nature non légale et informelle peut assurer le maintien de la ressource naturelle, à travers la mise en place de normes par la communauté responsable d'en réguler l'accès et l'utilisation (Ostrom et al., 1999 ; Ostrom et al., 1994 ; Ostrom, 2008). La propriété commune en est un exemple. Cette dernière est définie par Hess (2008) comme un type formel ou informel de régime juridique, qui permet l'allocation de certains droits, tels que l'accès et l'utilisation de la ressource, parmi les membres d'un certain groupe. Cette définition reste très large, mais elle nous permet de comprendre que l'objet d'étude est bien une certaine régulation, et non pas la ressource sous-jacente.

Les communs sont quant à eux des biens ayant les caractéristiques d'un bien collectif, bien non-rival et non-excluable, dont la gestion est assurée par un groupe d'agents appartenant à une communauté spécifique (Ostrom, 2009 ; Coriat, 2015 ; Boyle, 2003). Ils font l'objet de débats importants dans la littérature, l'intégration des biens communs dans les communs pouvant notamment s'interpréter comme le signe du caractère imparfaitement stabilisé des définitions, classifications et caractéristiques des deux catégories. Afin de clarifier le concept, Hess (2008) a proposé une cartographie des communs, qu'elle définit plus particulièrement en tant que « nouveaux communs », comme l'illustre la figure 1.3.

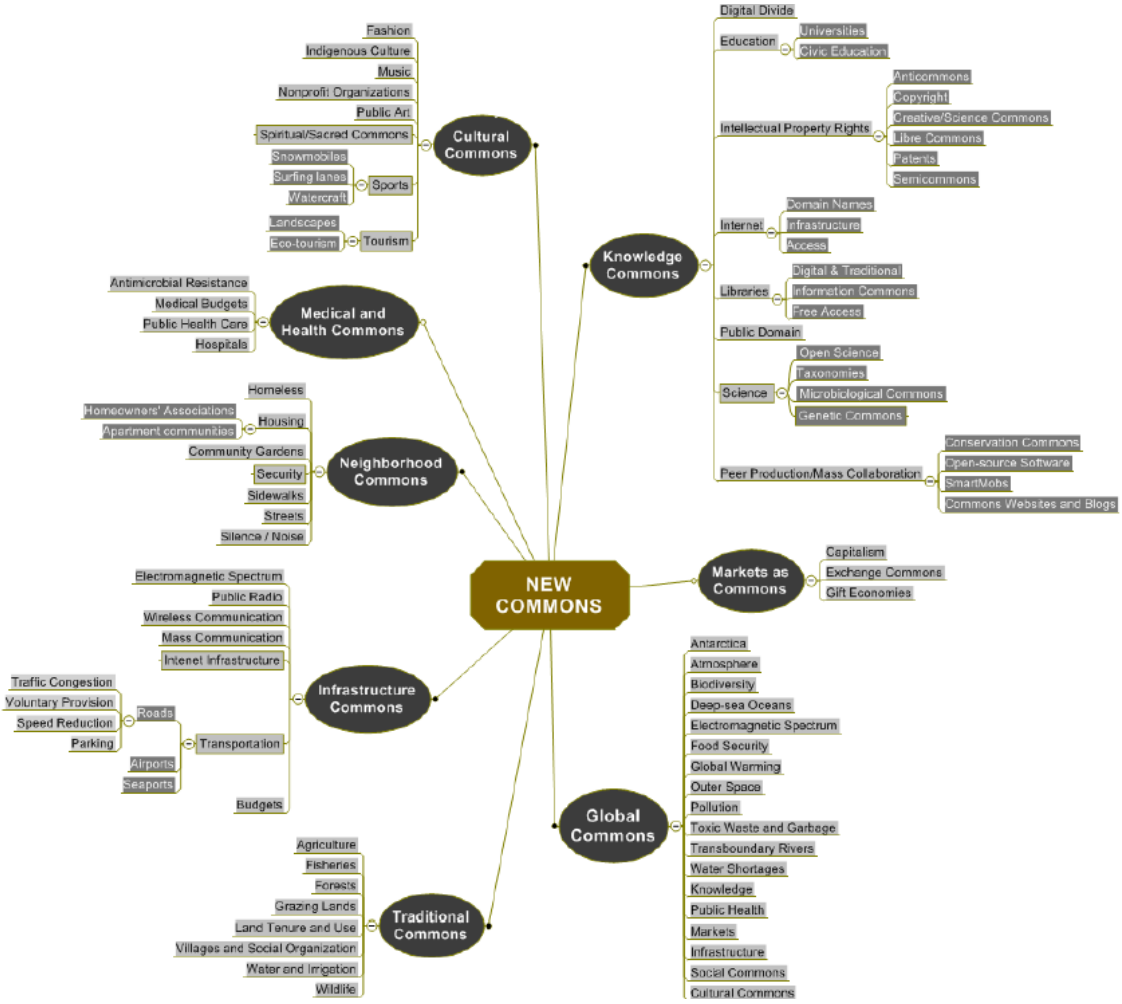

FIGURE 1.3 – Les nouveaux communs. Source : Hess (2008)

La « nouveauté » des « nouveaux communs » de Hess (2008) tient à deux raisons principales. D'une part, ils sont à l'origine d'une attention renouvelée vis-à-vis des biens communs traditionnels. Ainsi, non seulement la gestion des déchets (Anand, 2000), du trafic routier (Rosin, 1998), mais aussi la privatisation des infrastructures (Bowden et deJong, 2006) et la protection de l'environnement (Boda, 2003 ; Anderson et Grewell, 1999) font aujourd'hui partie de la nouvelle littérature sur les communs, justement en raison de l'intérêt renouvelé et croissant pour ces problématiques. D'autre part, selon Hess (2008), la nouveauté des communs peut tenir à ce qu'ils résultent des évolutions technologiques et de la numérisation. L'idée de communs numérique prend alors tout son sens. En effet, dans le domaine du savoir et de la connaissance, la numérisation a eu un impact profond et indéniable sur les processus de production et de dissémination. L'émergence de nouvelles collaborations

grâce à l'utilisation de l'internet pour la production du savoir (Bollier, 2001, 2007 ; Benkler, 2004), l'étude des conséquences de l'application du cadre juridique de la propriété intellectuelle à l'environnement ainsi reconfiguré par la numérisation (Boyle, 2002, 2003) et en général la capacité de l'Internet à faciliter la création d'activités communes invitent ainsi à analyser la connaissance comme un nouveau commun, en raison de la dématérialisation de supports papier et donc à la transformation du savoir en bien non-rival et difficilement excluable. C'est véritablement dans cette deuxième catégorie que la connaissance scientifique en tant que commun s'inscrit.

1.3.2 La gestion commune du savoir scientifique

L'étude de la connaissance scientifique en tant que commun est relativement nouvelle. Afin de comprendre en quoi la connaissance scientifique peut s'inscrire dans le cadre d'analyse des communs, nous allons en étudier les caractéristiques puis le régime de gestion auquel l'accès et le partage de la ressource sont soumis.

Les caractéristiques des communs scientifiques

Avant la révolution numérique, la connaissance scientifique était accessible soit à travers l'achat du support matériel à travers lequel la connaissance était transmise, étant ainsi un bien privé ou un bien de club, soit à travers un accès indirect, tel que l'accès assuré par les bibliothèques, ce qui faisait du savoir un bien commun. La numérisation permet de classifier la connaissance en tant que bien collectif et de concevoir des nouvelles modalités de partage du savoir, qui vont au-delà du régime de propriété privée. La séparation entre les caractéristiques de la ressource et le régime propriétaire qui en gouverne l'accès et le partage est l'un des apports majeurs de l'analyse économique des communs opérées par Ostrom (Ostrom et Ostrom, 1977 ; Coriat, 2015).

Si l'étude des communs occupe une place importante dans la recherche actuelle, son application au savoir scientifique reste un domaine de recherche très récent et en devenir, avec une première littérature qui ne voit le jour qu'avec le travail d'Hess et Ostrom en 2007. Cet ouvrage, intitulé *Understanding knowledge as a commons*, est le résultat de la collaboration entre les études d'Ostrom portant sur les biens communs et de Hess,

spécialiste des mécanismes de partage de l'information. Après avoir identifié et proposé pour les recherches futures un cadre d'analyse pour étudier la connaissance en tant que commun appelé *Institutional Analysis Developement framework* (IAD), cet ouvrage collectif se focalise sur les mécanismes juridiques et informels visant la protection du commun de la connaissance (Boyle, chapitre 4 ; Waters, chapitre 5), pour ensuite analyser non seulement les différentes formes de gestion commune de la connaissance, telles que les logiciels en *open source* (Schweik, chapitre 10), le rôle des bibliothèques dans un environnement numérisé (Lougee, chapitre 11), mais aussi le rôle de l'*Open Access* dans la gestion du savoir en tant que commun (Suber, chapitre 7).

La caractéristique centrale de ce nouveau type de bien devient alors la gouvernance opérée par une collectivité, qui assure une régulation de l'accès à la ressource et son utilisation, à travers des mécanismes autres que ceux du marché. Ce n'est pas l'absence de droits qui définit le commun, mais une gestion différente des droits attachés à la propriété intellectuelle (Hess et Ostrom, 2007). Cette nouvelle régulation des droits, souvent du type *bottom-up* (Lessig, 2002), est prévue par l'*Open Access* en tant que forme de régulation du partage de la connaissance, opérée par la communauté scientifique. En tant que modèle de publication, il se propose comme alternative à la marchandisation du savoir scientifique résultant du régime propriétaire du droit d'auteur (Dusollier, 2007 ; Suber, 2012). Cependant, en l'absence de régulations capables de réguler les nouvelles modalités du partage des résultats de la recherche et de limiter les effets négatifs du droit d'auteur à l'ère du numérique, l'*Open Access* risque de se traduire par un simple modèle alternatif au *Closed Access*, qui ne permettrait pas de profiter des bénéfices de la numérisation et d'atteindre une gestion commune du savoir scientifique.

Pour ce faire, il a fallu concevoir des nouvelles régulations, dont les licences créatives sont un exemple, qui complètent et organisent l'*Open Access* et permettent la mise en place du commun de la connaissance scientifique (voir la section « Les nouvelles normes légales pouur une gestion commune du savoir »).

Les régulations de la communauté scientifique

Une gestion pas totalement soumise au régime d'ordre public imposé par le droit d'auteur a depuis toujours caractérisé la gestion de la connaissance scientifique. Même avant l'introduction des licences créatives, et donc de la remise en cause du système d'ordre public du droit d'auteur, la communauté scientifique s'était dotée d'un système normatif capable d'assurer des incitations à la production et diffusion du savoir et de régler l'accès et la réutilisation des résultats de la recherche. Bien avant les premières études en économie de la science sur les incitations à la production individuelle d'un bien public dérivant des normes de la communauté scientifique, la sociologie s'était emparée de ce domaine d'étude à partir des années 1950. Dans *The Sociology of Science. Theoretical and Empirical Investigations*, Merton (1973) trace la structure organisationnelle et normative de la communauté scientifique, en enquêtant sur les motivations qui assurent la production et la diffusion de la science, en tant qu'ensemble de méthodes par lesquelles la connaissance est certifiée. Si Polanyi (1962) consacre son attention à l'ajustement mutuel qui coordonne les initiatives indépendantes de production du savoir scientifique, Merton se focalise sur la structure normative de la science, en expliquant que l'objectif commun de la diffusion de la connaissance certifiée de la communauté scientifique est assuré à travers un système d'incitations individuelles associées au désir d'affirmer la priorité de la découverte et de bénéficier ainsi de l'estime de ses pairs. Les journaux scientifiques, commerciaux ou non, deviennent dans ce cadre les instruments permettant de diffuser les nouveaux résultats et d'en certifier la véracité. Ce désir individuel de l'estime d'autrui devient l'élément essentiel de la théorie de l'estime développée par McAdams (1997), où la recherche de l'approbation par les agents appartenant à la même communauté amène à l'émergence de normes sociales qui influencent les comportements des membres de la communauté de référence. Les normes sociales sont ainsi conçues comme des mécanismes de coopération non légaux qui influencent les comportements individuels et permettent d'atteindre un objectif commun (Ellickson, 1991 ; Posner, 2000). En tant que mécanismes de contrôle et d'incitation à la conformité, elles acquièrent un rôle central, notamment dans les groupes sociaux où l'action collective et la reconnaissance réciproque ont une plus grande efficacité que la régulation verticale et centralisée (Ostrom, 2009).

A l'intérieur de la communauté scientifique, les normes sociales organisent la coopération entre les objectifs individuels, en vue de l'objectif commun de l'avancement de la science. Merton (1973) distingue ainsi des normes dites de principe, les normes CUDOS (*Communalism, Universalism, Disinterestedness, Organisd Skepticism*), qui orientent le comportement des chercheurs vers un modèle de développement de la science que l'on peut considérer comme précurseur de ce qu'on appelle aujourd'hui la science ouverte. Ces normes de principes assurent l'évaluation objective des résultats de la recherche, à travers des critères impersonnels, ce qu'on appelle l'universalisme de la science. D'un point de vue propriétaire, le seul droit reconnu au producteur du nouveau savoir est la reconnaissance de sa contribution à l'avancement de la science et l'estime des pairs, dans une optique de production et de propriété commune de la connaissance. Enfin, la recherche doit être désintéressée, dans la mesure où les convictions personnelles ou les objectifs individuels ne doivent pas fausser les résultats finaux, et guidée par un scepticisme organisé. C'est seulement à travers la validation des hypothèses par une analyse objective que l'avancement de la science peut être assuré. C'est à travers l'internalisation de ces normes que les scientifiques s'identifient comme membres de la communauté scientifique et collaborent pour atteindre l'objectif final d'avancement scientifique recherché. En outre, Merton explique que l'objectif de l'extension de la connaissance certifiée est mis en oeuvre à partir de ce qu'il appelle des normes techniques (Merton, 1973, p. 270), qui constituent des normes de cohérence logique et d'évidence empirique. Cette définition des normes techniques, que Merton ne développe pas autant que celle des normes de principe, n' explique cependant pas la mise en place d'un système de rémunération intellectuelle et de sanctions sociales qui participent également à la réalisation de l'objectif commun de développement scientifique. Ces mécanismes de contrôle social et de rémunération dans l'organisation scientifique sont représentés, par exemple, par la norme de crédit, qui oblige les chercheurs à attribuer la paternité de l'idée à son auteur, et la sanction sociale associée au plagiat, exprimée par différents degrés d'ostracisme en cas d'appropriation des idées d'autrui.

Ce système de normes participe de la régulation de la dissémination de la connaissance scientifique depuis le XVème siècle, à des degrés divers. Selon les époques, il a joué un rôle plus ou moins important, alternatif ou complémentaire au système juridique de gestion

des droits de propriété intellectuelle. Il fait aujourd'hui tout particulièrement écho à la problématique de l'*Open Access*, dont certaines institutions et régulations peuvent sembler inspirées par les normes sociales traditionnelles de la communauté scientifique.

Les nouvelles normes légales pour une gestion commune du savoir

Les possibilités de partage et de réutilisation du savoir scientifique sont mieux exploitables sous le régime d'*Open Access* que dans la situation de rareté artificielle créée par le droit d'auteur, l'*Open Access* permettant de subvertir le pouvoir de monopole assuré par le droit d'auteur et la relation entre droit à l'exclusivité et droit à l'exclusion. Sous le régime de gestion commune, « *[t]he exclusivity conferred by the intellectual property right is thus conceived not as an exclusionary power but as a liberty or monopoly to decide not to engage in exclusion* » (Dusollier, 2007). En ce sens là, le *copyright* ne donne pas un pouvoir pour exclure les utilisateurs de la jouissance du bien, mais il accorde plutôt le droit de décider de ne pas s'engager dans cette exclusion, notamment à travers l'utilisation de licences.

Le modèle en *Open Access* est alors l'élément préalable mais pas suffisant pour la mise en place du commun scientifique, et les normes endogènes qui règlent le fonctionnement de la communauté scientifique doivent être enrichies, afin d'adapter la gestion du partage du savoir à l'ère du numérique. La voie principale à travers laquelle la gestion commune de la connaissance scientifique peut être assurée est alors celle de la promotion de la réutilisation des résultats de la recherche à travers des mécanismes d'ordre privé, tels que les licences *copyleft*. Ces dernières se définissent comme des contrats privés qui disciplinent l'*an*, le *quomodo* et le *quantum* de l'utilisation de l'œuvre protégée par le droit d'auteur (Cogo, 2010). Appelées licences *copyleft* par opposition au *copyright*, elles permettent de créer des exceptions à la règle générale de « tous droits réservés » imposée par le droit d'auteur et de favoriser le partage et les utilisations secondaires de l'output intellectuel. Si les licences *copyleft* peuvent être utilisées pour renforcer le pouvoir assuré par les droits de propriété intellectuelle et prévenir toute utilisation de l'œuvre de l'esprit, elles peuvent aussi s'opposer à cette logique d'exclusion et promouvoir l'utilisation collective de la ressource.

Parmi les licences *copyleft*, les plus utilisées dans le milieu académique sont les licences *Creative Commons*. Ces contrats standardisés déterminent en amont les exceptions aux

interdictions imposées par le droit d'auteur, où tous les droits sont réservés au titulaire du droit. Elles agissent en tant que forces de désintermédiation (Carroll, 2006) et permettent la décentralisation de la gestion de la connaissance scientifique, assurée dans le système traditionnel par les revues scientifiques, à travers une négociation directe entre l'auteur et le consommateur final. Ces licences ont émergé lors de la Conférence de Budapest en 2002 en tant qu'instruments complémentaires au droit d'auteur pour assurer un accès ouvert et permanent au savoir scientifique. Conçues par L. Lessig en 2001, elles sont des licences à géométrie variable et peuvent donner lieu, à travers la combinaison de quatre clauses, à sept différents types de licences, avec une privatisation du bien sous-jacent plus ou moins étendue, comme le montre la figure 1.4.

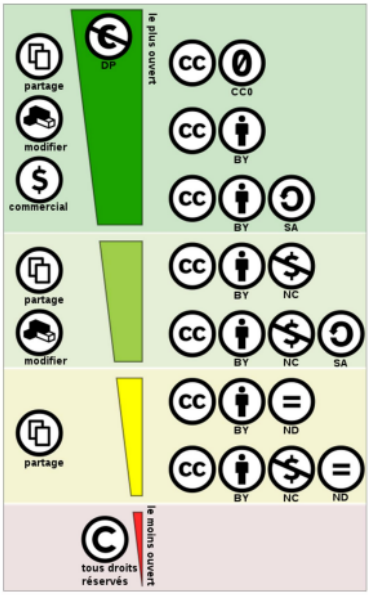

FIGURE 1.4 – Les licences créatives. Source Wikimedia Commons

Toutes ces licences ont comme objectif final le partage le plus large possible du travail protégé par le droit d'auteur, à travers des mécanismes d'ordre privé (Katz, 2006 ; Dusollier, 2007), tout en assurant l'attribution de la paternité à l'auteur de l'œuvre originale. L'idée des licences créatives est que l'auteur se prive de certains droits accordés par le *copyright* et donne son œuvre à la collectivité, comme une forme de production volontaire de savoir (Poynder, 2003 ; Cahir, 2004).

La première dérogation de la licence créative est qu'elle permet le partage automatique de la ressource intellectuelle sans autorisation spécifique et préalable de l'auteur. En ef-

fet, cette autorisation est contenue dans la licence même, qui permet à tout utilisateur de reproduire, d'executer, de communiquer et de distribuer des copies de l'œuvre originelle (Dusollier, 2007). Ce partage peut se faire pour des finalités commerciales ou pas. En pratique, la finalité commerciale reste une condition peu claire puisqu'elle n'a pas encore été définie ni par l'association des *Creative Commons* ni par la plupart les régimes juridiques (Hagedorn et al., 2011). La question de savoir si le partage doit être suivi par une compensation monétaire ou ne doit comporter aucun avantage commercial reste donc encore aujourd'hui une question ouverte (ibid.).

La deuxième forme de dérogation possible liée à l'utilisation des licences créatives est la modification de l'œuvre, qui permet de modifier la ressource originale, à condition de le faire pour des finalités autres que commerciales. Dans ce cas, l'auteur de l'œuvre originale peut imposer à l'auteur de l'œuvre dérivée de partager la nouvelle ressource sous les mêmes conditions que celle originale, en créant ainsi une propagation virale des œuvres dérivées (Dusollier, 2007). Une telle modification de l'œuvre serait impossible sous le régime standard assurée par le *copyright*. Cependant, son application au domaine scientifique a soulevé un certain nombre de questions, notamment en ce qui concerne l'attribution de la contribution à l'avancement de la science, comme souligné par Schweik (2007). Selon ce dernier, sous le régime de la licence *Creative Commons*, un papier scientifique théorique, étudiant les facteurs de changement d'affectation des terres dans une certaine région aux États-Unis, pourrait être modifié à la marge par d'autres auteurs et appliqué à une autre région. Dès lors, même en cas de maintien dans l'œuvre dérivée des auteurs du papier original, comment attribuer à chaque chercheur sa contribution à l'avancement de la science ? Ces questions, et plus généralement l'application des licences *Creative Commons* au domaine scientifique, restent ouvertes.

Dans leur forme la plus permissive, les *Creative Commons* permettent non seulement de partager et de modifier l'œuvre, mais aussi de le faire pour toute finalité commerciale. Dans ce cas, l'auteur de l'œuvre originale peut, à nouveau, déterminer les conditions de partage des œuvres dérivées. Une nouvelle forme de *Creative Commons*, appelée zéro ou *public domain,* a été introduite en ce sens. Sous cette licence, l'œuvre est considérée immédiatement en tant que bien collectif, sans attendre 70 ans après la mort de l'auteur

original, comme prévu par le droit d'auteur dans la plupart des pays. L'association des *Creative Commons* fournit à ce titre des exemples d'utilisation des licences CC0, comme Europeana, une bibliothèque numérique, qui utilise cette licence pour la mise à disposition de sa base de données, et figshare, archive en ligne pour chercheurs qui oblige les utilisateurs à partager leurs bases de données utilisées dans la rédaction de l'article archivé.

En définitive, seule la combinaison des *Creative Commons* avec l'*Open Access* en tant que modèle de dissémination du savoir permet d'inscrire la connaissance scientifique dans le domaine des communs. Elle permet ainsi à la communauté scientifique de promouvoir une nouvelle modalité de gestion du partage du savoir scientifique. Cette gestion en commun n'implique pas un vide juridique. Au contraire, elle établit ses bases sur le droit d'auteur même, en permettant d'exploiter les possibilités offertes par la numérisation de la connaissance.

Conclusion

La numérisation de l'information a permis l'émergence d'instruments de diffusion et de gestion de la connaissance scientifique autres que ceux fondés sur le régime d'ordre public assuré par le droit d'auteur. Outre le fait de permettre l'enchaînement du principe du partage de la connaissance scientifique, qui est le pilier fondateur de la science ouverte, les *Creative Commons* permettent de contourner les journaux scientifiques en tant qu'intermédiaires dans la gestion de la diffusion du savoir scientifique. Cette forme renouvelée de gestion commune du savoir, qui trouve ses racines dans les normes sociales de principe et les normes techniques gouvernant l'organisation de la communauté scientifique, devient nécessaire pour faire face aux défis liés à la numérisation du savoir scientifique et en exploiter l'ensemble des possibilités.

Même si le droit d'auteur a représenté pendant longtemps le meilleur compromis entre efficacité statique et dynamique, la rareté artificielle créée par cet instrument juridique ne semble plus être justifiée suite à la dématérialisation de la connaissance scientifique. L'analyse économique du droit d'auteur nous a ainsi permis de comprendre le fonctionnement du système traditionnel de diffusion du savoir et donc de la régulation de l'accès et du par-

tage de la ressource. Dans ce système, l'appropriation des droits par les maisons d'édition leur permet de bénéficier d'une situation de monopole et de rémunérer leur activité. En contrepartie, les chercheurs bénéficient de la dissémination des résultats de la recherche et de la transformation de leurs outputs en actifs incorporels, valorisables sur le marché du travail.

L'émergence de nouvelles méthodes de diffusion de la connaissance scientifique a remis en cause le modèle de publication en *Closed Access* et, de manière plus générale, l'efficacité du système d'ordre public dont le droit d'auteur est l'expression principale en matière de gestion du savoir scientifique. Défini comme un modèle de gestion de l'accès, du partage et de la réutilisation de la connaissance alternatif aux mécanismes de marché incarnés par le droit d'auteur, l'*Open Access* est aussi un modèle de partage du savoir, dûment analysé par la littérature en tant que tel.

Cependant, la remise en cause du modèle propriétaire doit s'inscrire dans une analyse institutionnelle plus large, en étudiant l'interdépendance existant entre les différentes normes, légales et sociales, qui règlent et coordonnent les comportements individuels dans la perspective de l'avancement de la science. Selon une grande partie de la littérature, la réforme du système de gestion de la connaissance, qui passe par une transition d'un régime essentiellement régulé par le droit d'auteur à une gestion commune des droits à l'accès et à la réutilisation des résultats de la recherche, s'inscrit dans un renouvellement de la gestion commune de la connaissance scientifique, où normes légales et sociales s'articulent pour faciliter le développement d'une science ouverte.

Bibliographie du chapitre 1

Anand, P. B. (2000). Co-operation and the urban environment : An exploration. *Journal of Development Studies 36*(5), 30–58. Taylor & Francis Journals.

Anderson, T. and J. B. Grewell (1999). Property Rights Solutions for the Global Commons : Bottom-Up or Top-Down ? *Duke Environmental Law & Policy Forum 10*(1), 73–102.

Antelman, K. (2017). Do Open-Access Articles Have a Greater Research Impact ? | Antelman | College & Research Libraries.

Archambault, , D. Amyot, P. Deschamps, A. Nicol, F. Provencher, L. Rebout, and G. Roberge (2014). Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels—1996–2013. *Copyright, Fair Use, Scholarly Communication, etc.*.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. NBER Chapters, National Bureau of Economic Research, Inc.

Ball, R. (2017). *An Introduction to Bibliometrics : New Development and Trends*. Chandos Publishing.

Beall, J. (2012). Predatory publishers are corrupting open access. *Nature 489*(7415), 179–179. Nature Publishing Group.

Benhamou, F. and J. Farchy (2009). *Droit d'auteur et copyright*. Repères. La Decouverte.

Benkler, Y. (2002). Intellectual property and the organization of information production. *International Review of Law and Economics 22*(1), 81–107. Elsevier.

Benkler, Y. (2004). Intellectual property. Commons-based strategies and the problems of patents. *Science (New York, N.Y.) 305*(5687), 1110–1111.

Bergstrom, T. C. (2001). Free Labour for Costly Journals ? *Journal of Economic Perspectives 15*(4), 183–198.

Bergstrom, T. C., P. N. Courant, R. P. McAfee, and M. A. Williams (2014). Evaluating big deal journal bundles. *Proceedings of the National Academy of Sciences of the United States of America 111*(26), 9425–9430.

Bernault, C. (2016). *Open access et droit d'auteur*. Larcier.

Björk, B.-C. (2014). Open access subject repositories : An overview. *Journal of the Association for Information Science and Technology 65*(4), 698–706.

Björk, B.-C. (2017). Scholarly journal publishing in transition- from restricted to open access. *Electronic Markets 27*(2), 101–109. Springer & IIM University of St. Gallen.

Björk, B.-C., M. Laakso, P. Welling, and P. Paetau (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology 65*(2), 237–250.

Björk, B.-C. and D. Solomon (2014). How research funders can finance APCs in full OA and hybrid journals. *Learned Publishing 27*(2), 93–103.

Björk, B.-C., P. Welling, M. Laakso, P. Majlender, T. Hedlund, and G. Guðnason (2010). Open Access to the Scientific Journal Literature : Situation 2009. *PLOS ONE 5*(6), e11273. Public Library of Science.

BODA, Z. (2003). GLOBAL ENVIRONMENTAL COMMONS AND THE NEED FOR ETHICS. *Society and Economy 25*(2), 213–224. Akadémiai Kiadó.

Bodó, B. (2016). Pirates in the Library – An Inquiry into the Guerilla Open Access Movement. SSRN Scholarly Paper ID 2816925, Social Science Research Network, Rochester, NY.

Bohannon, J. (2016). The frustrated science student behind Sci-Hub. *Science 352*(6285), 511–511. American Association for the Advancement of Science.

Bollier, D. (2001). *Public assets, private profits : Reclaiming the American commons in an age of market enclosure.* (New American Foundation. ed.).

Bollier, D. (2007). *The growth of the commons paradigm.* MIT Press.

Borgman, C. L. and J. Furner (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology 36*(1), 2–72.

Bourdieu, P. (1984). *Homo academicus.* Editions de Minuit.

Bowden, N. and M. De Jong (2006). Privatisation of seaport infrastructures : a framework for understanding the transfer of property rights. *International Journal of Critical Infrastructures 2*(2-3), 294–317. Inderscience Publishers.

Boyle, J. (2002). Fencing off Ideas : Enclosure & the Disappearance of the Public Domain. *Daedalus 131*(2), 13–25. The MIT Press.

Boyle, J. (2003). The Second Enclosure Movement and the Construction of the Public Domain. *Law and Contemporary Problems 66*(1/2), 33–74. Duke University School of Law.

Boyle, J. (2007). Mertonianism Unbound ? : Imagining Free, Decentralized Access to Most Cultural and Scientific Material. *Understanding Knowledge As A Commons*, 1–21.

Breyer, S. (1970). The Uneasy Case for Copyright : A Study of Copyright in Books, Photocopies, and Computer Programs. *Harvard Law Review 84*(2), 281–351. The Harvard Law Review Association.

Cabanac, G. (2016). Bibliogifts in LibGen ? A study of a text-sharing platform driven by biblioleaks and crowdsourcing. *Journal of the Association for Information Science and Technology 67*(4), 874–884.

Cahir, J. (2004). The Withering Away of Property : The Rise of the Internet Information Commons. *Oxford Journal of Legal Studies 24*(4), 619–641. Oxford University Press.

Carroll, M. W. (2006). Creative Commons and the New Intermediaries W(h)ither the Middleman : The Role and Future of Intermediaries in the Information Age. *Michigan State Law Review 2006*(1), 45–66.

Coase, R. H. (1960). The Problem of Social Cost. In C. Gopalakrishnan (Ed.), *Classic Papers in Natural Resource Economics*, pp. 87–137. London : Palgrave Macmillan UK.

Cogo, A. (2010). *I contratti di diritto d'autore nell'era digitale*. Giappichelli.

Coonin, B. and L. Younce (2009). Publishing in open access journals in the social sciences and humanities : Who's doing it and why. *ACRL fourteenth national conference*, 12–15.

Coriat, B. (2015). *Le retour des communs : la crise de l'idéologie propriétaire*. Liens qui libèrent.

Crawford, W. (2002). Free electronic refereed journals : getting past the arc of enthusiasm. *Learned Publishing 15*(2), 117–123.

Cullen, R. and B. Chawner (2011). Institutional Repositories, Open Access, and Scholarly Communication : A Study of Conflicting Paradigms. *The Journal of Academic Librarianship 37*(6), 460–470.

Dallmeier-Tiessen, S., R. Darby, B. Goerner, J. Hyppoelae, P. Igo-Kemenes, D. Kahn, S. Lambert, A. Lengenfelder, C. Leonard, S. Mele, M. Nowicka, P. Polydoratou, D. Ross, S. Ruiz-Perez, R. Schimmer, M. Swaisland, and W. van der Stelt (2011). Highlights from the SOAP project survey. What Scientists Think about Open Access Publishing. *arXiv :1101.5260 [cs]*.

Darby, R. M., C. M. Jones, L. D. Gilbert, and S. C. Lambert (2009). Increasing the Productivity of Interactions Between Subject and Institutional Repositories. *New Review of Information Networking 14*(2), 117–135. Routledge.

David, P. (1993). Intellectual property institutions and the panda's thumb : patents, Copyrights, and trade secrets in economic theory and history. In *Global dimensions of intellectual property rights in science and technology* (National Academy of Science ed.)., 19,29.

David, P. (2003). The economic logic of "open science" and the balance between private property rights and the public domain in scientific data and information : A primer. In J. Esanu and P. Uhlir (Eds.), *The role of the public domain in scientific and technical data and information* (National Research Council of the National Academies ed.)., pp. 19–34.

David, P. A. and D. Foray (2002). An introduction to the economy of the knowledge society. *International Social Science Journal 54*(171), 9–23.

Demsetz, H. (1967). Toward a Theory of Property Rights. *The American Economic Review 57*(2), 347–359. American Economic Association.

Dusollier, S. (2007). Sharing Access to Intellectual Property through Private Ordering. *Chicago-Kent Law Review 82*(3), 1391.

Edlin, A. S. and D. L. Rubinfeld (2004). Exclusion or Efficient Pricing - The Big Deal Bundling of Academic Journals Symposium : Integrating New Economic Learning with Antitrust Doctrine. *Antitrust Law Journal 72*(1), 119–158.

Edwards, R. and D. Shulenburger (2003). The High Cost of Scholarly Journals : (And What To Do About It). *Change : The Magazine of Higher Learning 35*(6), 10–19. Routledge.

Eger, T., M. Scheufen, and D. Meierrieks (2015). The determinants of open access publishing : survey evidence from Germany. *European Journal of Law and Economics 39*(3), 475–503.

Ellickson, R. (2021). *Order without Law*. Harvard University Press. Order without Law.

Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLOS Biology 4*(5), e157. Public Library of Science.

Feess, E. and M. Scheufen (2016). Academic copyright in the publishing game : a contest perspective. *European Journal of Law and Economics 42*(2), 263–294.

Foray, D. (2000). *L'économie de la connaissance* (La Découverte ed.). Repères.

Fortson, D. (2019). A Napster moment for ivory towers. *Sunday Times (Business Section)*.

Frazier, K. (2001). The librarians' dilemma : contemplating the costs of the. *D-Lib magazine 7*(03).

Frosio, G. (2014). Open Access Publishing : A Literature Review. SSRN Scholarly Paper ID 2697412, Social Science Research Network, Rochester, NY.

Fuchs, C. and M. Sandoval (2013). The Diamond Model of Open Access Publishing : Why Policy Makers, Scholars, Universities, Libraries, Labour Unions and the Publishing World Need to Take Non-Commercial, Non-Profit Open Access Serious. *tripleC : Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society 11*(2), 428–443.

Gadd, E. and T. Covey (2016). What it means to be Green : exploring publishers' changing approaches to Green open access.

Gardner, C. C. and G. J. Gardner (2017). Fast and Furious (at Publishers) : The Motivations behind Crowdsourced Research Sharing | Gardner | College & Research Libraries.

Gargouri, Y., C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, and S. Harnad (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLOS ONE 5*(10), e13636. Public Library of Science.

Ginsburg, J. (1990). A Tale of Two Copyrights : Literary Property in Revolutionary France and America. *Tul. L. Rev. 64*, 991.

Ginsparg, P. (2004). Scholarly Information Network. In E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai (Eds.), *Complex Networks*, Lecture Notes in Physics, pp. 313–336. Berlin, Heidelberg : Springer.

Goldstein, P. (2001). *International Copyright : Principles, Law, and Practice*. Oxford University Press.

Gordon, W. J. (1982). Fair Use as Market Failure : A Structural and Economic Analysis of the "Betamax" Case and Its Predecessors. *Columbia Law Review 82*(8), 1600–1657. Columbia Law Review Association, Inc.

Green, T. (2018). We're Still Failing to Deliver Open Access and Solve the Serials Crisis : To Succeed We Need a Digital Transformation of Scholarly Communication Using Internet-Era Principles. SSRN Scholarly Paper ID 3333907, Social Science Research Network, Rochester, NY.

Hagedorn, G., D. Mietchen, R. A. Morris, D. Agosti, L. Penev, W. G. Berendsohn, and D. Hobern (2011). Creative Commons licenses and the non-commercial condition : Implications for the re-use of biodiversity information. *ZooKeys* (150), 127–149.

Hajjem, C., S. Harnad, and Y. Gingras (2006). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *arXiv :cs/0606079*.

Hamermesh, D. S. and G. A. Pfann (2009). Markets for Reputation : Evidence on Quality and Quantity in Academe. Technical Report 15527, National Bureau of Economic Research, Inc. NBER Working Papers.

Handke, C. (2010). The Economics of Copyright and Digitalisation – A Report on the Literature and the Need for Further Research. SSRN Scholarly Paper ID 2707153, Social Science Research Network, Rochester, NY.

Hardin, G. (1968). The Tragedy of the Commons. *Science 162*(3859), 1243–1248. American Association for the Advancement of Science.

Haschak, P. (2007). The 'Platinum Route' to Open Access : A Case Study of *E-JASL : The Electronic Journal of Academic and Special Librarianship. E-JASL 1999-2009 (volumes 1-10)*.

Heckman, J. J. and S. Moktan (2020). Publishing and Promotion in Economics : The Tyranny of the Top Five. *Journal of Economic Literature 58*(2), 419–470.

Heller, M. A. and R. S. Eisenberg (1998). Can Patents Deter Innovation ? The Anticommons in Biomedical Research. *Science 280*(5364), 698–701. American Association for the Advancement of Science Section : Review.

Hess, C. (2008). Mapping the New Commons. SSRN Scholarly Paper ID 1356835, Social Science Research Network, Rochester, NY.

Hess, C. and E. Ostrom (Eds.) (2006). *Understanding Knowledge as a Commons : From Theory to Practice*. Cambridge, MA, USA : MIT Press.

Hess, C. and E. Ostrom (2007). *Understanding Knowledge as a Commons : From Theory to Practice*. The MIT Press.

Hilty, R., S. Krujatz, B. Bajon, A. Frueh, A. Kur, J. Drexl, C. Geiger, and N. Klass (2008). European Commission - Green Paper : Copyright in the Knowledge Economy - Comments by the Max Planck Institute for Intellectual Property, Competition and Tax Law. SSRN Scholarly Paper ID 1317730, Social Science Research Network, Rochester, NY.

Himmelstein, D. S., A. R. Romero, J. G. Levernier, T. A. Munro, S. R. McLaughlin, B. Greshake Tzovaras, and C. S. Greene (2018). Sci-Hub provides access to nearly all scholarly literature. *eLife 7*, e32822. eLife Sciences Publications, Ltd.

Houghton, J., B. Rasmussen, P. Sheehan, C. Oppenheim, A. Morris, C. Creaser, H. Greenwood, M. A. C. Summers, and A. Gourlay (2009). Economic implications of alternative scholarly publishing models : exploring the costs and benefits. JISC EI-ASPM Project. A report to the Joint Information Systems Committee (JISC). *Joint Information Systems Committee (JISC)*.

Howard, J. (2013). Posting your latest article ? You might have to take it down.

Hurt, R. M. and R. M. Schuchman (1966). The Economic Rationale of Copyright. *The American Economic Review 56*(1/2), 421–432. American Economic Association.

Jeon, D.-S. and D. Menicucci (2006). Bundling Electronic Journals and Competition among Publishers. *Journal of the European Economic Association 4*(5), 1038–1083.

Johnson, R., M. Fosci, A. Chiarelli, S. Pinfield, and M. Jubb (2017). Towards a Competitive and Sustainable OA Market in Europe - A Study of the Open Access Market and Policy Environment. Technical report, Research Consulting, OpenAire.

Jubb, M., S. Goldstein, M. Amin, A. Plume, M. el Aisati, S. Oeben, S. Pinfield, P. Bath, J. Salter, R. Johnson, and M. Fosci (2015). Monitoring the transition to open access : A report for Universities UK. *Research Information Network (RIN)*.

Katz, Z. (2005). Pitfalls of Open Licensing : An Analysis of Creative Commons Licensing. *IDEA : The Intellectual Property Law Review 46*(3), 391–414.

Kling, R. and G. McKim (2000). Not Just a Matter of Time : Field Differences and the Shaping of Electronic Media in Supporting Scientific Communication. *Journal of the American Society for Information Science 51*(14), 1306–1320.

Kurtz, M. J., G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, E. Henneken, and S. S. Murray (2005). The effect of use and access on citations. *Information Processing & Management 41*(6), 1395–1402.

Kyrillidou, M. and M. Young (2008). ARL Statistics 2005-2006.

Laakso, M. and B.-C. Björk (2012). Anatomy of open access publishing : a study of longitudinal development and internal structure. *BMC Medicine 10*(1), 124.

Laakso, M., P. Welling, H. Bukvova, L. Nyman, B.-C. Björk, and T. Hedlund (2011). The Development of Open Access Journal Publishing from 1993 to 2009. *PLOS ONE 6*(6), e20961. Public Library of Science.

Landes, W. and R. Posner (2003). Indefinitely Renewable Copyright. *University of Chicago Law Review 70*(2).

Landes, W. M. and R. A. Posner (1989). An Economic Analysis of Copyright Law. *The Journal of Legal Studies 18*(2), 325–363. University of Chicago Press, University of Chicago Law School.

Larivière, V., S. Haustein, and P. Mongeon (2015). The Oligopoly of Academic Publishers in the Digital Era. *PLOS ONE 10*(6), e0127502. Public Library of Science.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature 411*(6837), 521–521. Nature Publishing Group.

Lessig, L. (2002a). *The Future of Ideas : The Fate of the Commons in a Connected World.* Knopf Doubleday Publishing Group.

Lessig, L. (2002b). Innovating Copyright The Ninth Annual Herbert Tenzer Distinguished Lecture in Intellectual Property. *Cardozo Arts & Entertainment Law Journal 20*(3), 611–624.

Liebowitz, S. J. (1987). Some Puzzling Behavior by Owners of InTELlectual Products : An Analysis. *Contemporary Economic Policy 5*(3), 44–53.

Liebowitz, S. J. and S. E. Margolis (1982). Journals as Shared Goods : Comment. *The American Economic Review 72*(3), 597–602. American Economic Association.

Liebowitz, S. J. and R. Watt (2006). How to Best Ensure Remuneration for Creators in the Market for Music ? Copyright and Its Alternatives. *Journal of Economic Surveys 20*(4), 513–545.

Litman, J. (2006). The Economics of Open Access Law Publishing Symposium : Open Access Publishing and the Future of Legal Scholarship. *Lewis & Clark Law Review 10*(4), 779–796.

Lougee, W. P. (2007). Scholarly Communication and Libraries Unbound :. In C. Hess and E. Ostrom (Eds.), *Understanding Knowledge as a Commons*, From Theory to Practice, pp. 311–332. The MIT Press.

Machin-Mastromatteo, J. D., A. Uribe-Tirado, and M. E. Romero-Ortiz (2016). Piracy of scientific papers in Latin America : An analysis of Sci-Hub usage data. *Information Development 32*(5), 1806–1814. SAGE Publications Ltd.

Maurer, S. and S. Scotchmer (2004). A Primer for Nonlawyers on Intellectual Property. In *Innovation and incentives* (Scotchmer, S. ed.). MIT Press.

May, C. (2020). Academic publishing and open access : Costs, benefits and options for publishing research. *Politics 40*(1), 120–135. SAGE Publications Ltd.

McAdams, R. (1997). The Origin, Development, and Regulation of Norms. *Michigan Law Review 96*(2), 338–433.

McCabe, M. (2004). Information goods and endogenous pricing strategies : the case of academic journals. *Economics Bulletin 12*(10), 1–11. AccessEcon.

McCabe, M. J. and C. M. Snyder (2005). Open Access and Academic Journal Quality. *American Economic Review 95*(2), 453–459.

McCABE, M. J. and C. M. Snyder (2014). Identifying the Effect of Open Access on Citations Using a Panel of Science Journals. *Economic Inquiry 52*(4), 1284–1300.

McCabe, M. J. and C. M. Snyder (2018). Open Access as a Crude Solution to a Hold-Up Problem in the Two-Sided Market for Academic Journals. *The Journal of Industrial Economics 66*(2), 301–349.

McGuigan, G. and R. Russell (2008). The Business of Academic Publishing : A Strategic Analysis of the Academic Journal Publishing Industry and its Impact on the Future of Scholarly Publishing. *E-JASL 1999-2009 (volumes 1-10)*.

Menell, P. S. and S. Scotchmer (2007). Intellectual Property Law. In A. M. Polinsky and S. Shavell (Eds.), *Handbook of Law and Economics*, Volume 2, pp. 1473–1570. Elsevier.

Merrett, C. (2006). The expropriation of intellectual capital and the political economy of international academic publishing. *Critical Arts 20*(1), 96–111. Routledge.

Merton, R. K. (1957). Priorities in Scientific Discovery : A Chapter in the Sociology of Science. *American Sociological Review 22*(6), 635–659. American Sociological Association, Sage Publications, Inc.

Merton, R. K. (1973). *The Sociology of Science : Theoretical and Empirical Investigations*. University of Chicago Press.

Mueller-Langer, F. and M. Scheufen (2013). Academic publishing and open access. In *Chapters*, pp. 365–377. Edward Elgar Publishing.

Mueller-Langer, F. and R. Watt (2010). Copyright and Open Access for Academic Works. SSRN Scholarly Paper ID 1647586, Social Science Research Network, Rochester, NY.

Musgrave, R. A. (1959). *The theory of public finance : a study in public economy*. New York : McGraw-Hill.

Nelson, R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy 67*. University of Chicago Press.

Nevo, A., D. L. Rubinfeld, and M. McCabe (2005). Academic Journal Pricing and the Demand of Libraries. *American Economic Review 95*(2), 447–452.

Newton, I. and R. Hooke (1675). Isaac Newton letter to Robert Hooke.

Nordhaus, W. (1969). An Economic Theory of Technological Change. *American Economic Review 59*(2), 18–28. American Economic Association.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. Political Economy of Institutions and Decisions. Cambridge : Cambridge University Press.

OCDE (1996). The Knowledge-based Economy. Technical Report OCDE/GD(96), OCDE, Paris.

Odlyzko, A. M. (1995). Tragic loss or good riddance ? The impending demise of traditional scholarly journals. *International Journal of Human-Computer Studies 42*(1), 71–122.

O'Hare, M. (1985). Copyright : When Is Monopoly Efficient ? *Journal of Policy Analysis and Management 4*(3), 407–418. Wiley, Association for Public Policy Analysis and Management.

Origgi, G. (2010). Epistemic Vigilance and Epistemic Responsibility in the Liquid World of Scientific Publications. *Social Epistemology 24*(3), 149–159. Routledge.

Ostrom, E. (2008). The Challenge of Common-Pool Resources. *Environment : Science and Policy for Sustainable Development 50*(4), 8–21. Routledge.

Ostrom, E. (2009). The Institutional Analysis and Development Framework and the Commons Response. *Cornell Law Review 95*(4), 807–816.

Ostrom, E., J. Burger, C. B. Field, R. B. Norgaard, and D. Policansky (1999). Revisiting the commons : local lessons, global challenges. *Science (New York, N.Y.) 284*(5412), 278–282.

Ostrom, E., R. Gardner, J. Walker, J. M. Walker, and J. Walker (1994). *Rules, Games, and Common-pool Resources*. University of Michigan Press.

Ostrom, V. and E. Ostrom (1977). Public Goods and Public Choices. In *Alternatives for Delivering Public Services*. Routledge.

Ottaviani, J. (2016). The Post-Embargo Open Access Citation Advantage : It Exists (Probably), It's Modest (Usually), and the Rich Get Richer (of Course). *PLOS ONE 11*(8), e0159614. Public Library of Science.

Parker, C. A. (2007). Institutional Repositories and the Principle of Open Access : Changing the Way We Think about Legal Scholarship. *New Mexico Law Review 37*(2), 431–478.

Partha, D. and P. David (1994). Toward a new economics of science. *Research Policy 23*(5), 487–521. Elsevier.

Pinfield, S., J. Salter, and P. A. Bath (2016). The "total cost of publication" in a hybrid open-access environment : Institutional approaches to funding journal article-processing charges in combination with subscriptions. *Journal of the Association for Information Science and Technology 67*(7), 1751–1766.

Piwowar, H., J. Priem, V. Larivière, J. P. Alperin, L. Matthias, B. Norlander, A. Farley, J. West, and S. Haustein (2018). The state of OA : a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ 6*, e4375. PeerJ Inc.

Plant, A. (1934). The Economic Aspects of Copyright in Books. *Economica 1*(2), 167–195. Wiley.

Polanyi, M. (1962). The Republic of science. *Minerva 1*(1), 54–73.

Pollock, R. (2008). Optimal Copyright Over Time : Technological Change and the Stock of Works. SSRN Scholarly Paper ID 1144288, Social Science Research Network, Rochester, NY.

Posner, E. A. (2000). Law and Social Norms : The Case of Tax Compliance. *Virginia Law Review 86*(8), 1781–1819. Virginia Law Review.

Posner, R. A. (2005). Intellectual Property : The Law and Economics Approach. *Journal of Economic Perspectives 19*(2), 57–73.

Poynder, R. (2003). Reclaiming the Digital Commons. *Information Today 20*, 33–34.

Ramello, G. (2005). Intellectual Property and the Markets of Ideas. *Review of Network Economics 4*(2), 1–20. De Gruyter.

Ramello, G. B. (2010). Copyright and endogenous market structure : a glimpse from the journal-publishing market. Technical Report 146, Institute of Public Policy and Public Choice - POLIS. POLIS Working Papers.

Redhead, C. (2015). Growth of OA-only Journals Using a CC-BY License.

Reichman, J. and R. Okediji (2012). When Copyright Law and Science Collide : Empowering Digitally Integrated Research Methods on a Global Scale. *Minnesota Law Review 96*, 1362–1480.

Reichman, J. H., G. B. Dinwoodie, and P. Samuelson (2007). A Reverse Notice and Take-down Regime to Enable Public Interest Uses of Technically Protected Copyrighted Works. *Berkeley Technology Law Journal 22*(3), 981–1060. Temporary Publisher.

Rigamonti, C. P. (2007). The Conceptual Transformation of Moral Rights. *The American Journal of Comparative Law 55*(1), 67–122.

Rosin, T. (1998). The Street as Public Commons : A Cross-Cultural Comparative Framework for Studying Waste and Traffic in India. Crossing Boundaries, the Seventh Biennial Conference of the International Association for the Study of Common Property.

Rowley, J., F. Johnson, L. Sbaffi, W. Frass, and E. Devine (2017). Academics' behaviors and attitudes towards open access publishing in scholarly journals. *Journal of the Association for Information Science and Technology 68*(5), 1201–1211.

Sag, M. J. (2006). Beyond Abstraction : The Law and Economics of Copyright Scope and Doctrinal Efficiency. *Tulane Law Review 81*(1), 187–250.

Samuelson, P. A. (1954). The Pure Theory of Public Expenditure. *The Review of Economics and Statistics 36*(4), 387–389. The MIT Press.

Scheufen, M. (2015). *Copyright Versus Open Access : On the Organisation and International Political Economy of Access to Scientific Knowledge*. International Law and Economics. Springer International Publishing.

Schroter, S., L. Tite, and R. Smith (2005). Perceptions of open access publishing : interviews with journal authors. *BMJ (Clinical research ed.) 330*(7494), 756.

Schweik, C. M. (2007). Free/Open-Source Software as a Framework for Establishing Commons in Science. In C. Hess and E. Ostrom (Eds.), *Understanding Knowledge as a Commons*, From Theory to Practice, pp. 277–310. The MIT Press.

Shavell, S. (2009). Should Copyright of Academic Works be Abolished ? SSRN Scholarly Paper ID 1525667, Social Science Research Network, Rochester, NY.

Silva, F. and G. Ramello (2000). Sound recording market : the ambiguous case of copyright and piracy. *Industrial and Corporate Change 9*(3), 415–442.

Solomon, D. J. and B.-C. Björk (2012). A study of open access journals using article processing charges. *Journal of the American Society for Information Science and Technology 63*(8), 1485–1495.

Spezi, V., J. Fry, C. Creaser, S. Probets, and S. White (2013). Researchers' green open access practice : a cross-disciplinary analysis. *Journal of Documentation 69*(3), 334–359. Emerald Group Publishing Limited.

Stiglitz, J. E. (2007). Economic Foundations of Intellectual Property Rights Lecture. *Duke Law Journal 57*(6), 1693–1724.

Suber, P. Creating an Intellectual Commons through Open access. In *Understanding Knowledge as a Commons : From Theory to Practice* (The MIT Press ed.)., pp. 171–208. The MIT Press.

Suber, P. (2012). *Open Access*. MIT Press.

Swan, A. (2013). *Principes directeurs pour le développement et la promotion du libre accès*. Paris : Organisation des Nations Unies pour l'éducation, la Science et la culture.

Towse, R., C. Handke, and P. Stepan (2008). The Economics of Copyright Law : A Stocktake of the Literature. SSRN Scholarly Paper ID 1227762, Social Science Research Network, Rochester, NY.

Tran, D. (2016). Elsevier v. Sci-Hub : Piracy of Knowledge and the Jurisdictional Reach of U.S. Copyright Law. *Wake Forest Journal of Business and Intellectual Property Law 17*(2), 198–225.

Travis, J. (2016). In survey, most give thumbs-up to pirated papers.

Uhl, M. (2009). Survey on European Psychology Publication Issues. ZPID (Leibniz Institute for Psychology Information).

Van Noorden, R. (2013). Open access : The true cost of science publishing. *Nature 495*(7442), 426–429.

Varian, H. R. (2000). Buying, Sharing and Renting Information Goods. *The Journal of Industrial Economics 48*(4), 473–488.

Vinkler, P. (2004). Characterization of the impact of sets of scientific papers : The Garfield (impact) factor. *Journal of the American Society for Information Science and Technology 55*(5), 431–435.

Ware, M. and M. Mabe (2015). The STM Report : An overview of scientific and scholarly journal publishing. *Copyright, Fair Use, Scholarly Communication, etc.*.

Warlick, S. E. and K. Vaughan (2007). Factors influencing publication choice : why faculty choose open access. *Biomedical Digital Libraries 4*(1), 1.

Waters, D. J. (2007). Preserving the Knowledge Commons. In C. Hess and E. Ostrom (Eds.), *Understanding Knowledge as a Commons*, From Theory to Practice, pp. 145–168. The MIT Press.

Watt, R. (2000). *Copyright and Economic Theory*. Edward Elgar Publishing.

Watt, R. (2008). The Past and the Future of the Economics of Copyright. SSRN Scholarly Paper ID 1143022, Social Science Research Network, Rochester, NY.

Watt, R. (2014). Handbook on the Economics of Copyright. Books, Edward Elgar Publishing.

Weingart, P. (2005). Impact of bibliometrics upon the science system : Inadvertent consequences ? *Scientometrics 62*(1), 117–131.

Willinsky, J. (2009). The Stratified Economics of Open Access. *Economic Analysis and Policy 39*(1), 53–70.

Woolston, C. (2016a). Paper piracy sparks online debate. *Nature*.

Woolston, C. (2016b). Paper piracy sparks online debate. *Nature*.

Wu, T. (2006). Intellectual Property, Innovation, and Decentralized Decisions. *Va. L. Rev. 92*, 123.

Xia, J. (2008). A Comparison of Subject and Institutional Repositories in Self-archiving Practices. *The Journal of Academic Librarianship 34*(6), 489–495.

Xia, J. and K. Nakanishi (2012). Self-selection and the citation advantage of open access articles. *Online Information Review 36*(1), 40–51. Emerald Group Publishing Limited.

Annexe

REVUE DE LA LITTERATURE	
Axe de recherche	**Questions de recherche**
Economie de la connaissance	Relation entre production de connaissance et croissance économique
Economie du numérique	Etude des activités économiques fondées sur le numérique
Economie de la science	Etude des institutions dans le monde scientifique
Analyse économique du droit d'auteur	Analyse des coûts et bénéfices dérivants de l'octroi d'un pouvoir de monopole
Analyse économique de l'Open Access	Analyse de l'impact sur le marché de la publication scientifique de la mise en place de l'OA
Analyse économique des communs	Analyse de la connaissance scientifique en tant que bien collectif, dont la gestion est assurée par la communauté scientifique

TABLE 1.5 – Axes de recherche de la revue de la littérature portant sur l'*Open Access*

En réalité, ces branches de l'économie sont fortement liées les unes aux autres. À titre d'exemple, on ne pourrait pas appliquer l'économie du numérique au marché de la publication scientifique sans utiliser l'analyse économique de l'*Open Access*.

Chapitre 2

Rule vs. Communality Norm in the Scholarly Publishing Market : Institutional Overlap as a Barrier to Publishing in Open Access journals ?

Abstract [1]

The scientific publishing market is governed by a complex set of both norms- and law-based intellectual property systems. In the literature these two systems are strictly counterposed, with legal rules equated to Closed Access publishing, and the social norms of the scientific community corresponding to Open Access. This article provides new insights into scholars' decisions to publish in open vs. Closed Access, arguing that the co-evolution of these two intellectual property systems has shaped scholarly publishing over time. In particular, norms-based and law-based intellectual property systems influence scholars' behaviors, which in turn shape the content of those very norms and rules. Today, the predominance of copyright has caused knowledge enclosure–in the form of closed-access publication–to become accepted as the new norm. This shift in the social norm to align with copyright explains the limited success of legal framework reforms aimed at promoting wider use of Open Access publishing, and the reluctance of most researchers to choose Open Access tools.

Introduction

Open Access (OA) publishing denotes a new organizational model for managing scholarly journals that enables researchers to circulate peer-reviewed scientific articles with fewer restrictions than the currently dominant Closed Access (CA) model, which instead relies upon the usual price system, excluding those unable to bear the cost. The expected gains from OA are at least twofold. First, if a journal adopts the OA model, then its potential audience is enlarged. Second, OA promotes the wide dissemination of scientific knowledge at minimal cost, making scholarly communication very cheap. Despite these advantages, which on the whole make OA publishing a disruptive innovation, the empirical evidence shows that CA remains the dominant model of scholarly publishing : The number of scholarly papers published in OA journals is still low, and the majority of works continue to be published in copyrighted journals (Johnson et al., 2017). According to various studies, 80% of scientific articles are still published on CA journals (Green, 2018).

1. Ce chapitre a été coécrit avec les Professeurs S. Harnay et G.B. Ramello

One might legitimately wonder why researchers do not overwhelmingly adopt the OA model to publish their work, since it seems to be in both their individual and collective interest to do so. The question is all the more puzzling given the proliferation of private and public initiatives aimed at encouraging the scientific community to shift towards an OA-based publishing model. In recent years, more than 700 policies to promote OA, relying on both mandatory and voluntary approaches, have been developed by public organizations around the world [2]. Several groups of scholars have also been actively supporting the idea of *Open Science* by creating legal tools ancillary to OA-based publishing (*e.g.*, Creative Commons licenses). Yet all these activities have so far had only modest effects, and the dominant model in scholarly publishing remains CA. Such an outcome seems to be at odds with scholars' individual and collective motivations. According to the sociology of science literature, citations are the "pellets of peer recognition that aggregate into reputational wealth" (Merton 1988, p.620). Economics of science, building upon this idea, has shown that a scholar's individual reputation increases with the number of citations received by his or her work. Since OA increases the availability of a scholar's work to others, and hence the likelihood of being cited, one would expect researchers to publish in OA journals. A similar conclusion is reached if we consider the collective viewpoint of the scientific community : Since OA enhances the dissemination of new knowledge and internal debate, scholars ought to be in favor of it.

If the OA system is aligned with researchers' individual and collective interest, its low adoption in the scientific community poses a conundrum. We believe that the solution to this puzzle lies in understanding the institutional inertia affecting scholarly publishing, which has locked the scientific community into the CA model. More precisely, today's overwhelming dominance of CA publishing is the outcome of a dialectic process within the academic world, pitting norms-based intellectual property against its law-based counterpart (mainly copyright).

The aim of this article is to discuss the dynamic interplay between norms- and law-based intellectual property systems within scholarly publishing, and how it has affected researchers' publishing choices. To do this we build upon three strands of literature, namely

2. For an overview about the policies that promote the adoption of the self-archiving practice, see OpenDoar. Available at : `https://v2.sherpa.ac.uk/opendoar/`. Last access 01/01/2021.

economics of science, law and economics of intellectual property, and law and social norms. In particular, as compared to the previous literature that strictly counterposes law-based vs. norms-based systems, we show how in this case the two are not mutually exclusive but instead jointly contribute to shape the publishing choices of scholars. In this regard, we argue that norms and legal rules evolve over time in response to both external factors and mutual influence. This enables us to analyze scholars' choices to publish in OA vs. CA journals as the outcome of an ongoing evolutionary interplay between social norms and legal rules in scholarly communication. Finally, we contend that changes in the academic landscape in recent decades can be reasonably ascribed to the increased salience of copyright at the expense of social norms. More specifically, we suggest that, within scholarly communication, this salience of copyright may have fueled the idea of knowledge enclosure as a new norm, with publishing in CA journals becoming the "normal" – expected – behavior associated with this new norm. Increased salience of copyright rules thus helped to establish a new social norm for publication–one more consistent with the legally prevailing view of privatized knowledge, at the expense of the previous norm of communality. This shift in social norm to align with the formal rules of copyright in its turn triggered wider adherence among scholars to the closed-access logic, which ultimately explains why the scientific community at large does not fully take advantage of today's new legal tools encouraging researchers to publish in OA journals.

The outline of the article is as follows : Section 1 describes the complex interplay and overlapping of IP legal rules and social norms in the scholarly publishing market, and discusses the current shift towards Open Access in this context. Using this analytical framework, Section 2 reassesses the historical shifts in the governance of scientific intellectual property, showing how various institutional equilibria observed over the centuries can be interpreted as the outcome of a changing mix and interplay of law- and norms-based intellectual property systems. Section 3 argues that the formal rules of IP and social norms mutually influence each other, and that this reciprocal influence accounts for the current reluctance of most scholars to publish in OA journals. Section 4 provides further comments on the complementarity between the two systems while Section 5 concludes.

2.1 Intellectual property and Open Access in the scholarly publishing market : a complex interplay of legal rules and social norms

2.1.1 A tale of two coexisting systems

For a long time, norms-based intellectual property (IP) governed scientists' behavior, while law-based intellectual property mostly applied to printers and booksellers. Nowadays, the two systems tend increasingly to overlap, with both formal intellectual property laws and the informal norms that exist within the scholarly community being jointly used by academic scientists.

Law-based IP, for its part, has been amended to foster knowledge production and its disclosure. According to the mainstream rationale, IP rights grant the holder a temporary exclusive right that promotes the production and the dissemination of the subject matters, striking the optimal balance between the private interest of innovators and social welfare (Posner, 2005). Along this line, intellectual property rights (IPRs)–*i.e.*, patent, copyright, trademarks and trade secrets—are the formal tools available to innovators to reap economic benefits from their creative effort, in the form of monetary and non-monetary rewards. They therefore ensure that the production of knowledge will be profitable for rational agents, as the latter will benefit from the legal monopoly over successful knowledge items and in turn will cover fixed production (Caves, 2000) [3].

In the scholarly publishing market, the relevant legal rule is copyright, and the usual pattern is that authors transfer rights over their intellectual output to publishers for free, in exchange for the publication of their scientific work. Publishers accordingly become monopolist and disseminate the scientific publication–until recently in the form of physical books and journals. Copyright is thus the pivotal element for financing the dissemination of scientific knowledge in a context where individuals are unwilling to directly perform this task. Publishers also become entitled to sue unauthorized use by third parties before courts ; the legal sanctions that punish infringement of copyright thus lead to monetary

3. This viewpoint has been increasingly challenged as IPRs are not the only weapons to permit appropriability and they sometime clash with the values of sustainability (Castaldi, 2021).

compensation for publishers. However, scientific authors also derive gains from this enforcement of copyright : it enables them to uphold a paternity right on the expression of their ideas and retain the attribution right, or the right to be cited when third parties reproduce the original expression of the idea in derivative works. Hence, even though authors no longer have a direct economic interest in copyright, they still benefit from the protection of reputational gain afforded by its enforcement.

However, within the bounds of the scientific community, the production and circulation of knowledge was traditionally governed by a set of norms–frequently referred to as the "ethos of science"– which operate differently on the domain of IP (Merton, 1957, 1973 ; Ziman, 2000 ; Kalleberg, 2007) [4]. These norms, which we term 'norms-based IP', emerged spontaneously within the scientific community and for a long time provided the institutional framework that governed researchers' behavior. Like law-based IP, albeit in a different manner, norms-based IP plays a key role in ensuring that individuals will produce scientific knowledge and disseminate their contributions to scientific advances. In particular, they include credit norms for scientific advances which enable individuals to obtain a reward, in the form of reputational gain, when they disclose and circulate their findings. The protection of credit and attribution then serves to shield individuals from the risk of plagiarism, defined as "the appropriation of another person's ideas, processes, results, or words without giving appropriate credit" (Office of Science and Technology Policy, 2000, p. 76263). It can be either plain copying of another's text or plundering of another's idea [5]. In similar vein to law-based IP, norms-based IP not only tackles a problem of appropriability of benefits, but also makes it profitable for individuals to disclose and circulate their findings. Hence it ensures not just that the new idea will be cited, but also that its infringement will be sanctioned as a violation of the collegial trust among scientists (Green, 2002 ; Berk, 1991 ;

4. The ethos of science is a set of norms, composed of the CUDOS norms (communalism, universalism, disinterestedness and organized skepticism), first described by Merton (1942), and of technical norms, for which Merton never gave a detailed definition. The universalism criterion implies that knowledge should be evaluated and accepted according to impersonal criteria. Communalism refers to the boundaries of property in science–limited to the recognition of one's own contribution to the advancement of knowledge and the esteem of peers. Disinterestedness in science stands for the fact that its laborers direct themselves toward the extension of certified knowledge. Finally, organized skepticism refers to a methodological and institutional mandate.

5. The term plagiarism comes from the Latin word plagiarius, which means a person who abducts the child or slave belonging to another. It thus straightforwardly connotes the intimate sphere of the author, to which reputation is connected.

Masic, 2014). Sanctions take the form of extra-legal (*e.g.*, non-monetary, reputational) costs incurred by offenders. This in turn serves the public interest as it increases the collective stock of knowledge.

It is easy to see that the norms-based IP system functionally plays in scholarly publishing a role similar to copyright, setting up behavioral incentives and informal punishments. This kind of spontaneous ordering can also be found in many professional communities with the specific aim of regulating the production and circulation of specialized knowledge. Prominent examples include fan fiction communities (Tushnet, 2007), the French and Italian haute cuisine community (Fauchart & von Hippel, 2008 ; Di Stefano et al., 2014), magicians (Loshin, 2007), stand-up comedians (Oliar and Sprigman, 2008), roller derbies (Fagundes, 2011), tattoos (Perzanowski, 2013, 2017), pornography (Darling, 2013), and street artists (Iljadica, 2017) [6]. As is the case with the scientific community, all these norms-based IP systems protect the output of creative processes within their professional fields. They also enable professionals to establish and enforce some types of appropriation of their economic advantage (*e.g.*, associated with monetary and non-monetary rewards, including reputational gains). However, quite interestingly, all these professional communities have chosen to almost exclusively rely only on norms-based systems to govern the circulation of professional knowledge among members, and have never amended formal legal rules. It is only within the scientific field that law-based IP and norms-based IP coexist, although their overlap came about somewhat accidentally, as a by-product of the current regulation of the publishing industry.

2.1.2 The complex interplay of norms-based and law-based IP in scholarly publishing

Law-based and norms-based IP systems are complementary when they reinforce each other. This is typically the case today when copyright strengthens the implementation and enforcement of the scientific community's norms-based IP system. In other words, the threat of legal sanctions for copyright infringement can actually increase compliance with the norms-based IP of the scientific community. For example, when plagiarism takes the

6. Further examples of norms-based IP systems serving some of the same functions as law-based IP systems are documented by Darling and Perzanowki (2017) and Raustalia and Sprigman (2019).

form of plain copying, the claim of copyright infringement, intended as the appropriation of an expression of an idea, also protects the paternity and hence the credit associated to this idea, thereby echoing and reinforcing the IP-norm of credit. It is easy to see that the use of copyright rules by publishers is thus a tool oriented not only directly towards their own profit, but also indirectly towards the protection of scientists' IP. This alignment of the financial interest of publishers with the reputational interest of scientists accounts for the strong adherence of the scientific community to copyright. In this respect, copyright thus provides a comprehensive framework in which agents can create a private ordering and interact « in the shadow of the law » (Mnookin & Kornhauser, 1978).

However, the law-based and norms-based IP systems may not always reinforce each other. For example, one might have expected the scientific community's long-standing norm of commonality and tradition of free circulation of knowledge to pave the way for wide acceptance among scholars of recent legislation promoting OA publishing. Both systems indeed have the same objectives. They aim to promote free, unrestricted access to knowledge for all, and also seemingly serve researchers' interests, insofar as wider access to scientific work increases the number of potential citations and so benefits individual reputations. Yet the recent legislative IP reforms to promote OA, despite being well aligned with the demands and norms of research communities [7], have had little effect so far, and publishing in CA journals remains today the most frequent choice.

The law-based and norms-based IP systems can also play in opposite directions. Simultaneous use then dilutes their effectiveness, compared to a situation where one is used in isolation. An illustration is provided by piracy. Although using pirated copies is at odds with copyright rules, it does not necessarily conflict with the scientific community's norms of free knowledge sharing, provided the authors of the illegal copy are cited. Piracy then creates a decoupling between the monetary interest of publishers, defended by copyright, and the reputational utility of authors, supported by the social norms of the scientific community. In the same fashion, the use of an author's idea in a manner consistent with copyright rules could nevertheless conflict with the norms of the community if it takes the form of a biased, partial, or disrespectful use of the original idea.

7. *E.g.* Budapest Open Access Initiative in 2002 as well as the application of the Creative Commons licenses in the academic field.

Finally, legal rules of IP and the norms of the scientific community can also act as substitutes for each other. Generally, informal norms are considered to substitute for legal rules when the latter are either inexistent, ineffective or have a limited salience, *e.g.* due to historical, practical, cultural or environmental reasons (*e.g.*, Macaulay 1963 ; Ellickson 1991 ; Huang and Wu 1994). Empirical evidence has also shown that, in some settings where both formal rules and social norms coexist, legal rules can sometimes become unimportant relative to norms in shaping social interactions (Ellickson, 1991) [8]. In the domain of circulation of knowledge, it has been argued that substitutability will occur when there are "negative spaces" of IP–defined as creative and innovative issues not addressed by IP rules, and which must therefore be managed by social norms (Raustiala & Sprigman, 2006). An example of this substitutability occurs when law-based IP does not perform its expected role efficiently. For instance, when plagiarism takes the form of plundering–that is, appropriating another's' ideas while changing the text–it is not covered by copyright rules because these only protect the "expression of ideas fixed on a medium", and not the ideas themselves [9]. Then, such plagiarism is better tackled by the social norms of the scientific community because they are concerned with originality, which is regarded as one of the most valuable criteria for assessing scientific research.

Conversely, in other settings formal rules can substitute for social norms. From a general standpoint, it has been shown that self-organized norms-based systems are effective within small, stable, and homogeneous communities where the participants interact frequently and are strongly connected (See *e.g.* Hardin, 1982 ; Coleman, 1990 ; Gintis, 2000). These conditions ensure that compliance with norms will be observed and achieved at low cost, resulting in the socially desirable outcome. When such conditions are not met, compliance with social norms may be insufficient, thus making it necessary to apply formal rules to monitor undesirable behaviors. This view provides an explanation for the partial

8. Ellickson (1991) for example, found that the cattle ranchers of Shasta County, California, developed and enforced a set of social norms regarding responsibility for straying cattle and most often settled disputes and allocated the costs of maintaining the fences that separated their properties without paying any attention to formal property rules and institutions applying to their disputes. On the contrary, they most often followed boundary-fence maintenance norms differing from the law in a number of respects.

9. *E.g.* the US Code at title 17 Section 101 defining copyrighted items talks about an "embodiment in a copy or phonorecord, by or under the authority of the author, that is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration. A work consisting of sounds, images, or both, that are being transmitted, is 'fixed' [. . .] if a fixation of the work is being made simultaneously with its transmission."

replacement of "ethos of science" norms by formal IP regulations in the second half of the 20th century. As the scientific community increased in size, becoming more globalized and hyper-specialized, its social connectedness was gradually undermined. Norms thus became less effective at monitoring deviant behavior and IP infringements, thereby justifying their replacement with new IP regimes [10].

2.1.3 Open Access at the crossroads between norms-based and law-based IP

Until recently, copyright was the pivotal element of the scientific publishing industry, while norms-based IP mostly dealt with scientists' behaviors within the scientific community and its job market. Today, an easy return to the ethos of science could be expected to occur, since OA technology readily permits unrestricted access to scientific knowledge. Yet, despite the disruptive potential of the OA publishing model, it has not yet overtaken the old system. This despite the numerous actions that have sought to promote OA and introduce exceptions to copyright laws in recent decades. For example, the European Horizon 2020 project requires researchers to publish their work under the OA model. Obviously, such legislative changes endorse the scientific norms of commonality and free unrestricted access to knowledge, and supplement private bottom-up initiatives (*e.g.* Creative Commons licenses) challenging copyright. However, since they do so under the shadow of copyright, they do not simply challenge law-based IP, they also, in a sense, further confirm it. In other words, recent legislative changes are fostering the emergence of hybridized property regimes which, on the one hand, seem to push toward the ethos of science, while on the other hand fully recognizing the legitimacy of copyright.

These changes should encourage researchers to publish their work under the OA regime without any doubt or stigma of infringing law-based IP in some way. Yet, the current picture is much more blurred, with very modest use and unequal reception of OA among different scientific disciplines (Cullen & Chawner, 2011 ; Dellmeier-Tisessen et al., 2011). The fact that OA publishing is now encouraged by many public authorities and research institutions

10. See for example Kleinman, D.L., (1998). 'Pervasive Influence : Intellectual Property, Industrial History, and University Science'. Science and Public Policy 25 (2) : 95–102 ; Moore, K., Kleinman, D.L., Hess, D., and Frickel, S. (2011). 'Science and Neoliberal Globalization : A Political Sociological Approach'. Theory and Society 40 (5) : 505–32.

makes its low uptake all the more puzzling, given that OA also aligns with the individual objectives of researchers, offering wider access to their work that increases their number of citations and thus their individual reputation. The current picture reflects the inertia of the CA model. The large-scale migration toward OA publishing everywhere, if it ever occurs, is still to come.

2.2 The evolving institutional dynamics in scholarly publishing : an historical overview

The current governance is the outcome of a lengthy evolutionary process, in which the scientific community's norms and law-based IP can hardly be considered in isolation. A more appropriate perspective is to interpret the institutional equilibria observed at various periods in time as specific resultants of an interplay between social norms and legal rules, which emerged in response to the conditions for knowledge production and peculiarities and needs of the scientific and social environment.

2.2.1 The long-lasting prevalence of social norms in the governance of scientific exchanges

Until the late 19th century, the circulation of scientific knowledge was largely regulated by the communality norms governing the scientific community. Knowledge advancement and scientific research were activities reserved for independently wealthy men from the middle and upper classes, with private resources for financing their studies (Fyfe et al., 2017). Thus, appropriability in science was regulated by gentlemanly norms, essentially linked to an author's honor and reputation (Chartier, 2013). Communication of results was essentially based on books, and private correspondence was the most common means of discussing and disseminating researches. Compensation was then based on reputation and did not rely upon market mechanisms. The validation of scientific results as we know it today did not exist. Documents that we call "scientific" were considered truthful if the name of the author was indicated. The social position lent authority to the newly produced knowledge, because what "scientists" said was without consideration of remunera-

tion [11](Foucault, 1969).

A more structured system for scientific communication started only with the birth of the first scientific journals, promoted by learned societies, *e.g.* the Philosophical Transactions of the Royal Society of London (1665) and the Journal des Sçavants of the Académie Royale of Paris [12](1665) . Learned societies, in the form of associations and non-profit organizations, were mainly dedicated to the promotion and widespread circulation of science without financial barriers. Their objectives as publishers were thus aligned with those of scientists, as the latter were interested in publishing their scientific results because publication was one way for enriching their social prestige (Morrel & Thackray, 1981). The emergence of learned societies and the dissemination of knowledge through journals thus deeply changed the validation process of scientific results that were "[. . .] accepted on their own merits and positioned within an anonymous and coherent conceptual system of established truths and methods of verification" (Foucault, 1969). Accordingly, the former system of authorship as a means for certifying the truthfulness of scientific results was replaced with a validation by journals' publishers, who acted as defenders of "good taste" (Nash, 1963).

IP laws played no role for a long time. The ancestors of copyright and other IPRs had entered the scene in the 14th century, with the concession of privilegii, a form of monopolies granted by the central authority (Ulmer, 1978 ; David, 1993). Copyright officially appeared with the Statute of Anne (1710), which aimed at encouraging "learned men to compose and write useful books". For instance, the popularity of the Journal des Sçavants owed much to the flourishing number of pirated copies that were widely distributed in France, openly infringing the printing privileges. In England, the publication of the Philosophical Transactions, which was crucial to disseminating the work of many scholars including Newton, Boyle, Leibniz, Cassini and Halley, flourished thanks to a royal exemption from the Stationers' Company control (Johns, 2000 ; Nicita & Ramello, 2007). Hence, copyright law played hardly any role in academic publishing until the 20th century, and most scientific activities were exclusively governed by scientific social norms. Furthermore, it focused originally on books and "[p]eriodicals were not explicitly included in copyright legislation

11. The defense of the honor and reputation of gentlemanly norms was similar to the moral integrity protected in ancient Rome by the action iniuriam aestimatoria (Zimmermann, 1998).

12. The Journal des Sçavants was first published by Denis de Sallo in 1665, who joined the Académie Royale of Paris in 1667 (Vittu, 2002).

in the United Kingdom" (Fyfe et al., 2018, p. 598). The publishing costs of scientific journals were directly supported by authors and members of learned societies, and sold for free or at reasonable price.

The 19th century was characterized by dramatic changes in the field of institutions governing scientific exchanges, as a consequence of several industrial revolutions, the development of positive philosophy, and a growing belief in scientific, economic and political progress. This generated major structural and cultural changes in the organization of scientific research. In particular, the expansion of universities and the professionalization of scientists prompted the development of an academic job market aimed at improving university recruitment processes and the selection of high-quality scholars. This in turn exposed the need for an assessment of the academic community. From a system based on social reputation, the evaluation of intellectual merit shifted to a system based on a list of published books and articles for establishing the scientific reputation (Babbage, 1830). The publishing imperative thus became an obligation to disseminate scientific results, fundamental for the production of new certified knowledge (Bernal, 1939).

If publishers initially had to convince authors to publish their works, by the late 19th century journals became the standard way to demonstrate a successful research career (Csiszar, 2017). This new environment led to the gradual replacement of the gentlemanly norms that previously governed the scientific community with a more sophisticated system, in which authorship became the focal point for academic prestige (Biagioli & Galison, 2013 ; Moxham & Fyfe, 2018). Within this new institutional framework, the general principles and guidelines for the production of valuable new knowledge were set by IP norms, issued and enforced by the scientific community itself, and strong enough to guarantee that authors would derive sufficient reputational gains from their contribution to scientific advancement. They were thus able to impose (non-legal) sanctions on non-compliant behaviors. At the same time, despite not being written down or explicitly discussed, these norms were effective enough to perform their expected role of regulating the circulation of ideas within the scientific community. As a result, the majority of scholars complied with the established norms, which fueled a self-reinforcing process of adherence. Hence, there was little need to implement a costly statutory IP system and, in practice, copyright

was unnecessary to the fair regulation of publishing within the scientific community. The norms-based and the law-based IP systems factually co-existed, but the actual governance of relations between scientists was mostly ensured by the informal system of norms.

Of utmost importance in this institutional process is the fact that the scientific journals published by learned societies–which proliferated in the 19th century, notably due to the development of specialized fields of scientific expertise – were able to benefit from extant norms within the scientific community to manage an increasing volume of publications. The pre-existing norms were thus largely incorporated into the new authorship system, of which they became an essential part. For instance, because the previous communality norms remained perfectly relevant to the needs of actors in the new environment, they were accepted by both the publishers and scientists and could therefore survive [13]. In this view, institutional changes in the field of scientific exchanges are more accurately described as a process of stratification – or sedimentation – rather than substitution of one IP system for another.

2.2.2 The takeover of copyright on norms-based IP

At the end of the 19th century, a set of formal IP rules was established in various countries on a more systematic basis (see *e.g.* the French legislation of 1866 and the Berne Convention of 1886). Because these legal rules were explicitly oriented towards the interests of "public science", they were congruent with the private interests of journals and researchers, as well as with the wider goals of the scientific community, and did not come into conflict with the pre-existing communality norms. On the contrary, IP norms and legal rules were used to distinctly regulate particular fields, with copyright law seen as inapplicable to journals and devoted only to the protection of books, while the social norms of the scientific community were used to govern scientific exchanges not covered by the formal rules. They were also applied for distinct purposes, with copyright protecting only

13. Learned societies were still the main producers of scientific journals, which were not substantial enough from an economic perspective to justify the attention of commercial publishers. The market was regarded, with justification, as unprofitable owing to the high uncertainty. For example, the Royal Society's Philosophical Transactions issues were available as bound volumes only once all the issues were printed, and this took several months. Hence, accessing a full volume was a hard task, to the point that in 1875 the Royal Society went to a London based bookseller, Trübner, in order to sell the Transactions in the book trade market. The new retailer began to unbind the issues and to sell them. Yet, the reception was negligible, and in 1883 Trübner declared that no more than ten copies had been sold (Howsam, 1998).

the expression of ideas. while social norms protected ideas per se.

Within this new framework, copyright started to gradually gain ground throughout the 20th century. In the aftermath of WW2, the structure of the scientific environment, its conditions, and the organization of research were dramatically transformed by the new political and economic goals ascribed to research in the Cold War context, and the ensuing increase in funds for R&D (Ziman, 2000 ; Kellog, 2006 ; Kalleberg, 2007 ; Berman, 2012). These changes, called into question the former governance of scientific exchanges. The "ethos of science" gradually lost ground and was partly replaced by the new institutional arrangements of "post-academic science", with reference to how science was increasingly driven by business agendas and the search for profit, rather than by the Mertonian norms (Ziman, 1996). In the market for scientific publication, this shift was accompanied by the growth of the publishing industry, mainly due to the growing size of the scientific community coupled with a marked increase in the number of manuscripts. The latter was due to not just to the enlargement of the research community, but also to the increasing pressure to publish placed on researchers. The "publish or perish" era, officially starting in 1928 [14], operated a major transformation in scientific activity by making scholarly publications the "output" of researchers, and the pivotal element on which to build their careers. If until then scholarly publications were a means for circulating new ideas, they soon became the basis on which to assess academic performance. The age of bibliometrics was beginning, with researchers mainly seen as producers of papers, the units on which their activity was evaluated. A system of metrics – starting from the impact factor - was also invented in order to create a complex of values through which to measure individual and institutional productivity (Origgi, 2016 ; Osterloh & Frey, 2020).

The industrialization of science was accompanied by increased reliance on the law-based IP system. The amendment of the Bayh-Dole Act (1980) in the USA is traditionally seen as a milestone in the move towards a proprietary system in the field of academic science : it authorized federal contractors, including universities and non-profit organizations, to retain ownership of inventions made with federal government-funded research and to seek revenues from their scientists' IP. Outside the United States, similar legal evolutions were

14. See Case, C. M. (1928). 'Scholarhip in Sociology'. Sociology and Social Research, 323–40.

observed in many other countries, with the enactment of various legislations endorsing a proprietary view of science. Indeed, a law-based IP system, providing agents with a well-defined legal framework and rights, was seen as a prerequisite for developing an efficient market for scientific results. To foster an institutional environment tailored to this new market logic, universities and research institutes established new governance entities to support the academic entrepreneurship and the marketization process through the development of in-house expertise in IP management, patenting, and licensing (*e.g.* university-industry technology transfer "TT" offices) (Berman, 2012 ; Etzkowitz, 2004, 2013).

From the institutional point of view, a shift was underway, from a paradigm based on free communication of new scientific ideas and discoveries to another which exploited market exchange by defining and strictly enforcing IPRs over scientific knowledge. The increasingly dominant market logic, with the privatization and commercialization of science, was de facto at odds with the idea of free, unrestricted access to knowledge for all promoted by the traditional norms of the scientific community, and the latter gradually lost ground. In the publishing market, these changes also had an impact on scholarly journals. Within the new context of privatized science, copyright did not just serve the interests of publishers in the usual way, permitting them to reap profits from printing and distributing journals ; it also provided researchers and research entities with legal tools tailored to the new private property environment and to the "spirit of academic capitalism", by enabling them to stockpile a number of published papers. At the same time, the former role of scientific publication was not entirely denied, because the circulation of articles continued to produce the usual benefits for scholars, that is to say citations and credit. Like the "old" norms of free knowledge sharing in the scientific community, copyright not only provided a mutual citation system, but also specific benefits in the form of the services of publishing houses to the scientific community (control of quality, knowledge diffusion).

Ultimately, in no more than three decades, the interaction between law-based and norms-based IP systems worked in favor of the privatization and commercialization of science and against the ethos of science. Today, according to financial analysts, one of the most lucrative industries in the world is scholarly publishing. In 2011, the largest scholarly publishing company, Reed-Elsevier, had operating profit margins of 36%, much

higher than those of Apple, Google or Amazon for the same year (Buranyi, 2017). With few exceptions, copyright rules have become the main institutional arrangement to govern the scientific publication market, in accordance with the dominant logic of profit-seeking and commercialization of scientific results.

Of course, in the scientific field, copyright operates de facto – albeit not de jure – differently than it does in other creative industries. Most of the original copyright holders, *i.e.* scientific authors, are not directly concerned about the monetary rewards associated with copyright, but they supply "free labor" in exchange for the reputational rewards that they will derive from being published in well-ranked journals. Therefore, they agree to transfer their rights to publishers for free, as an unavoidable prerequisite to the publication of their work. Within this business model, publishers take advantage of the exclusivity of original discoveries to produce and sell "costly journals" to the worldwide scientific community (Bergstrom, 2001). Strategies of bundling such as the so-called Big Deal allow them to extract users' surpluses and to increase their profits, with the further consequence of skewing the market structure toward a large prevalence of commercial players (Edlin & Rubinfeld, 2004 ; Migheli & Ramello, 2018).

The growing trend towards the privatization and commercialization of science does not, however, mean that the scientific community's norms have fully disappeared. Although the norm of communalism was strongly challenged by the new logic of Closed Access and knowledge enclosures, some IP norms nevertheless persist today. For instance, the peculiar use of copyright that is made by scientists today remains largely consistent with the previous norm of disinterestedness : since they transfer their copyright to the publishers free of charge and thus forego the related monetary gains, the authors remain faithful to the former norm of disinterestedness, guaranteeing that that their only personal gains will be reputational. Yet a noteworthy nuance is that the former model of free disclosure to the community for gaining credit and reputation has turned into one of licensing copyright to publishers to gain visibility, and thereby credit and reputation.

2.3 The salience of copyright as a barrier for the migration to OA publishing

For a long time, copyright played almost no role in circulating scientific knowledge, until the previously described shift occurred, from a paradigm of free unrestricted access to knowledge to a new conception of knowledge enclosure better suited to the market environment. These changes caused copyright to gain salience at the expense of norms-based IP–according to the usual dynamic observed in many contributions that, when the specific context and social group in which agents have to act and cooperate focuses on a particular norm, this creates a convergence toward it (Bicchieri, 2000 ; Andrighetto et al., 2010). In particular, when a source of information becomes more prominent, individuals' responses will change in a way that is consistent with the new framework, and people tend to internalize the new behavior (Cialdini et al., 1990).

2.3.1 The rise of copyright

With the growing prevalence of copyright, scientists started to adapt their publishing behavior to this newly salient institution. The fact that publishing in CA copyright-based journals was the expected behavior then tended to make the set of alternative options more limited and less profitable, further reinforcing the salience of copyright. At the same time, the increased role of markets in scholarly publishing worked to erode the salience of norms-based IP, as agents realized that the violation of the old communalism norm, through the practice of knowledge enclosure, would not incur any penalty. Interestingly, copyright's salience did not just alter scholars' behavior but also the very nature of the publishing imperative as the social norm for knowledge disclosure. As we underlined in the previous section, scientific journals were conceived by the early scientific community as means for enhancing the circulation of scientific knowledge, without any intention to make profits. Far from being a lucrative business, the main aim was to facilitate and enlarge access to new scientific results, serving the communalism norm of science. Publishing results accordingly became an imperative, as a moral obligation to disclose scientific advances and allow the members of the scientific community to access and use them for the purposes of science's

cumulative development (Bernal, 1939).

Since the second half of the 20th century, the picture has changed and scholars have essentially become producers of copyright-protected papers (Origgi, 2016). This goes hand-in-hand with a shift in the publishing imperative–from a moral obligation to disclose knowledge, to a means for individual gain through the use of knowledge enclosure. By producing new copyrighted items, scholars absolve their obligation to the scientific market and feed the main inputs of the publishing industry. Copyright first, and citations only afterwards, have therefore become the new "pellets" of peer recognition–the currency convertible into money in the academic market (Osterloh & Frey, 2020). Copyright has also had another distinct effect on researchers' behavior. According to expressive law theory [15], law has the ability to affect agents' behavior as it signals the attitudes that are expected and valued in a particular community, independently of the sanctions imposed on deviant behavior. By this reasoning, the growing salience of copyright has brought about a new consensus – or focal point – in the scientific community, by changing the very meaning of publication from "contribution to common knowledge" to "piece of private knowledge".

To some extent, the described dynamics can be assimilated to a sort of "normification" of copyright. That is to say, beyond the regular protection it provides within the market of ideas, copyright has opened the way to commodification and marketization of various other spheres of the scientific community's functioning, ranging from access to the community itself to career prestige within it. Accordingly, agents understand that knowledge enclosure, as promoted by law, is the model most accepted and approved by the community, even if it is at odds with the ethos of science. This leads them to update their beliefs about the approval patterns of the scientific community, and to consequently expect that violation of the communalism will not trigger any social disapproval or stigmatization from the scientific community. By contrast, they will interpret non-adherence to knowledge enclosure as the sign of a researcher's inability to fully belong to the scientific community, as this researcher will be considered as infringing on the rules accepted by the whole community. Ultimately, it can be argued that the increase in the salience of copyright rules and their expressive

15. See *e.g.* Cooter, R. (1998). 'Expressive Law And Economics'. The Journal of Legal Studies 27 (S2) : 585–607 ; McAdams, R.H. (2000). 'An Attitudinal Theory of Expressive Law Norms Theory'. Oregon Law Review 79 (2) : 339–90.

function has led to a new institutional equilibrium within the scientific field. Although the change in equilibrium has resulted in a change in social norms, individual values have not necessarily followed suit (Cooter, 1997).

2.3.2 The new copyright era and its stickiness

If copyright seems today to have largely displaced the ethos of science, the change in equilibrium did not completely eradicated norms-based IP from the scientific community. Those norms, which were pushed into the background by the growing salience of copyright, have been brought back to the fore and reinvigorated by new technological opportunities that open the door to OA publishing within the copyright system. A growing number of OA initiatives provide examples of communalism in science. Starting from the Budapest Open Access Initiative in 2002, various actions have been launched to promote the adoption of OA to the detriment of CA, and this has somehow breathed new life into the ethos of science. These spontaneous efforts have even been supplemented by political acts specifically aimed at further boosting the adoption of OA publishing[16] . It is hoped that this "gentle nudge" approach, backed up by public authorities, will change the 'meaning of actions'–that is, the perception of copyright and its salience in academic publication (Sustein, 1996 ; Kahan, 2000)–to trigger a shift back towards the old social norms of the ethos of science.

Notwithstanding these efforts, the OA publishing model–as previously noted—has had only limited success to date. The reason for this low uptake is the inertia of the scholarly community, which resists abandoning CA publishing in favor of OA journals. The strong salience of copyright within the scientific community has overshadowed the former norms-based IP system, which is at best only partially embodied in the current legal system. The legal framework itself makes the way back to norms, though possible, tricky. Because legal rules were conceived for a world whose leitmotiv was exclusion by default, an additional legal layer was needed to essentially nullify the rigid limits established by IP-laws. Indeed, since copyright applies by default to any expression of idea, authors must act at the IP-laws level to create a countervailing legal institution that frees the ground for the norms

16. For example, in 2016 the French government passed the Pour une république numérique (For a digital republic) law, that creates exceptions to the all right reserved prevision ensured to the copyright holder, and allows authors to upload published papers into open archivesTrans..

of science. This is precisely the purpose of Creative Commons (CC) licenses, a contractual arrangement that depowers, through a specifically written proviso, all or significant parts of the exclusive rights otherwise automatically and *de jure* entitled by copyright [17]. In this respect, CC was an elegant answer set up by a handful of distinguished scholars from world-renowned universities to avoid copyright damages in the digital society, and relying upon a clever legal response. It is worth repeating that, in the world of copyright, the free circulation and reuse of ideas requires disarming the default proprietary system which is automatically applied by the dominating legal framework.

The salience of copyright is also revealed by initiatives that apparently serve communalism but do so by infringing copyright–and so somehow reconfirm its salience. A simple way promoting a de-facto OA was found by exploiting the opportunities offered by new technologies to create "pirate" repositories. The most famous websites, Libgen and Sci-hub, are currently a reference and work tool in both rich and poor countries (Bohannon, 2016). Their existence suggests that, on the one hand, copyright infringement is needed to go back towards openness, and on the other hand that law infringement in this specific context is not immediately perceived as an illegal act by a large share of the scientific community, nor it does challenge copyright salience. Of course, there have been various legal reactions and lawsuits by stakeholders in the market place–*i.e.*, publishers–such as Elsevier Inc. v. Sci-Hub (1 :15-cv-04282). Yet it is interesting to note that, even in this era of "normified" copyright, there have so far been no visible consequences or unanimous public condemnation by scholars, and "pirate" scientific repositories are still in operation within the scientific community. The new "normification" of copyright appears to have differentiated its role : while the proprietary system of copyright is exploited for career advancement and selling journals, infringement is instead framed in terms of the traditional norms-based IP regulation, without incurring any social stigma or psychological cost for the infringer.

17. Creative Commons licenses were ideated by Laurence Lessing, Hal Abelsom and Eric Eldred, founders of a US non-profit organization, in order to free ideas from the restrictions imposed by the copyright. CC do not cancel copyright, but they apply in addition to the it, making legal acts that otherwise would not be. In the Open Access publishing case, it is up to the author to choose under which license the work will be published.

2.4 The complementarity between the two IP systems : further comments

This two-way institutional complementarity between the new formal IP and its norm-based counterpart, with its apparent contradictions, is well-documented in the literature which reveals varying degrees of open-access compliance among scholars, depending on the field to which they belong. As observed by Severin et al. (2020), for certain disciplines, especially in the STM (Science, Technology and Mathematics) domain, the Open Access practice is becoming a "natural continuation of publishing culture", whereas in the HSS (Humanities and Social Sciences) domain implementation of Open Access still requires a cultural change. The different value placed on the communalism norm by scholars belonging to different communities may depend on the idiosyncratic evolution of the peer groups, and confirms that behaviors are the resultant of distinct forces, leading to disparate outcomes. In communities where the salience of copyright and marketization of knowledge persists, it may hamper the implementation of the OA publishing model.

The complementarity between the two IP systems is then fundamental. The functional performance of an institution is bounded by the presence of the other institution and vice versa (Höpner, 2005a, 2005b ; Amable 2000). Policies for OA implementation and exceptions posed to copyright aim to establish a complementarity between communalism as a social IP-norm and the renewed legal IP system. The reforms that shape the boundaries of copyright application exploit communalism to force agents to comply with the law, and at the same time reinforce the application of communalism as social norms. The complementarity between the two forms of IP should in this way grant a new stable reference frame for scholars' actions, where Open Access publishing becomes the strategy that complies with both legal and informal IP systems. In fact, the adoption of OA should reinforce communalism, exploited by the legal system to obtain scholars' compliance with the reformed law-based IP framework, legitimizing the choice of OA through a self-reinforcing process.

The reinforcement over time of a set of norms conceived for enclosing scientific knowledge, coupled with the strong complementarity between copyright and the marketization of the scientific community's functioning, has today produced a certain normative structure

in which enclosure of knowledge still dominates. This kind of complementarity weakens the new coherence pursued by governments and leaders of scientific community organizations, focused on communalism and *Open Science*. The current anchoring to knowledge-enclosure norms limits the salience of communalism. In this way, the publishing norm–enshrined as the ultimate goal of scientific activity, rather than a means for contributing to the communalism of science–destabilizes the coherence of the new institutional architecture, threatening the desired complementarity between social and law-based IP institutions.

Conclusion

Technological change in the digital era has engendered a new disruptive publishing model, OA, for the circulation of knowledge. In particular, the OA publishing system enables scholars to widely disseminate certified scientific articles with fewer restrictions than the CA model, at zero marginal cost. Despite the multitudinous anticipated benefits of this new publishing opportunity, OA is still underused, especially by researchers. A reasonable explanation for this state of affairs–which seems counter to what one would expect from a scientific community dedicated to promoting the circulation of knowledge and scientific debate in every way–is connected to the interplay between the institutions in question, namely norms-based IP and law-based IP, and the present-day salience of copyright. Copyright, the most relevant law-based IP in scholarly publishing, originally played essentially no role within the scientific community. Scientific communication was almost entirely governed by the norms system. At some point, the two IP systems partially acted as complementary tools, reinforcing each other. An example is provided by plagiarism, where copyright was a tool that protected the financial interest of publishers, but also of scholars by protecting the credit and reputation, which ensure indirect monetary rewards for scientists. In this case, the alignment of publishers' interest with those of scientists accounts for a stronger adherence of the scientific community to the legal framework.

With the professionalization of the scientific community, norms-based IP was no longer able to meet the needs of the publishing industry and the market. Starting from WW2, a partial take over by copyright occurred. Copyright slowly edged out the norms-based spirit

of communalism in favor of stricter privatization, at first for the purpose of selling journals within the market. At the same time, it gained salience among the scientific community, for whose members publishing shifted from being a 'means for circulating knowledge' to being an ultimate end in itself. As copyrighted works–*i.e.*, articles–have become the fundamental unit of the "publish or perish" system, and then the metrics by which researchers are assessed, the production of copyrights and copyrighted items, rather than production of ideas, has become the main output of the scientific community. Articles – and their perceived quality given by bibliometrics – are the new "pellets" of individual reputation that scientists in the academic market can use to attract resources and advance their careers. Even though copyright, through the publishing industry, continues to incidentally disseminate knowledge, the spirit has substantively changed, and it is now the means for disseminating a "piece of private knowledge", to the detriment of free, unrestrained circulation.

It is worth noting that copyright does not require complete deletion of previous values. In many cases it has produced a sort of hybridization of the modern copyright system, in which elements of the previous IP institutions co-exist. So, for instance, while researchers are well accustomed to licensing copyright to journals without complaint, they equally accept the existence of pirate repositories without complaint. This can reasonably be ascribed to the perception that the latter are legitimated on the grounds of the pre-existing norms, that somehow did not entirely disappear within the copyright world. That said, with difference nuances scholars remain essentially refractory to adopting OA publishing and discarding the dominating CA model. It is interesting to note that things might be poised for change, because the current dynamics are bringing the publishing industry closer to OA.

Bibliography of chapter 2

Amable, B. Institutional complementarity and diversity of social systems of innovation and production. *Review of International Political Economy 7*(4).

Andrighetto, G., D. Villatoro, and R. Conte. Norm internalization in artificial societies. *AI Communications 23*(4).

Babbage, C. (1830). *Reflections on the Decline of Science in England : And on Some of Its Causes, by Charles Babbage (1830). To which is Added On the Alleged Decline of Science in England, by a Foreigner (Gerard Moll) with a Foreword by Michael Faraday (1831).* B. Fellowes.

Bergstrom, T. C. Free Labour for Costly Journals ? *Journal of Economic Perspectives 15*(4).

Berk, R. N. Is plagiarism ever insignificant ? *AJR. American journal of roentgenology 157*(3).

Berman, E. P. *Creating the Market University : How Academic Science Became an Economic Engine*. Princeton University Press.

Bernal, J. D. (1939). The Social Function of Science. *The Social Function of Science.*. George Routledge & Sons Ltd.

Biagioli, M. and P. Galison (2013). *Scientific Authorship : Credit and Intellectual Property in Science*. Routledge.

Bicchieri, C. (2000). Words and Deeds : A Focus Theory of Norms. In J. Nida-Rümelin and W. Spohn (Eds.), *Rationality, Rules, and Structure*, Theory and Decision Library, pp. 153–184. Dordrecht : Springer Netherlands.

Bohannon, J. Who's downloading pirated papers ? Everyone.

Buranyi, S. Is the staggeringly profitable business of scientific publishing bad for science ?

Case, C. M. (1928). Scholarhip in Sociology. *Sociology and Social Research*, 323–340.

Castaldi, C. (2021). Sustainable innovation and intellectual property rights : Friends, foes or perfect strangers ? In *Sustainable innovation : Strategy, Process and Impact*. Routledge.

Caves, R. E. (2000). *Creative Industries : Contracts Between Art and Commerce*. Harvard University Press.

Chartier, G. (2013). *Anarchy and Legal Order : Law and Politics for a Stateless Society*. Cambridge University Press.

Cialdini, R. B., R. R. Reno, and C. A. Kallgren (1990). A focus theory of normative conduct : Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology 58*(6), 1015–1026. American Psychological Association.

Coleman, J. S. (1990). *Foundations of social theory.* Cambridge, Mass : Belknap Press of Harvard University Press.

Cooter, R. Expressive Law And Economics. *The Journal of Legal Studies 27*(S2).

Csiszar, A. (2017). How lives became lists and scientific papers became data : cataloguing authorship during the nineteenth century. *The British Journal for the History of Science 50*(1), 23–60. Cambridge University Press.

Cullen, R. and B. Chawner. Institutional Repositories, Open Access, and Scholarly Communication : A Study of Conflicting Paradigms. *The Journal of Academic Librarianship 37*(6).

Dallmeier-Tiessen, S., R. Darby, B. Goerner, J. Hyppoelae, P. Igo-Kemenes, D. Kahn, S. Lambert, A. Lengenfelder, C. Leonard, S. Mele, M. Nowicka, P. Polydoratou, D. Ross, S. Ruiz-Perez, R. Schimmer, M. Swaisland, and W. van der Stelt (2011). Highlights from the SOAP project survey. What Scientists Think about Open Access Publishing. *arXiv :1101.5260 [cs]*.

Darling, K. (2013). IP without IP : A Study of the Online Adult Entertainment Industry. *Stanford Technology Law Review 17*, 709–772.

Darling, K. and A. Perzanowski. *Creativity Without Law : Challenging the Assumptions of Intellectual Property.* NYU Press.

David, P. (1993). Intellectual Property Institutions and the Panda's Thumb : Patents, Copyrights, and Trade Secrets in Economic Theory and History. In N. R. Council (Ed.), *Global Dimensions of Intellectual Property Rights in Science and Technology* (The National Academies Press ed.).

Di Stefano, G., A. A. King, and G. Verona (2014). Kitchen confidential ? Norms for the use of transferred knowledge in gourmet cuisine. *Strategic Management Journal 35*(11), 1645–1670.

Edlin, A. S. and D. L. Rubinfeld (2004). Exclusion or Efficient Pricing - The Big Deal Bundling of Academic Journals Symposium : Integrating New Economic Learning with Antitrust Doctrine. *Antitrust Law Journal 72*(1), 119–158.

Ellickson, R. C. (1991). *Order without Law : How Neighbors Settle Disputes.* Harvard University Press.

Etzkowitz, H. Anatomy of the entrepreneurial university. *Social Science Information 52*(3).

Etzkowitz, H. The evolution of the entrepreneurial university. *International Journal of Technology and Globalisation 1*(1).

Fagundes, D. (2011). Talk Derby to Me : Intellectual Property Norms Governing Roller Derby Pseudonyms. *Texas Law Review 90*(5), 1093–1152.

Fauchart, E. and E. von Hippel (2008). Norms-Based Intellectual Property Systems : The Case of French Chefs. *Organization Science 19*(2), 187–201. INFORMS.

Foucault, M. (1969). Qu'est-Ce Qu'un Auteur ? *Société Française de Philosophie, Bulletin 63*(3), 73. Librairie Philosophique J. Vrin.

Fyfe, A., K. Coate, S. Curry, S. Lawson, N. Moxham, and C. M. Røstvik (2017). Untangling Academic Publishing : A history of the relationship between commercial interests, academic prestige and the circulation of research. Technical report, Zenodo.

Fyfe, A., J. McDougall-Waters, and N. Moxham (2018). Credit, Copyright, and the Circulation of Scientific Knowledge : The Royal Society in the Long Nineteenth Century. *Victorian Periodicals Review 51*(4), 597–615. Johns Hopkins University Press.

Gintis, H. *Game Theory Evolving*. Princeton University Press.

Green, S. P. (2002). Plagiarism, Norms, and the Limits of Theft Law : Some Observations on the Use of Criminal Sanctions in Enforcing Intellectual Property Rights. *Hastings Law Journal 54*(1), 167–242.

Green, T. We're Still Failing to Deliver Open Access and Solve the Serials Crisis : To Succeed We Need a Digital Transformation of Scholarly Communication Using Internet-Era Principles. SSRN Scholarly Paper ID 3333907, Social Science Research Network, Rochester, NY.

Hardin, R. (1982). *Collective Action*. New York : RFF Press.

Howsam, L. (1998). *Kegan Paul - A Victorian Imprint : Publishers, Books, and Cultural History*. University of Toronto Press.

Huang, P. H. and H.-M. Wu (1994). More Order without More Law : A Theory of Social Norms and Organizational Cultures. *Journal of Law, Economics, and Organization 10*(2), 390–406.

Höpner, M. What connects industrial relations and corporate governance ? Explaining institutional complementarity. *Socio-Economic Review 3*(2).

Höpner, M. (2005). Epilogue to Explaining institutional complementarity : What have we learnt ? Complementarity, coherence and institutional change. *Socio-Economic Review 3*(2).

Iljadica, M. (2017). *5. Painting on walls : street art without copyright ?* New York University Press. p. 118-141, Creativity without Law.

Johns, A. Miscellaneous methods : authors, societies and journals in early modern England. *The British Journal for the History of Science 33*(2).

Johnson, R., M. Fosci, A. Chiarelli, S. Pinfield, and M. Jubb (2017). Towards a Competitive and Sustainable OA Market in Europe - A Study of the Open Access Market and Policy Environment. Technical report, Research Consulting, OpenAire.

Kahan, D. M. (2000). Gentle Nudges vs. Hard Shoves : Solving the Sticky Norms Problem. *The University of Chicago Law Review 67*(3), 607–645. University of Chicago Law Review.

Kalleberg, R. A Reconstruction of the Ethos of Science. *Journal of Classical Sociology 7*(2).

Kellogg, D. Toward a Post-Academic Science Policy : Scientific Communication and the Collapse of the Mertonian Norms. SSRN Scholarly Paper ID 900042, Social Science Research Network, Rochester, NY.

Kleinman, D. L. (1998). Pervasive influence : Intellectual property, industrial history, and university science. *Science and Public Policy 25*(2), 95–102.

Loshin, J. (2007). Secrets Revealed : How Magicians Protect Intellectual Property without Law. SSRN Scholarly Paper ID 1005564, Social Science Research Network, Rochester, NY.

Macaulay, S. (1963). Non-Contractual Relations in Business : A Preliminary Study. *American Sociological Review 28*(1), 55–67. American Sociological Association, Sage Publications, Inc.

Masic, I. Plagiarism in scientific research and publications and how to prevent it. *Materia Socio-Medica 26*(2).

McAdams, R. H. (2000). An Attitudinal Theory of Expressive Law Norms Theory. *Oregon Law Review 79*(2), 339–390.

Merton, R. K. (1942). A Note on Science and Democracy. *Journal of Legal and Political Sociology 1*(1 and 2), 115–126.

Merton, R. K. (1957). Priorities in Scientific Discovery : A Chapter in the Sociology of Science. *American Sociological Review 22*(6), 635–659. American Sociological Association, Sage Publications, Inc.

Merton, R. K. (1973). *The Sociology of Science : Theoretical and Empirical Investigations*. University of Chicago Press.

Merton, R. K. (1988). The Matthew Effect in Science, II : Cumulative Advantage and the Symbolism of Intellectual Property. *Isis 79*(4), 606–623. The University of Chicago Press, The History of Science Society.

Migheli, M. and G. B. Ramello. The market of academic attention. *Scientometrics 114*(1).

Mnookin, R. H. and L. Kornhauser (1978). Bargaining in the Shadow of the Law : The Case of Divorce Dispute Resolution. *Yale Law Journal 88*(5), 950–997.

Moore, K., D. L. Kleinman, D. Hess, and S. Frickel (2011). Science and neoliberal globalization : a political sociological approach. *Theory and Society 40*(5), 505–532.

Morrell, J. and A. Thackray (1981). *Gentlemen of Science : Early Years of the British Association for the Advancement of Science*. Clarendon Press.

Moxham, N. and A. Fyfe (2018). THE ROYAL SOCIETY AND THE PREHISTORY OF PEER REVIEW, 1665–1965. *The Historical Journal 61*(4), 863–889. Cambridge University Press.

Nash, L. K. (1963). *The Nature of the Natural Sciences*. Little, Brown.

Nicita, A. and G. B. Ramello (2007). Property, Liability and Market Power : The Antitrust Side of Copyright. *Review of Law & Economics 3*(3), 767–791. De Gruyter.

of Science and Technology Policy, O. (2000). Federal Research Misconduct Policy.

Oliar, D. and C. Sprigman (2008). There's No Free Laugh (Anymore) : The Emergence of Intellectual Property Norms and the Transformation of Stand-Up Comedy. *Virginia Law Review 94*(8), 1787–1867. Virginia Law Review.

Origgi, G. La réputation scientifique, une course infernale. *Sciences Humaines N° 278*(2).

Osterloh, M. and B. S. Frey. How to avoid borrowed plumes in academia. *Research Policy 49*(1).

Perzanowski, A. (2013). Tattoos & IP Norms. *Minnesota Law Review 98*(2), 511–591.

Perzanowski, A. (2017). *4. Owning the body : creative norms in the tattoo industry*. New York University Press. p. : 89-117, Title : Creativity without Law.

Posner, R. A. Intellectual Property : The Law and Economics Approach. *Journal of Economic Perspectives 19*(2).

Raustiala, K. and C. Sprigman (2006). The Piracy and Paradox : Innovation and Intellectual Property in Fashion Design. *Virginia Law Review 92*(8), 1687–1778.

Raustiala, K. and C. J. Sprigman (2019). When are IP rights necessary ? Evidence from innovation in IP's negative space. *Research Handbook on the Economics of Intellectual Property Law*. Edward Elgar Publishing, Research Handbook on the Economics of Intellectual Property Law.

Severin, A., M. Egger, M. P. Eve, and D. Hürlimann. Discipline-specific open access publishing practices and barriers to change : an evidence-based review. *F1000Research 7*.

Sunstein, C. R. (1996). Social Norms and Social Roles. *Columbia Law Review 96*(4), 903–968.

Tushnet, R. (2007). Payment in Credit : Copyright Law and Subcultural Creativity. *Law and Contemporary Problems 70*(2), 135–174. Duke University School of Law.

Ulmer, E. (1978). *Intellectual Property Rights and the Conflict of Laws : A Study Carried Out for the Commission of the European Communities, Directorate-General for Internal Market and Industrial Affairs*. Kluwer.

Vittu, J.-P. (2002). La formation d'une institution scientifique : le Journal des Savants de 1665 à 1714. *Journal des Savants 2*(1), 349–377. Persée - Portail des revues scientifiques en SHS.

Ziman, J. (1996). Is science losing its objectivity ? *Nature 382*(6594), 751–754. Nature Publishing Group.

Ziman, J. (2000). *Real Science : What it Is and What it Means*. Cambridge : Cambridge University Press.

Zimmermann, J.-B. (1998). Un régime de droit d'auteur : la propriété intellectuelle du logiciel. *Réseaux. Communication - Technologie - Société 16*(88), 91–105. Persée - Portail des revues scientifiques en SHS.

Chapitre 3

Who wants Open Access ? Copyright Reform and Scientific Knowledge Dissemination : Empirical Evidences Based on French Data

Abstract

The attitude of scholars to Open Access publishing has been much discussed in the literature, showing that copyright is the main barrier to the development of Open Access. In 2016, the French Government reformed copyright, giving scholars the right to upload their scientific work, even after publication. Using data from the French open archive HAL, we call into question whether copyright is indeed the sole barrier to the transition from the traditional publishing system to the open one. In addition to testing whether the copyright reform has increased the number of published articles on free access, we also focus our attention on the time lag between articles being published and uploaded into the open archive. In fact, the French reform imposes a statutory embargo period, which should reduce the time lag for uploading Open Access articles.

Our results suggest that in most cases, the law has not increased the number of uploaded files, but there are differences among disciplines. Nevertheless, the French reform has had a positive impact on the time lag, reducing the period between the publication of scientific articles in journals and the upload to HAL.

Introduction

The digitisation of knowledge has profoundly changed the structure of scholarly communication, allowing publishers to reduce production costs. Nonetheless, journal subscription prices have continued to increase and this has long been a subject for discussion in the literature. Tenopir and King's (2000) study on commercial journals highlights an increase in subscription prices of +310% between 1975-1995. While new titles in libraries increased by only 1.9% from 1989 to 2005, expenditure for their provision increased by as much as 302% (Kyrillidou, & Young, 2008). The breaking point was reached in 2018, when Couperin (the unified French consortium of academic and research institutions for access to digital scientific publications) failed to reach an agreement with the publisher Springer, preventing access to the scientific literature for several months.

Alternative solutions for accessing newly produced knowledge should be taken seriously, and Open Access (OA) could be one of the suitable options. This knowledge dissemination

model aims to remove all legal and economic barriers to access to scientific literature (Suber, 2012). In particular, Green Open Access (hereinafter referred to as GOA) may be the most suitable solution for ensuring the transition from the traditional publishing system (also called Closed Access - CA) to the open one. It consists in self-archiving scientific documents, which may or may not have been previously published in scientific journals, in open repositories. Neither the author nor the consumer has to bear costs for respectively uploading and downloading papers, and authors will not suffer any reputational loss. In fact, unlike the Gold Open Access, which is truly a publishing model, authors can still benefit from the reputational gains resulting from Closed Access publications [1]. GOA is then promoted as the most economical solution to comply with Open Access policies [2] (Zhang & Watson, 2017).

Nevertheless, there are two main prerequisites for the implementation of GOA. Firstly, the storage of scientific documents requires the creation of digital infrastructure. Open archive costs vary a great deal, depending on their dimensions and storage capabilities. While Swan (2010) estimates that these costs range between $32,000 and $258,000, Willinsky (2009) argues that they can reach $1.5M, but it is usually up to public institutions to bear these costs. The second challenge that has to be faced in promoting the adoption of the Green route is the legal system, which protects intellectual property over intellectual works. Because Open Access eliminates economic barriers, authors could increase their visibility by uploading their papers into open archives, after publication in peer-reviewed CA journals, which certify the quality of the newly produced knowledge. Green Open Access is then an alternative way to transform a Closed Access article into an open one. The problem is that once the article is published, the publisher becomes the new copyright holder, and it has the power to allow or disallow published articles to be uploaded into open archives. The

1. Gold Open Access is a publishing model that does not charge consumers. In order to cover production costs, publishers have developed different business models, such as the APC model (Article Processing Charges) and the hybrid model, where Open Access papers submitted to APC are published in Closed Access journals. Although the Gold model is the main alternative to CA, its implementation requires coordination among the members of the scientific community. Because citations are a valuable asset for tenure and promotions (Cronin, 2005 ; Osterloh & Frey, 2020), publishing in new Gold Open Access journal is a riskier choice, due to the journal's lack of reputation (Warlik, Vaughan, 2007 ; Dallmeier-Tiessen et al., 2011). Moreover, this model requires the creation of funds intended to finance publication costs, usually imposed by established journals, while continuing to finance access to traditional journals, until the end of the transition from the closed to the open model.

2. There are more than 700 policies worldwide for the promotion of the Open Access model. For an exhaustive list, see RoarMap (https ://roarmap.eprints.org. Last access 03/06/2020.

fear of copyright infringement has been identified as one of the most important barriers to Open Access development (Lawal, 2002 ; Creaser, 2010 ; Schöpfel et al., 2016). Even when publishers allow the upload, their policies are different, leaving authors uncertain about their right to disseminate scientific results through different channels. Therefore, the Green route is virtually dependent on journals' policies on self-archiving.

In order to face the two main challenges mentioned above - of Green Open Access implementation, the French Government reformed copyright law and created a national Open Access repository. In 2001, the French National Center for Scientific Research (CNRS) launched Hyper Articles en ligne (HAL), to promote "a coordinated approach on a national level to open archiving of scientific output" (Baruch, 2007). With more than 700,000 scientific documents, HAL is the biggest database in the French scientific landscape. Copyright concerns have been faced through copyright reform. The "Pour une république numérique" law (French law for a digital republic, n. 2016-1321, 7/10/2016) allows authors who are at least half-financed by public funds, to self-archive post-print papers after a statutory embargo period of six months in the field of STM sciences (Science, Technology, and Medicine), and twelve months for SHS (Social and Human Sciences). The law has a double effect. Firstly, it grants authors the right to upload the published version of their articles, creating an exception to the copyright law. Secondly, the law standardises policies among publishers, reducing the uncertainty linked to the embargo period, which is the time that has to run between publication and the article upload. The law does not have a binding effect for authors, who can decide whether to enforce the right or not, but it obliges editors to accept the dissemination of intellectual works, over which they have a property right, through institutional repositories, once the embargo period has expired. This elimination of the main barrier to the development of Green Open Access should then be followed by an increase in post-print paper uploads.

The aim of this paper is to evaluate the impact of the copyright reform on scholars' practices, analysing both the volume of uploaded post-print papers and the time lag between article publication and its upload into an open archive, a question generally ignored by the literature (Rowley et al., 2017). As far as the author knows, the study provides the first empirical investigation of the HAL database. In fact, scholars' attitudes towards the

Open Access choice have generally been studied through surveys, looking at scholars' desire to contribute to Open Access development (Hendler, 2007 ; Kim, 2007 ; Creaser, 2010). Given the difference in Open Access practices among disciplines (Dellmeier-Tissen et al., 2011 ; Cooning & Younce, 2010 ; Spezi et al., 2013 ; Rodriguez, 2014), we test the effect of the law in different scientific domains. This approach allows us to analyse whether the "one-size-fits-all" policy (Eger et al., 2015) adopted by the French Government is the most suitable legal instrument for promoting the Green route.

The paper is structured as follows : Section 3.2 presents the literature review on Green Open Access studies, underlining the value of this instrument in allowing the transition from the closed knowledge dissemination system to the open one. In Section 3.3, we detail the empirical strategy, while in sections 3.4 and 3.5 the results are presented and discussed respectively. We discuss policy implications about the copyright reform in section 3.6.

3.1 The economic analysis of barriers to the development of Green Open Access

3.1.1 The Green Open Access model

Green Open Access is not a publishing model. It allows authors to self-archive research outputs into a digital environment, called repositories or archives, without imposing any charges on authors or consumers. Despite the lack of quality control, which only the journals can provide, GOA still has a central role in scientific knowledge dissemination for different reasons. Because it eliminates legal and economic barriers to access to scientific knowledge, it fosters the advancement of science, knowledge and technology (Willinsky, 2005), contributing to the innovation process and then to economic growth (Houghton & Sheehan, 2009).

As reported by the Directory of Open Access Repositories (DOAR) [3], Green Open Access is a growing phenomenon. From 86 repositories registered in 2005, today the DOAR has more than 4,300 repositories, with an increase of 1.64% in the last year. In France, we

3. "OpenDOAR is the quality-assured global directory of academic Open Access repositories. It enables the identification, browsing and search for repositories, based on a range of features, such as location, software or type of material held". Source : http ://v2.sherpa.ac.uk/opendoar/. Last access 03/06/2020.

find 142 different repositories, 78% of them are institutional, 13.4% disciplinary and the others either governmental or aggregating. More than 90% of these repositories are linked to HAL, which makes it the most important repository in the French research landscape.

As Houghton and Swan (2013) highlight, Green Open Access is the most cost-effective and affordable way for institutions (*e.g.* universities, libraries, research centres), authors and consumers to move towards Open Access. Consumers can access published papers, uploaded into an open archive, without paying the price imposed by journals. Academic institutions and research centres have the possibility to create coalitions to bear the cost of repository management and to avoid paying subscriptions to journals for maintaining their access to previous knowledge. In fact, with the digitisation of knowledge, consumers have to pay a price not only for accessing newly produced knowledge, but also for maintaining digital access to journals' previous volumes, which was not the case before the digital era, when consumers bought the paper version. In the latter case, consumers acquired a property right over the journal copy, while in the former they have only a right to access, with a limited duration in time. Repositories then become an alternative for accessing journals' knowledge stock and they allow a partial independence from the costly digital archiving service provided by publishers.

For authors, the Green model serves two main functions. On the one hand, it allows the dissemination of pre-print works, providing the opportunity to discuss results before their publication and then improving the quality of the research. On the other hand, Green Open Access implementation is less expensive in terms of coordination costs than the Gold Open Access model (Björk et al., 2014). New Open Access journals have not yet built up a reputation and publishing in OA journals is riskier for researchers than publishing in traditional journals, because consumers may have doubts about the quality of the newly produced knowledge. They might then avoid using and citing these journals (Warlik, Vaughan, 2007 ; Dallmeier-Tiessen et al., 2011). The fear of citation loss could then discourage the adoption of the Gold model. Instead, the Green way gives authors the possibility to upload a published paper and then to enlarge their audience, without fear of reputational loss and without paying any charges, as is usually the case for Gold Open Access publications.

The adoption of the Green strategy will benefit not only consumers, granting them free access to scientific knowledge, but also authors, who can increase the visibility of their scientific results, and then the probability of being cited by peers. However, the literature underlines the low rate of deposits in Open Access archives (Hendler, 2007 ; Kim, 2007 ; Creaser, 2010).

3.1.2 Copyright infringement as a barrier to Green Open Access

What seems to be a form of carelessness on the part of the scientific community is in fact a fear of copyright infringement (Swan & Brown, 2005 ; Schöpfel et al., 2016 ; Rowley et al., 2017). Once a paper is published, authors are no longer the copyright holders. Publishers are the sole actors entitled to control the dissemination of the intellectual works, and as a result, the Green model could be incompatible with the intellectual property rights system.

In much of the literature, the fear of copyright infringement is considered as the main barrier for Open Access implementation (Van Westrienen & Lynch, 2005 ; Spezi et al., 2013 ; Lawal, 2002 ; Creaser, 2010 ; Swan & Brown, 2005 ; Schöpfel et al., 2016). Nevertheless, Closed Access publishing and Green Open Access can coexist, but it depends on publishers' policies on copyright exceptions. These policies determine when, where and in what format third parties, usually the authors of the intellectual output, can disseminate a published document. Policies vary across publishers, making it costly for authors to acquire information about the conditions under which they can disseminate their published intellectual output. Unlike Gold Open Access [4], Green Open Access does not add any value to publishers' income, on the contrary, it is a valuable alternative instrument for the storage of scientific knowledge. However, publishers' policies are usually not so restrictive, and the embargo period (*i.e.* the time that has to run between the publication date and upload into an archive of the published work) is shorter than what authors may believe. Only a minority prohibit the Green Open Access option (Björk et al., 2014). Laakso (2013)

4. Journals are more lenient in promoting the Gold model, because of the fees that authors usually have to pay for publishing their scientific results. For publishers, even the Gold and the Green models are rival models to some extent. In fact, Green Open Access may deter authors from choosing the Gold model. Authors can continue publishing under the traditional model, and then upload their work into a repository, ensuring that the scientific community has free access. Analysing the policies of 107 publishers listed on the SHERPA/RoMEO from 2004 through 2015, Gadd and Covey (2016) find a significant correlation between the increase in self-archive restriction policies (*i.e.* Green Open Access) and the introduction of the Gold Open Access author-pays model.

highlights that in 4/5 cases, journals allow intellectual works to be self-archived, and in 2/3 cases without imposing any embargo period.

Nonetheless, in the same study, Laakso (2013) reports that while publishers are quite liberal in allowing the distribution of the accepted paper (81%), policies are more restrictive for the upload of the published version (11%). Similar results have been confirmed by Björk et al. (2014) and Gadd and Covey (2016). This paradox – because the knowledge and then the intellectual good are almost the same – can be explained by scholars' perception of the pre-print version of papers. Mullingam and Mabe (2011) show that only 16% of their survey respondents consider the final manuscript to be as valuable as the published version. This behaviour can be explained by the fact that only the published version attests immediately to the certification of the newly produced knowledge. The accepted and the published versions may differ to some degree and, due to time constraints, scholars want to be sure they have access to certified new knowledge, and not to results that may undergo changes. Scholars therefore prefer peer-reviewed papers (Creaser, 2010 ; Coonin & Younce, 2010 ; Coonin, 2011 ; Nicholas et al., 2015), which are the accepted papers modified following peer-reviewers' remarks and comments.

While the number of publishers allowing some form of self-archiving has increased in the past decade, other restrictions have been imposed on authors. Gadd and Covey (2016) look at the trend of self-archiving policies, using data from 107 publishers listed on the SHERPA/RoMEO database from 2004 to 2015, and they underline the fact that that restrictions about how, where and when self-archiving is allowed by publishers' policies have increased by 199%, 190% and 1000% respectively. Checking the policy of a single publisher in order not to incur copyright infringement, then becomes another barrier to Green Open Access use (Spezi et al., 2013 ; Rowley et al., 2017).

3.1.3 Scholars' attitudes towards Green Open Access

Repositories have been the focal point of a number of studies. In particular, researchers question whether contributing to the development of Green Open Access depends on authors' belonging to a particular scientific discipline. Moreover, even scholars in the very same discipline may have different behaviours, depending on the country in which they

operate.

Studies usually use surveys to examine self-archiving practices and they constitute a rich literature source. Lawal (2002) chooses a sample from a population of 240,000 researchers, representative of 9 different disciplines from colleges and universities in the USA and Canada. 82% of the respondents declare that they do not use repositories and no one posts articles accepted for publication, with the sole exception of biological sciences. Björk et al. (2010) look at the Green Open Access option for post-print papers. In this study, which analyses 1,837 peer-reviewed articles published in 2008, authors find that areas related to medicine have a rate of post-print paper uploading of 4.6%, compared to 25.9% in earth sciences, considered as the most pro-self-archiving discipline. Together with mathematics, earth sciences have been confirmed to have the highest Green Open Access paper rate, also by Gargouri et al. (2012). Using two samples of articles published in journals indexed by Thomson-Routers-ISI between 1998-2006 and 2005-2010 from 14 different disciplines, the authors find 43% of Green papers in mathematics, 35.9% in earth sciences, while biomedical research has the lowest rate of 5.5%. Social and human sciences have generally been considered by the literature to have a slower penetration of Open Access (Coonin & Youce, 2010) and a modest use of repositories (Cullen & Chawner, 2011 ; Heath et al., 2008). However, the recent study by Rowley et al. (2017) shows that scholars' attitudes towards Open Access are "relatively consistent across the academic community".

These differences in uploading practices may be explained by the researchers' awareness of Open Access as a complementary tool for knowledge dissemination. In a survey of 3,000 respondents, Creaser et al. (2010) find that even if the Open Access is still a neglected instrument for scientific communication, there are differences among disciplines. In particular, physical and medical sciences appear to be more aware of the Open Access option. This survey has been extended by Spezi et al. (2013). Compared to Lawal (2002), they register an increase in the number of post-print papers, in particular in life and medical sciences, while it remains a niche phenomenon for physical and social sciences. However, in a survey conducted by Cullen and Chawner (2011) in New Zealand, health sciences record the lowest rate of deposits.

One hypothesis is that the practice of self-archiving differs not only at disciplinary

level, but also among countries. This point has been developed by Eger et al. (2016), who compare Green Open Access among different countries. In a survey conducted in 2012 in Germany, Eger et al. (2015) find a higher rate of self-archiving for physics and mathematics, while health, chemistry, biology and life sciences, arts and architecture, agriculture and food sciences have a rate of self-archiving lower than 10%. When Eger et al. (2016) compare these results to other countries, they find different discipline effects. While mathematics remains one of the most prolific disciplines using the Green way in France and Italy (more than 70%), the discipline effect was different in Spain and Portugal. For example, in Spain more than 60% of respondents in mathematics and physics upload papers into open repositories, but other disciplines, such as law, political sciences, art and history, have a self-archiving rate higher than 50%.

Because different studies show that commitment to Open Access implementation varies among disciplines (Egen et al., 2015 ; Björk et al., 2010 ; Gargouri et al., 2012), we test the effect of the law for different scientific subfields. We expect that, in line with the literature, the copyright reform may have a non-uniform impact on scholars' behaviour, and that the "one-size-fits-all" approach may not be the best policy initiative for reforming academic publishing (Egen et al., 2015). Moreover, to limit any interference from different practices among countries, which suggest the existence of a country dimension in the scientific community, and not an international one, we choose to focus our attention on the French landscape.

3.2 Empirical strategy

3.2.1 The French Law "*Pour une république numérique*"

Through the Commission Recommendation C (2012) 4890 "Access to and preservation of scientific information", the European Commission encourages Member States to support the diffusion into the public domain of publicly funded research results. The European Commission recommends policies for the dissemination of scientific publications, the dissemination of research data and the preservation of results, through the promotion of interstate actions. For this reason, in 2016, France adopted the law "Pour une République

numérique" (Law n. 2016-1321, 7/10/2016), which sets out the European Commission's recommendation on Open Access development (article 30) [5].

Without imposing a mandatory policy, which would oblige authors to upload their intellectual work into an open repository, the French Government opted for a voluntary approach to developing Green Open Access [6]. Article 30 of the "Pour une république numérique" [7] law allows authors who are at least half-financed by a public agency, to upload the final version of their article accepted for publication into an institutional repository, after the expiration of a statutory embargo period. The law has a mandatory effect for all publishers, regardless of whether they operate in France or in other countries, and they cannot complain if, after the end of the statutory embargo, authors upload the final paper into an institutional repository. The law has two main consequences. Firstly, it creates an exception to the copyright law, allowing authors, as non-copyright holders, to redistribute their intellectual work. Secondly, it harmonizes publishers' policies, reducing the cost of searching and checking an individual publisher's policy to almost zero.

Because authors now have the right to upload their post-print paper, and following the main hypothesis of the literature, we should expect an increase in the probability of finding a post-print paper after the introduction of the law. The law also imposes a statutory embargo period, which no longer differs from one publisher to another. After six months for STM and twelve for SHS, authors have the right to upload their intellectual published works into an open archive. We expect that the law will reduce the time lag between the publication and the upload of the published intellectual work.

5. We decided to focus on the French case for different reasons. Firstly, France is one of the rare countries that has adopted a centralized archive for the storage of scientific papers, whereas most other countries have different and unlinked archives, which make data collection difficult. Secondly, previous research conducted by the CNRS (Schöpfel et al., 2016) clearly demonstrates that the limited use of Green Open Access is partly explained by copyright concerns about documents uploaded into the French scientific landscape. Thirdly, France was one of the first EU States to amend its copyright law. It will be very interesting indeed for future researchers to compare the French strategy and results with other countries that have reformed the Copyright Law to promote the Green Open Access choice, such as Finland. (For an analysis of EU States' compliance with the Commission Recommendation C (2012) 4890, see "Access to and Preservation of Scientific Information in Europe. Report on the implementation of Commission Recommendation C(2012) 4890." (2015)).

6. In order to limit constraints imposed by publishers' policies, academic and funding institutions have also adopted policies for promoting Open Access publishing. The European Union has 463 different policies, which represent more than 60% of policies worldwide. Policymakers are funders, research organisations or sub-units of research organisations (see ROARMAP for further details : `http://roarmap.eprints.org/view/policymaker_type/`, last access 01/06/2020).

7. For a detailed analysis of Art.30 of the "Pour une République numérique" law, see Annex I.

3.2.2 Data collection and descriptive statistics

As reported by Schöpfel et al. (2016), HAL is the dominant French digital archive for the storage of scientific knowledge. We built the database using the web scraping method, starting from application programming interfaces (API) provided by HAL [8]. However, to obtain some variables, we were obliged to inspect the source code for bugs or missing information, such as the upload date or different access to uploaded documents. Despite the fact that HAL provides information about the type of journal in which the document is published (Open Access or not), we cross this information using three other sources, namely Web of Science, Scimago Journal Ranking and the Directory of Open Access Journals databases, in order to avoid errors.

After cleaning the data, the database contains 788 770 references, created on HAL from 01/01/2001 – date of HAL's creation – until 02/04/2020 – date of data download. A reference is the bibliographic information of a scientific work stored on HAL. In our case, all the notices refer to published articles, as they are the only scientific documents concerned by the copyright reform. As Figure 3.1 shows, the number of references stored on HAL has increased from the very beginning. For example, in 2020, where data refer only to the first four months, the number of new notices is almost the same as for the whole of 2015. Figure 3.2 reports the average number of notices providing access to a copy of the published paper. However, upload rates differ between SHS and STM sciences. While STM have an average upload rate for published documents of about 55.7%, the SHS average rate is only about 26.7%. Moreover, almost 26% of STM notices provide a copy of the article in a secondary repository (*e.g.* ArXive, PubmedCentral, etc.), while SHS accounts for only 6.9%. These preliminary observations seem to confirm the existence of the discipline effect, well documented by the literature, in the development of Open Access engagement and in the use of open archives.

Looking at the time lag, Figure 3.3 shows a decreasing trend. After the introduction of the law, the decreasing time lag has characterized both STM and SHS sciences. For all disciplines, the average time lag decreased by more than three years. This could perhaps be attributed to the law, but also to a general trend, which started before the copyright

8. `https://api.archives-ouvertes.fr/docs/search/schema/field-types`

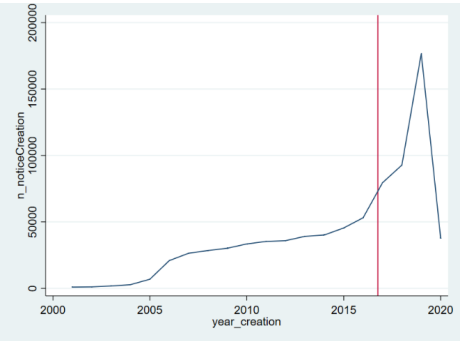

FIGURE 3.1 – Number of notices per year

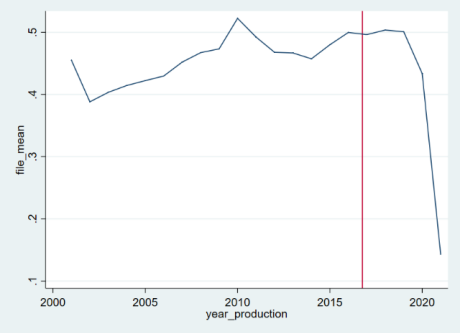

FIGURE 3.2 – Average of uploaded files

reform, with increasing attention being paid by the scientific community to Green Open Access as a complementary instrument for scientific knowledge dissemination.

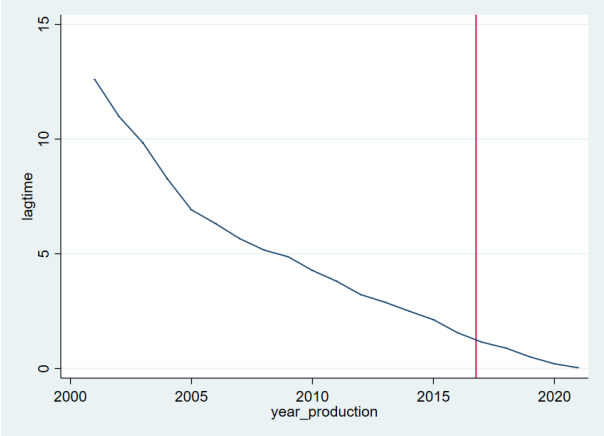

FIGURE 3.3 – Time lag

As well as reforming copyright, the French law standardises publishers' policies. Documents stored on HAL have been published by more than 10,000 publishers, in about 22,000 different journals. Policies vary among publishers, who decide when and which version the author can upload into a repository. This information is collected and analysed by HAL using the Sherpa RoMEO classification. It is a partnership among different institutions, which aims to promote the use of institutional repositories. Journal policies are grouped into four categories formulated by Sherpa, plus a grey category, conceived by HAL for non-transparent or unknown policies, as summarised in Table 3.1.

More than 50% of the notices referring to a published document have the right to upload the post-print version into an open archive, whereas the upload of post-print papers is prohibited for about 20% of the notices. Uncertainty about publishers' policies affects

Sherpa Colour	Policy
Blue	Can archive post print (*i.e.* final draft post-refereeing) or publisher's version
Grey	Policy unknown
Green	Can archive pre-print, post-print or publisher's version
Withe	Archiving not formally supported
Yellow	Can archive pre-print version (*i.e.* pre-refereeing)

TABLE 3.1 – Sherpa Romeo Policies

nearly 25% of notices stored on HAL.

The introduction of the law should eliminate this uncertainty and promote the practice of self-archiving, despite the colour of publishers' policies. Ignorance about the Green Open Access publishing agreement creates a distorted perception of publishers' Open Access policies. Looking at notices on HAL, we see that the most frequent policy adopted by Elsevier and Springer-Nature, responsible together for almost 30% of published notices stored on HAL, is the Green one. Almost all the first ten publishers recorded on HAL, which account for more than 50% of notices referring to published articles on HAL, authorize the upload of both pre- and post-print versions. There is an exception for Wiley-Blackwell, which in most cases authorizes the upload of the pre-print version only, and for the American Chemical Society, which does not formally support self-archiving. The case of Dalloz is controversial. It is a French publisher specialised in law, and its policies have not yet been classified.

Publishers' policies found on HAL seem to confirm the results of Tennant et al. (2016), who show that there is no evidence that these policies completely limit the possibility of uploading published papers into open repositories. However, the Green policy is not always the dominant one. There are important differences among disciplines, which reinforce our decision to study the effect of the law in each scientific sub-domain.

Another factor that may influence the adoption of the Green route is scholars' willingness to contribute to Open Access development in general, choosing, for example the Gold model. When we look at notices referring to articles published in Open Access journals, we see that it is up to the notice's creator to establish whether or not the paper has been published in an Open Access journal. The problem is that we think that there is very little knowledge about the Open Access model. Many papers are registered as published in an Open Access journal. When we cross-reference HAL data with data from Web of Sciences,

Scimago Journal Ranking and Directory of Open Access Journal, we find discrepancies. We consider a paper to have been published in an Open Access journal if the journal is listed in one of the above-mentioned databases. Almost 11.48% of the scientific article notices stored on HAL are published in an Open Access journal. The rate drops to 6.8% when we look at Open Access publications in the SHS domain suggesting, once again, different publication choices among disciplines.

3.2.3 The theoretical model

Following the literature hypothesis, the reform of the copyright law in the French scientific landscape should have a double impact on scholars' behaviour. Firstly, we should expect an increase in the notices of scientific published articles providing access to a copy of the work. Secondly, because of the harmonization of publishers' policies, the time lag between the publication of the paper and its upload into HAL should decrease.

Let $Y_{kt} = 1$ be an article notice providing access to a copy of the work and $Y_{kt} = 0$ otherwise, where :

- $t = P$ if the notice refers to an article published after the entry of the law into force, because we assume that the law is not retroactive (See Annex I for a detailed explanation).
- $t = B$ otherwise
- $k = 1$ if the paper is published under the "open" regime, which means after the entry of the law into force
- $k = 0$ if the paper is published under the "closed" regime, *i.e.* before the copyright reform.

The difference between t and k is that while t captures some period effect, which is independent from the entry of the law into force but could have influenced scholars' behaviour, the variable k aims to explain to what extent Y has been influenced by the copyright reform. The aim is to measure the impact of the law reform, or $Y_{1t} - Y_{0t}$. The outcome could be different in the period B or P, but the copyright reform took place in 2016. Given that, we are looking for $Y_{1P} - Y_{0P}$, or the difference at the same period between the two regimes. However, we are only able to observe $Y_{1P} - Y_{0B}$, given that all the papers

published in period P are under the "open" regime and all the papers published in period B fall under the "closed" regime. Moreover, articles published at $t = P$ are not the same as those that can be observed at $t = B$.

The fact that a notice does or does not provide a copy of the paper then depends on the period t, the regime k, and the unobserved determinants, depending on the period, the regime and the article :

$$Y_{ikt} = a_k + b_t + c_{kt} + \varepsilon_{ikt} \tag{3.1}$$

In an ideal world, where we have the possibility to observe the same notice, at the same period but under two different regimes, we could estimate :

$$\begin{aligned} \Delta_{iP} = Y_{i1P} - Y_{i0P} &= a_1 + b_P + c_{1P} + \varepsilon_{i1P} - (a_0 + b_P + c_{0P} + \varepsilon_{i0P}) \\ &= (a_1 - a_0) + (c_{1P} - c_{0P}) + (\varepsilon_{i1P} - \varepsilon_{i0P}) \end{aligned} \tag{3.2}$$

Where $(a_1 - a_0)$ is the systemic effect due to the regime change, $(c_{1P} - c_{0P})$ is the regime change effect specific to the period P and $(\varepsilon_{i1P} - \varepsilon_{i0P})$ is the effect of the regime change, specific not only to the period P, but also to the notice i. If the effect of the regime change is homogeneous (*i.e.* $\varepsilon_{i1P} = \varepsilon_{i0P}$) then the effect of the law only depends on the observable determinants $[Y_{i1P} - Y_{i0P} = (a_1 - a_0) + (c_{1P} - c_{0P})]$, otherwise we should also consider $(\varepsilon_{i1P} - \varepsilon_{i0P})$. Because of the difference in i, we can only estimate the average effect for heterogeneous unobservable characteristics or :

$$\Delta_{iP} = (a_1 - a_0) + (c_{1P} - c_{0P}) + (\frac{1}{n_{1P}} \sum \varepsilon_{i1P} - \frac{1}{n_{0P}} \sum \varepsilon_{i0P}) \tag{3.3}$$

In our case study, we have two different notice groups : group 1 (G1), for all the article notices published before 2016, and group 2 (G2), for all the article notices published after 2016. The outcome for a notice in G1 should be :

$$Y_i = Y_{i0B} = a_0 + b_B + c_{0B} + \varepsilon_{i0B} \tag{3.4}$$

Because of the difference in i, we can estimate the average impact of the law :

$$\frac{1}{n_{0B}} \sum Y_i = a_0 + b_B + c_{0B} + \frac{1}{n_{0B}} \sum \varepsilon_{i0B} \tag{3.5}$$

For group 2 we obtain :

$$Y_i = Y_{i1P} = a_1 + b_P + c_{1P} + \varepsilon_{i1P} \tag{3.6}$$

Because of the difference in i, we can estimate the average impact of the law :

$$\frac{1}{n_{1P}} \sum Y_i = a_1 + b_P + c_{1P} + \frac{1}{n_{1P}} \sum \varepsilon_{i1P} \tag{3.7}$$

Then, we can estimate the expected average impact of the copyright reform as :

$$\Delta = (a_1 - a_0) + (b_P - b_B) + (c_{1P} - c_{0B}) + (\frac{1}{n_{1P}} \sum \varepsilon_{i1P} - \frac{1}{n_{0B}} \sum \varepsilon_{i0B}) \tag{3.8}$$

The impact of the law can be estimated by comparing the notices before and after the entry of the law into force if :

$$\begin{aligned} &(a_1 - a_0) + (b_P - b_B) + (c_{1P} - c_{0B}) + (\frac{1}{n_{1P}} \sum \varepsilon_{i1P} - \frac{1}{n_{0B}} \sum \varepsilon_{i0B}) \\ &= (a_1 - a_0) + (c_{1P} - c_{0P}) + (\frac{1}{n_{1P}} \sum \varepsilon_{i1P} - \frac{1}{n_{0P}} \sum \varepsilon_{i0P}) \end{aligned} \tag{3.9}$$

The equation 3.9 is satisfied under the following hypothesis [9] :

- $b_P = b_B$
- $c_{1P} - c_{0P} = c_{1P} - c_{0B}$
- $\frac{1}{n_{0B}} \sum \varepsilon_{i0B} = \frac{1}{n_{0P}} \sum \varepsilon_{i0P}$

The first hypothesis $b_P = bB$ states that there is no "period" effect, which means that the year 2016 has had no impact on scholars' behaviour. The second hypothesis $c_{1P} - c_{0P} = c_{1P} - c_{0B}$ means that there is no regime change specific to the period P. These hypotheses are quite realistic, if we consider the situation in which the law entered into force. No technological changes had affected the use of Green Open Access among scholars. In fact, HAL had already been developed and no major technical adjustment had taken place. However, it is true that even before the law some laboratories had adopted specific policies for increasing the engagement of the scientific community through Green Open Access. Because HAL provides access to the research centre to which authors belong, we are

9. In order to strengthen our results, in Annex V we propose a difference-in-differences model, to test the effect of the law once again.

able to control for these changes in the empirical analysis. The third hypothesis states that an unobserved determinant has on average the same outcomes under the "closed" regime, independently from the periods B and P. As we said, we cannot observe determinant ε_{i0P}, because all papers published in P are automatically under the "open" regime. However, no particular event occurred at this time, apart from the copyright reform, and we control for journal policy changes and for research centre policy changes. For these reasons, we assume that the notices that we observe at time P are representative of those observed at time B. We can then assume that $\frac{1}{n_{0B}} \sum \varepsilon_{i0B} | i \in G1$ gives an unbiased estimate of $E(\frac{1}{n_{0P}} \sum \varepsilon_{i0P} | i \in G2)$.

3.2.4 The empirical model

Following Eger et al. (2015), we use a logit model to estimate the impact of the law reform on scholars' behaviour to upload a copy of a published paper into HAL due to the copyright reform. The specification of the logit model is as follows :

$$Y_i^* = \alpha_0 + \beta_1 + X_i + \epsilon_i \tag{3.10}$$

The first outcome of interest Y_i^* is then a dummy equal to 1 if and only if the notice provides a copy of the document after its publication. β_1 is then our variable of interest, which measures the impact of the law on the probability of finding a notice providing access to a copy of the paper. It is a dummy variable equal to 1 if the notice refers to a paper published after the entry of the law into force, and to 0 otherwise.

X_i is a vector of control variables. We control for 18 different disciplines (see Annex II for a detailed list), due to the heterogeneous approach highlighted by the literature on the adoption of Open Access. Another variable that could help explain scholars' behaviour is the Open Access policy adopted by research laboratories or the project under which the paper is produced. In fact, some institutions with projects financed by the European Union require authors to put their articles under the Open Access model (see Annex III for research institutions for which we control). The fact that a document has been previously published in an Open Access journal may be a sign that the scholar has a higher propensity

to use Open Access development. The same scholar may be more sensitive to the Open Access issue, and then upload the paper into an open archive. Using data from Web of Sciences, Scimago Journal Ranking and the Directory of Open Access journal, we are able to control for this propensity to adopt Open Access. As we said, the law imposes a statutory embargo period, which differs from one discipline to another, therefore we control for any violation of the copyright law. Finally, the pandemic that spread in 2020 may have affected scholars' behaviour. It is for this reason that we introduce a control variable for those documents that have been uploaded during the pandemic crisis.

As mentioned previously, the law does not only allow the uploading of a copy of a post-print paper, but it also imposes a statutory embargo period. To measure the impact of the law over the time lag, we use the following linear regression model :

$$Y_i^* = \alpha_0 + \beta_1 + X_i + \epsilon_i \tag{3.11}$$

Where Y_i^* is the time lag expressed in years, and β_1 measures the impact of the law over the period that elapses between the publication of a scientific document and its upload into an open archive. The aim is to understand whether the law β_1 diminishes the time lag, for those scholars who choose to comply with the law. For this reason, for this second outcome, we consider only notices which provide access to a document in our database. All the other control variables are equal to the previous model.

3.3 Results

3.3.1 The effect of the copyright reform on increasing free access to published papers

The tables below show estimates of equation 3.10, the effect of the copyright reform β_1 in finding a notice referring to a published paper providing a copy of it. The first general result that emerges from Table 3.2 is that estimates of β_1 seem to suggest that the law has not produced the forecast results highlighted by the literature, namely that copyright is a serious barrier to the development of Green Open Access. When analysing all the

disciplines together, the estimate is not significant ; it is positively significant, but only at 10%, for SHS and negatively significant at 5% for STM. A deeper analysis of disciplines is therefore required, given the immediate difference that can be seen in the general analysis performed, and between the SHS and STM domains.

VARIABLES	General	STM	SHS
β_1	-0.00398	-0.0158**	0.0211*
	(0.00604)	(0.00688)	(0.0127)
Constant	-0.928***	-0.0565***	-0.667***
	(0.00795)	(0.00394)	(0.0113)
Observations	788,77	564,403	224,367
Control	YES	YES	YES
Pseudo R2	0.116	0.0676	0.0771

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.2 – Estimates of equation 3.10 for general, STM and SHS disciplines

Table 3.3 reports results for each discipline analysed in our sample for SHS. For most of the SHS disciplines, it appears that, after the law came into force, the odds of finding a notice providing access to a copy of the paper have decreased. This is the case for psychology, political sciences and economics, for which the odds of finding a paper decrease by 0.7 times on average compared to before the copyright reform.

VARIABLES	psy	scipo	philo	art	eco	droit	socio	histo	litt
β_1	-0.369***	-0.275***	-0.109	0.199***	-0.163***	0.618***	0.138***	-0.00672	-0.0161
	(0.0873)	(0.0720)	(0.0752)	(0.0723)	(0.0316)	(0.0544)	(0.0429)	(0.0355)	(0.0376)
Constant	-0.572***	-0.325***	-0.0541	-0.457***	-0.453***	-1.418***	-0.341***	-0.924***	-0.563***
	(0.0559)	(0.0702)	(0.0679)	(0.0958)	(0.0202)	(0.166)	(0.0386)	(0.0489)	(0.0498)
Observations	3,654	5,965	5,363	4,866	30,633	61,644	16,343	22,12	19,527
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R2	0.0623	0.0222	0.0291	0.0264	0.0518	0.0446	0.0144	0.0656	0.0147

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.3 – Estimates of equation 3.10 for SHS disciplines

While for history, philosophy and literature the law seems not to have changed scholars' attitudes toward Green Open Access, estimates are significant and positive in three disciplines, arts, law and sociology. Here, the odds of finding a paper after the law came into force increased by 1.85 times for law, by 1.22 for art and by 1.15 times for sociology.

For all these first results, we assume that the law produces an immediate effect, starting from the day of its entry into force. However, its effect may be delayed. In fact, it could

take time for scholars to adapt their behaviour, to understand the functioning of HAL, to learn how to upload a paper and, most of all, to become familiar with the content of the law and their new rights. For these reasons, we perform different tests. In the first one (see Annex IV.I.I, Table 3.11), we drop observations of documents produced between 07/10/2016 and 07/10/2017, *i.e.* for one entire year after the entry of the law into force. As a second check on the robustness of our results, we run the very same regression without dropping observations but setting the variable β_1, which represents the entry into force of the law as equal to one if the notice refers to a paper published after one year of the law coming into force, *i.e.* after 07/10/2017 (see Table 3.14 in Annex IV.I.II).

For those disciplines where estimates were already negative, namely psychology, political sciences and economics, results are confirmed by both robustness check models and the odds are very similar to previous results.

However, once we drop one-year observations, in those disciplines for which estimates suggested a non-effect of the law, namely philosophy and literature, results differ. For literature, the non-effect of the law is confirmed for both robustness check models, whereas estimates for philosophy are negative. Specifically, the odds of finding a paper are 17.5% lower for post-law papers in the field of philosophy than for papers published before the copyright reform, although estimates are significant only at a 5% confidence interval. Results are consistent even when we postpone the entry of the law into force by one year, showing very similar percentage changes in the odds.

Once the one-year window is introduced, the positive impact of the law that we found in the disciplines of law, sociology and art is still confirmed, showing that the odds of finding a paper after the copyright reform have increased by 1.9, 1.1 and 1.2 times respectively compared to finding a paper previously. When we postpone the entry of the law into force by one year, results do not change. However, for art and sociology, estimates are significant only at 5%.

VARIABLES	spi	sdv	sdu	sde	phys	scco	chim	info	mat
β_1	0.150***	0.00949	0.0753***	0.0230	0.0927***	0.0538	0.0518***	0.121***	-0.324***
	(0.0198)	(0.0130)	(0.0281)	(0.0324)	(0.0181)	(0.0507)	(0.0193)	(0.0248)	(0.0277)
Constant	-0.451***	-0.141***	0.316***	-0.289***	0.299***	-0.736***	-1.054***	-0.0794***	0.328***
	(0.0115)	(0.00888)	(0.0137)	(0.0230)	(0.00888)	(0.0309)	(0.0133)	(0.0157)	(0.0163)
Observations	59,897	159,573	43,458	23,761	89,552	10,78	71,755	40,317	31,206
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R2	0.0475	0.124	0.0853	0.0960	0.0378	0.135	0.0860	0.0502	0.0174

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.4 – Estimates of equation 3.10 for STM disciplines

Table 3.4 provides estimates for the STM disciplines. Only mathematics shows negative estimates of the effect of the law, for which the odds of finding a notice providing access to a paper are 27.6% lower than for notices of papers published before the law. Results are consistent even when both robustness checks are performed, as shown in Table 3.12 (Annex IV.I.I) and Table 3.15 (Annex IV.I.II).

For life science, environmental sciences and cognitive sciences, on the other hand, the law seems not to have changed scholars' behaviour. Once we drop observations of notices published one year after the law, results do not change for cognitive sciences and environmental sciences and they remain consistent even when we postpone the effect of the law by one year. We find similar results for life science. In fact, the non-effect of the law is confirmed when its effect is postponed by one year, and the small positive effect that we find when we drop the one-year observation, an increase in odds of about 2.8%, is significant only at 10%.

At first glance, small but significant positive estimates emerge more frequently for STM, than for SHS disciplines. This is the case for informatics, chemistry, physics, astrophysics and engineering, where all the estimates are significant at 1%. However, when we run the robustness checks, results change for some disciplines. Estimates for chemistry become non-significant both when we drop the one-year observation and when we postpone the entry of the law into force. For astrophysics, the odds of finding a notice providing access to a paper after the copyright reform are 1.07 times the odds for a notice in the pre-copyright reform period. However, when we drop the one-year observation the odds are similar but the estimate is significant only at 10% and the significance disappears when we postpone the entry of the law into force. For physics, informatics and engineering, the positive effect

of the law is confirmed in both the robustness checks, even if for informatics the coefficient of interest is significant at 5% when we postpone the entry of the law into force. On average, the odds of finding a notice providing access to a paper after the copyright reform are 14.6%, 10.1% and 7.1% higher (for engineering, physics and informatics respectively) than for notices of papers published before the law reform.

3.3.2 Time lag

The second aim of the law is to reduce the time lag between the publication date of a scientific document and its upload into an open archive. In order to test this second effect, we reduce our sample to observations that provide access to a scientific document. At first glance, estimates of equation 3.11 provided in Table 3.5 show a decreasing time lag for both SHS and STM disciplines, for papers published after the copyright reform. For SHS disciplines, the time lag has been reduced by more than 4 and a half years, and for STM disciplines by more than 3 and a half years.

VARIABLES	General	STM	SHS
β_1	-3.809***	-3.601***	-4.674***
	(0.0101)	(0.0110)	(0.0244)
Constant	5.489***	5.121***	5.271***
	(0.0231)	(0.0127)	(0.0373)
Observations	221,141	176,198	44,943
R-squared	0.293	0.289	0.287
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.5 – OLS estimates of equation 3.11 time lag for all disciplines, STM and SHS fields

When we look at both SHS and STM in more in detail, the decreased time lag emerges for both fields, although SHS seems to have experienced a greater decrease in time lag than STM, as reported in Tables 3.6 and 3.7. This is also because publishers usually imposed a longer embargo period on SHS than on STM disciplines.

VARIABLES	psy	scipo	philo	art	eco	droit	socio	histo	litt
β_1	-4.131***	-4.668***	-4.360***	-4.787***	-3.702***	-4.925***	-4.612***	-5.770***	-4.683***
	(0.136)	(0.119)	(0.114)	(0.134)	(0.0563)	(0.123)	(0.0649)	(0.0706)	(0.0656)
Constant	5.404***	5.731***	5.320***	4.764***	4.472***	6.224***	5.622***	6.402***	5.124***
	(0.179)	(0.175)	(0.178)	(0.254)	(0.0612)	(0.491)	(0.113)	(0.163)	(0.138)
Observations	1,197	1,907	1,974	1,53	6,9	1,833	5,848	6,353	6,274
R-squared	0.263	0.237	0.273	0.330	0.256	0.378	0.266	0.333	0.260
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.6 – OLS estimates of equation 3.11 for STM disciplines

VARIABLES	spi	sdv	sdu	sde	phys	scco	chim	info	mat
β_1	-3.612***	-3.232***	-3.691***	-2.739***	-3.772***	-3.719***	-3.422***	-2.807***	-2.729***
	(0.0363)	(0.0213)	(0.0353)	(0.0545)	(0.0324)	(0.0900)	(0.0501)	(0.0311)	(0.0370)
Constant	4.820***	5.739***	4.992***	4.510***	4.471***	4.874***	4.220***	3.773***	3.521***
	(0.0408)	(0.0333)	(0.0363)	(0.0839)	(0.0312)	(0.118)	(0.0568)	(0.0404)	(0.0430)
Observations	19,131	50,337	16,786	6,303	24,348	2,631	11,403	16,908	11,668
R-squared	0.239	0.325	0.296	0.297	0.242	0.317	0.293	0.247	0.178
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.7 – OLS estimates of equation 3.11 for STM disciplines

In Annex IV.II.I, we provide OLS estimates of regression 3.11, after dropping one-year observations (Tables 3.17 and 3.18) and in Annex IV.II.II we introduce the postponed effect of the law (Tables 3.20 and 3.21). Even after these robustness checks, results remain consistent, showing an average decrease of about four and a half years for SHS disciplines and more than three years for STM.

3.3.3 Results discussion

While the main hypothesis of the literature, *i.e.* the underdevelopment of Green Open Access due to copyright concerns, seems not to be validated by our results, a discipline effect has emerged.

For most disciplines, in both STM and SHS domains, the copyright reform has had no effect on scholars' behaviour, as this is the case not only for history and literature, but also for astrophysics, life, cognitive and environmental sciences and chemistry, suggesting that there may be other reasons to explain the underuse of GOA.

One astonishing observation is the negative effect of the law, meaning that after the copyright reform, the odds of finding a notice providing access to a paper are lower than for notices referring to papers published before the copyright reform. However, we do

not believe that the law has had a negative impact on scholars' behaviour, discouraging the upload of published papers. We do not believe that the law has created disincentives to upload post-print papers, but simply that, in general, it has not changed scholars' behaviour. The negative effect could be explained through opposite forces that regulate the scientific community. In fact, part of the scientific community is promoting the use of HAL and the OA movement overall as a main repository for scientific documents. This is explained by the huge increase in bibliographic information on HAL, as reported in Figure 3.1. This task may be carried out by laboratory administrative staff, which could explain the increase in bibliographic information stored on HAL. However, they cannot provide a copy of the paper, because only authors have the right to upload documents to HAL. In order for administrative staff to be able to enrich HAL with bibliographic information about papers produced by laboratory members, financial resources have been allocated to support this activity. However, although this investment reduces the cost of GOA adoption for scholars, given the time that they save in not having to carry out this task themselves, it has not been followed by an increase in the number of papers uploaded, even after the copyright reform.

Counter to the main claim in the literature, copyright concerns are not the only barrier to Green Open Access implementation and development. As we have seen, the law has not so far changed scholars' behaviour. Either researchers have still not been informed about their rights to upload post-print papers into an open archive, or the importance of Open Access, so widely discussed at the highest levels of the scientific community [10], has not yet been accepted by all.

For some disciplines, the effect is positive and further research should be carried out to find the reasons why the law has had a positive impact on those scholars. For those disciplines where we find that the law has had a positive impact, STM shows an average increase in the odds of about 12%, against +42% for SHS sciences, with a particularly positive effect for the law discipline. The major impact of the copyright reform in the law domain may be explained as a positive bias specific to this sub-scientific community. Scientists in law may be more attentive to legislation changes, and could thus have a deeper

10. As demonstrated by different research centres' initiatives for promoting the adoption of GOA, listed in Annex III.

understanding of their new right to upload a paper, making an exception to the general copyright discipline.

While the effect of the law on changing scholars' behaviour regarding the uploading of their post-print papers into HAL seems not to have produced the expected results, our analysis suggests that it has a positive impact on reducing the time lag, the period between the publication of a paper and its upload into an open archive. We have to stress the fact that in order to study this second effect of the law, we had to take into account notices that already provided a copy of the paper, even before the law came into force. This means that our study group is composed of scholars who were already sensitive to the Open Access cause, and so they may also be more attentive to all the reforms that extend their rights for developing Open Access practices. In this case, we can hypothesise that the copyright was a barrier for uploading a paper, since the author had to wait until the end of the embargo period. In fact, once the law came into force, the delay was considerably reduced.

Conclusion

The literature about the development of Open Access publishing reports commitment on the part of the scientific community, but this is hindered by copyright. In fact, the low upload rate of post-print scientific documents has been attributed to copyright infringement concerns and to overly restrictive publishers' policies. To ensure the transition from the traditional publishing system to the open one and the development of Green Open Access as a complementary tool to standard scientific communication, two elements are required. Firstly, the creation of a centralized repository for collecting scientific works, and secondly, the reform of copyright. In 2001, France created HAL, an online archive which is today the most important digital repository in the French scientific landscape (Schöpfel et al., 2016). The introduction of the French law *Pour une république numérique*, and the resulting exceptions imposed on the application of the copyright to post-print papers, was intended to increase the upload of printed documents and hence the development of Green Open Access.

Starting from HAL data, we test whether or not a copyright reform will really change

scholars' behaviour, increasing the upload of post-print papers, as highlighted by the literature. Our findings suggest that the copyright exceptions are a necessary but insufficient step for promoting Open Access in most of the disciplines analysed here. In fact, the pre-reform copyright system was not aligned with the possibilities offered by the digitisation of information and the new knowledge dissemination system.

Moreover, the different impact of the copyright reform on the different disciplines analysed confirms that the "one-size-fits-all approach", already questioned by Eger et al. (2015), is not the best approach for promoting the development of Green Open Access, given the existing differences within the scientific community. For some disciplines, the law has a non-significant effect, especially in the SHS domain, suggesting that copyright is not the only barrier to the upload of post-print documents. Either the law is still unknown to scholars working in these fields of science, or their commitment to Open Access development is low. At this time, there are still no incentives to push scholars towards Open Access, and a voluntary approach to regulation may not be the most effective solution. Authors have the right and not the duty to contribute to the development of Open Access and no incentives have been adopted to promote this new practice.

We find that the negative effect of the law may be the result of the increase in notices stored on HAL, which was not followed by documents being uploaded at the same rate. The increase in notices on HAL is also due to the fact that, while bibliographic information is usually uploaded by research laboratory staff, authors are the only authorised agents for uploading their work. This situation clearly explains the tension that exists within the scientific community. Heads of research centres are trying to establish funds and policies to simplify the procedure for scholars to contribute to the transition from a closed to an open publishing system. However, for affiliated researchers the development of Open Access is still not a priority. Scholars may not be aware of the importance of Open Access publishing for the development of *Open Science*, or they do not have enough incentives to contribute to this new practice.

Positive results emerge for some social science disciplines, but the increase in uploads is generally weak, showing once again that copyright concerns are not the main barrier to Open Access development. The surprising increase in uploaded papers in the law discipline

suggests that awareness of the copyright reform can contribute to GOA development.

The law may not have produced the expected results for promoting the upload of papers after publication, however, it has had a positive impact on the time lag that occurs between publication date and the upload of documents to HAL. All other things being equal, results show a decrease in time lag of about three years, with differences among disciplines. Thus the law has had some effect, and the reduction of the time lag is particularly important for ensuring access to published papers, via a channel other than traditional journals. Because Green Open Access has been conceived for the storage of scientific documents, thus avoiding the fees imposed by traditional publishers for accessing online private archives, the increase in notices but not in papers limits its usefulness. Changes for promoting the adoption of Open Access should embrace not only legal constraints, such as copyright, but also social ones, for example, implementing incentives for using GOA. As we suggest, the evaluation of scientific output and the allocation of funds, could be based on scientific output reported in HAL. This study has produced some diverse results, highlighting the differences in Open Access practices among disciplines, which have then different norms to regulate the publication, dissemination and then the commitment to Open Access publishing development. For these reasons, a tailored reform of internal barriers specific to each sub-domain should be considered as a future strategy for the promotion of alternative publishing practices.

Bibliography of chapter 3

Law n°2016-1321, 7/10/2016, Pour une république numérique.

Baruch, P. (2007). Open access developments in France : the HAL Open Archives system. *Learned Publishing 20*(4), 267–282.

Björk, B., M. Laakso, P. Welling, and P. Paetau (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology 65*(2), 237–250.

Björk, B., P. Welling, M. Laakso, P. Majlender, T. Hedlund, and G. Guðnason (2010). Open access to the scientific journal literature : situation 2009. *PloS one 5*(6), 11273.

Coonin, B. (2011). Open access publishing in business research : The authors' perspective. *Journal of Business & Finance Librarianship 16*(3), 193–212.

Coonin, B. and L. Younce (2010). Publishing in open access education journals : The authors' perspectives. *Behavioral & Social Sciences Librarian 29*(2), 118–132.

Creaser, C. (2010). Open access to research outputs—institutional policies and researchers' views : results from two complementary surveys. *New Review of Academic Librarianship 16*(1), 4–25.

Creaser, C., J. Fry, H. Greenwood, C. Oppenheim, S. Probets, V. Spezi, and S. White (2010). Authors' awareness and attitudes toward open access repositories. *New Review of Academic Librarianship 16*(S1), 145–161.

Cronin, B. (2005). *The hand of science : Academic writing and its rewards.* Scarecrow Press.

Cullen, R. and B. Chawner (2011). Institutional repositories, open access, and scholarly communication : a study of conflicting paradigms". *The Journal of Academic Librarianship 37*(6), 460–470.

Dallmeier-Tiessen, S., R. Darby, B. Goerner, J. Hyppoelae, P. Igo-Kemenes, D. Kahn, and M. Nowicka (2011). Highlights from the SOAP project survey. What scientists think about open access publishing.

Eger, T., M. Scheufen, and D. Meierrieks (2015). The determinants of open access publishing : Survey evidence from Germany. *European Journal of Law and Economics 39*(3), 475–503.

Eger, T., M. Scheufen, and D. Meierrieks (2016). The determinants of open access publishing : Survey evidence from countries in the Mediterranean open access network (medoanet. *Economia e Politica Industriale 43*(4), 463–489.

Gadd, E. and D. Troll Covey (2016). What it means to be Green : exploring publishers' changing approaches to Green open access. Impact of Social Sciences Blog.

Gargouri, Y., C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, and S. Harnad (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PloS one 5*(10), 13636.

Heath, M., M. Jubb, and D. Robey (2008). E-Publication and open access in the arts and humanities in the UK, Ariadne, 54.

Hendler, J. (2007). Reinventing academic publishing – Part 1. *IEEE Intelligent Systems 22*(5), 2–3.

Houghton, J. and P. Sheehan (2009). Estimating the potential impacts of open access to research findings. *Economic Analysis and Policy 39*(1), 127–142.

Houghton, J. and A. Swan (2013). Planting the green seeds for a golden harvest : Comments and clarifications on "Going for Gold". *D-Lib Magazine 19*(1/2).

Kim, J. (2007). Motivating and impeding factors affecting faculty contribution to institutional repositories. *Journal of Digital Information 8*(2).

Kyrillidou, M. and M. Young (2008). ARL Statistics, 2005-06 : A Compilation of Statistics from the One Hundred and Twenty-Three Members of the Association of Research Libraries. *Association of Research Libraries*.

Laakso, M. (2014). Green open access policies of scholarly journal publishers : a study of what, when, and where self-archiving is allowed. *Scientometrics 99*(2), 475–494.

Lawal, I. (2002). Scholarly communication : The use and non-use of e-print archives for the dissemination of scientific information. *Issues in Science and Technology Librarianship*.

Mulligan, A. and M. Mabe (2011). The effect of the internet on researcher motivations, behaviour and attitudes. *Journal of Documentation 67*(2), 290–311.

Nicholas, D., A. Watkinson, H. Jamali, E. Herman, C. Tenopir, R. Volentine, and K. Levine (2015). Peer review : still king in the digital age. *Learned Publishing 28*(1), 15–21.

Osterloh, M. and B. Frey (2020). How to avoid borrowed plumes in academia. *Research Policy 49*(1), 103831.

Rodriguez, J. (2014). Awareness and attitudes about open access publishing : A glance at generational differences. *The Journal of Academic Librarianship 40*(6), 604–610.

Rowley, J., F. Johnson, L. Sbaffi, W. Frass, and E. Devine (2017). Academics' behaviors and attitudes towards open access publishing in scholarly journals. *Journal of the Association for Information Science and Technology 68*(5), 1201–1211.

Schöpfel, J., C. Ferrant, F. Andre, and R. Fabre (2016). Ready for the future ? A survey on open access with scientists from the French National Research Center (CNRS. *Interlending & document supply 44*(4), 141–149.

Spezi, V., J. Fry, C. Creaser, S. Probets, and S. White (2013). Researchers' green open access practice : A cross-disciplinary analysis. *Journal of Documentation 69*(3), 334–359.

Suber, P. (2012). Open Access, the MIT Press essential Knowledge series.

Swan, A. (2010). The Open Access citation advantage : Studies and results to date. Technical Report,, School of Electronics & Computer Science, University of Southampton.

Swan, A. and S. Brown (2005). *Open access self-archiving : An author study.* UK FE and HE funding councils.

Tennant, J., F. Waldner, D. Jacques, P. Masuzzo, L. Collister, and C. Hartgerink (2016). The academic, economic and societal impacts of Open Access : an evidence-based review. *F1000Research 5.*

Tenopir, C. and D. King (2000). *Towards electronic journals : realities for scientists, librarians, and publishers.* Washington, DC : Special Libraries Association.

Van Westrienen, G. and C. Lynch (2005). Academic institutional repositories. *D-lib Magazine 11*(9), 1082–9873.

Warlick, S. and K. Vaughan (2007). Factors influencing publication choice : why faculty choose open access. *Biomedical digital libraries 4*(1), 1–12.

Willinsky, J. (2005). The unacknowledged convergence of open source, open access, and open science. *First Monday 10.*

Willinsky, J. (2009). The stratified economics of open access. *Economic Analysis and Policy 39*(1), 53–70.

Zhang, L. and E. Watson (2017). Measuring the impact of gold and green open access. *The journal of academic librarianship 43*(4), 337–345.

Annex

I. Article 30 of the French Pour une république numérique law

In 2016, the French Government passed the *Pour une république numérique* law (law n. 2016-1321, 07/10/2016). It was designed to face the challenges of the digital transition, and promote an open digital economy and society. Hereafter, the text of Article 30 :

"When a scientific article, result of a research activity funded for at least half by the State, local authorities or public institutions, by national agencies or by European Union grant, is published in a journal which comes out at least once a year, his author has the right to provide, even if he has granted an exclusive right to a publisher, a free availability in an open format, via digital channels, subject to the agreement of possible co-authors, to all the successive versions of the manuscript till the final version accepted for publication, as soon as it is freely made available by the publisher via digital channels, or, failing that, within a set period starting from the first publication date. That period is a maximum of 6 months for a publication in sciences, technical sciences and medicine and 12 in humanities and social sciences. The version made available in accordance with the first subparagraph cannot be commercially exploited" [11].

The last paragraph states that : "The provisions of this article are of *ordre public* [12] and any clause contrary to them is deemed unwritten". In its broader meaning, *ordre public* covers the essential values of the social and legal system in France. During the preparatory work on the law, the Council of State delivered an opinion on the choice of the *ordre public* clause. For the Council, the nature of public order can only be enforceable on French territory, while the effect of knowledge dissemination through the internet is worldwide [13]. Discussion about the scope of the geographical validity of the law is still

11. Trad. Dazy, COUPERIN and French OpenAIRE National Open Access, at `https://www.openaire.eu/blogs/new-french-digital-republic-law-boosts-support-for-oa-and-tdm-1`, last access 02/07/2020.

12. The French text refers to the notion of "*ordre public*". As declared by Bernard Strin, president of the litigation Section of the Council of State, there is still no correct English translation for this concept (L'ordre public : regards croisés du Conseil d'État et de la Cour de cassation, 6th March, 2017. Available at `https://www.conseil-etat.fr/actualites/discours-et-interventions/l-ordre-public-regards-croises-du-conseil-d-etat-et-de-la-cour-de-cassation2`. Last access 07/07/2020).

13. « En ce qui concerne la mise à disposition gratuite sur l'internet des résultats de recherches financées sur fonds publics, prévue par l'article 14, le Conseil d'État a relevé que l'impact d'une telle mesure sur les contrats futurs entre éditeurs et auteurs tenait à son caractère d'ordre public, lequel ne peut jouer que sur le territoire français, alors que l'effet de la diffusion sur l'internet est mondial. Cette incohérence lui a paru

discussed today through informal channels. The *Centre pour la Communication Scientifique Directe*, created in 2000 by the National Centre for Scientific Research (CNRS), explains that the law produces effects regardless of the contract stipulated between the author and the publisher. In the comments, lawyers debate the validity of the law for foreign publishers. A.L. Stérin, legal consultant, explains that "nothing, neither foreign law nor contract [...] may contravene the application of this measure". For B. Moron-Puech, associate professor at University Paris II, it is a matter of private international law. Then, as set out in Article 4 of the Rome Convention, "the applicable law is the law with which the contract is most closely connected" [14], leaving the matter open to further discussion. The point is that if the copyright is infringed, due to the unauthorised upload of a paper published by a foreign publisher, we can propose two scenarios. In the first one, the contract is concluded under French law, so there will be no copyright infringement, because no contract can contravene the law in the internal legal system. In the second case, if the legal system chosen by the parties is not the French one, it is up to the foreign court to judge whether the copyright infringement subsists. Even in this case, the contract clause, which assigns the right to judge future issues to the foreign court, can be considered as an unfair term [15]. Then, the right to examine the offence would be assigned to the French court, which would probably be considered as a non-copyright violation. For these reasons, we assume that the law is binding for national and foreign contracts.

II. Disciplines

Because the literature has found different attitudes to Open Access on the part of scholars, and in particular a more committed approach by STM disciplines, we test the effect of the law over 18 different disciplines, listed in the table below.

faire obstacle à l'adoption de cette mesure » Conseil d'État Assemblée générale, Séance du jeudi 3 décembre 2015 N. 390741, available at urlhttps ://www.conseil-etat.fr/ressources/avis-aux-pouvoirs-publics/derniers-avis-publies/republique-numerique, last access 10/05/2020.

14. Source : `https://www.ccsd.cnrs.fr/2016/10/vos-depots-dans-hal-ce-qui-change-avec-la-loi-pour-une-repu` Last access 10/05/2020.

15. Ordonnance du 05 Mars 2015 -Tribunal de Grande Instance de Paris – RG n. 12/12401

Discipline ID	Discipline
Art	Art/architecture
Chim	Chemistry
Droit	Law
Eco	Economics & Finance and Econometrics
Histo	History
Info	Informatics
Litt	Language/Literature
Math	Mathematics
Philo	Philosophy
Phys	Physics
Psy	Psychology
Scco	Cognitive Sciences
Scipo	Political Sciences
Sde	Environmental Sciences
Sdu	Astrophysics
Sdv	Life Sciences
Socio	Sociology
Spi	Engineering

TABLE 3.8 – Disciplines analysed

III. Institutions' Open Access policies

Some institutions require free access to their scientific results. We target these research centres using data from RoarMap (Registry of Open Repository Mandates and Policies [16]), a registry of Open Access policies adopted by research institutions worldwide. The European Union counts 463 different policies, which represent more than 60% of policies all around the world. Policymakers are funders, research organisations or sub-units of research organisations. In 2017, France had 23 different policies, 16 promoted by research organisations, 4 by sub-unit or multiple research organisations and 3 by project funders. However, not all the institutions have adopted a mandatory approach. Because some institutions force their members to upload their research results into an open repository, the increase in the number of uploaded papers may be the result of these policies and not of the copyright reform. For these reasons, we control for the following research centres, listed in Table 3.9, and we assume that each policy has no retroactive effect.

16. `http://roarmap.eprints.org/view/policymaker_type/`. Last access 03/06/2020

Structure	Policy	Content type	Policy date
ANR	Requested	Not specified	12/08/2007
ANR Humanities & Social Sciences Branch	Required	Peer-reviewed manuscripts, ETDs, Books, Books Sections-Author's final peer-reviewed version	01/07/2008
Arts et Métiers Paris Tech	Required	Peer-reviewed manuscripts, Books, Book Sections	01/01/2013
CIRAD	Required	Peer-reviewed manuscripts, Books, Book Sections - Author's final peer-reviewed version	01/12/2008
CNRS	Requested	Peer-reviewed manuscripts, ETDs, Books, Book Sections	21/06/2006
Ecole des Ponts Paris Tech	Required	Peer-reviewed manuscripts, Other - Author's final peer-reviewed version	1/03/2017
INRA	Required	Peer-reviewed manuscripts, ETDs	1/1/2011
INRA Department of Animal Physiology and Livestock Systems	Required	Peer-reviewed manuscripts - Author's final peer-reviewed version	01/04/2004
INRIA (National Institute for Research in Computer Sciences & Control)	Required	Peer-reviewed manuscripts, ETDs, Books, Book Sections	27/04/2005
INSERM (Institut national de la santé et de la recherche médical)	Required	Peer-reviewed manuscripts	01/01/2008
IFSTTAR (Institut français des sciences et technologies des transports, de l'aménagement et des réseaux)	Required	Peer-reviewed manuscripts, ETDs, Books, Book Sections - Author's final peer-reviewed version	18/02/2013
Institut Jean Nicod	Required	Peer-reviewed manuscripts - Author's final peer-reviewed version	9/02/2004
Institut Pasteur	Required	Peer-reviewed manuscripts, Books, Book Sections - Author's final peer-reviewed version	23/09/2016
IFREMER (Institut français de recherche pour l'exploitation de la mer)	Required	Peer-reviewed manuscripts, Books, Book Sections - Author's final peer-reviewed version	7/09/2010
Laboratoire de psychologie et neurosciences cognitives	Required	Peer-reviewed manuscripts - Author's final peer-reviewed version	24/01/2007
Sciences Po	Required	Not specified	01/12/2015
Université d'Angers	Required	Peer-reviewed manuscripts	01/09/2014
Université de Lorraine	Required	ETDs	24/08/2006
École Normale Supérieure	Required	Peer-reviewed manuscripts	01/09/2016
Horizon 2020	Required	published version or final peer-reviewed manuscript accepted for publication	01/01/2014

TABLE 3.9 – Policies by research laboratory

IV. Robustness test

In this section, we provide some robustness tests for checking previous results. In section IV.I, we provide robustness checks for equation 3.10, *i.e.* the effect of the copyright reform in increasing the upload of papers into an open archive, while in section IV.II results refer to equation 3.11, which measures the impact of the copyright reform in reducing the time lag.

IV.I Copyright reform for increasing the number of uploaded post-print papers

IV.I.I One-year observations dropped

As previously described, we run multiple robustness tests, for both STM and SHS disciplines. Firstly, we assume a one-year temporal window, to allow scholars to comply with the law. In doing so, we decide to adopt a double approach. In Table 3.10, we report results for model 3.10, where we drop observations from the time the law comes into force until one year later.

VARIABLES	General	STM	SHS
β_1	-0.00753	-0.0158*	0.00150
	(0.00719)	(0.00812)	(0.0156)
Constant	-0.909***	-0.0487***	-0.669***
	(0.00821)	(0.00400)	(0.0116)
Observations	728,53	521,43	207,1
Control	YES	YES	YES
Pseudo R2	0.114	0.0653	0.0756

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.10 – Estimates of equation 3.10 for all disciplines, SHS and STM domains

VARIABLES	psy	scipo	philo	art	eco	droit	socio	histo	litt
β_1	-0.473***	-0.417***	-0.193**	0.246***	-0.248***	0.671***	0.138***	-0.0197	-0.0435
	(0.110)	(0.0918)	(0.0942)	(0.0880)	(0.0383)	(0.0661)	(0.0532)	(0.0427)	(0.0467)
Constant	-0.574***	-0.343***	-0.0492	-0.473***	-0.446***	-1.442***	-0.347***	-0.960***	-0.560***
	(0.0573)	(0.0725)	(0.0701)	(0.0996)	(0.0206)	(0.186)	(0.0396)	(0.0512)	(0.0517)
Observations	3,317	5,462	4,959	4,402	28,381	56,729	15,197	20,339	17,937
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R2	0.0597	0.0214	0.0309	0.0252	0.0532	0.0426	0.0144	0.0637	0.0141

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.11 – Estimates of equation 3.10 for SHS

VARIABLES	spi	sdv	sdu	sde	phys	scco	chim	info	mat
β_1	0.159***	0.0277*	0.0623*	-0.0319	0.107***	0.0541	0.0215	0.0869***	-0.346***
	(0.0233)	(0.0154)	(0.0328)	(0.0379)	(0.0215)	(0.0604)	(0.0227)	(0.0293)	(0.0326)
Constant	-0.443***	-0.125***	0.320***	-0.285***	0.301***	-0.740***	-1.077***	-0.0668***	0.329***
	(0.0116)	(0.00908)	(0.0138)	(0.0235)	(0.00897)	(0.0316)	(0.0137)	(0.0159)	(0.0166)
Observations	54,912	144,344	41,075	21,666	83,637	9,849	65,879	37,167	28,847
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R2	0.0458	0.119	0.0871	0.0951	0.0375	0.130	0.0897	0.0491	0.0172

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.12 – Estimates of equation 3.10 for STM

IV.I.II Entry into force of the law postponed by one year

For the second robustness test, we do not drop any observations. Instead, we consider that the effect of the law is postponed by one year. This is to take into account the fact that the change in scholars' behaviour may not be immediate.

VARIABLES	General	STM	SHS
β_1	-0.0130*	-0.0208***	-0.00478
	(0.00713)	(0.00804)	(0.0155)
Constant	-0.927***	-0.0569***	-0.663***
	(0.00792)	(0.00387)	(0.0112)
Observations	788,77	564,403	224,367
Control	YES	YES	YES
Pseudo R2	0.116	0.0677	0.0771

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.13 – Estimates of equation 3.10 for all disciplines, SHS and STM domains

VARIABLES	psy	scipo	philo	art	eco	droit	socio	histo	litt
β_1	-0.455***	-0.412***	-0.186**	0.219**	-0.247***	0.597***	0.135**	-0.0257	-0.0420
	(0.109)	(0.0904)	(0.0930)	(0.0868)	(0.0378)	(0.0651)	(0.0526)	(0.0424)	(0.0463)
Constant	-0.600***	-0.331***	-0.0532	-0.445***	-0.455***	-1.328***	-0.334***	-0.922***	-0.562***
	(0.0540)	(0.0697)	(0.0677)	(0.0954)	(0.0198)	(0.166)	(0.0384)	(0.0488)	(0.0495)
Observations	3,654	5,965	5,363	4,866	30,633	61,644	16,343	22,12	19,527
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R2	0.0622	0.0231	0.0294	0.0262	0.0522	0.0419	0.0142	0.0656	0.0147

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.14 – Estimates of equation 3.10 for SHS

VARIABLES	spi	sdv	sdu	sde	phys	scco	chim	info	mat
β_1	0.137***	0.0213	0.0509	-0.0465	0.0958***	0.0462	0.00697	0.0682**	-0.320***
	(0.0231)	(0.0153)	(0.0325)	(0.0374)	(0.0213)	(0.0599)	(0.0222)	(0.0291)	(0.0324)
Constant	-0.438***	-0.142***	0.322***	-0.279***	0.304***	-0.732***	-1.044***	-0.0624***	0.306***
	(0.0111)	(0.00872)	(0.0135)	(0.0226)	(0.00869)	(0.0303)	(0.0130)	(0.0153)	(0.0159)
Observations	59,897	159,573	43,458	23,761	89,552	10,78	71,755	40,317	31,206
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R2	0.0473	0.124	0.0852	0.0960	0.0377	0.135	0.0859	0.0498	0.0165

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.15 – Estimates of equation 3.10 for STM

IV.II Time lag

Even for the time lag, we check the robustness of results, using a one-year temporal window. As before, we drop observations in one case (IV.II.I), and for the second approach, we postpone the entry of the law into force by one year (IV.II.II).

IV.II.I One-year observations dropped

VARIABLES	General	STM	SHS
β_1	-3.944***	-3.731***	-4.915***
	(0.0107)	(0.0116)	(0.0259)
Constant	5.519***	5.163***	5.244***
	(0.0249)	(0.0130)	(0.0393)
Observations	202,963	161,652	41,311
R-squared	0.267	0.264	0.253
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.16 – OLS estimates of equation 3.11 for all disciplines, SHS and STM domains

VARIABLES	psy	scipo	philo	art	eco	droit	socio	histo	litt
β_1	-4.209***	-4.890***	-4.597***	-5.038***	-3.823***	-5.164***	-4.835***	-6.129***	-4.916***
	(0.150)	(0.131)	(0.128)	(0.141)	(0.0609)	(0.124)	(0.0684)	(0.0745)	(0.0700)
Constant	5.383***	5.716***	5.326***	4.686***	4.459***	6.232***	5.626***	6.342***	5.068***
	(0.187)	(0.184)	(0.189)	(0.276)	(0.0630)	(0.616)	(0.119)	(0.178)	(0.148)
Observations	1,1	1,765	1,804	1,375	6,408	1,614	5,417	5,866	5,745
R-squared	0.218	0.194	0.230	0.296	0.228	0.342	0.230	0.303	0.220
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.17 – OLS estimates of equation 3.11 for SHS

VARIABLES	spi	sdv	sdu	sde	phys	scco	chim	info	mat
β_1	-3.745***	-3.382***	-3.759***	-2.900***	-3.943***	-3.922***	-3.566***	-2.909***	-2.832***
	(0.0370)	(0.0228)	(0.0384)	(0.0574)	(0.0339)	(0.0944)	(0.0504)	(0.0327)	(0.0391)
Constant	4.849***	5.838***	5.013***	4.594***	4.468***	4.923***	4.216***	3.784***	3.521***
	(0.0416)	(0.0346)	(0.0372)	(0.0885)	(0.0318)	(0.124)	(0.0589)	(0.0419)	(0.0441)
Observations	17,31	45,196	15,889	5,621	22,408	2,382	9,88	15,42	10,878
R-squared	0.216	0.311	0.264	0.290	0.218	0.297	0.276	0.223	0.155
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.18 – OLS estimates of equation 3.11 for STM

IV.II.II Entry of the law into force after one year

VARIABLES	General	STM	SHS
β_1	-3.568***	-3.373***	-4.468***
	(0.0103)	(0.0112)	(0.0251)
Constant	5.268***	4.886***	4.953***
	(0.0240)	(0.0127)	(0.0379)
Observations	221,141	176,198	44,943
R-squared	0.232	0.231	0.214
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.19 – OLS estimates of equation 3.11 for all disciplines, SHS and STM domains

VARIABLES	psy	scipo	philo	art	eco	droit	socio	histo	litt
β_1	-3.797***	-4.489***	-4.153***	-4.464***	-3.471***	-4.421***	-4.442***	-5.609***	-4.487***
	(0.148)	(0.129)	(0.127)	(0.135)	(0.0586)	(0.116)	(0.0669)	(0.0740)	(0.0675)
Constant	4.977***	5.467***	5.115***	4.312***	4.273***	5.095***	5.349***	5.881***	4.721***
	(0.178)	(0.181)	(0.182)	(0.255)	(0.0613)	(0.549)	(0.115)	(0.170)	(0.140)
Observations	1,197	1,907	1,974	1,53	6,9	1,833	5,848	6,353	6,274
R-squared	0.189	0.168	0.202	0.244	0.199	0.255	0.197	0.257	0.182
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.20 – OLS estimates of equation 3.11 for SHS

VARIABLES	spi	sdv	sdu	sde	phys	scco	chim	info	mat
β_1	-3.319***	-2.987***	-3.489***	-2.558***	-3.570***	-3.521***	-2.898***	-2.622***	-2.624***
	(0.0344)	(0.0214)	(0.0380)	(0.0523)	(0.0319)	(0.0898)	(0.0434)	(0.0309)	(0.0377)
Constant	4.495***	5.514***	4.805***	4.211***	4.206***	4.592***	3.662***	3.521***	3.348***
	(0.0394)	(0.0334)	(0.0365)	(0.0822)	(0.0304)	(0.119)	(0.0527)	(0.0396)	(0.0422)
Observations	19,131	50,337	16,786	6,303	24,348	2,631	11,403	16,908	11,668
R-squared	0.181	0.277	0.242	0.247	0.186	0.256	0.200	0.196	0.137
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.21 – OLS estimates of equation 3.11 for STM

V. The difference-in-differences model

V.I Theoretical model

A possible criticism of our model is that there is a difference between the two periods t, depending on B and P. A possible solution is then to perform a difference-in-differences model, in order to compare the treated group with a control group, on which the law has no impact, even after the entry of the law into force. As described above, all the papers stored on HAL are affected by the law, so we propose to create a control group, starting with all the scientific documents not affected by the law (see Annex V.III for a detailed explanation of the identification of the treated and control groups).

As for the previous model, in an ideal world, where we have the possibility to observe the same notices but under two different regimes, we could estimate :

$$\begin{aligned}\Delta_{iB} = Y_{i1B} - Y_{i0B} &= \alpha_1 + \beta_B + \gamma_{1B} + \mu_{i1B} - (\alpha_0 + \beta_B + \gamma_{0B} + \mu_{i0B}) \\ &= (\alpha_1 - \alpha_0) + (\gamma_{1B} - \gamma_{0B}) + (\mu_{i1B} - \mu_{i0B})\end{aligned} \tag{3.12}$$

However, because of the difference in i, we can only estimate the average effect for heterogeneous unobservable characteristics or :

$$\Delta_{iB} = (\alpha_1 - \alpha_0) + (\gamma_{1B} - \gamma_{0B}) + (\frac{1}{n_{1B}} \sum \mu_{i1B} - \frac{1}{n_{0B}} \sum \mu_{i0B}) \tag{3.13}$$

As in the case of scientific articles, we create two different groups for the control group : group 3 (G3), which includes all scientific papers not affected by the law (hereafter WP for the sake of simplicity) created before the law came into force, and group 4 (G4), for all WP created after the law.

The outcome for a notice in G3 should be :

$$Y_i = Z_{i0B} = \alpha_0 + \beta_B + \gamma_{0B} + \mu_{i0B} \tag{3.14}$$

Because of the difference in i, we can estimate the average impact of the law :

$$\frac{1}{n_{0B}} \sum Y_i = \alpha_0 + \beta_B + \gamma_{0B} + \frac{1}{n_{0B}} \sum \mu_{i0B} \tag{3.15}$$

Specifically, for group 4 we obtain :

$$Y_i = Z_{i0P} = \alpha_0 + \beta_P + \gamma_{0P} + \mu_{i0P} \tag{3.16}$$

However, we can only estimate :

$$\frac{1}{n_{0P}} \sum Y_i = \alpha_0 + \beta_P + \gamma_{0P} + \frac{1}{n_{0P}} \sum \mu_{i0P} \tag{3.17}$$

Then, we can estimate the expected average impact of the copyright reform as :

$$\Delta = (\alpha_0 - \alpha_0) + (\beta_P - \beta_B) + (\gamma_{0P} - \gamma_{0B}) + (\frac{1}{n_{0P}} \sum \mu_{i0P} - \frac{1}{n_{0B}} \sum \mu_{i0B}) \tag{3.18}$$

The impact of the law can be estimated comparing the notices before and after the entry of the law into force if :

$$\begin{aligned} &(\alpha_0 - \alpha_0) + (\beta_P - \beta_B) + (\gamma_{0P} - \gamma_{0B}) + (\frac{1}{n_{0P}} \sum \mu_{i0P} - \frac{1}{n_{0B}} \sum \mu_{i0B}) \\ &= (\alpha_1 - \alpha_0) + (\gamma_{1B} - \gamma_{0B}) + (\frac{1}{n_{1B}} \sum \mu_{i1B} - \frac{1}{n_{0B}} \sum \mu_{i0B}) \end{aligned} \tag{3.19}$$

Which is satisfied under the following hypothesis :

- $\alpha_0 = \alpha_1$
- $(\beta_P = \beta_B)$
- $\gamma_{0P} - \gamma_{0B} = \gamma_{1B} - \gamma_{0B}$
- $\frac{1}{n_{0P}} \sum \mu_{i0P} = \frac{1}{n_{1B}} \sum \mu_{i1B}$

The first hypothesis requires that there is no change in the regime, which is the case, given that authors already had the possibility to upload WP even before the law came into force. The second hypothesis requires that we should not have a period effect. As was the case for papers, even for WP there were no major technological changes. Once again, only policy changes could concern editors or research centre policies, for which we control. The third hypothesis states that there would be no change in the outcome for WP, even after the law change at period P. As before, no major technological changes occurred during the two periods so scholars' behaviour should not have changed. Finally, the last hypothesis states that an unobserved determinant has on average the same outcomes under both regimes, independently of periods B and P. As we said, we cannot observe determinant

μ_{i1P}, because the law does not apply to WP. However, no particular event occurred at this time, apart from the copyright reform and we control for journal policy changes and research centre policy changes, even for WP. For these reasons, we assume that the notices that we observe at time P under the "open" regime are representative of those observed at time B under the hypothetical regime 1. We can then assume that $\frac{1}{n_{0B}} \sum \mu_{i0B} | i \in G3$ gives an unbiased estimate of $E(\frac{1}{n_{1P}} \sum \mu_{i1P} | i \in G4)$.

V.II Empirical model

We use a difference-in-differences model (DID) as a quasi-experimental setting for our analysis. DID measures policy changes that do not affect everybody at the same time and in the same way. Equation 3.20 reports our difference-in-differences logit model :

$$Y_i = \alpha + \beta_0 treated_i + \beta_1 time_t + \beta_2 treated_i * time_t + X_{it} + \epsilon_{it} \tag{3.20}$$

The dependent variableY_i defines two outcomes. Firstly, we are interested in the effect of the law in terms of the likelihood of uploading a paper. The first outcome of interest is then a dummy, whose value depends on the treated or the control group. For the treated group, $Y_i = 1$ if and only if the notice provides a copy of the document after its publication, while for the control group, $Y_i = 1$ if there is a paper.

The treated group ($treated_i = 1$) includes all the references eligible by the law to upload the document after its publication. Specifically, we consider all the notices for which a journal-identifier (ISSN or EISSN) is provided. The control group ($treated_i = 0$) includes all the other references for which the copyright reform has no impact, and specifically working papers, books and book chapters. These scientific communication instruments may differ among disciplines, but because we analyse the effect of the law for each discipline, this difference is taken into account. The variable $time_t$ is a dummy equal to one if the document has been published after the entry of the law into force. As for the first model, we consider that the law has no retroactive effect. For unprinted documents, the date of production refers to the presentation of the paper at a conference or to the date that the author has chosen as the date on which the first draft was produced. The variable $treated_i * time_t$, renamed in the following tables as did, is our variable of interest, which

captures the effect of the copyright reform. X_{it} is the same vector of control variables used for model 3.10.

The second outcome of interest is the time lag, expressed in years, which measures the impact of the law over the period that elapses between the publication of a scientific document and its upload into an open archive. The aim is to understand whether the law reduces the time lag, for those scholars who choose to comply with the law. For this reason, when we analyse $Y_i = timelag$ expressed in years, we consider only notices which provide access to a document in our database and, for the treated group, only if the document has been uploaded after publication. For measuring the impact of the law on the time lag, we use the following linear regression model :

$$Y_i = \alpha + \beta_0 treated_i + \beta_1 time_t + \beta_2 treated_i * time_t + X_{it} + \epsilon_{it} \tag{3.21}$$

Once again, the variable $treated_i * time_t$, renamed in the following tables as *did*, is our variable of interest, which captures the effect of the copyright reform for decreasing the time lag. X_{it} it is the same vector of control variables used for model 3.20.

V.III Control and treated groups : justification

In this section, we explain the reasons for our choice of composition of the control and treated groups. We must first describe the composition and the nature of documents stored on HAL. We stress that it is the responsibility of the author (or the person responsible for creating the notice) to choose the type of document to which the notice refers. It can happen that authors misclassify their intellectual work putting an unpublished article in the "article" group, for example, and not in the correct class, which would be "working papers". This is why we cannot use the classification proposed by HAL directly, and compare articles on the one hand, as the treated group, versus working papers, on the other, as the control group. In this section, we explain how we create our treated group, *i.e.* how we identify scientific works concerned by the French copyright reform, and how we create the control group.

HAL provides access to four different document groups : publications, unpublished works, academic works and research data. Hereinafter is the composition and description

of each group.

Publications :

a) Articles (ART) : intellectual works published in a scientific review (peer-reviewed or not). Sometimes the notice does not provide the journal's HAL-code, an identification number assigned by HAL, nor the id-code (EISSN for online journals or ISSN for printed journals).

b) Communications at scientific congresses (COMM & PRESCONF) : works published in special issues, after a conference/workshop etc. Sometimes the notice does not provide the journal's HAL-code, an identification number assigned by HAL, nor the id-code (EISSN for online journals or ISSN for printed journals).

c) Book chapters (COUV) : articles published in a collective work, such as a book.

d) Books (OUV)

e) Direction d'ouvrages (DOUV) : proceedings, introduction of special issues managed by the authors

f) POSTER : poster of scientific papers, published or not in scientific reviews

g) Patents (PATENT)

h) Summaries (SYNTHESE) : Reports, executive summaries and slide presentations, mostly related to public administration projects.

i) Other (OTHER) : other works, published or not in a journal. Here we can find documents that do not belong to the above-mentioned groups, such as book reviews, but also works that are included in other classes, such as working papers. HAL splits this class into 4 subgroups :

   - Articles published in non-scientific journals or summaries of workshops and conferences ;
   - Articles from blogs ;
   - Book reviews and commentaries ;
   - Contributions to encyclopaedias and dictionaries ;
   - Article translations.

Unpublished works :

a) Working papers (UNDEFINED) : Pre-print and working papers ;

b) Reports (REPORT) : Report sponsored by the Ministry of Research, but also scientific reports – and then scientific standard works – published in journals ;

c) Other reports (OTHERREPORT) : Research and technical reports, laboratory scientific reports, not published in journals ;

d) Activity reports (REPACT) : project reports or laboratory reports, about their annual or multiannual research activity.

Academic works :

a) Doctoral dissertations (THESE)

b) Habilitation à diriger des recherches (HDR) : thesis dissertation required to become a PhD supervisor.

c) Bachelor and Master dissertation (MEM)

d) Class lecture (LECTURE)

Research data :

a) Images (IMG)

b) Videos (VIDEO)

c) Map (MAP)

d) Software (SOFTWARE)

e) Sounds (SON)

As we have said, it is up to each researcher to decide to which category his work belongs. As we can see, some works that are classified as unpublished, have in fact been published in a journal.

For this reason, we cannot create our treated group from the HAL classification, but we have to verify for each category whether or not the work was published in a periodical. In doing this, we decided to create our treated group as follows :

1. Articles : we considered only articles for which a journal-id is provided (ISSN, EISSN). In this way, we are sure that the work has indeed been published in a journal.
2. Communications at scientific congresses : not all the papers presented at a conference are published. Sometimes special issues of periodicals are created to highlight the most relevant paper in a specific domain. In this class, HAL assigns a journal-id

even if the scientific document is not published in a journal, but merely presented in conferences. For this reason, we considered only documents for which the ISSN or the EISSN are provided.

3. Books, book chapters : because these works are not published in journals, they are not concerned by the copyright reform. They are therefore excluded from our treated group. Due to mistakes made by the person responsible for creating the notice, we can find documents in this category that are not books, but articles. Once again, if there is a journal-id, we include these notices in our treated group. Because HAL assigns a HAL-id to the journal, even if the source is a book, we do not take into account notices which only provide the HAL-id.
4. Direction d'ouvrages : These works may be published in a book or as a special issue in a periodical journal. For this reason, once again, we take into account only documents for which a journal-id is provided (ISSN, EISSN). Because HAL assigns a HAL-id to the journal, even if the source is a book, we do not take into account notices which only provide the HAL-id.
5. Poster : even if a small minority of the posters stored on HAL are published (0.65%), we decided not to consider this class. In fact, in scientific communications, posters are not comparable to articles. They are a preliminary step for the drafting of the article. We therefore do not include posters in our database.
6. Patent : patents are excluded from our sample, due to the specificity of this category. A patent is never published in its original version (*i.e.* as a patent application for obtaining IP protection). Access to the patent application is subject to a statutory delay, which depends on the national/international law from which protection is sought. When patents are published in journals, they become articles, which are included in our first category. Otherwise, they are only administrative works, which do not fall into our analysis of tools for academic communications.
7. Summaries : because these documents are not published in journals, they are excluded from our treated group (none of the observations has a journal-id).
8. Other : in this case, due to the heterogeneity of the resources, we take into account only notices which provide the journal-id (ISSN, EISSN) and not the HAL-id. In

fact, in this class there are different kinds of scientific work, and HAL assigns a journal-id even if the scientific document is not published in a journal, but merely presented at conferences.

9. Working paper and reports : these are works that are listed in unpublished works, under the working paper or report categories, but they are published in a journal. Once again, if the notice provides the journal-id (except for the HAL journal-id), we consider these works in our treated group.
10. Other reports : they are not published in journals, so we exclude them from our treated group.
11. Activity reports : they do not refer to scientific production, because they are administrative documents, so we exclude them from our dataset.
12. Doctoral dissertations, bachelor and masters' dissertations and Habilitation à diriger des recherches : these documents are unpublished, so they are not included in our treated group.
13. Lecture : lectures are excluded from our sample, because they are not inherent to scientific communication.
14. Research data : research data are not publishable per se, so we exclude them from our sample.

As for the treated group, the choice of documents used to create our control group also has to be discussed and analysed, due to inaccuracies in the classification proposed by HAL. Our control group brings together all the references which are not concerned by the copyright reform. Hereafter the list of documents which constitute our control group.

1. Articles : we assume that if an article does not provide the journal-id it is because it has not yet been published.
2. Communication at scientific conferences : scholars usually present working papers at conferences, in order to improve their work. If the notice does not provide the journal-id, or if there is only the journal-id provided by HAL, we consider these works as working papers, so unpublished articles.
3. Book chapters : they are similar to articles, but published in books rather than journals. They are therefore not concerned by the copyright reform and they are

included in the control group.

4. Direction d'ouvrages : as we said, these works may be published in a book or as a special issue in a periodical journal. We took into account only documents for which there is no journal-id, meaning that they are published in books
5. Books : books are not concerned by the copyright reform. In order to avoid mistakes made by authors, we exclude books which do not provide the ISBN. We could have decided not to consider books at all, but we believe that, especially for some disciplines, books are valuable intangible assets for a scholar's reputation, the same as articles.
6. Poster : as we said before, we do not include posters in our database
7. Summaries : we avoid considering the summary category, due to its heterogeneity and its lack of links with standard scientific communications.
8. Other : in this category, there is a risk of finding both working papers, thesis/dissertations and other non-scientific documents. We analyse the description of the notice, and we only take into account notices which refer to working papers. More specifically :
   a. Articles published in commercial journals or summaries of workshops and conferences : they are not considered in our control group, because they do not represent a standard form of scientific communication
   b. Articles from blogs : they are not considered in our control group, because they do not represent a standard form of scientific communication
   c. Book reviews and commentaries : we keep them in our control group, because they represent a form of scientific communication.
   d. Contribution to encyclopaedias and dictionaries : they are not considered in our control group, because they do not represent a standard form of scientific communication
   e. Article translations : we keep them in our control group, because they represent a form of scientific communication. The same article can be published in a second journal in another language.
9. Working papers : they are listed in the control group if they do not provide a journal-id.

10. Reports : if the notice does not provide the journal-id (except for the HAL journal-id), we consider these works in our control group.
11. Other reports : in this category, there is a risk of finding both working papers, thesis/dissertations and other non-scientific documents. We analyse the description of the notice, and only take into account notices which refer to working papers.
12. Bachelor and masters' dissertations : we do not consider these documents in our sample, because we consider them to be a non-conventional means of knowledge dissemination. The work is either published or stored in HAL as a working paper.
13. Doctoral dissertation : the problem with doctoral theses is that, even if they are unpublished, some of the chapters could have been published in a journal. If this is the case, there will be an indirect effect on the upload of the PhD dissertation on HAL, due to the copyright reform, even if the thesis has not been published. For these reasons, we prefer to exclude PhD dissertations from our sample.
14. Habilitation à diriger des recherches : the format of this document does not respect scientific communication standards. In fact, more than one section of these documents is preceded by the author's CV and his academic career. We assume that if the aim of the author was to disseminate only the research sections, he would have uploaded the HDR as a working paper. For this reason, we do not consider HDRs in our sample.

HAL Code	Type of document	Treated	Control
ART	Articles published in scientific journals	YES, if the notice provides the journal-id	YES, if there is no journal-id
COMM & PRESCONF	Document referring to workshop, conferences etc.	YES, if the notice provides the journal-id (not HAL-id)	YES, if there is no journal-id or only the HAL-id
COUV	Book chapters	YES, if the notice provides the journal-id (not HAL-id)	YES, if there is no journal-id or only the HAL-id
OUV	Books	YES, if the notice provides the journal-id (not HAL-id)	YES, but we exclude books for which only the HAL-id is provided
DOUV	Proceedings, introduction of journals' special issues	YES, if the notice provides the journal-id (not HAL-id)	YES, if there is no journal-id or only the HAL-id
POSTER	Poster	YES, if the notice provides the journal-id (not HAL-id)	YES, if there is no journal-id
PATENT	Patent	NOT	NOT
SYNTHESE	Reports and executive summaries related to research for the public administration	NOT	NOT
OTHER	Different documents, which may or may not belong to the other categories	YES, if there is the journal-id (not HAL-id)	YES, if there is no journal-id (or only the HAL-id) and if they belong to class 2 and 4.
UNDEFINED	Working papers	YES, if there is the journal-id (not HAL-id)	YES, if there is no journal-id (or only the HAL-id)
REPORT	Reports sponsored by the Ministry of Research, working papers and published articles	YES, if there is the journal-id (not HAL-id)	YES, if there is no journal-id (or only the HAL-id)
OTHERREPORT	Working papers and other unpublished works	NOT	YES, if there is no journal-id (or only the HAL-id) and if they are working papers
REPACT	Administrative papers	NOT	NOT
THESE	PhD dissertations	NOT	NOT
HDR	Dissertations for becoming a PhD supervisor	NOT	NOT
MEM	Bachelor and Master thesis	NOT	NOT
LECTURE	Lectures	NOT	NOT
IMG	Images	NOT	NOT
VIDEO	Video	NOT	NOT
MAP	Maps	NOT	NOT
SOFTWARE	Software	NOT	NOT
SON	Sounds	NOT	NOT

TABLE 3.22 – Type of notice for the treated and control group.

VI. Results and discussion

After cleaning data, and testing for the common trend assumption, the database for the difference-in-differences model is based on 1 207 566 observations. In section VI.I, we provide results of equation 3.20, or the effect of the law reform on increasing the upload of post-print papers, while section VI.II shows results of the copyright reform for reducing the time lag.

VI.I Copyright reform for increasing the number of uploaded post-print papers

The tables below show estimates for equation 3.20, *i.e.* how the copyright reform (*did* variable) affects scholars' behaviour in providing a copy of their published works. Table 3.23 reports estimates for all disciplines (General). The estimate coefficient of did shows that, all other things being equal, the policy change has a negative impact on uploading a copy of the paper. However, due to discipline differences that we have already underlined, we perform the same regression for each sub-scientific domain.

VARIABLES	General	STM	SHS
treated	0.486***	0.429***	0.708***
	(0.00542)	(0.00629)	(0.0111)
time	0.325***	0.466***	0.0300***
	(0.00611)	(0.00775)	(0.0112)
did	-0.450***	-0.591***	-0.139***
	(0.00874)	(0.0106)	(0.0164)
Constant	-1.506***	-0.372***	-1.500***
	(0.00467)	(0.00446)	(0.00561)
Observations	1,207,566	718,251	489,315
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.23 – Estimates of equation 3.20 for all disciplines, SHS and STM domains

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	0.995***	1.014***	1.385***	0.842***	0.185***	-1.070***	0.941***	0.975***	0.834***
	(0.0676)	(0.0519)	(0.0505)	(0.0594)	(0.0266)	(0.0675)	(0.0484)	(0.0299)	(0.0377)
time	-0.124	-0.192***	0.242***	-0.0979*	0.0285	0.184***	0.174***	-0.215***	-0.0215
	(0.116)	(0.0637)	(0.0587)	(0.0547)	(0.0266)	(0.0509)	(0.0432)	(0.0323)	(0.0328)
did	-0.260*	-0.150*	-0.434***	0.285***	-0.251***	0.0896	-0.128**	0.200***	0.0415
	(0.139)	(0.0896)	(0.0907)	(0.0861)	(0.0415)	(0.0757)	(0.0629)	(0.0463)	(0.0513)
Constant	-1.485***	-1.260***	-1.420***	-1.429***	-0.613***	-2.398***	-1.290***	-1.784***	-1.526***
	(0.0517)	(0.0235)	(0.0252)	(0.0229)	(0.0155)	(0.0335)	(0.0284)	(0.0142)	(0.0203)
Observations	6,962	19,855	19,32	21,768	50,849	61,706	21,674	79,433	43,171
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.24 – Estimates of equation 3.20 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Info	Math
treated	0.348***	1.462***	1.233***	0.983***	0.306***	1.164***	-0.231***	-0.601***
	(0.0188)	(0.0234)	(0.0400)	(0.0127)	(0.0516)	(0.0388)	(0.0182)	(0.0294)
time	0.166***	0.399***	0.0686	1.083***	0.190***	0.328***	0.280***	0.724***
	(0.0190)	(0.0313)	(0.0498)	(0.0177)	(0.0617)	(0.0523)	(0.0150)	(0.0359)
did	-0.0751***	-0.443***	-0.0111	-1.051***	-0.277***	-0.259***	-0.314***	-0.978***
	(0.0288)	(0.0346)	(0.0590)	(0.0247)	(0.0820)	(0.0564)	(0.0299)	(0.0465)
Constant	-0.780***	-1.586***	-0.970***	-0.674***	-0.900***	-2.060***	0.375***	0.950***
	(0.0122)	(0.0228)	(0.0340)	(0.00942)	(0.0400)	(0.0352)	(0.00847)	(0.0221)
Observations	92,219	128,869	27,311	161,996	12,814	58,514	134,786	40,805
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.25 – Estimates of equation 3.20 for STM

The negative effect of the law that we found for some SHS disciplines is confirmed. This is the case for economics, psychology, philosophy and political sciences. However, for psychology, the effect is less strong when we perform the robustness tests (see Table 3.30 in Annex VII.I.I and Table 33 in Annex VI.I.II). As already mentioned, in our control group we include books and book chapters. It is possible that in some disciplines these instruments may not be considered as a standard way to communicate scientific results. When we drop books and book chapters from our observations (see Tables 3.42, 3.45 and 3.48 in Annex VII.III), estimates for psychology become non-significant. Once again, these results suggest strong inter-disciplinary differences in scientific knowledge dissemination practices. The non-effect of the law is confirmed for the literature discipline, where results do not change even when we perform the robustness checks provided in Tables 3.30 and 3.33 in Annex VII.I and when we perform the very same robustness check once observations referring to books and books chapters are dropped (Tables 3.42, 3.45 and 3.48 in Annex VII.III).

In the first model we propose, the effect of the law on sociology was very small once we introduce the robustness check, reducing the odds of finding a paper from 12% to 5% for notices affected by the law. This very small change in scholars' behaviour then calls the effectiveness of the law into question. When we use the DID model, the general result that emerges is the non-effect of the law, (see Table 3.30 and 3.33 in Annex VII.I for general robustness check and Tables 3.42, 3.45 and 3.48 in Annex VII.III, referring to the dataset without books and book chapters).

The results for law and history are ambiguous and they change when we test our

hypothesis using the different-in-differences model. The positive effect of the copyright law on the law discipline disappears, while estimates for history become significant. In the first case, it may be that the law coming into force has also encouraged the upload of working papers, thus making the difference between our control group and treated group before and after the copyright reform not significant. For history, the opposite results may be the consequence of the sample variation. In fact, for the first model, we have a sample of about 20,000 observations, against more than 75,000 for the second model. This is a very preliminary study, and the doubtful effect of the law for some disciplines opens the path to new research. In particular, it would be interesting to find disciplinary repositories not affected by the copyright reform for these two scientific fields and compare them to data from HAL.

Deeper insights are also required for the STM disciplines. The negative impact of the law is confirmed for mathematics and non-significant estimates are validated for astrophysics, as reported in Table 3.25, and in the tables for robustness checks (see Annex VII.I, Tables 3.31 and 3.34 for variation in the entry into force of the law and Tables 3.43, 3.46 and 3.49 in Annex VII.III for results referring to the sample without books and book chapters). The inefficacy of the law that we found in chemistry, cognitive and life sciences is even reinforced when using the difference-in-differences models, where a negative effect emerges.

However, as was the case for some SHS disciplines, even in the STM domain the positive effect of the law that we found using model 3.10 disappears. This is the case for engineering, physics and informatics. While it is true that previous results do not demonstrate a complete change in scholars' practices in using HAL, results from difference-in-differences require further investigation. A possible solution is to use ArXive as a control group, to understand whether the copyright reform has, at least for these three disciplines, changed scholars' attitudes towards Green Open Access.

VI.II The copyright reform for reducing the time lag

The second aim of the law is to reduce the time lag between the publication date of a scientific document and its upload into an open archive. We reduce our sample to observations that provide access to a scientific document and, for the treated group, we consider only notices that provide access to a post-print paper.

VARIABLES	General	STM	SHS
treated	1.100***	1.161***	0.849***
	(0.0140)	(0.0149)	(0.0350)
time	-2.361***	-2.058***	-3.168***
	(0.00893)	(0.00956)	(0.0192)
did	-0.771***	-0.788***	-1.041***
	(0.0140)	(0.0147)	(0.0322)
Constant	3.980***	3.033***	4.154***
	(0.0134)	(0.00937)	(0.0185)
Observations	343,711	250,715	92,996
R-squared	0.291	0.292	0.257
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.26 – OLS Estimates of equation 3.21 for all disciplines, SHS and STM

Results show a general decrease in the time lag, compared to the pre-copyright reform period. Estimates reported in Table 26 explain that, all other things being equal, the policy change reduces the time lag for post-print published documents by more than 8 months. The difference is greater for SHS, where the reported upload decreases by one year, confirming results from the model 3.11.

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	1.285***	0.940***	0.687***	0.423**	0.835***	1.176***	0.357***	1.548***	0.259***
	(0.223)	(0.167)	(0.145)	(0.189)	(0.0521)	(0.120)	(0.0622)	(0.107)	(0.0617)
time	-3.146***	-3.540***	-3.671***	-4.135***	-1.423***	-1.881***	-1.504***	-4.485***	-1.894***
	(0.174)	(0.0991)	(0.0830)	(0.0909)	(0.0261)	(0.0735)	(0.0442)	(0.0561)	(0.0387)
did	-1.009***	-1.195***	-0.757***	-0.639***	-0.791***	-1.311***	-0.416***	-1.369***	-0.425***
	(0.219)	(0.151)	(0.138)	(0.162)	(0.0525)	(0.112)	(0.0639)	(0.0894)	(0.0579)
Constant	4.104***	4.621***	4.684***	4.949***	2.111***	2.901***	2.392***	5.658***	2.776***
	(0.156)	(0.0782)	(0.0797)	(0.0829)	(0.0243)	(0.0731)	(0.0417)	(0.0505)	(0.0351)
Observations	1,795	4,776	4,816	4,694	15,077	3,385	6,097	14,669	9,371
R-squared	0.267	0.195	0.249	0.232	0.262	0.392	0.351	0.285	0.324
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.27 – OLS Estimates of equation 3.21 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Math	Info
treated	0.384***	0.0873*	0.334***	0.787***	0.825***	-0.296***	0.727***	-0.156***
	(0.0299)	(0.0455)	(0.0623)	(0.0457)	(0.0985)	(0.0957)	(0.0340)	(0.0253)
time	-1.377***	-1.580***	-1.191***	-2.933***	-1.620***	-1.532***	-0.815***	-1.851***
	(0.0199)	(0.0459)	(0.0584)	(0.0397)	(0.0698)	(0.0973)	(0.0219)	(0.0109)
did	-0.276***	-0.0346	-0.162**	-0.830***	-0.778***	-0.181*	-0.525***	0.299***
	(0.0316)	(0.0481)	(0.0663)	(0.0494)	(0.0971)	(0.102)	(0.0355)	(0.0261)
Constant	2.095***	2.538***	1.773***	3.683***	2.362***	2.679***	1.340***	2.751***
	(0.0188)	(0.0436)	(0.0539)	(0.0336)	(0.0682)	(0.0900)	(0.0206)	(0.0114)
Observations	30,654	38,238	8,111	37,918	3,636	10,481	21,221	76,351
R-squared	0.227	0.303	0.250	0.229	0.330	0.316	0.209	0.244
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.28 – OLS Estimates of equation 3.21 for STM

Once again, the impact of the law differs among disciplines, showing a greater impact on social and human sciences. Results for life sciences and informatics are ambiguous too, with an increase in the time lag, but only of about two months and one month respectively. Results are confirmed even when we introduce the effect of postponing the law (see Annex VII.II Tables 3.37 and 3.40) and when we exclude books and book chapters from our sample (see Annex VII.IV Tables 3.52, 3.55 and 3.58). The law has a strong impact on SHS disciplines, with the decrease in the time lag ranging from almost eight months to one and a half years. Compared to the results from model 3.11, coefficients of the difference-in-differences model show a decrease in the time lag which is still significant although smaller. Even for the STM disciplines, the effect of the law is still negative, meaning that the law reduces the time lag between the publication of the paper and its upload into HAL. However, as in the case of SHS, the effect is less important, ranging from almost one year for the cognitive sciences to two months for engineering. The results for mathematics are controversial, as the copyright reform seems to have increased the time lag, whereas in the model 3.11, we found a negative effect. This result may be an indirect effect of the law on working papers, which, since the copyright reform, have been uploaded more rapidly than previously.

Once again, we want to stress the fact that further research needs to look into the effect of the copyright reform on scholars' behaviour for a better understanding of the extent to which the legal framework is a barrier to the development of the Green Open Access model. Our results show in general that the law has not changed scholars' attitude

towards Open Access substantially, and that there are other barriers limiting the use of this model. However, for those scholars who are already familiar with self-archiving, the law has reduced the time lag, suggesting that a possible explanation for the underuse of the model is a lack of awareness of the Green model and of the importance in general of alternative forms of scientific knowledge dissemination, other than traditional scientific journals.

VII. Robustness test

In this section, we provide different robustness tests for models 3.20 and 3.21. Section VII.I provides results of robustness tests for equation 3.20, namely the effect of the law in increasing the number of post-print papers, while section VII.II refers to model 3.21, or the effect of the law for decreasing the time lag. Moreover, in sections VII.III and VII.IV we provide robustness tests for models 3.20 and 3.21 respectively, once books and book chapters are eliminated from the control group.

VII.I Copyright reform for increasing the number of uploaded post-print papers

As for equation 3.10, in this section we propose the same robustness checks for model 3.10, namely the elimination of the one-year observation (VII.I.I) and the postponed entry into force of the law (VII.I.II).

VII.I.I One-year observations dropped

VARIABLES	General	STM	SHS
treated	0.493***	0.432***	0.711***
	(0.00548)	(0.00633)	(0.0113)
time	0.374***	0.525***	0.0486***
	(0.00715)	(0.00909)	(0.0133)
did	-0.523***	-0.669***	-0.204***
	(0.0101)	(0.0122)	(0.0192)
Constant	-1.494***	-0.371***	-1.500***
	(0.00476)	(0.00449)	(0.00563)
Observations	1,089,307	643,485	445,822
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.29 – Estimates of equation 3.20 for all disciplines, SHS and STM domains

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	0.994***	1.017***	1.399***	0.838***	0.191***	-1.141***	0.935***	0.979***	0.852***
	(0.0680)	(0.0529)	(0.0514)	(0.0608)	(0.0269)	(0.0684)	(0.0496)	(0.0304)	(0.0388)
time	-0.187	-0.262***	0.266***	-0.0692	0.0742**	0.165***	0.151***	-0.266***	-0.0465
	(0.143)	(0.0803)	(0.0732)	(0.0678)	(0.0316)	(0.0590)	(0.0512)	(0.0396)	(0.0396)
did	-0.297*	-0.246**	-0.562***	0.288***	-0.408***	0.0947	-0.0949	0.236***	0.0373
	(0.168)	(0.110)	(0.110)	(0.103)	(0.0477)	(0.0885)	(0.0726)	(0.0550)	(0.0603)
Constant	-1.481***	-1.260***	-1.417***	-1.429***	-0.612***	-2.417***	-1.296***	-1.784***	-1.523***
	(0.0517)	(0.0235)	(0.0253)	(0.0230)	(0.0156)	(0.0344)	(0.0287)	(0.0142)	(0.0205)
Observations	6,384	18,551	18,128	20,131	45,286	54,473	18,657	73,687	37,786
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.30 – Estimates of equation 3.20 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Info	Math
treated	0.345***	1.464***	1.229***	0.982***	0.304***	1.152***	-0.228***	-0.602***
	(0.0189)	(0.0235)	(0.0401)	(0.0128)	(0.0518)	(0.0391)	(0.0183)	(0.0295)
time	0.140***	0.555***	0.0463	1.202***	0.258***	0.374***	0.324***	0.816***
	(0.0221)	(0.0358)	(0.0563)	(0.0207)	(0.0706)	(0.0583)	(0.0180)	(0.0423)
did	-0.0542	-0.598***	-0.0121	-1.185***	-0.361***	-0.336***	-0.407***	-1.112***
	(0.0330)	(0.0395)	(0.0666)	(0.0286)	(0.0939)	(0.0630)	(0.0344)	(0.0533)
Constant	-0.774***	-1.580***	-0.969***	-0.672***	-0.894***	-2.062***	0.384***	0.950***
	(0.0123)	(0.0232)	(0.0342)	(0.00942)	(0.0404)	(0.0353)	(0.00856)	(0.0222)
Observations	80,35	109,848	23,746	150,042	11,286	51,018	121,461	35,564
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.31 – Estimates of equation 3.20 for STM

VI.I.II Entry of the law into force postponed by one year

VARIABLES	General	STM	SHS
treated	0.449***	0.374***	0.707***
	(0.00514)	(0.00592)	(0.0107)
time	0.345***	0.479***	0.0494***
	(0.00706)	(0.00896)	(0.0132)
did	-0.488***	-0.616***	-0.203***
	(0.00992)	(0.0120)	(0.0190)
Constant	-1.480***	-0.326***	-1.501***
	(0.00451)	(0.00419)	(0.00537)
Observations	1,207,566	718,251	489,315
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.32 – Estimates of equation 3.20 for all disciplines, SHS and STM domains

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	0.972***	1.014***	1.367***	0.871***	0.197***	-1.056***	0.905***	0.988***	0.843***
	(0.0648)	(0.0505)	(0.0492)	(0.0576)	(0.0250)	(0.0637)	(0.0450)	(0.0292)	(0.0358)
time	-0.185	-0.260***	0.247***	-0.0658	0.0839***	0.127**	0.111**	-0.256***	-0.0429
	(0.142)	(0.0797)	(0.0730)	(0.0674)	(0.0310)	(0.0576)	(0.0496)	(0.0394)	(0.0389)
did	-0.277*	-0.242**	-0.542***	0.263***	-0.419***	0.0795	-0.0564	0.223***	0.0262
	(0.166)	(0.110)	(0.110)	(0.102)	(0.0466)	(0.0861)	(0.0701)	(0.0547)	(0.0588)
Constant	-1.490***	-1.266***	-1.407***	-1.438***	-0.622***	-2.365***	-1.250***	-1.793***	-1.525***
	(0.0496)	(0.0229)	(0.0245)	(0.0223)	(0.0146)	(0.0315)	(0.0258)	(0.0137)	(0.0189)
Observations	6,962	19,855	19,32	21,768	50,849	61,706	21,674	79,433	43,171
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.33 – Estimates of equation 3.20 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Info	Math
treated	0.330***	1.431***	1.229***	0.900***	0.290***	1.147***	-0.244***	-0.728***
	(0.0171)	(0.0210)	(0.0358)	(0.0121)	(0.0478)	(0.0351)	(0.0171)	(0.0270)
time	0.104***	0.534***	0.0215	1.110***	0.248***	0.330***	0.301***	0.716***
	(0.0216)	(0.0345)	(0.0541)	(0.0204)	(0.0692)	(0.0562)	(0.0179)	(0.0414)
did	-0.0413	-0.571***	-0.00806	-1.105***	-0.345***	-0.321***	-0.394***	-0.986***
	(0.0320)	(0.0382)	(0.0641)	(0.0283)	(0.0921)	(0.0607)	(0.0339)	(0.0519)
Constant	-0.744***	-1.563***	-0.946***	-0.586***	-0.891***	-2.018***	0.399***	1.054***
	(0.0112)	(0.0210)	(0.0304)	(0.00882)	(0.0372)	(0.0316)	(0.00807)	(0.0202)
Observations	92,219	128,869	27,311	161,996	12,814	58,514	134,786	40,805
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.34 – Estimates of equation 3.20 for STM

VII.II Copyright reform for decreasing the time lag

VII.II.I One-year observations dropped

VARIABLES	General	STM	SHS
treated	1.111***	1.167***	0.834***
	(0.0142)	(0.0151)	(0.0361)
time	-2.504***	-2.200***	-3.362***
	(0.00930)	(0.00991)	(0.0199)
did	-0.768***	-0.779***	-1.111***
	(0.0144)	(0.0150)	(0.0332)
Constant	4.038***	3.065***	4.162***
	(0.0141)	(0.00948)	(0.0186)
Observations	307,519	222,822	84,697
R-squared	0.274	0.277	0.234
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.35 – OLS Estimates of equation 3.21 for all disciplines, SHS and STM

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	1.255***	0.893***	0.690***	0.395**	0.825***	1.131***	0.341***	1.574***	0.224***
	(0.227)	(0.173)	(0.148)	(0.197)	(0.0532)	(0.125)	(0.0634)	(0.112)	(0.0641)
time	-3.333***	-3.734***	-3.896***	-4.396***	-1.544***	-2.093***	-1.715***	-4.805***	-2.093***
	(0.175)	(0.108)	(0.0894)	(0.0971)	(0.0259)	(0.0746)	(0.0439)	(0.0600)	(0.0392)
did	-0.928***	-1.219***	-0.759***	-0.665***	-0.847***	-1.353***	-0.456***	-1.429***	-0.469***
	(0.224)	(0.161)	(0.147)	(0.172)	(0.0531)	(0.113)	(0.0631)	(0.0933)	(0.0584)
Constant	4.111***	4.624***	4.703***	4.950***	2.111***	2.916***	2.388***	5.663***	2.763***
	(0.157)	(0.0785)	(0.0809)	(0.0835)	(0.0245)	(0.0747)	(0.0423)	(0.0509)	(0.0356)
Observations	1,655	4,483	4,459	4,338	13,49	2,92	5,188	13,631	8,158
R-squared	0.232	0.166	0.213	0.202	0.251	0.386	0.360	0.256	0.312
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.36 – OLS Estimates of equation 3.21 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Math	Info
treated	0.377***	0.0719	0.333***	0.781***	0.822***	-0.319***	0.720***	-0.154***
	(0.0301)	(0.0458)	(0.0626)	(0.0461)	(0.101)	(0.0960)	(0.0343)	(0.0256)
time	-1.541***	-1.740***	-1.265***	-3.102***	-1.753***	-1.862***	-0.980***	-1.943***
	(0.0201)	(0.0456)	(0.0585)	(0.0415)	(0.0709)	(0.0951)	(0.0217)	(0.0113)
did	-0.269***	-0.0238	-0.201***	-0.830***	-0.840***	-0.00998	-0.455***	0.286***
	(0.0315)	(0.0478)	(0.0663)	(0.0519)	(0.0976)	(0.0997)	(0.0352)	(0.0265)
Constant	2.104***	2.553***	1.772***	3.683***	2.379***	2.671***	1.326***	2.769***
	(0.0189)	(0.0439)	(0.0541)	(0.0337)	(0.0693)	(0.0901)	(0.0207)	(0.0117)
Observations	26,391	32,603	6,919	34,795	3,212	8,787	18,331	68,399
R-squared	0.226	0.308	0.255	0.207	0.321	0.337	0.226	0.222
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.37 – OLS Estimates of equation 3.21 for STM

VII.II.II Entry of the law into force postponed by one year

VARIABLES	Gen	STM	SHS
treated	1.036***	1.094***	0.742***
	(0.0131)	(0.0139)	(0.0337)
time	-2.223***	-1.933***	-3.055***
	(0.00861)	(0.00906)	(0.0188)
did	-0.639***	-0.664***	-0.929***
	(0.0133)	(0.0137)	(0.0313)
Constant	3.781***	2.798***	3.896***
	(0.0132)	(0.00855)	(0.0175)
Observations	343,711	250,715	92,996
R-squared	0.233	0.231	0.197
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.38 – OLS Estimates of equation 3.21 for all disciplines, SHS and STM

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	1.119***	0.889***	0.617***	0.266	0.759***	0.804***	0.298***	1.482***	0.178***
	(0.215)	(0.166)	(0.141)	(0.185)	(0.0491)	(0.112)	(0.0570)	(0.107)	(0.0592)
time	-3.089***	-3.456***	-3.620***	-4.078***	-1.367***	-1.802***	-1.458***	-4.459***	-1.841***
	(0.169)	(0.106)	(0.0888)	(0.0931)	(0.0237)	(0.0669)	(0.0390)	(0.0582)	(0.0365)
did	-0.756***	-1.081***	-0.588***	-0.430***	-0.711***	-1.072***	-0.400***	-1.250***	-0.367***
	(0.218)	(0.158)	(0.143)	(0.163)	(0.0486)	(0.103)	(0.0565)	(0.0902)	(0.0543)
Constant	3.870***	4.477***	4.505***	4.712***	1.958***	2.611***	2.136***	5.396***	2.523***
	(0.149)	(0.0762)	(0.0781)	(0.0802)	(0.0223)	(0.0656)	(0.0365)	(0.0493)	(0.0322)
Observations	1,795	4,776	4,816	4,694	15,077	3,385	6,097	14,669	9,371
R-squared	0.201	0.145	0.187	0.170	0.212	0.295	0.283	0.219	0.243
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.39 – OLS Estimates of equation 3.21 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Math	Info
treated	0.329***	0.0864**	0.337***	0.753***	0.717***	-0.454***	0.662***	-0.145***
	(0.0259)	(0.0390)	(0.0523)	(0.0434)	(0.0907)	(0.0749)	(0.0297)	(0.0228)
time	-1.338***	-1.459***	-1.013***	-2.778***	-1.554***	-1.657***	-0.868***	-1.728***
	(0.0177)	(0.0399)	(0.0503)	(0.0375)	(0.0646)	(0.0762)	(0.0180)	(0.0105)
did	-0.192***	-0.00767	-0.203***	-0.774***	-0.697***	0.193**	-0.382***	0.271***
	(0.0275)	(0.0417)	(0.0570)	(0.0476)	(0.0889)	(0.0800)	(0.0310)	(0.0242)
Constant	1.895***	2.278***	1.526***	3.451***	2.191***	2.466***	1.223***	2.542***
	(0.0163)	(0.0374)	(0.0449)	(0.0315)	(0.0622)	(0.0696)	(0.0166)	(0.0109)
Observations	30,654	38,238	8,111	37,918	3,636	10,481	21,221	76,351
R-squared	0.178	0.238	0.190	0.175	0.266	0.248	0.188	0.185
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.40 – OLS Estimates of equation 3.21 for STM

VII.III Copyright reform for increasing the number of uploaded post-print papers- No books

A possible remark is that in our sample, we include books and book chapters in the control group. In fact, these instruments may not be considered in some disciplines as a standard way of communicating scientific results. For this reason, we decided to run equation 3.20 reformulating the control group and excluding books and book chapters from our observations.

VARIABLES	General	STM	SHS
treated	0.404***	0.366***	0.713***
	(0.00575)	(0.00644)	(0.0129)
time	0.366***	0.484***	0.0366***
	(0.00667)	(0.00803)	(0.0133)
did	-0.487***	-0.616***	-0.133***
	(0.00910)	(0.0108)	(0.0179)
Constant	-1.377***	-0.327***	-1.355***
	(0.00524)	(0.00457)	(0.00658)
Observations	1,067,335	696,865	370,47
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.41 – Estimates of equation 3.20 for all disciplines, SHS and STM domains

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	0.803***	0.991***	1.368***	0.906***	-0.120***	0.996***	1.199***	0.991***	1.131***
	(0.0778)	(0.0650)	(0.0657)	(0.0839)	(0.0292)	(0.184)	(0.0630)	(0.0395)	(0.0522)
time	-0.166	-0.160**	0.250***	-0.144**	0.0296	0.340***	0.193***	-0.332***	-0.0389
	(0.131)	(0.0784)	(0.0774)	(0.0677)	(0.0291)	(0.0680)	(0.0562)	(0.0432)	(0.0436)
did	-0.205	-0.200**	-0.389***	0.359***	-0.235***	-0.0343	-0.151**	0.311***	0.0515
	(0.151)	(0.101)	(0.103)	(0.0944)	(0.0426)	(0.0881)	(0.0722)	(0.0542)	(0.0585)
Constant	-1.348***	-1.138***	-1.269***	-1.188***	-0.382***	-2.432***	-1.309***	-1.639***	-1.478***
	(0.0578)	(0.0282)	(0.0305)	(0.0267)	(0.0168)	(0.0427)	(0.0362)	(0.0176)	(0.0255)
Observations	6,046	14,174	12,943	14,872	44,455	52,813	16,011	53,089	29,139
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.42 – Estimates of equation 3.20 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Info	Math
treated	0.292***	1.358***	1.184***	0.964***	0.180***	1.122***	-0.323***	-0.770***
	(0.0192)	(0.0246)	(0.0415)	(0.0129)	(0.0562)	(0.0405)	(0.0185)	(0.0310)
time	0.168***	0.366***	0.0605	1.109***	0.213***	0.339***	0.285***	0.802***
	(0.0194)	(0.0334)	(0.0517)	(0.0181)	(0.0679)	(0.0547)	(0.0156)	(0.0399)
did	-0.0845***	-0.409***	-0.00330	-1.081***	-0.296***	-0.272***	-0.318***	-1.067***
	(0.0291)	(0.0365)	(0.0607)	(0.0249)	(0.0867)	(0.0585)	(0.0299)	(0.0493)
Constant	-0.750***	-1.500***	-0.926***	-0.660***	-0.799***	-2.026***	0.430***	1.078***
	(0.0124)	(0.0238)	(0.0352)	(0.00956)	(0.0439)	(0.0365)	(0.00871)	(0.0238)
Observations	89,606	124,699	26,564	159,586	11,706	57,306	128,995	39,155
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.43 – Estimates of equation 3.20 for STM

VII.III.I One-year observations dropped

VARIABLES	General	STM	SHS
treated	0.414***	0.372***	0.705***
	(0.00581)	(0.00649)	(0.0132)
time	0.428***	0.547***	0.0914***
	(0.00783)	(0.00944)	(0.0158)
did	-0.570***	-0.695***	-0.231***
	(0.0105)	(0.0124)	(0.0208)
Constant	-1.362***	-0.327***	-1.355***
	(0.00535)	(0.00460)	(0.00660)
Observations	961,579	624,193	337,386
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.44 – Estimates of equation 3.20 for all disciplines, SHS and STM domains

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	0.799***	0.975***	1.379***	0.904***	-0.113***	0.933***	1.199***	0.967***	1.137***
	(0.0786)	(0.0667)	(0.0672)	(0.0859)	(0.0297)	(0.206)	(0.0651)	(0.0404)	(0.0543)
time	-0.228	-0.139	0.328***	-0.0701	0.0929***	0.328***	0.177***	-0.371***	-0.00588
	(0.159)	(0.0991)	(0.0961)	(0.0830)	(0.0347)	(0.0779)	(0.0666)	(0.0528)	(0.0518)
did	-0.243	-0.379***	-0.561***	0.323***	-0.397***	-0.0111	-0.132	0.334***	-0.0118
	(0.180)	(0.123)	(0.125)	(0.112)	(0.0489)	(0.102)	(0.0833)	(0.0646)	(0.0682)
Constant	-1.347***	-1.138***	-1.272***	-1.185***	-0.382***	-2.425***	-1.313***	-1.640***	-1.483***
	(0.0579)	(0.0282)	(0.0307)	(0.0267)	(0.0169)	(0.0427)	(0.0363)	(0.0176)	(0.0256)
Observations	5,539	13,218	12,143	13,769	39,568	46,639	13,715	49,2	25,466
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.45 – Estimates of equation 3.20 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Info	Math
treated	0.294***	1.364***	1.180***	0.963***	0.168***	1.108***	-0.316***	-0.771***
	(0.0193)	(0.0248)	(0.0416)	(0.0130)	(0.0566)	(0.0408)	(0.0187)	(0.0312)
time	0.137***	0.533***	0.0237	1.230***	0.310***	0.381***	0.337***	0.908***
	(0.0226)	(0.0383)	(0.0588)	(0.0211)	(0.0773)	(0.0611)	(0.0189)	(0.0472)
did	-0.0578*	-0.574***	0.00960	-1.217***	-0.405***	-0.344***	-0.418***	-1.212***
	(0.0333)	(0.0418)	(0.0687)	(0.0288)	(0.0989)	(0.0656)	(0.0345)	(0.0567)
Constant	-0.746***	-1.497***	-0.925***	-0.658***	-0.790***	-2.027***	0.438***	1.076***
	(0.0125)	(0.0242)	(0.0353)	(0.00956)	(0.0442)	(0.0365)	(0.00881)	(0.0239)
Observations	78,054	106,264	23,076	147,778	10,302	49,955	116,139	34,144
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.46 – Estimates of equation 3.20 for STM

VII.III.I Entry of the law into force postponed by one year

VARIABLES	General	STM	SHS
treated	0.365***	0.308***	0.720***
	(0.00547)	(0.00607)	(0.0126)
time	0.397***	0.500***	0.0983***
	(0.00773)	(0.00930)	(0.0157)
did	-0.533***	-0.641***	-0.237***
	(0.0103)	(0.0122)	(0.0206)
Constant	-1.348***	-0.280***	-1.362***
	(0.00507)	(0.00429)	(0.00629)
Observations	1,067,335	696,865	370,47
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.47 – Estimates of equation 3.20 for all disciplines, SHS and STM domains

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	0.783***	0.997***	1.362***	0.941***	-0.103***	0.993***	1.159***	1.013***	1.157***
	(0.0747)	(0.0640)	(0.0647)	(0.0829)	(0.0276)	(0.184)	(0.0596)	(0.0388)	(0.0504)
time	-0.222	-0.136	0.317***	-0.0683	0.107***	0.262***	0.133**	-0.351***	0.0129
	(0.158)	(0.0989)	(0.0958)	(0.0827)	(0.0341)	(0.0758)	(0.0644)	(0.0526)	(0.0508)
did	-0.228	-0.379***	-0.548***	0.290***	-0.412***	-0.000277	-0.0909	0.311***	-0.0380
	(0.178)	(0.123)	(0.125)	(0.111)	(0.0479)	(0.0991)	(0.0804)	(0.0643)	(0.0666)
Constant	-1.356***	-1.150***	-1.262***	-1.204***	-0.396***	-2.372***	-1.266***	-1.658***	-1.494***
	(0.0556)	(0.0275)	(0.0297)	(0.0259)	(0.0158)	(0.0385)	(0.0323)	(0.0170)	(0.0237)
Observations	6,046	14,174	12,943	14,872	44,455	52,813	16,011	53,089	29,139
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.48 – Estimates of equation 3.20 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Info	Math
treated	0.270***	1.336***	1.176***	0.876***	0.171***	1.099***	-0.334***	-0.907***
	(0.0175)	(0.0223)	(0.0371)	(0.0123)	(0.0523)	(0.0366)	(0.0174)	(0.0286)
time	0.0991***	0.521***	-0.00559	1.135***	0.310***	0.334***	0.314***	0.799***
	(0.0220)	(0.0370)	(0.0564)	(0.0208)	(0.0757)	(0.0589)	(0.0187)	(0.0463)
did	-0.0414	-0.556***	0.0179	-1.134***	-0.399***	-0.325***	-0.406***	-1.075***
	(0.0324)	(0.0404)	(0.0661)	(0.0285)	(0.0969)	(0.0632)	(0.0339)	(0.0554)
Constant	-0.712***	-1.487***	-0.899***	-0.568***	-0.798***	-1.980***	0.453***	1.189***
	(0.0113)	(0.0219)	(0.0313)	(0.00895)	(0.0408)	(0.0327)	(0.00830)	(0.0218)
Observations	89,606	124,699	26,564	159,586	11,706	57,306	128,995	39,155
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.49 – Estimates of equation 3.20 for STM

VII.IV Copyright reform for decreasing the time lag – No books Even for the time lag, we check the robustness of results, using a one-year temporal window. As before, we drop observations in one case (VII.IV.I), and for the second approach, we postpone the entry of the law into force by one year (VI.IV.II).

VARIABLES	General	STM	SHS
treated	1.195***	1.182***	0.820***
	(0.0141)	(0.0150)	(0.0422)
time	-2.251***	-2.059***	-2.950***
	(0.00915)	(0.00964)	(0.0222)
did	-0.885***	-0.780***	-1.273***
	(0.0141)	(0.0148)	(0.0343)
Constant	3.810***	3.025***	3.953***
	(0.0143)	(0.00945)	(0.0211)
Observations	323,85	246,745	77,105
R-squared	0.294	0.294	0.268
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.50 – OLS Estimates of equation 3.21 for all disciplines, SHS and STM

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	1.776***	1.191***	0.840***	0.340	0.916***	1.188***	0.785***	0.734***	0.654***
	(0.231)	(0.196)	(0.190)	(0.251)	(0.0531)	(0.257)	(0.0837)	(0.168)	(0.0944)
time	-2.644***	-3.246***	-3.445***	-3.856***	-1.381***	-1.891***	-1.227***	-4.601***	-1.554***
	(0.176)	(0.118)	(0.104)	(0.106)	(0.0265)	(0.0989)	(0.0574)	(0.0732)	(0.0483)
did	-1.513***	-1.485***	-0.990***	-0.920***	-0.826***	-1.307***	-0.694***	-1.276***	-0.769***
	(0.220)	(0.164)	(0.152)	(0.172)	(0.0527)	(0.130)	(0.0732)	(0.101)	(0.0645)
Constant	3.651***	4.509***	4.361***	4.658***	2.064***	2.752***	2.091***	5.672***	2.420***
	(0.158)	(0.0947)	(0.0944)	(0.0932)	(0.0246)	(0.0927)	(0.0521)	(0.0632)	(0.0430)
Observations	1,662	3,847	3,758	3,766	14,418	2,525	4,735	11,32	6,861
R-squared	0.285	0.201	0.249	0.238	0.267	0.441	0.361	0.298	0.348
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.51 – OLS Estimates of equation 3.21 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Math	Info
treated	0.410***	0.196***	0.358***	0.836***	1.034***	-0.294***	0.729***	-0.167***
	(0.0300)	(0.0470)	(0.0636)	(0.0455)	(0.104)	(0.0989)	(0.0342)	(0.0256)
time	-1.362***	-1.510***	-1.189***	-2.876***	-1.465***	-1.502***	-0.803***	-1.855***
	(0.0200)	(0.0478)	(0.0598)	(0.0396)	(0.0735)	(0.101)	(0.0221)	(0.0111)
did	-0.287***	-0.103**	-0.165**	-0.878***	-0.930***	-0.211**	-0.538***	0.305***
	(0.0317)	(0.0499)	(0.0675)	(0.0495)	(0.0999)	(0.106)	(0.0357)	(0.0261)
Constant	2.078***	2.436***	1.752***	3.644***	2.204***	2.675***	1.336***	2.752***
	(0.0189)	(0.0452)	(0.0553)	(0.0336)	(0.0730)	(0.0933)	(0.0208)	(0.0116)
Observations	30,219	37,783	8,003	37,529	3,366	10,374	20,694	74,693
R-squared	0.226	0.302	0.252	0.228	0.344	0.316	0.214	0.244
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.52 – OLS Estimates of equation 3.21 for STM

VII.IV.I One-year observations dropped

VARIABLES	General	STM	SHS
treated	1.200***	1.191***	0.774***
	(0.0144)	(0.0152)	(0.0442)
time	-2.388***	-2.200***	-3.104***
	(0.00950)	(0.00998)	(0.0229)
did	-0.888***	-0.771***	-1.382***
	(0.0145)	(0.0151)	(0.0354)
Constant	3.870***	3.056***	3.965***
	(0.0151)	(0.00957)	(0.0211)
Observations	289,635	219,305	70,33
R-squared	0.277	0.278	0.245
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.53 – OLS Estimates of equation 3.21 for all disciplines, SHS and STM

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	1.741***	1.161***	0.803***	0.248	0.910***	1.006***	0.790***	0.666***	0.605***
	(0.237)	(0.203)	(0.198)	(0.269)	(0.0543)	(0.311)	(0.0874)	(0.180)	(0.102)
time	-2.788***	-3.414***	-3.624***	-4.096***	-1.504***	-2.036***	-1.413***	-4.894***	-1.718***
	(0.175)	(0.130)	(0.111)	(0.113)	(0.0262)	(0.0995)	(0.0567)	(0.0789)	(0.0484)
did	-1.467***	-1.523***	-1.035***	-0.964***	-0.880***	-1.403***	-0.760***	-1.376***	-0.847***
	(0.224)	(0.176)	(0.160)	(0.183)	(0.0532)	(0.131)	(0.0722)	(0.108)	(0.0648)
Constant	3.663***	4.521***	4.378***	4.666***	2.065***	2.769***	2.090***	5.673***	2.420***
	(0.158)	(0.0948)	(0.0950)	(0.0933)	(0.0248)	(0.0930)	(0.0523)	(0.0633)	(0.0432)
Observations	1,534	3,611	3,488	3,491	12,899	2,17	4,017	10,541	5,984
R-squared	0.252	0.172	0.215	0.212	0.257	0.431	0.372	0.269	0.338
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.54 – OLS Estimates of equation 3.21 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Math	Info
treated	0.403***	0.181***	0.357***	0.830***	1.029***	-0.317***	0.721***	-0.167***
	(0.0302)	(0.0473)	(0.0639)	(0.0458)	(0.107)	(0.0992)	(0.0345)	(0.0260)
time	-1.525***	-1.658***	-1.268***	-3.048***	-1.575***	-1.842***	-0.970***	-1.945***
	(0.0202)	(0.0475)	(0.0597)	(0.0413)	(0.0745)	(0.0987)	(0.0218)	(0.0114)
did	-0.281***	-0.105**	-0.198***	-0.874***	-1.015***	-0.0294	-0.466***	0.292***
	(0.0316)	(0.0496)	(0.0674)	(0.0518)	(0.100)	(0.103)	(0.0352)	(0.0266)
Constant	2.087***	2.451***	1.751***	3.646***	2.228***	2.670***	1.323***	2.771***
	(0.0190)	(0.0455)	(0.0554)	(0.0336)	(0.0742)	(0.0933)	(0.0209)	(0.0118)
Observations	26,005	32,225	6,817	34,43	2,978	8,694	17,891	66,901
R-squared	0.226	0.307	0.257	0.206	0.337	0.337	0.231	0.222
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.55 – OLS Estimates of equation 3.21 for STM

VII.IV.II Entry of the law into force postponed by one year

VARIABLES	General	STM	SHS
treated	1.123***	1.121***	0.665***
	(0.0133)	(0.0140)	(0.0415)
time	-2.117***	-1.934***	-2.832***
	(0.00879)	(0.00912)	(0.0217)
did	-0.748***	-0.657***	-1.164***
	(0.0134)	(0.0138)	(0.0334)
Constant	3.625***	2.788***	3.717***
	(0.0141)	(0.00862)	(0.0199)
Observations	323,85	246,745	77,105
R-squared	0.235	0.233	0.207
Control	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.56 – OLS Estimates of equation 3.21 for all disciplines, SHS and STM

VARIABLES	Psycho	Po_Sc	Philo	Arts	Eco	Law	Socio	Hist	Lit
treated	1.521***	1.048***	0.744***	0.0682	0.847***	0.652**	0.742***	0.454***	0.497***
	(0.224)	(0.198)	(0.190)	(0.249)	(0.0504)	(0.270)	(0.0787)	(0.173)	(0.0927)
time	-2.598***	-3.187***	-3.409***	-3.858***	-1.331***	-1.713***	-1.182***	-4.590***	-1.519***
	(0.169)	(0.129)	(0.110)	(0.109)	(0.0240)	(0.0890)	(0.0495)	(0.0775)	(0.0446)
did	-1.245***	-1.331***	-0.809***	-0.651***	-0.743***	-1.152***	-0.678***	-1.155***	-0.693***
	(0.218)	(0.174)	(0.157)	(0.175)	(0.0486)	(0.118)	(0.0638)	(0.105)	(0.0600)
Constant	3.474***	4.380***	4.203***	4.467***	1.913***	2.455***	1.864***	5.423***	2.214***
	(0.150)	(0.0919)	(0.0917)	(0.0898)	(0.0225)	(0.0815)	(0.0442)	(0.0614)	(0.0388)
Observations	1,662	3,847	3,758	3,766	14,418	2,525	4,735	11,32	6,861
R-squared	0.220	0.151	0.187	0.180	0.218	0.328	0.296	0.231	0.268
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.57 – OLS Estimates of equation 3.21 for SHS

VARIABLES	Spi	Sdv	Sdu	Phys	Ssco	Chem	Math	Info
treated	0.359***	0.180***	0.368***	0.804***	0.894***	-0.455***	0.662***	-0.155***
	(0.0260)	(0.0405)	(0.0531)	(0.0431)	(0.0966)	(0.0773)	(0.0300)	(0.0232)
time	-1.324***	-1.396***	-1.009***	-2.727***	-1.396***	-1.643***	-0.861***	-1.729***
	(0.0178)	(0.0417)	(0.0510)	(0.0373)	(0.0676)	(0.0790)	(0.0180)	(0.0106)
did	-0.203***	-0.0695	-0.208***	-0.816***	-0.853***	0.180**	-0.389***	0.275***
	(0.0275)	(0.0435)	(0.0576)	(0.0476)	(0.0912)	(0.0827)	(0.0309)	(0.0242)
Constant	1.878***	2.191***	1.499***	3.414***	2.058***	2.466***	1.223***	2.543***
	(0.0164)	(0.0389)	(0.0457)	(0.0314)	(0.0666)	(0.0719)	(0.0168)	(0.0110)
Observations	30,219	37,783	8,003	37,529	3,366	10,374	20,694	74,693
R-squared	0.178	0.237	0.192	0.174	0.281	0.247	0.193	0.185
Control	YES	YES	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 3.58 – OLS Estimates of equation 3.21 for STM

Chapitre 4

The star effect of self-archiving practice : how journal rank affects green Open Access choices

Abstract

The purpose of green Open Access is to eliminate financial barriers to access to scientific literature. Despite the greater access that this instrument does ensure, use of the green model is still underdeveloped. In this paper, we question the influence of journal rank on scholars' attitudes to adopting the green model. In order to test our hypothesis, we focused our attention on the economics discipline and used data from HAL, the largest archive in the French scientific landscape. The ranks of both national and international journals were used to enrich our dataset.

Given the utility functions of the various scholars, our hypothesis is that green Open Access is not only a complementary tool for ensuring access knowledge, but also a substitute one, depending on the knowledge dissemination strategy adopted by the researchers. The results suggest that, compared to their non-ranked counterparts, top-ranked papers are more likely to be uploaded into HAL after publication. Only by understanding the reasons that lead researchers to adopt this green strategy can a new bundle of incentives be defined to promote the green route in the scientific community.

Introduction

Since the middle of the 20th century, peer recognition has become a valuable intangible asset for scholars. The "publish or perish" imperative pushes researchers to disseminate their scientific results through scientific journals and the more recent "impact or perish" paradigm (Biagioli, 2016 ; Biagioli and Lippman, 2020) explains the central role of journal rankings in scholars' strategies for publishing their results. In fact, with the professionalization of the scientific community, scholars' careers and prestige have come to depend on the quantitative and qualitative evaluation of their intellectual production (Heckman and Moktan 2020).

Researchers' influence in the advancement of science is measured by citations – defined as the peer recognition of the new knowledge value produced – which are a strong predictor of their salaries and prestige in the scientific community (Hilmer et al., 2012). The evaluation of scholars' productivity is based nowadays on bibliometrics, a discipline that evaluates not only the quantity of scientific production, but also its quality. While counting

the number of written papers and books is an easy task, qualitative evaluation is trickier. The main rule that guides the bibliometric approach is that "a publication is all the more important the more it is perceived" (Ball, 2017, p. 2). This perception is quantified using citations : the more a publication is cited, the more important it is for the scientific community. Citations are thus a proxy of the quality of the new knowledge produced. Even though the variety of indicators of this perception is overwhelming, they are all based on the assumption that citations matter, and that the frequency by which a paper is cited determines its relevance.

Bibliometrics has found a promising field of study with the arrival of the Open Access (OA) model. OA is a knowledge dissemination model that aims to remove all legal and financial barriers to access to scientific results (Suber, 2012). While the traditional "Closed Access" (CA) model imposes a price for accessing scientific results, Open Access eliminates the price wall, granting consumers free access. The question of whether Open Access can increase the relevance of scientific results has been investigated in depth in the literature. The positive effect of free access to scientific literature on citations has been widely documented by the research community (Lawrence, 2001 ; Harnad and Brody, 2004 ; Eysenbach, 2006 ; Tang et al., 2017 ; Patel et al., 2019). However, this citation advantage may not be uniform, affecting top-ranked and low-ranked Open Access papers differently. In this sense, some studies find a skewed citation advantage for top-ranked papers (Gargouri et al., 2010 ; McCabe and Snyder, 2014), while for other authors the citation advantage is stronger for the low-ranked ones (Xia and Nakanishi, 2012).

Publishing in Open Access journals, or Gold Open Access, is not the only way to ensure free access to scientific outputs. Green Open Access (GOA) is another solution and consists in self-archiving pre- or post-published papers in an open archive, a digital environment conceived for article storage. It is not a publishing model, because it does not benefit from a peer-review process that checks article quality, but it does eliminate the price wall for accessing articles published under the closed access model.

Use of this model may be influenced by the quality of the article [1] a scholar wants to

1. In this paper, we refer to science quality from the "inside" usage (Mazlish, 1982), *i.e.* as measured by the scientific community though different indices. See annex 1 for a detailed discussion about the quality of the knowledge.

upload into an open archive. The green model could be the only way of ensuring access to lower- or non-ranked papers which would never be accessible otherwise, due to library budget constraints (Shavell, 2010). In this sense, access to scientific knowledge could be granted by the green model, making GOA a substitute tool for disseminating and accessing new scientific results. By contrast, the literature explains that authors who publish papers in top-ranked journals use green Open Access as a marketing tool for highlighting the importance of their publication, generating the so-called trophy effect (Wren, 2005). In this second case, the green model can be seen as a complementary tool to closed access, increasing scholars' awareness of the latest scientific results. Because a scholar's career is based upon peer evaluation, as measured by citations, it is in their interest to make their scientific output freely accessible to consumers, in order to open up access to all and thus increase the probability of being cited. Nevertheless, this publishing model is still less preferred than the traditional one. 80% of the knowledge produced still remains behind a paywall (Green, 2018 ; Fortson, 2019) and only a minority of the scientific community self-archives its works in repositories (Creaser, 2010 ; Creaser et al., 2010 ; Hendler, 2007 ; Kim, 2007).

Understanding the reasons that push researchers to adopt the green model (or not) is the first step towards solving the research paradox. It is a fact that the scientific community is calling for greater access to scientific literature, but most researchers are not using the Open Access public devices intended to help them (such as GOA) and which are at their disposal. Ultimately, study of scholars' behaviour towards the green choice may help convince them to adopt the OA model.

Different causes have been found by the literature to explain underuse of the green model, such as copyright concerns (Cullen and Chawner, 2011 ; Rowley *et al.*, 2017 ; Creaser, 2010), and engagement in Open Access development in different disciplines (Spezi et al., 2013 ; Cullen and Chawner, 2011 ; Eger *et al.*, 2015 ; Rodriguez, 2014). This paper aims to contribute to this literature by analysing another possible factor that may influence scholars' adoption of the green Open Access. We argue that, because the quality of the paper largely determines ease of access to it, it may also create different strategies for adopting self-archiving practices and thus influence scholars in their choice to adopt green Open

Access to enlarge their audience.

We examine how journal rank affects the likelihood of a paper being uploaded, compared to non-ranked papers. Are high- and low- or non-ranked articles more likely to be uploaded into open archives, showing that journal quality may influence scholars' behaviour, even in their green Open Access choice ? If this is actually the case, we can identify two different uses of the green route. For top-ranked papers, the green model becomes a complementary tool to the traditional publishing system, enabling authors to highlight their best works, and using this model as an advertising tool. For lower-ranked or non-ranked papers, meanwhile, green Open Access is a substitute model for granting consumers access to their works. In fact, due to library budget constraints, the green route may even be the only way of accessing these scientific results.

Using the web scraping method, we downloaded data from HAL, the largest open archive in the French scientific landscape (Schöpfel *et al.*, 2016), to test whether the journal quality of published articles influences development of the green route in the economics discipline. As far as we know, it is the first time that HAL has been used for studying scholars' attitudes towards green Open Access. Journal quality was analysed using different national and international rankings, which help scholars to guide their consumption choices. Journals are not the only distributors of scientific knowledge, but their brand is a signal of the quality of articles they publish, thus guiding the choices of consumers (Bomsel, 2010, 2014), especially through the impact factor (IF) [2] of the journal, the most common index for measuring journal quality.

We first present the different interests that have been identified as inducing researchers to adopt green Open Access in the literature on Open Access quality concerns (4.2). We then build a theoretical model for explaining the different scholars' utility functions in adopting the green strategy (4.3). In section 4.4, we present data from the French open archive HAL. The logistic model studying the impact of journal quality on scholars' behaviour in adopting the green strategy is presented in section 4.5, followed by a discussion of

2. The impact factor, which is the ratio of the number of citations received at time t by articles published during the preceding two years over the number of articles published in the journal during the same period, is one of the measure of journal quality. Although IF is not the best instrument for measuring journal and article quality, "[. . .] there is nothing better [. . .] and is therefore a good technique for scientific evaluation" (Hoeffel, 1998).

our results (4.6) and policy implications (4.7).

4.1 Literature review

4.1.1 Citation advantage and the Open Access choice

The scientific community's call for greater access to scientific knowledge is not new. OA grants researchers-consumers access to all scientific results, while for researchers-producers OA can increase access to their scientific results and thus the probability of being cited. In fact, once the budget constraint on access has been eliminated, only the quality of the article will determine whether the new knowledge that is produced is reused and its importance in the cumulative process of advancing science. As Jaffe et al. (2002, p. 100) stated, "the citations function is the product of the usefulness of old ideas and the probability of having seen them". The Open Access model thus increases the likelihood of seeing new ideas, because it allows free access and eliminates all financial barriers.

The bibliometric discipline has inquired deeply into the impact of the Open Access model on the probability of being cited. In his pioneering work, Lawrence (2001) found that, among 199,924 articles analysed in computer science and related disciplines, free online articles were cited more on average. A huge literature has developed around the question of whether Open Access, in both green and gold formats, might increase the citations received by a paper and, in general, evidence suggests that there is a positive effect. Hajjem et al. (2006) replicated the Lawrence study in physics and found the very same result. Using a multidisciplinary approach, Antelman (2004) confirmed the greater research impact of Open Access articles and in 2006, Eysenbach found that Open Access papers were twice as likely to be published as non-Open Access ones published in the same journal [3].

The effect of the Open Access choice on received citations has been called into question by some authors, however. For example, Davis et al. (2008) did not find any citation

3. For an exhaustive literature review of the Open Access citation advantage, see The Open Citation Project for articles published until 2013 (available at http ://opcit.eprints.org/oacitation-biblio.html. Last access 25/01/20201). For an update on Open Access citations impact until 2015, see The Open Access Citation Advantage Service, provided by SPARC Europe (available at `https://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/`. Last access 25/01/2021).

advantage in the first year after publication, suggesting that other causes may explain the citation advantage found previously by the literature. However, the authors did not propose any explanations for the positive impact of Open Access highlighted by the literature.

The literature then turned its attention to the quality of the paper, suggesting that papers published in top-ranked and low-ranked journals may benefit from Open Access in different ways. Specifically, the focus has been on how journal quality contributes to receiving more citations when a paper is freely accessible. Gargouri *et al.* (2010) found that the Open Access choice boosted more citable articles, because of their quality advantage. Using a sample of 27,197 articles, they found that articles that have been uploaded into an open archive after their publication in a traditional journal receive more citations. But they also found that the Open Access advantage was skewed in favour of articles that had already been highly cited, even before adopting the green Open Access strategy. A positive impact of OA on top-ranked journals was confirmed by McCabe and Snyder (2014). They confirmed the general increase in citations due to an Open Access strategy, by 8% on average. Astronomy papers archived in arXiv were cited almost twice as often as non-archived papers, but papers published in Nature and Science appeared to be cited 5 times more than non-Open Access ones (Metcalfe, 2005). The author suggested that the extra citations benefit may result from the "brand recognition" of these high-ranked journals, which gives, once again, a stronger Open Access citation advantage to top-ranked papers.

This quality effect crossed with the Open Access strategy was studied by Xia and Nakanishi (2012) too. Focusing on the anthropology literature, they highlighted that, although in general Open Access articles (Green or Gold) receive more citations, lower-ranked journals benefit from a greater increase in their rate of citations when they are freely accessible. Similar results were confirmed by Harnad and Brody (2004). Using a sample of 14 million articles, they suggested that despite the fact that access to a scientific article is a necessary condition for increasing the number of received citations, it is not sufficient. They added that Open Access increases the number of potential users, and in particular, those who cannot afford access to the journal in which papers are published, due to their institution's budget constraints.

4.1.2 Journal quality and the Open Access choice

Because scholars' careers depend strongly on peer recognition, as expressed through citations, it is in their interest to adopt one of the Open Access strategies. As we have seen, the citation function is based on the probability of having access to the scientific knowledge, but also on its quality. The usefulness of a paper is determined in the first place by the journal, which is a sign of the quality that it publishes, a signal that guides the consumer choice (Bomsel, 2010). This signal attributed by the journal reduces uncertainty about the quality of the new knowledge, thereby increasing its value. The importance of a paper thus strongly depends on the journal in which the article is published.

Scholars have two main options for granting consumers free access to their scientific results, and thus increasing the probability of being cited. The first strategy is to publish in a journal that provides free access to the scientific literature. Gold Open Access journals are journals that only publish Open Access articles, while hybrid journals follow the traditional model, meaning closed access, providing scholars with the choice of publishing their articles in Open Access or not. Gold journals usually do not have high quality, due to their novelty, and are rarely listed in the institutional journal rankings that are used as the main tool for evaluating researchers' careers. For example, when we look at the CNRS (Centre National de la Recherche Scientifique) [4] ranking, only 16 out of 989 listed journals are categorised as Open Access. For these reasons, publishing in a gold Open Access journal could be a risky choice for an individual, due to the weak quality signal of the journal.

A less risky choice to make a paper freely accessible is to adopt the Green model. It consists in self-archiving papers that have been published previously in a closed access journal (or not) in an open repository, where everyone can have access to them without paying any fees. Because researchers incur no potential reputational loss, given the previous publication in a well-established traditional journal, all scholars could be expected to adopt the green model to broaden access to their works.

Two different types of behaviour can therefore be expected, depending on the motivations that push scholars to adopt the green route. Studying the impact of APC (article

4. The CNRS is the French national centre for scientific research, which aims to promote the production of the scientific knowledge. The board is composed of scientists and scholars.

processing charges) on publication choices, Shavell (2010) underlined that high APC would discourage authors of poor quality from adopting the Gold model, because of the price they have to pay. However, poor-quality articles "could and probably would be posted on an Open Access repository". Access to scientific literature is mostly provided by university libraries and their subscriptions to lower-quality journals may be dropped due to budget constraints, as they prefer to ensure access to top-quality journals instead. In this way, green Open Access increases access to lower-ranked articles, eliminating the economic barriers that prevent the scientific community from reading the article. For lower-ranked papers, green Open Access is a tool that eliminates financial barriers, allowing access to scientific results that are usually not included in libraries' catalogues. Choosing green thus becomes a substitute tool for allowing access to this scientific knowledge, given the budget constraints that limit access under the closed model.

For top-ranked papers, meanwhile, the literature highlights that the green choice is a form of advertising to increase scientific community awareness of the most important advances and discoveries. Wren (2005) suggested the existence of a trophy effect, which leads high-ranked scholars to upload their scientific works in order to benefit from the elimination of financial barriers that limit access to their works. The trophy effect thus explains that most prominent journals and best-ranked papers are more likely to be posted into open archives (Moed, 2007 ; Eysenbach, 2006). In this latter case, green Open Access is not used for eliminating the economic barrier that is imposed by the closed access system. The best-ranked journals are usually included in the libraries' subscriptions and their access is usually granted. Green Open Access here is a way of increasing awareness of the latest and most important scientific results among researchers in the scientific community. In this case, green Open Access is exploited as an advertisement tool for stressing the importance of a new scientific result. This trophy effect is supported by Kurtz et al. (2005), who showed that authors may use the self-archive option just for promoting the most citable and important articles. In this sense, the green choice is a complementary tool to the closed access model, because it does not enlarge access to this scientific knowledge, actually, but increases awareness of the latest discoveries among readers.

4.2 Theoretical model

In this section, we propose a theoretical model for explaining how different levels of paper quality, and then different journal quality levels, may influence scholars in their choice to adopt green Open Access. Because we were only interested in the quality of the paper, we assumed that there is only one generation of ideas. In this respect, we focused our attention on how the quality of the paper, certified by the journal in which the article is published, influences the researcher's choice to upload the paper, *ceteris paribus*. As we have seen through the literature review, scholars may be interested in adopting the green strategy either to broaden access to their scientific results or increase awareness of the latest scientific discoveries in the scientific community.

The probability of seeing a paper depends on access to the journal (o_j) where the article is published. Access to an article (o_a) is thus equal to access to the journal ($o_j = o_a$). The more accessible a journal, the greater the probability of the paper being cited. Because the green choice eliminates all economic barriers to access to the scientific literature, we assume that access to an article uploaded into an open repository (o_r) is greater than the access to a closed access article ($o_r > o_j$).

The usefulness of ideas is not known until the consumer reads the paper, and the quality of the paper is known only after its consumption. However, reading is a costly activity and scholars cannot afford to read all the papers they are interested in. They therefore select the papers that they will read according to their expected quality. Quality can be known ex ante before "consumption", however, through the quality certification given by publication in a journal. Through the peer-review system, journals guarantee that the article corresponds to a certain quality level, which reflects the quality of the journal in which the paper is published. It is in this way that the quality of an article is not perceived as the same if the paper is published in a top-ranked journal or in a lower-ranked one. Scientific journals reduce the information asymmetries that characterize scientific knowledge, by signalling to the reader the quality of the article (q_a) that, for the sake of simplicity, we assume to be the same as that of the journal (q_j). The scholar utility function for having published an

article into a closed access journal is then :

$$U_s^{CA} = o_j q_j \tag{4.1}$$

Where U_s^{CA} is the utility function of a scholar s who publishes his paper in a closed access journal (CA).

Because intellectual outputs are stored together in open repositories without any quality control or distinction, the quality of a closed access article may not be perceived with the same precision by the reader when it is stored in an open archive, as some information is usually lost [5]. Hence, once the article is stored under the green Open Access model, agents may fear a signal quality loss. Contrary to journal quality, which is ranked by several institutions, repositories may not be able to signal the quality of the uploaded paper (*e.g.* Which is the uploaded version ? Is the archive only for pre-published works ? How may the archive make sure it has the last and most important scientific advancements ?). We assume that $\alpha \in [0, 1)$ is the ability of the repository to highlight the quality difference existing between the repository, defined as the average quality of all documents stored in the repository ($q_r = \sum_{n=1}^{N} \frac{q_n}{N}$, and the quality of the published article q_j. When $\alpha = 0$, the repository is not able to signal the quality difference between the article and the repository. We exclude $\alpha = 1$ because the quality signal perceived in accessing the article through the journal website is stronger than through the repository (*e.g.* the format of the publisher under which the paper is presented immediately gives the quality signal of the paper).

As we said previously, the incentives for adopting the green Open Access differ among scholars, depending on the quality of their intellectual output. We focused our attention first on cases where the quality of the article is higher than repository quality ($q_j \geq q_r$), named case 1, and after on articles of lower quality than the repository quality ($q_r \geq q_j$),

5. For example, the format imposed by the publisher, which gives an immediate and visible signal that the article has been published in a particular journal, is not provided by an article stored in the repository. This simple situation may create doubts as to the version of the paper (*e.g.* Is the uploaded version the same as the one published in the journal or is it an earlier one ? Has the uploaded version been reviewed by peers or is it the version which still needs some modifications ?). All these situations may diminish the quality of the archived article, compared to the published one. In this way, the usefulness of the open archive is brought into question, and users may have to turn to the original publisher to check the version. This situation is less likely for low-quality papers, to which access may be possible only by open archive, due to the libraries' budget constraint previously mentioned. It is thus a kind of quality concern that affects the green choice.

referred to as case 2.

In case 1, researchers make the green choice to creating a trophy effect, as discussed above. For the $(q_j \geq q_r)$ scholar type, the utility of adopting green Open Access is :

$$U_s^{OA} = o_j q_j + o_r q_r - o_j q_r + \alpha(o_r - o_j)(q_j - q_r) \tag{4.2}$$

The utility of adopting GOA is maximized with respect to q_j and o_j. Solving for the first order condition gives :

$$\frac{\partial U}{\partial q_j} = o_j + \alpha o_r - \alpha o_j \tag{4.3}$$

$$\frac{\partial U}{\partial o_j} = q_j - q_r + \alpha q_r - \alpha q_j \tag{4.4}$$

And solving for the equilibrium yields :

$$q_j^* = \frac{o_j(1-\alpha) + \alpha o_r + q_r(1-\alpha)}{1-\alpha} \tag{4.5}$$

$$o_j^* = \frac{q_j(1-\alpha) - \alpha o_r + q_r(1-\alpha)}{1-\alpha} \tag{4.6}$$

In case 2, for authors of lower-rank papers, the incentives in adopting green Open Access are increased access to their work, and sometimes the possibility of making their scientific results accessible.

When the paper quality is lower than that of the repository $(q_r \geq q_j))$, the scholar's utility function is :

$$U_s^{OA} = o_r q_r - o_j q_j - \alpha o_r(q_r - q_j) \tag{4.7}$$

And solving for the equilibrium yields :

$$q_j^* = -\alpha o_r + o_j \tag{4.8}$$

$$o_j^* = \alpha o_r + q_j \tag{4.9}$$

In case 1, the incentive for adopting green Open Access as a trophy strategy increases with the openness of the repository. It means that top-ranked papers have strong incentives to

adopt this strategy because the archive increases awareness of the latest discoveries in the scientific community. This result summarizes the trophy effect, in which the best papers have higher incentives to adopt the green strategy in order to make the scientific community aware of their latest results. For low-ranked articles, unless the repository signal quality is completely inexistent, incentives for adopting the green strategy decrease with the openness of the repository. In this case, incentives to upload post-print papers are stronger, the lower the quality of the new knowledge produced. This result is consistent with the literature, which states that the lower the quality of an article, the lower the probability of being seen by the scientific community, due to the budget constraint of libraries.

In brief, our model shows that open archives will become repositories for the best and worst papers, because of the trophy effect in case 1 described above, and because of the possibility of boosting otherwise very scarce access in case 2. These results summarize the behaviour of the various scholars and their different incentives to adopt Open Access, as previously found in the literature.

This leads us to our empirical research question : are the best and low-ranked or non-ranked articles really more likely to be uploaded into open archives, showing that journal quality may influence scholars' behaviour even in the green Open Access choice ? If it is the case, we can identify two different uses of the green route. For top-ranked papers, the green model becomes a complementary tool to the traditional publishing system, enabling authors to highlight their best works and use this model as an advertising tool. For non-ranked or low-ranked papers, meanwhile, green Open Access is a substitute model for granting consumers access to their works. In fact, due to library budget constraints, the green route may be the only way for members of the scientific community to have access to these scientific results.

4.3 Data analysis

In order to test whether and to what extent the quality of the journal in which the paper is published is a good predictor of the choice to adopt green Open Access, we used data from the HAL (Hyper Archive en Ligne) archive. HAL is considered the largest open archive

in the French scientific landscape (Schöpfel *et al.*, 2016). Due to our research question, we chose to focus our attention on published papers in the economics discipline [6]. We downloaded data with the API provided by HAL [7], using the web scraping method. Anyone can create the bibliographic information of the scientific document to which it relates, but only the author has the right to upload a document into the open archive. Because publishers are the copyright holders of the intellectual good, they have the right to dictate the conditions under which the paper can be released in such repositories. Specific policies have been adopted by established journals in order to regulate the issue. While forbidding uploads is an isolated practice, restrictions about when, where and which version can be uploaded are very frequent. Using the Sherpa Romeo [8] data, HAL classifies journal policies into four main categories, depending on the restrictions imposed on authors. However, in order to decrease the cost of checking each journal policy and derogate from copyright holders' right to limit the use of green Open Access, the French government passed the *Pour une République Numérique* law (n° 2016-1321 7/10/2016). The law gives authors the right to upload the final version of their works when they have been accepted for the publication, after a statutory embargo period. In order to promote use of the green route, some research centres have also adopted mandatory policies for their members, obliging them to adopt one of the Open Access options [9].

Data was collected from 01/01/2014 [10] to 31/12/2019, representing a total of 12,647 bibliographic information items. Figures 4.1 and 4.2 show the amount of bibliographic information created on HAL every year and the number of files uploaded per year, respectively. We can notice that green Open Access in an increasing phenomenon.

6. We decided to focus our attention on economics because of the abundance of rankings provided by public institutions in France. As highlighted by Pontille and Torny (2010), rankings are widely used in the economics field – where there is a bigger literature about measuring scientists' performances via rankings than in other disciplines – for measuring university and laboratory performances (Baden-Fuller et al., 2000) and for evaluating scholars (Chung *et al.*, 1993). The HAL open archive is used more in France than elsewhere, and we therefore decided to test our hypothesis in a discipline where both international rankings and French institutional rankings apply. This allowed us to consider different indicators, and analyse a discipline where the quality of the journal strongly determines the future career of a researcher. In particular, institutional rankings are so important for researchers in economics, due to their role for obtaining the Qualification, a necessary title required for entering the French academic job market (The CNRS ranking was abolished in December 2020).
7. `https://api.archives-ouvertes.fr/docs/search/schema/field-types`.
8. `https://v2.sherpa.ac.uk/romeo/`
9. See `https://roarmap.eprints.org/foranexhaustivelist`.
10. The time choice depends on journal ranking availability for each period.

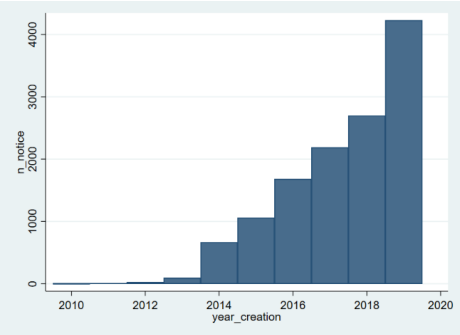

FIGURE 4.1 – Number of notices per year

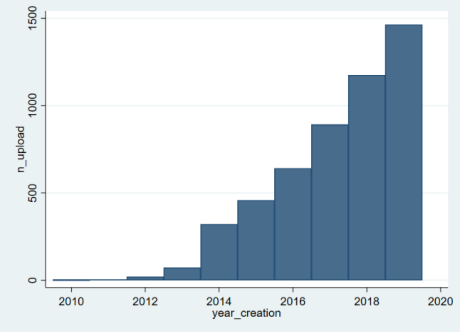

FIGURE 4.2 – Number of uploaded published papers per year

Only 5.65% of the notices refer to papers published in an Open Access journal, while the others were published in traditional closed access journals, suggesting slow penetration of Open Access in the publishing market for economics. Figure 4 shows that uploading papers is an increasing phenomenon. However, only 39% of notices provide access to the document.

Because the aim of the paper is to determine whether the quality of the journal in which articles are published may influence scholars' choice to adopt the green route – either by providing the scientific community with access to their scientific results which would not be accessible otherwise or for signalling the publication of a star paper – we chose different journal ranks to analyse our question. To do so, we used the three most important national journal rankings in the French scientific landscape, and two international journal rankings, as reported in Table 4.1.

Journal Ranking	Description
Institutional Ranking	
CNRS	French National Centre for Scientific Research
HCERES	High Council for the Evaluation of Research and Higher Education
FNEGE	National Foundation for COmpanies Management Academic Education
International Ranking	
JCR	Journal Citation Report Clarivate
SJR	Scimago Journal Rank

TABLE 4.1 – Journal ranking indices

We took different journal rankings into account for testing our hypothesis. National rankings generally grant greater value to national journals, which are usually excluded

from international rankings. The CNRS journal ranking is conceived by the *French Centre National de la Recherche Scientifique*, after consultation with representatives of the French scientific community. The first list was published in 2003 and it is updated every year. Journals are ranked from 1, excellent journals, to 4, journals where young researchers publish their first works (Marco, 2018). The second national index is the HCERES, provided by the French Haut Conseil for the evaluation of the research and the higher education. This serves as a reference for evaluating scholars' career advancement, among other indicators (Rubinstein, 2016). The last index is the FNEGE (*Fondation Nationale pour l'Enseignement de la Gestion des Entreprises*), which does not limit its analysis to economics, and also stresses the importance of journals in the management and business fields. This index is a mix of both national rankings (FNEGE, CNRS, HCERES) and international ones (such as Financial Times, ABS, EJL, etc.). All these indices differ from the international ones by the importance they give to national journals, ranking national journals that are not listed in the international lists due to their low impact factors. International journal rankings do not make any such discrimination in favour of one discipline or another or by country. They are based on citation counts, weighting them in different ways. We decided to look at the Clarivate Journal Citation Report (JCR), the prominent index based on Garfield's (1955) impact factor. The second international index is the Scimago Journal Rank (SJR), which differs from the JCR because it accounts for both the number of citations received by a journal and the importance of the journal those citations came from.

Figure 5 reports the breakdown of bibliographic quality information stored on HAL, by different journal rankings, where Q0 represents non-ranked journals.

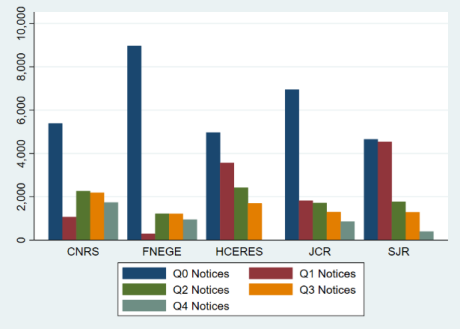

FIGURE 4.3 – Quality of bibliographic notices on HAL

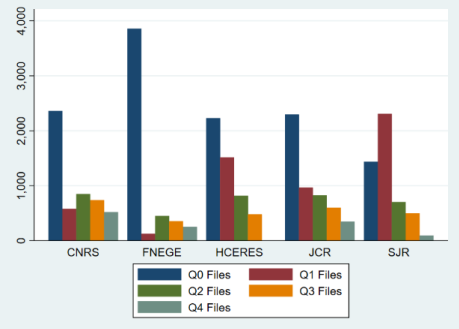

FIGURE 4.4 – Quality of articles stored on HAL

At first glance, it seems that HAL provides bibliographic information and access to non-ranked documents, *i.e.* whose quality is not accepted by the scientific community. Because access to non-ranked papers is limited by library budget constraints, it seems that GOA is mostly used as a tool for eliminating the price barrier on access to non-ranked papers. In this sense, the green model makes access to these scientific results easier, becoming a complementary tool to the closed access model and increasing awareness of the importance of these scientific articles in the scientific community.

4.4 The empirical model

In this section, we present the model we used to test whether journal quality influences the behaviour of scholars. In particular, we provide an analysis of whether there is a star effect, confirming case 1 in the theoretical model, by which the best papers are more likely to be uploaded, or whether scholars instead prefer to ensure access to articles that are less easily available through institutional channels, as described in case 2.

To do so, we used the logit model used by Eger *et al.* (2015) to examine the variables that affect the probability of publishing in Open Access journals or uploading the paper into an open repository. Our dependent variable was a dummy equal to 1 if the bibliographic notice stored on HAL provides access to the document, 0 otherwise :

$$Y_i = \alpha + \beta_0 X_i^{'} + \chi_i + \varepsilon_i \tag{4.10}$$

$X_i^{'}$ is a vector of the different categories listed by each journal ranking. Each ranking provides different categories. For example the CNRS rank splits journals into five different classes, while the HCERES uses only three.

χ_i is a vector of control variables. In particular, we controlled for the copyright reform introduced by the French government in 2016, which now allows authors to upload their post-print papers into an open archive, derogating from copyright rules. Also in line with the copyright legal framework, we controlled for the policy adopted by the laboratories to which the authors belong. Some research centres have adopted mandatory policies obliging scholars to provide a copy of their works via green Open Access. This category also included

documents financed by institutions and organizations, such as the European Horizon 2020 project, which request free access to be provided to the final work. The last variable concerned by copyright infringement was journal policy. Because journals are the copyright holders to the scientific intellectual works, we added the policy adopted by the journal to our control variables, and specifically whether they allow uploading of the published work into an open archive, and under what conditions. As we believe that authors who publish in Open Access journals are more responsive to the Open Access cause, we therefore included Open Access publication among our control variables. Other control variables were the version of the uploaded paper, *i.e.* a pre-published document or a file uploaded after the paper publication, and the copyright infringement. As we said, we included all papers whose notice had been created before 31/12/2019 in our dataset. We then controlled for papers that were in our dataset but could not be uploaded on account of the embargo period, meaning the time that must pass between the publication date and upload of the file into an open archive.

4.5 Results

Using the model presented in the previous section, we tested the two main hypotheses of the literature, i.e. whether journal rank influences the choice of scholars to provide a copy of their scientific results using the green Open Access model, or not. If it is the case, our study will contribute to determining whether star papers are more likely to be uploaded into open archives than non-ranked ones.

4.5.1 National journal ranking indices

National journal rankings differ from international ones by the attention they pay to national journals, which are usually less well ranked in international rankings. Non-ranked papers were considered as the omitted variable. Table 4.2 reports the coefficients for the CNRS ranking. The CNRS index ranks journals in 5 classes, namely excellence, grouped together under the Q1a variable. This is followed by the Q1 class, where we find high-quality journals, and then the Q2, Q3 and Q4 journals in descending order of quality.

Non-significant coefficients for the best-published papers seem to suggest that there is no difference in the odds of having a bibliographic notice giving access to a top ranked paper (Q1 and Q1a) stored on HAL compared to a non-ranked paper. Negative and significant coefficients for the other journal quality levels (Q2, Q3 and Q4) suggest that use of green Open Access as an instrument for granting access to articles is hardly possible.

This qualitatively different use of HAL was confirmed when data was analysed using the FNEGE ranking, explained in Table 4.3, which uses the same classes as the CNRS ranking. For the best-ranked articles (Q1a), the odds of finding them are 3.1 times higher than of finding a non-ranked paper, when controlling for all other variables. These papers are defined as the excellence of scientific research. We can consider Q1 papers as the top-ranked scientific outputs, and the fact that the coefficient is not significant, while it is negative and significant for the other categories, confirms the previous results.

The third journal ranking used for our study was the HCERES index, which differs from the other national indices by the different quality categories it provides. As reported in Table 4.4, HCERES does not split top journals into excellence (Q1a) and top-ranked journals (Q1), and places them in a single category. Lower-ranked journals are also grouped together in a single class (Q3). In this case, because the coefficient of the top-ranked papers is significant but negative, the trophy effect does not emerge. Instead, non-ranked papers are more likely to be uploaded. All the coefficients are negative and significant, meaning that the odds of finding a bibliographic notice providing access to non-ranked journals are higher than for ranked papers. Specifically, the odds of finding a paper are 36.5% lower for top-ranked papers than for non-ranked ones. Odds become more unfavourable for the lower-ranked articles, where they are 0.15 times those of a non-ranked paper. Although the trophy effect was not found, among ranked papers, the top ones are more likely to provide access to their scientific results than the other classes.

VARIABLES	CNRS
CNRSQ1a	0.468
	(0.523)
CNRSQ1	-0.0276
	(0.150)
CNRSQ2	-0.426***
	(0.126)
CNRSQ3	-1.171***
	(0.154)
CNRSQ4	-1.770***
	(0.247)
Constant	-2.202***
	(0.108)
Observations	12,647
Control	YES
Pseudo R2	0.798

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 4.2 – Estimates of equation 4.10 for CNRS ranking

VARIABLES	FNEGE
FNEGEQ1a	1.138**
	(0.564)
FNEGEQ1	-0.239
	(0.270)
FNEGEQ2	-0.867***
	(0.189)
FNEGEQ3	-1.000***
	(0.199)
FNEGEQ4	-1.869***
	(0.312)
Constant	-2.361***
	(0.0930)
Observations	12,647
Control	YES
Pseudo R2	0.796

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 4.3 – Estimates of equation 4.10 for FNEGE ranking

VARIABLES	HCERES
HCERESQ1	-0.455***
	(0.110)
HCERESQ2	-1.219***
	(0.152)
HCERESQ3	-1.876***
	(0.248)
Constant	-2.103***
	(0.107)
Observations	12,647
Control	YES
Pseudo R2	0.798

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 4.4 – Estimates of equation 4.10 for HCERES ranking

4.5.2 International journal ranking indices

National indices balance journals' impact factors with other criteria, in order to boost the importance of national scientific journals, which are mostly penalised by international rankings. Conversely, international rankings, such as the Scimago Journal Rank and Journal Citation Reports, are based solely on bibliometrics. However, every ranking has a

different way of evaluating the impact of a journal. For example, the SJR index weights the number of citations received by the importance of the journal in which the citation appears. A simpler evaluation is proposed by JCR, which uses the well-known journal impact factor equal to the ratio of the number of citations received by articles published in the journal in the two previous years over the number of articles published in the same period (See annex 2 for the calculation of the two indices).

Looking at the SJR journal ranking (Table 4.5), the trophy effect is still present. In fact, for the fourth quartile, no differences in finding a notice providing access to a scientific document exist between these this class and non-ranked journals. Conversely, articles published in the first two quartiles are more likely to provide a copy of the document on HAL and the difference is greatest for top-ranked papers. For articles ranked in the second quartile, the odds of finding a bibliographic information followed by the document are 3 times higher than finding a paper for not-ranked papers, all else equals. The odds are greater for papers published in journals ranked in the first quartile, for which the odds are 5 times greater than for non-ranked articles.

The trophy effect is still present when we look at the JCR index, as reported in Table 4.6. It is true that all four quartiles report a significant and positive coefficient, but top ranked articles are, once again, more likely to provide a copy of their document, compared to the other quartiles. For articles published in the second, third and fourth quartile of JCR, the odds of finding a paper on HAL range between 4.1 to 4.7 times those of finding a paper for non-ranked articles. Top ranked articles report the highest odds, five times higher than for non-ranked articles.

VARIABLES	SJR
SJRQ1	1.988***
	(0.224)
SJRQ2	1.398***
	(0.239)
SJRQ3	0.588**
	(0.272)
SJRQ4	0.449
	(0.452)
Constant	-4.265***
	(0.228)
Observations	12,647
Control	YES
Pseudo R2	0.798

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 4.5 – Estimates of equation 4.10 for SJR ranking

VARIABLES	JCR
JCRQ1	1.798***
	(0.173)
JCRQ2	1.692***
	(0.176)
JCRQ3	1.709***
	(0.180)
JCRQ4	1.514***
	(0.194)
Constant	-4.039***
	(0.162)
Observations	12,647
Control	YES
Pseudo R2	0.800

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

TABLE 4.6 – Estimates of equation 4.10 for JCR ranking

4.6 Discussion and policy implications

Open Access implementation as a publishing model requires profound reconfiguration of the scientific publishing market, transferring funds allocated to access to scientific knowledge to article processing charges imposed by journals for publishing scientific results. However, the green model does not require such reconfiguration of the financial system of the scientific publishing market. GOA promotes a different way for disclosing scientific outputs that, once published in a closed access journal, can be stored in open archives, where no price is imposed for accessing the new produced knowledge.

The reasons that push scholars to adopt the GOA model have already been investigated by the literature, which in particular questions whether journal rankings may influence the green choice. Our first result is that the quality of the journal matters and influences scholars' choice to adopt the green model.

As suggested by the literature, two different types of scholars may have greater incentives to adopt the model. First, authors of top-ranked papers or, as the literature defines them, stars, may be more likely to self-archive their work in open archives to underscore the importance of their discoveries. In this way, green Open Access becomes a tool

for increasing scholars' renown and highlighting their role in the advancement of scientific knowledge. Referring then to the CNRS ranking, because the odds of finding a paper do not differ between non-ranked and top-ranked papers, the trophy effect seems to be confirmed, showing that scholars are more prone to upload well-ranked articles compared to middle and low-ranked papers. Results for the FNEGE ranking highlight, once again, the existence of the trophy effect, reinforced by the positive and significant effect found for Q1a papers. For the two international indices, the trophy effect appears clearly. Our results clearly underscore the existence of the trophy effect, showing at both national and international level that top-ranked papers are more likely to be stored in open archives, compared to non-ranked or low-ranked ones. They confirm the existing literature showing a greater likelihood of top-ranked papers being uploaded into an open archive, in order to increase awareness of the most important contributions to scientific knowledge advancement in the scientific community.

In this case, green Open Access can be considered a marketing tool to increase awareness of the latest and most important scientific results among consumers. The green choice is more about increasing authors' reputation and prestige than ensuring access to their scientific results. In fact, research centres and universities usually grant their members access to the best scientific journals, to the detriment of low-ranked and non-ranked reviews due to limited budgets. Green Open Access serves the best scholars' interest as a marketing tool to emphasise their importance and achievements.

The second category of scholars that has a stronger interest in adopting green Open Access is authors of non-ranked papers to which access is almost impossible via the closed access model. In fact, due to library budget constraints, subscriptions to these journals are usually dropped in order to grant access to the most relevant scientific literature, usually published in top-ranked journals. Green Open Access is thus an alternative way of disclosing new scientific results, especially for low- and non-ranked papers, to which access through the traditional publishing system may be severely limited. As we said, national journal rankings usually give more importance to national journals, which are often not included in international rankings. Referring to the CNRS ranking, bibliographic information on non-ranked papers is more likely to provide access to files than middle and

low-quality articles, making the green route a substitute for granting consumers access to these papers. Results for the FNEGE ranking confirm the non-ranked effect. As for the CNRS ranking, for high-quality (Q1) and non-ranked papers there is no significant difference in providing the notice with a file, while middle and low-ranked papers are less prone to provide access to their scientific outputs.

Non-ranked articles should be published following the green Open Access model to grant access that would be otherwise denied to researchers, due to library budget constraints. Looking at national rankings, it seems that authors consider green Open Access to be an alternative tool to traditional journals for granting consumers access to scientific results. The fact that, when we look at international indices, non-ranked articles are less uploaded into HAL than ranked ones calls into question the real vocation of researchers to make their scientific results accessible. While less renowned journals are ranked for national indices, because of their importance at that level, international rankings are based on journal impact, measured by the number of citations. The same scientific recognition granted to national journals is not matched in international rankings. The fact that ranked papers are uploaded into HAL more than non-ranked ones may show that when a publication has not been recognized by the scientific community as a worthy one, authors have no interest in making it accessible to other members of the scientific community. Publication thus becomes the end for the scholar and not the means for allowing dissemination of his scientific results and access to them.

For star papers, the green choice has less impact on the transition from the traditional publishing system to the open one. In fact, because the upload into an open archive has to respect an embargo period, libraries have to keep their subscriptions to ensure immediate access to the new produced knowledge, which cannot be delayed by one year. The green model is particularly important for that scientific knowledge which may not be accessible through the standard scientific channels – *i.e.* scientific papers – due to the limited budgets of libraries. As we said, we chose to analyse the impact of journal rankings on the choice of green Open Access in economics for the central role played by article rankings in the evaluation of scholars. Last year, the CNRS decided to eliminate its journals ranking and it is still not clear how the research centre will evaluate the importance of the new knowledge

that is produced and then the advancement of scholars' careers. For the open transition, the risk is that growing importance will be given to international rankings. In this scenario, we see that journals which are considered valuable nationally are less well ranked or not ranked at all at international level. The risk is that international rankings could become the reference for scholars, and that scholars might thus become less likely to upload non-ranked papers, for which the green open system is usually the only access possible.

Even if these scenarios are merely hypotheses for the moment, what is clear is that, even in such green choices, journal rankings do influence scholars' behaviour towards Open Access. Instead of eliminating the ranking, the CNRS could base scholar evaluations on HAL, thus giving scholars an incentive to upload their scientific results into the open repository, no matter what the journal ranking. In the longer term, this practice might be internalized by members of the scientific community, transforming green Open Access from an individual marketing strategy into a practice for the common advancement of the scientific knowledge.

Conclusion

The quality of journals in which papers are published is a central matter for researchers. Journals are ranked using different indices, but all of them take citations as the unit of measure. The more a journal is cited, through the citations received by articles published in it, the higher its quality. Because scholars' careers depend strongly on the quantitative and qualitative evaluation of their intellectual production, the quality of the journal is a fundamental element in their choice of where to publish the final output of their research.

With the digitization of information, new models of knowledge disclosure have come to characterize the scientific publishing industry. Open Access allows the elimination of legal and economic barriers to access to, and reuse of, scientific literature. By eliminating the price, consumers have free access to new scientific results. Different studies have shown that the greater access provided by OA increases the probability of an author being cited.

Green Open Access is one of the OA models that scholars can choose to ensure free access to scientific knowledge. It allows authors to self-archive their scientific works into an

open repository, even after publication in a closed access journal. Because of the underuse of GOA, we question whether the quality of the journal in which the article is published may influence scholars' choice to adopt the green model.

Depending on the quality of the published paper, which is based on the journal in which the article is published, scholars have different incentives for using the GOA. For the non-ranked and lower-ranked papers, GOA could be the sole solution for ensuring access, due to the budget constraints faced by libraries. GOA thus becomes an alternative tool for granting access to scientific results. By contrast, for top-ranked papers, or stars, as the literature defines them, the green choice may be a marketing tool for highlighting the importance of their publications, generating the so-called trophy effect. GOA is thus a complementary instrument for disseminating the latest scientific results in this case.

Our study shows that journal quality matters and influences scholars in their choice to adopt green Open Access. This use for the first time of HAL, the largest open archive in the French scientific landscape, shows how the trophy effect emerges clearly at international level, at the expense of non-ranked papers. When we take national quality indices into account, the trophy effect still emerges. However, while non-ranked papers are uploaded as much as top-ranked ones, lower-ranked are uploaded less. This difference should be investigated in future research. This issue is all of the more important when we consider the elimination of the CNRS ranking for scientific evaluation in 2020.

The point is that, rather than a marketing tool, GOA was born as an instrument to allow access to scientific papers. As we said, for low-ranked and non-ranked papers, the price imposed by journals on access to scientific knowledge is a genuine barrier for consumers. Nowadays, GOA is still underused as a tool for granting consumers access to scientific results and its use as a practice for the common advancement of the knowledge is called into question.

Bibliography of chapter 4

Antelman, K. (2017). Do Open-Access Articles Have a Greater Research Impact ? | Antelman | College & Research Libraries.

Baden-Fuller, C., F. Ravazzolo, and T. Schweizer (2000). Making and Measuring Reputations : The Research Ranking of European Business Schools. *Long Range Planning 33*(5), 621–650.

Ball, R. (2017). *An Introduction to Bibliometrics : New Development and Trends*. Chandos Publishing.

Biagioli, M. (2016). Watch out for cheats in citation game. *Nature 535*(7611), 201–201.

Biagioli, M. and A. Lippman (2020). *Gaming the Metrics : Misconduct and Manipulation in Academic Research*. MIT Press.

Bomsel, O. (2010). Copyright et droit des marques à l'ère numérique. *Propriété industrielle*, 18.

Bomsel, O. (2014). Copyright et économie des revues académiques. *Revue européenne des sciences sociales. European Journal of Social Sciences* (52-1), 93–110. Librairie Droz.

Chung, K. H., R. A. K. Cox, and A. A. Okunade (1993). Publishing Behavior of Individuals and Most Prolific Authors in the Economics Literature. *Quarterly Journal of Business and Economics 32*(3), 32–42. University of Nebraska-Lincoln College of Business Administration.

Cole, J. R. and S. Cole (1973). *Social stratification in science*. Chicago : University of Chicago Press.

Creaser, C. (2010). Open Access to Research Outputs—Institutional Policies and Researchers' Views : Results From Two Complementary Surveys. *New Review of Academic Librarianship 16*(1), 4–25. Routledge.

Creaser, C., J. Fry, H. Greenwood, C. Oppenheim, S. Probets, V. Spezi, and S. White (2010). Authors' Awareness and Attitudes Toward Open Access Repositories. *New Review of Academic Librarianship 16*(sup1), 145–161. Routledge.

Cullen, R. and B. Chawner (2011). Institutional Repositories, Open Access, and Scholarly Communication : A Study of Conflicting Paradigms. *The Journal of Academic Librarianship 37*(6), 460–470.

Davis, P. M., B. V. Lewenstein, D. H. Simon, J. G. Booth, and M. J. L. Connolly. Open access publishing, article downloads, and citations : randomised controlled trial. *BMJ (Clinical research ed.) 337*, a568.

Eger, T., M. Scheufen, and D. Meierrieks (2015). The determinants of open access publishing : survey evidence from Germany. *European Journal of Law and Economics 39*(3), 475–503.

Eysenbach, G. A Napster moment for ivory towers. *A Napster moment for ivory towers. Sunday Times (Business Section), 17, 8.*.

Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLOS Biology 4*(5), e157. Public Library of Science.

Garfield, E. (1955). Citation Indexes for Science : A New Dimension in Documentation through Association of Ideas. *Science 122*(3159), 108–111. American Association for the Advancement of Science.

Gargouri, Y., C. Hajjem, V. Larivière, Y. Gingras, L. Carr, T. Brody, and S. Harnad (2010). Self-Selected or Mandated, Open Access Increases Citation Impact for Higher Quality Research. *PLOS ONE 5*(10), e13636. Public Library of Science.

Green, T. (2018). We're Still Failing to Deliver Open Access and Solve the Serials Crisis : To Succeed We Need a Digital Transformation of Scholarly Communication Using Internet-Era Principles. SSRN Scholarly Paper ID 3333907, Social Science Research Network, Rochester, NY.

Hajjem, C., S. Harnad, and Y. Gingras (2006). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. *arXiv :cs/0606079*.

Harnad, S. and T. Brody (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine 10*(6).

Heckman, J. J. and S. Moktan (2020). Publishing and Promotion in Economics : The Tyranny of the Top Five. *Journal of Economic Literature 58*(2), 419–470.

Hendler, J. (2007). Reinventing Academic Publishing, Part 2. *IEEE Intelligent Systems 22*(6), 2–3. IEEE Intelligent Systems.

Hilmer, M. J., M. R. Ransom, and C. E. Hilmer (2015). Fame and the fortune of academic economists : How the market rewards influential research in economics. *Southern Economic Journal 82*(2), 430–452.

Hoeffel, C. (1998). journal impact factors. *Allergy 53*(12), 1225–1225.

Jaffe, A. B. and M. Trajtenberg (2002). *Patents, Citations, and Innovations : A Window on the Knowledge Economy*. MIT Press.

Kim, J. (2007). Motivating and Impeding Factors Affecting Faculty Contribution to Institutional Repositories. *J. Digit. Inf.*.

Kurtz, M. J., G. Eichhorn, A. Accomazzi, C. Grant, M. Demleitner, E. Henneken, and S. S. Murray (2005). The effect of use and access on citations. *Information Processing & Management 41*(6), 1395–1402.

Lawani, S. M. (1986). Some bibliometric correlates of quality in scientific research. *Scientometrics 9*(1-2), 13–25.

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature 411*(6837), 521–521. Nature Publishing Group.

Marco, L. (2018). Le nouveau classement des revues d'économie-gestion du CNRS. *Vie sciences de l'entreprise N° 205*(1), 154–157. ANDESE.

Mazlish, B. (1982). The Quality of 'The Quality of Science' : An Evaluation. *Science, Technology, & Human Values 7*(1), 42–52. SAGE Publications Inc.

McCABE, M. J. and C. M. Snyder (2014). Identifying the Effect of Open Access on Citations Using a Panel of Science Journals. *Economic Inquiry 52*(4), 1284–1300.

Metcalfe, T. S. (2006). The Citation Impact of Digital Preprint Archives for Solar Physics Papers. *Solar Physics 239*(1), 549–553.

Mingers, J. and L. Yang (2017). Evaluating journal quality : A review of journal citation indicators and ranking in business and management. *European Journal of Operational Research 257*(1), 323–337.

Moed, H. F. (2007). The effect of "open access" on citation impact : An analysis of ArXiv's condensed matter section. *Journal of the American Society for Information Science and Technology 58*(13), 2047–2054.

Patel, R. B., M. Vaduganathan, R. C. Mosarla, R. V. Venkateswaran, D. L. Bhatt, and R. O. Bonow (2019). Open Access Publishing and Subsequent Citations Among Articles in Major Cardiovascular Journals. *The American Journal of Medicine 132*(9), 1103–1105.

Plume, A. and D. van Weijen. Publish or perish ? The rise of the fractional author.

Pontille, D. and D. Torny (2010). Revues qui comptent, revues qu'on compte :produire des classements en économie et gestion. *Revue de la régulation. Capitalisme, institutions, pouvoirs* (8). Association Recherche & Régulation.

Rodriguez, J. E. (2014). Awareness and Attitudes about Open Access Publishing : A Glance at Generational Differences. *The Journal of Academic Librarianship 40*(6), 604–610.

Rowley, J., F. Johnson, L. Sbaffi, W. Frass, and E. Devine (2017). Academics' behaviors and attitudes towards open access publishing in scholarly journals. *Journal of the Association for Information Science and Technology 68*(5), 1201–1211.

Rubinstein, M. (2016). Comment on écrit l'économie. *L'Economie politique N° 72*(4), 65–79. Alternatives économiques.

Schöpfel, J., C. Ferrant, F. André, and R. Fabre (2016). Ready for the future ? A survey on open access with scientists from the French National Research Center (CNRS). *Interlending & Document Supply 44*(4), 141–149. Emerald Group Publishing Limited.

Sengupta, I. N. (1992). Bibliometrics, Informetrics, Scientometrics and Librametrics : An Overview. *Libri 42*(2), 75–98.

Shavell, S. (2009). Should Copyright of Academic Works be Abolished ? SSRN Scholarly Paper ID 1525667, Social Science Research Network, Rochester, NY.

Spezi, V., J. Fry, C. Creaser, S. Probets, and S. White (2013). Researchers' green open access practice : a cross-disciplinary analysis. *Journal of Documentation 69*(3), 334–359. Emerald Group Publishing Limited.

Suber, P. (2012). *Open Access*. MIT Press.

Tang, M., J. D. Bever, and F.-H. Yu (2017). Open access increases citations of papers in ecology. *Ecosphere 8*(7), e01887.

Wren, J. D. (2005). Open access and openly accessible : a study of scientific publications shared via the internet. *BMJ 330*(7500), 1128. British Medical Journal Publishing Group Section : Information in practice.

Xia, J. and K. Nakanishi (2012). Self-selection and the citation advantage of open access articles. *Online Information Review 36*(1), 40–51. Emerald Group Publishing Limited.

Annexe

1. Knowledge Quality

What is knowledge quality ? The quality of knowledge has different connotations, involving moral, philosophic, social and ethical dimensions (Lawani, 1984). However, these connotations refer mostly to the "outside" judgment of the science. In trying to define the quality of the science, Mazlish (1982) proposed a double approach. From the "outside", the quality of the new scientific advancements is judged in terms of the impact of science – and technology – on society. In this sense, knowledge involves the different spheres mentioned above. We can consider, for example, the discovery of nuclear energy and its relation to the quality of life for the society.

The "inside" usage of science quality is the establishment of quality though quantitative indicators. With this term, Mazlish (ibid.) refers to the insider point of view of scientists in estimating the value of a piece of knowledge. Scientists have developed different approaches for measuring the quality of knowledge, such as bibliometrics, informetrics, scientometrics and librametrics (Sengupta, 1992), which nowadays constitute new fields of scientific research. These approaches are very similar in their scope, and the distinction lies in the application of statistical/mathematical measuring techniques, as summarised in the table above.

Discipline	**Statistical/mathematical techniques applied to :**
Bibliometrics	Bibliographies
Librametrics	Libraries activities
scientometrics	Science organization
Infometrics	Information systems

TABLE 4.7 – Knowledge quality evaluation - data retrieved from Sengupta, 1992

As we said, bibliometrics allows a quantitative evaluation of the quality of newly-produced knowledge. Although there are different indices for counting and weighting the frequency with which a paper is cited, all the evaluative tools take citations as a proxy of the quality of the new knowledge produced. As stated by Cole and Cole (1973) "The number of citations is taken to represent the relative scientific significance or 'quality' of papers". Using bibliometrics, and then the different indices proposed for counting citations, public

and private organizations have started to rank journals to give an immediate interpretation of their quality. In fact, the number of citations received per journal differs considerably among disciplines, making it difficult to compare journal quality between different scientific domains. Top-ranked journals thus receive a greater number of citations than lower-ranked ones, making their quality greater.

Obviously, it would be quite a hard task to estimate the quality of each new scientific article, considering that more than 2.5M papers are produced every year (Plume and Van Weijen, 2014). Moreover, due to time constraints, scholars would like to know the quality of the new knowledge that is produced before "consuming" it, meaning before reading the new paper. The quality of new articles is thus determined in two steps. Ex ante, though acceptance for publication in a specific journal, and basing its evaluation on the citations received by the journal, and ex post, using indices that evaluate the citations received only by the article (such as the H-index, which measures the performance of a scholar, starting from the citations their works have received). Nowadays, ex ante evaluation, which consists in judging the paper using the quality of the journal in which it has been published, is a common practice among the scientific community (Mingers and Liying, 2017). In this way, the quality of an article is not perceived as the same if the paper is published in a top-ranked journal or in a low-ranked one, even before having read the article.

Publishing in scientific journal is only the preliminary step in making a paper more highly cited. Once the paper is published, scholars can boost their audience by putting their work into an open repository – referred to as green Open Access publishing. Because this knowledge dissemination model eliminates the price on access to the scientific literature, there are no longer any financial barriers to access and thus to citation. Bearing in mind that citations are a valuable intangible asset for scholars, due to their impact on researchers' careers, the scientific community thus has a strong incentive to upload post-print papers into open archives.

2. Impact Factor and SJR ranking methods

Impact factor calculation :

$$IF_y = \frac{Citations_y}{Publications_{y-1+Publications_{y-2}}} \tag{4.11}$$

Scimago Journal Rank calculation The SJR index is more complex than the IF, because it weights the number of citations received by the importance of the journal in which the citation appears. It is based on a transfer of prestige among journals [11]

$$SJR_i = \frac{1-d-e}{N} + e\frac{Art_i}{\sum_{j=1}^{N} Art_j} + d\sum_{j=1}^{N} \frac{C_{ji}SJR_j}{C_j} * \frac{1-(\sum_{k\in\{Danglingnodes\}} SJR_k}{\sum_{h=1}^{N}\sum_{k=1}^{N} \frac{C_{kh}SJR_k}{C_k}} + d[\sum_{k\in\{Danglingnodes\}} SJR_k] * \frac{Art_i}{\sum_{j=1}^{N} Art_j} \tag{4.12}$$

Where :

- SJR_i is the Scimago Journal Rank of the Journal i.
- e is a constant, normally 0.1.
- d is a constant, normally 0.85.
- N is the number of journals.
- Art_i is the number of articles of journal j.
- C_{ji} is the citation from journal j to journal i.
- C_j is the number of total references of journal j.
- Dandling nodes are journals that do not have references to any other journals.

11. From `https://www.scimagojr.com/SCImagoJournalRank.pdf`

Chapitre 5

La gouvernance du marché de la publication scientifique à l'ère du numérique : une analyse de la connaissance scientifique en tant que *commons*

Résumé [1]

La nature de la connaissance en tant que bien économique est déterminée à partir de la rivalité et excluabilité du bien, mais aussi du faisceau de droits qui en règlent l'accès et l'utilisation. Si la numérisation a permis à la connaissance de devenir un bien non-rival, c'est la transition d'un modèle de gestion de type public order à une gestion de type private ordering qui pourrait aboutir à la transformation de la connaissance en tant que commun. Cette ressource ouverte et partagée, dont la gestion est assurée par une communauté, a profondément changé la structure du marché de la publication scientifique.

À travers l'*Institutional Analysis Development framework* nous proposons une analyse du marché de la publication scientifique, afin de comprendre quel a été l'impact de la numérisation de l'information sur les différents acteurs du marché, leurs intérêts et, surtout sur la chaîne de valeur du commun de la connaissance scientifique.

Introduction

La connaissance a longtemps été analysée en tant que bien économique, à savoir non-rival et non-excluable. Cette classification, qui trouve ses racines dans la théorie des dépenses publiques de Samuelson (1954), a ensuite évolué, en donnant naissance aux biens publics globaux (Stiglitz, 1999), jouissant de la propriété d'ubiquité. Si cette classification des biens économiques se fonde sur les caractéristiques de la ressource, l'un des apports majeurs des études sur les biens communs et ensuite sur les communs (ou *commons*) de E. Ostrom a été sa capacité de distinguer entre les caractéristiques de la ressource et le régime propriétaire qui en assure la gouvernance (Coriat, 2015). Ainsi, la connaissance n'est ni un bien public ni un bien privé, mais une combinaison des deux (Pavitt, 1987 ; Nelson, 1992 ; Callon, 1994). De ce fait, la nature de la connaissance doit être déterminée à partir des droits qui en règlent l'accès et l'utilisation et des caractéristiques intrinsèques du bien. La classification dichotomique proposée par Samuelson (1954) et Musgrave (1959), reposant sur l'exclusion à l'accès, s'enrichit avec le degré de rivalité du bien, introduit par Ostrom

1. Une première version de ce chapitre à été publiée comme suit : Bechis, J. (2020). La gouvernance du marché de la publication scientifique à l'ère du numérique : une analyse de la connaissance scientifique en tant que commons. Innovations, (3), 95-123.

et Ostrom en 1977. En suivant cette classification duale, la connaissance peut ainsi être considérée comme un bien privé, un bien public, un bien commun ou un bien club, selon le régime juridique qui en détermine l'utilisation et l'accès (Figure 5.1).

		SUBTRACTABILITY	
		Low	*High*
EXCLUSION	*Difficult*	**Public goods** Useful knowledge Sunsets	**Common-pool resources** Libraries Irrigation systems
	Easy	**Toll or club goods** Journal subscriptions Day-care centers	**Private goods** Personal computers Doughnuts

FIGURE 5.1 – Types de biens (Source : Hess et Ostrom 2007, p.9)

La numérisation de l'information a de ce fait permis l'émergence d'une nouvelle catégorie de biens (Hess, 2007), se plaçant à l'intersection entre les biens publics et les biens communs. Cette dématérialisation a engendré l'apparition de nouvelles modalités pour l'organisation de la production du savoir, qui passe d'un régime de propriété privée à un régime de « common-based peer-production » (Benkler, 2002). Ce système se fonde sur le partage de la ressource entre ses membres, qui en assurent une production très décentralisée et collaborative (Benkler, 2006), à travers des formes d'organisations autres que le marché.

Les communs sont des biens ayant en général les caractéristiques de la non-rivalité et de la non-excluabilité typiques des biens publics, mais soumis à la gestion d'une communauté spécifique (Boyle, 2003 ; Hess et Ostrom, 2007 ; Coriat, 2015), qui doit faire face à un dilemme social. Avant la numérisation de l'information, ce dilemme était représenté, dans le cas de la circulation du savoir scientifique, par la sous-production de la connaissance sous forme d'articles scientifiques. Cette situation de déséquilibre est le résultat de l'existence d'externalités dérivant de la non-excluabilité du bien, qui empêche le producteur de s'approprier des profits dérivant de la vente du bien dans un marché en concurrence pure et parfaite. La gestion commune de la ressource, et donc l'établissement de règles complémentaires ou substituts aux règles formelles imposées à travers les droits de la propriété intellectuelle qui en déterminent l'accès et l'utilisation, empêchent ainsi la dégradation de

la ressource décrite dans la tragédie des biens communs d'Hardin (1968). Si la numérisation permet d'éviter l'épuisement de la ressource, puisque la dématérialisation du support papier exclut la rivalité du bien, l'objectif principal de la gestion commune est d'assurer le maintien de la connaissance, son développement et son enrichissement, à travers la régulation de l'accès et l'utilisation de la ressource (Lukasik, 2000 ; Cooper, 2006).

On le voit, le marché de la publication scientifique [2], où la connaissance scientifique constitue l'input primaire, a subi des transformations importantes à la suite de la numérisation de l'information. Si une gestion fondée sur les droits de la propriété intellectuelle justifiait l'octroi d'un monopole pour financer les coûts fixes élevés pour la production des articles en format papier, la création de revues numériques remet en cause ce régime de type propriétaire.

À partir de là se pose la question de savoir quels sont les effets d'une transition d'un régime de propriété privée, dominé par le droit d'auteur, à une gestion commune du savoir dans le marché de la publication scientifique. Plus précisément, cet article vise à contribuer à la compréhension de la connaissance scientifique en tant que commun et à définir quelles sont les conséquences d'une gestion commune du savoir sur les différentes parties prenantes impliquées dans le marché de la diffusion de la connaissance scientifique. En se plaçant à un niveau d'analyse différent de celui adopté par Hess et Ostrom (2007), nous proposons une analyse qui vise à comprendre les défis imposés par la numérisation de l'information pour le passage à une gestion commune de la connaissance. La communauté scientifique n'est pas la seule à devoir répondre à ces défis. Elle doit être accompagnée par l'ensemble des acteurs économiques qui opèrent sur le marché de la publication scientifique. Une telle analyse vise à montrer que la numérisation de l'information n'a pas seulement impacté la structure des coûts du marché de la publication scientifique (Tenopir et King, 2000), mais aussi le rôle des différents acteurs dans la chaîne de valeur et donc la chaîne elle-même.

Pour comprendre comment les différents acteurs interagissent, dans un cadre où des normes exogènes (telles que le *copyright*) et endogènes (dont les normes CUDOS proposées par Merton (1973) et les licences *Creative Commons*) à la communauté scientifique

2. Dans le cadre de cette analyse, nous ne considérons que la connaissance scientifique exprimée sous forme d'articles scientifiques, considérés comme l'instrument standard pour la communication scientifique (Origgi, 2010).

contraignent les choix des agents économiques, l'*Institutional Analysis Development framework* (IAD) est utilisé en tant qu'outil de diagnostic (Kiser et Ostrom, 1982 ; Ostrom, 2009). L'IAD est le cadre d'analyse théorique le plus général ; il permet d'identifier les relations et les éléments fondamentaux et de fournir un ensemble de variables qui doivent être considérées pour étudier toute forme d'arrangement institutionnel. En tant que forme spécifique d'organisation, et donc d'arrangement institutionnel pour la production/diffusion de la connaissance, le marché de la publication scientifique est analysé à l'aide de l'IAD, apte à comprendre comment différentes formes organisationnelles permettent aux agents de résoudre des problèmes de coordination (Ostrom, 2011). Il est donc devenu l'instrument le plus utilisé pour l'analyse de la connaissance en tant que commun, car capable de conjuguer dans une seule analyse les différentes parties prenantes et les différents cadres institutionnels dans lesquels elles opèrent (Gibson et al., 2000 ; Moran et Ostrom, 2005). Comme cela a été souligné par Hess et Ostrom (2007), la flexibilité de l'IAD permet l'application à des contextes où le changement technologique a particulièrement affecté la gestion de la ressource, tel que c'est le cas dans le marché de la circulation du savoir scientifique. Sa capacité à faire émerger les situations de path dependence qui caractérisent les interactions entre agents (Oakerson, 1978) participants à la chaîne de valeur de la connaissance scientifique en justifie une fois de plus notre choix en tant qu'instrument primaire pour la présente étude.

Puisque l'IAD est caractérisé par un certain degré de flexibilité, qui en permet une application à un large éventail d'arrangements sociaux, ainsi qu'à différents niveaux de choix (Cole, 2017), il sera appliqué point par point au marché de la publication scientifique. Après avoir éclairci les caractéristiques de la connaissance en tant que commun, et identifié les acteurs impliqués dans le marché de la publication scientifique (I), l'article analyse les enjeux de la transition d'une régulation de type propriétaire à une gestion commune du savoir scientifique. À travers l'analyse de l'arène de l'action (II), les intérêts de chacun des agents économiques sont étudiés dans une perspective de coûts-bénéfices, à partir des différents business models proposés par les maisons d'édition. Les interactions entre les agents seront enfin placées dans la chaîne de valeur de la connaissance scientifique (III), qui montre comment des stratégies d'adaptation et d'innovation ont profondément changé

sa structure. Toutefois, les intérêts qui semblent être rivaux sur le marché de la publication scientifique pourraient être ramenés à l'équilibre, en considérant les différents services qui caractérisent ce marché à l'heure actuelle comme complémentaires.

5.1 La connaissance scientifique à l'aune de l'*Institutional Analysis Development framework*

Parmi les outils les plus utilisés pour l'analyse des communs se trouve l'IAD (Weible, Sabatier, 2018). Il permet d'étudier un système dans lequel les acteurs économiques interagissent à plusieurs reprises, ce qui est le cas d'un marché. L'IAD peut être utilisé pour analyser des situations dynamiques, où le changement technologique modifie la gestion de la ressource analysée. Sa généralité en permet une application à différents domaines d'étude, comme la création d'une nouvelle forme d'organisation pour la connaissance scientifique à travers les archives ouvertes (Hess e Ostrom, 2007 ; Beagles, 1999 ; Duncan, 1998), les limites d'une organisation de type propriétaire à l'ère du numérique (Boyle, 1996 ; Nelson, 2004 ; Cook-Deegan, 2007), les nouvelles formes de gestion décentralisée (Benkler, 2004 ; Benkler et Nissenbaum, 2006) ou le rôle des plateformes dans la chaîne de valeur du marché (Thiele et al., 2011).

Même si l'IAD n'a pas une structure rigide – il a subi des modifications dans le temps – il se compose généralement de trois piliers principaux, permettant l'analyse de certaines variables et processus (Cole et al., 2019), comme le montre la figure 5.2.

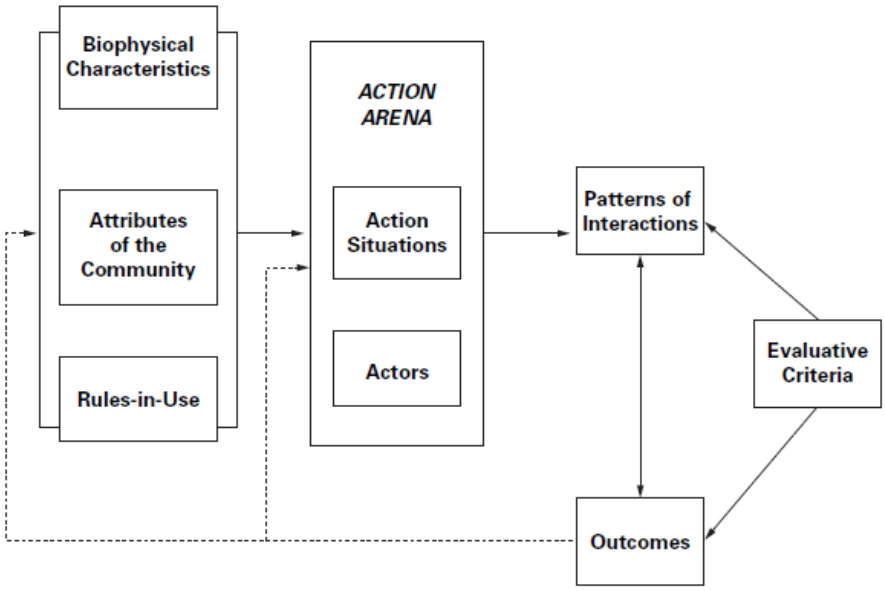

FIGURE 5.2 – *IAD framework* (Source : Hess et Ostrom, 2007, p.44)

Dans la première section, se trouve l'analyse des caractéristiques de la ressource, à savoir ses propriétés physiques, qui peuvent être affectées par un changement technologique, d'où la nécessité d'en déterminer des règles nouvelles pour en contrôler l'accès et l'utilisation. L'analyse se poursuit avec l'individuation des différents agents impliqués dans le contexte analysé. La communauté de référence est pourtant analysée comme un ensemble d'acteurs ayant un rôle très différent dans la gestion de la ressource. Ce cadre nous permet d'identifier les utilisateurs et les producteurs de la ressource, mais aussi le cadre normatif dans lequel ils interagissent. Ces normes, de nature étatique mais aussi sociale, contraignent le choix et les comportements des acteurs en déterminant les conditions d'accès, d'utilisation, de production et de maintien de la ressource. Ce premier cadre théorique est développé dans la section suivante, qui vise à comprendre les changements des caractéristiques de la connaissance scientifique, exprimée sous forme d'articles scientifiques, à la suite de la numérisation de l'information. Après avoir présenté les différents acteurs concernés par la gestion en commun du savoir, nous poursuivons notre analyse en identifiant les normes qui constituent le cadre d'action des agents.

Le cœur de l'IAD est l'arène de l'action, un ensemble de variables qui permettent d'analyser la situation où les différents agents se trouvent à interagir (Cole et McGinniss, 2017). Les acteurs sont analysés comme des individus ou des groupes, ayant des intérêts qui

peuvent être en contraste avec ceux des autres agents impliqués dans l'arène de l'action. Les agents interagissent dans l'action situation, abstraction de l'environnement de la prise de décision. Selon le degré de complémentarité entre intérêts, différents types d'outcomes peuvent résulter de l'interaction entre les acteurs. L'arène permet d'isoler l'émergence de comportements réguliers et de nouvelles stratégies d'interaction suite à une modification des caractéristiques de la ressource ou des incitations qui déterminent l'action des acteurs dans le marché analysé. Pourtant, le marché de la publication scientifique sous la perspective de l'arène de l'action sera expliqué à travers l'interaction entre les différents acteurs, interactions à géométrie variable et dépendant du business model adopté par les maisons d'édition.

C'est seulement à partir de la compréhension des objectifs des agents et de leur concertation que les outcomes peuvent être analysés. Les outcomes désignent les résultats de l'interaction des acteurs concernés dans l'analyse qui, grâce à la flexibilité de l'IAD, peuvent être analysés à différents niveaux de choix. Dans le cas de la connaissance scientifique, l'analyse des résultats peut concerner autant la création d'une archive ouverte que les différents modèles de diffusion présents à l'heure actuelle sur le marché de la publication ou encore l'impact de la numérisation sur la chaîne de la valeur de la connaissance scientifique. C'est véritablement dans ce dernier cas que notre analyse se situe.

5.2 De la numérisation de l'information à la connaissance scientifique en tant que commun

Comme il a été souligné par Hess (2007), la numérisation a permis l'émergence de nouveaux communs. L'identification des caractéristiques de la connaissance scientifique, mais aussi des agents impliqués dans sa gestion et du cadre normatif dans lequel ils opèrent représente l'étape préalable pour comprendre comment la gestion de la connaissance a évolué à la suite de la dématérialisation des articles scientifiques.

5.2.1 Les conditions de l'existence de la connaissance scientifique en tant que communs

L'objet de l'analyse ici proposée est la connaissance scientifique en tant que commun, à savoir une ressource non–excluable et non-rivale, dont la gestion est assurée par une communauté spécifique (Ostrom, 2009 ; Coriat, 2015 ; Boyle, 2003). Même si une riche littérature a pour objet l'impact de la numérisation de la connaissance (Bollier, 2001 ; Boyle, 2002 ; Lessing, 2001), l'étude de la connaissance en tant que commun dérive surtout de la collaboration entre Ostrom et Hess. Leur contribution principale est d'avoir appliqué les études sur le management et la gouvernance des common-pool ressources, au domaine des biens intellectuels.

La connaissance en tant que commun est une ressource ouverte et partagée, dans la mesure où aucune barrière n'en empêche l'entrée et lorsque sa gestion est soumise à une collectivité qui en détermine les modalités d'utilisation. La production même de la connaissance scientifique est un processus collectif (Cooke, 2001 ; Gibbson, 2004). Les agents économiques impliqués dans la production ont mis en place des normes qui permettent le développement de la ressource, tout en rémunérant, par le biais de paiements non-monétaires, le travail individuel (Merton, 1973). La numérisation a permis une circulation plus aisée de la connaissance, en dissociant la diffusion du savoir du support matériel. Cependant, elle n'est pas un flux dans l'air (Pénin, 2005), mais un ensemble d'idées qui circulent à travers une infrastructure numérique qui permet sa diffusion (Hess, Ostrom, 2003).

La circulation de l'information est toujours dépendante d'une structure physique et ce qui nous intéresse particulièrement ici est ce que Benkler (2000) appelle le content layer. Il s'agit du contenu des systèmes de communication, correspondant dans notre analyse à la connaissance scientifique. Si la numérisation a permis à la connaissance scientifique de devenir un bien non-rival, c'est la gestion des droits déterminant l'accès et l'utilisation qui permettrait de la définir comme un commun. Par conséquent, un nouveau système de gestion des monopoles octroyés aux producteurs des œuvres scientifiques est demandé afin d'assurer une gestion collective de la ressource. Cette gestion doit assurer un accès libre aux membres de la communauté et une utilisation partagée, où chaque membre est libre d'utiliser les biens sous-jacents.

5.2.2 Quelle communauté pour la gestion de la connaissance scientifique ?

La connaissance en tant que commun dépend de la gestion de la ressource, et l'identification des intérêts des parties prenantes impliquées, à savoir les producteurs, les consommateurs et les intermédiaires, devient l'étape préalable pour comprendre la régulation commune du savoir scientifique.

Les chercheurs peuvent être analysés comme des prosumers, étant donnée leur participation dans la chaîne de valeur en tant que producteurs et consommateurs du savoir scientifique. En qualité de producteurs, leur objectif principal est une diffusion du savoir la plus vaste possible, apportant comme rémunération du travail intellectuel une reconnaissance par les pairs, exprimée sous forme de citations. La publication devient ainsi un actif incorporel : selon la qualité de la revue dans laquelle les résultats sont publiés, les chercheurs peuvent être titularisés et obtenir des promotions (Heckman et Moktan, 2018 ; Osterloh et Frey, 2020). Les maisons d'édition occupent un rôle fondamental dans la chaîne de valeur de la connaissance scientifique. D'un côté, elles assurent la coordination de l'activité de certification de la connaissance, en centralisant le processus d'évaluation du savoir opéré par les rapporteurs. De l'autre, elles fournissent un service de circulation du savoir à travers la publication des articles. Leur objectif principal est la maximisation du profit, qui repose en partie sur la réputation. Les revues sont des intermédiaires du savoir scientifique et le business model adopté, fondé sur la détention d'un droit exclusif pour l'exploitation du bien intellectuel, leur a permis d'atteindre des super profits autour de 25-30% (Larivière et al., 2015). Leur activité a longtemps été justifiée comme une forme de valeur sociale pour la communauté scientifique, à travers la certification de la connaissance. Le processus de certification requiert la participation d'autres acteurs. En premier lieu, les éditeurs des revues, souvent membres de la communauté scientifique, dont l'expertise est reconnue dans un domaine précis, gèrent le processus de communication entre les auteurs et les rapporteurs, en donnant aussi un avis préalable à l'évaluation. Si l'article est en ligne avec les thématiques de la revue et suffisamment solide du point de vue théorique, il est transmis aux rapporteurs qui en évaluent le contenu. Une telle activité engendre des gains en termes de réputation qui peut ensuite être valorisée sur le marché du travail académique

par les éditeurs et les rapporteurs.

L'accès au savoir scientifique pour les chercheurs, en tant que consommateurs de connaissance, dépend principalement de leur université. Les universités représentent les consommateurs principaux de connaissance sur le marché de la publication (McCabe, 2000). Avant la numérisation de l'information, la cause du contrat entre bibliothèques et maisons d'édition était un véritable droit de propriété sur le bien sous-jacent, pour se transformer ensuite en un droit à l'accès, ayant une durée limitée. Ce n'est pas seulement le titre juridique qui a changé, mais l'objet lui-même. Avec la numérisation, l'objet du contrat n'est plus une et une seule revue scientifique, mais un bouquet de revues, dont la qualité est très variable. Avec ce système, les maisons d'édition peuvent utiliser les revues avec une réputation solide et établie comme levier, et enrichir le bouquet avec des nouveaux journaux, dont la qualité est incertaine. Une telle stratégie permet de créer une demande immédiate pour des biens dont l'achat serait autrement différé dans le temps. Elle permet en outre d'avoir un environnement moins risqué (Phillips, 2009) et d'affiner les stratégies de discrimination des prix (Edlin et Rubinfeld, 2004). Afin d'avoir un pouvoir de négociation plus fort, les bibliothèques ont créé des consortia (*e.g.* Couperin), qui organisent en concertation l'accès au savoir scientifique. Puisqu'on passe d'un droit de propriété à un droit à l'accès, les consommateurs se trouvent dans une situation de lock-in, où l'accès à la connaissance antérieure est soumis au paiement continu aux maisons d'édition (Merret, 2006), situation qui témoigne d'une appropriation privée d'un bien public (Pirie, 2009).

5.2.3 La régulation du marché de la publication scientifique : entre public order et private ordering

L'existence même de la connaissance comme commun dépend de la mise en place d'une structure de gouvernance collective visant à faire face au dilemme social auquel la nature de bien public de la connaissance est soumise. La gestion est ainsi entendue comme un ensemble de règles qui déterminent l'accès, la contribution, l'exclusion et l'aliénation du bien, et qui sont capables d'assurer l'alignement des différents intérêts des parties prenantes – qui peuvent être opposées – et de créer des équilibres résultants dans une instabilité à long terme. Une reconfiguration des règles régissant la gestion commune de la

connaissance devient nécessaire pour atteindre les objectifs des participants au commun. L'élément principal du modèle standard de la publication scientifique est la détention d'un droit de propriété intellectuelle, qui permet l'exclusion à la ressource et l'imposition d'un prix à l'accès. Le droit d'auteur a été fondamental pour la diffusion du savoir scientifique. Sa centralisation au sein d'une même organisation, de type maison d'édition, a permis une circulation du savoir qui aurait été impossible autrement, du fait des coûts fixes très élevés liés à la production et à la gestion logistique du support matériel. Avec la réduction importante des coûts de publication et l'émergence d'instruments alternatifs pour la diffusion de la connaissance, une gestion de type propriétaire ne semble plus être le modèle le plus adéquat. Au contraire, elle empêcherait une gestion commune du savoir. Une telle situation amène inévitablement à une concurrence entre les intérêts des consommateurs de la connaissance, à savoir la maximisation de l'accès et de son utilisation, et ceux des producteurs, c'est-à-dire entre chercheurs et maisons d'édition. Comme on l'a vu, ces dernières visent la maximisation de leurs profits, assurée par l'existence d'une barrière à l'accès, tandis que les premiers souhaitent une diffusion de leurs résultats la plus vaste possible [3]. Cependant, la nature libre de la connaissance ne doit pas être entendue comme un « *free beer* », mais comme un « *free speech* » (Lessing, 2006). Bien que l'existence d'un prix à l'accès limite la diffusion du savoir, si le prix reste non-discriminatoire, la ressource peut continuer à être considérée comme une ressource libre. Ce qui menace l'openess de la ressource, et donc l'établissement du commun, est l'existence du droit exclusif entre les mains d'un seul membre de la communauté. Si les caractéristiques économiques de la connaissance ont changé avec la numérisation, pour la mise en place d'un commun il est nécessaire de reconfigurer les modes de gouvernance du bien sous-jacent. Sans une telle réforme, la numérisation pourrait avoir l'effet opposé, à savoir une second enclosure (Boyle, 2003). Dans un régime de type propriétaire, les maisons d'édition sont en même temps

3. L'objectif de la maximisation de la visibilité d'un chercheur, à travers la diffusion de ses résultats la plus vaste possible, s'explique par le système de rémunération indirectement monétaire régissant la communauté scientifique. Ce système, fondé sur des normes endogènes de la communauté scientifique, pousse le chercheur à céder le *copyright* à une maison d'édition contre la publication de ses résultats scientifiques. La publication constitue ainsi un actif incorporel, qui sera valorisé dans le marché du travail (Heckman et Moktan, 2018). Pour une explication approfondie des normes endogènes de la communauté scientifique voir Merton, Robert K. *The sociology of science : Theoretical and empirical investigations.* University of Chicago press, 1973 et Rai, A. K., Regulating scientific research : Intellectual property rights and the norms of science. *Nw. UL Rev.* 94 (1999) : 77 ».

les titulaires de la composante économique du droit d'auteur, titre qui leur permet de déterminer les conditions à l'accès et à l'utilisation de la ressource, et les intermédiaires de connaissance entre auteurs et consommateurs. C'est ainsi que les licences *copyleft* ont trouvé leur application dans le marché de la publication scientifique, agissant en tant que forces de désintermédiation et favorisant les transactions end-to-end (Carroll, 2006). Bien que dans le contexte académique des exceptions à la réutilisation du savoir scientifique existent déjà (*e.g. fair use*), les licences *copyleft* (dont les licences *Creative Commons* sont un exemple) deviennent essentielles vis-à-vis des nouvelles possibilités d'exploitation du savoir scientifique dérivant des nouvelles technologies. Ainsi, au lieu de transférer le droit d'auteur aux maisons d'édition, les auteurs restent titulaires du droit de la propriété intellectuelle et attribuent aux œuvres de l'esprit une licence *Creative Commons* (CC) qui, dans sa forme la plus « ouverte » (CC-BY), permet une exploitation du bien auquel la licence est liée avec la seule limite d'attribuer à l'auteur la paternité de l'œuvre originale. Ces licences permettent de réduire les coûts de transaction, en autorisant les consommateurs à « lire, télécharger, copier, distribuer, imprimer, rechercher, créer un lien vers le texte de ces articles, les explorer pour les indexer, les transmettre en tant que données à des logiciels, ou de les utiliser pour tout autre fin licite, sans barrières financières, juridiques ou techniques autres que celles inséparables de l'accès à Internet lui-même » (Budapest Open Access Initiative, 2002). L'utilisation de la ressource originale fait référence à ce que Wiley (2017) appelle les 5R, à savoir la réutilisation (droit d'utiliser le contenu de diverses manières), la révision (droit d'adapter, d'ajuster, de modifier ou d'altérer le contenu lui-même), la redistribution (droit de partager des copies du contenu original), le remix (droit de combiner le contenu original avec d'autres contenus) et la conservation (droit de posséder et contrôler les copies du contenu de l'œuvre) du bien. Une telle utilisation de la connaissance ouvre différentes possibilités à sa valorisation socio-économique. En effet, si librement accessible, la connaissance pourrait augmenter le bien-être global (May, 2019 ; Houghton et Sheehan, 2009).

L'utilisation de la connaissance scientifique, selon les termes ci-dessus énoncés, impose des exceptions à l'exclusivité garantie par le *copyright*, tout en empêchant la tragédie des anticommons décrite par Heller (1998), où la présence d'un droit de propriété exclusif

limite le développement des communs, en créant des barrières à la création d'innovation. Les licences CC, exemple d'une régulation *bottom-up* (Lessing, 2002), sont des contrats de nature privée qui fondent leur existence sur un instrument de public order – le droit d'auteur – et qui rendent licites des actes qui ne les seraient pas autrement. Les exceptions à un droit exclusif trouvent leur justification dans le fait que les incitations offertes par un droit de propriété ne sont plus indispensables pour atteindre le taux optimal de créativité et d'innovation, puisque les bases sociales et technologiques de la créativité même ont changé.

5.3 L'arène de l'action : coordination ou compétition entre intérêts ?

L'interaction entre les agents ci-dessus analysés est au cœur de l'arène de l'action de l'IAD. Cette interaction dépend des formes d'organisation sous-jacentes à la gestion de la dissémination du savoir scientifique. Si avant l'ère du numérique le régime propriétaire exprimé à travers le modèle *Closed Access* monopolisait le marché de la publication scientifique, la dématérialisation du support papier a promu l'émergence de nouvelles formes d'organisation de la circulation du savoir scientifique. Sous le modèle *Open Access*, on retrouve différents modèles de business, ayant chacun des formes de gestion du savoir spécifiques.

5.3.1 Le *Closed Access* à l'ère du numérique

Avant la numérisation de l'information, le modèle de type propriétaire a été capable d'offrir une première solution, bien que de second best (Arrow, 1962), à la création privée d'un bien public. L'existence d'un droit de propriété permet aux maisons d'édition de rémunérer leurs investissements et de couvrir les coûts fixes élevés, impossible dans un marché en concurrence pure et parfaite, en imposant un mark-up price. Une telle circulation des résultats scientifiques aurait été autrement impossible si soutenue au niveau individuel par les chercheurs, qui peuvent ainsi maximiser leur visibilité. D'autre part, grâce au signal de qualité offert par les revues scientifiques, les chercheurs-consommateurs peuvent discriminer la qualité de l'information, et avoir accès à la connaissance qualitativement

importante.

Ce schéma de fonctionnement de la circulation du savoir, qui se fonde sur l'exclusion à l'accès au savoir, a été appelé *Closed Access* (CA). À partir de la numérisation de l'information et de la réduction des coûts de diffusion de la connaissance, une restriction à l'accès via l'imposition d'un prix, ne semble plus être justifiée. Les super-profits des maisons d'édition dérivent avant tout de la gratuité des inputs, qui dans d'autres secteurs de l'industrie de la publication représentent 12-15% des coûts totaux (Buranyi, 2017). Ce contenu est ensuite soumis au système de peer-review, lui aussi rémunéré entièrement en termes de réputation. Si les coûts de rédaction étaient à la charge des maisons d'édition, avec la numérisation ils sont presque entièrement couverts par les auteurs. Cependant, les maisons d'édition continuent à revendiquer leur apport majeur dans la commercialisation de la connaissance, en ciblant les consommateurs et les disciplines, tâche qui serait impossible à accomplir par les seuls éditeurs (May, 2019). Ce modèle se traduit par une visibilité mineure pour les auteurs, une situation de lock-in pour les bibliothèques, vis-à-vis de la connaissance antérieure, dont l'accès est soumis à un abonnement, et par un double paiement de l'activité de recherche pour la société.

5.3.2 Différents modèles pour différents intérêts : Les modèles *Open Access*

De nouvelles formes de circulation du savoir ont émergé à partir de la numérisation de l'information, ce qui a permis au savoir scientifique de s'inscrire potentiellement dans la sphère des communs. Le modèle qui s'oppose au CA est l'*Open Access* (OA), qui trouve son expression sous différentes formes (Figure 5.3), ayant toutes comme objectif primaire l'élimination de toute barrière à l'accès et à l'utilisation de la ressource, comme établit dans la Conférence de Budapest (2002).

Bien que l'objectif de chaque modèle soit le même, les modalités adoptées pour l'atteindre et pour satisfaire les différents intérêts des parties prenantes diffèrent. Le *Black Open Access* (Björk, 2017) est un modèle d'accès à la littérature scientifique illégal, fondé sur l'appropriation frauduleuse d'œuvres protégées par le droit d'auteur (*e.g.* Sci-hub, LibGen). Ce modèle offre désormais accès à la plupart de la littérature scientifique existante

(Himmelstein et al., 2018). Il n'est donc pas un véritable modèle d'OA, mais il représente l'outil de piratage le plus utilisé par la communauté scientifique et la cible des autorités publiques du monde entier (Elsevier Inc. v. Sci-Hub 1 :15-cv-04282 ; TGI Paris, 3e ch. 4e sect. jugement en la forme des référés le 7 mars 2019). Un tel modèle porte une atteinte très limitée aux maisons d'édition, car les bibliothèques continuent à maintenir leurs abonnements aux revues scientifiques. Si pour les auteurs ce modèle semble augmenter leur visibilité (Ferrucci et al., 2018), et que les consommateurs ont une dissuasion faible à l'utiliser, du fait de la faible probabilité d'être incriminés, ce modèle représente un coût pour la société. En plus du financement de l'accès au savoir, s'ajoute le coût du contrôle et de la répression des comportements illicites. Au final, le *Black Open Access* ne constitue pas une véritable alternative au modèle *Closed Access* ; au contraire, il provoque une augmentation des free-riders parmi les membres gérants le commun.

Un autre modèle qui ne se propose pas comme substitut direct du CA, mais plutôt comme un outil complémentaire à la diffusion du savoir est le *Green Open Access*. Ce modèle permet l'accès au savoir scientifique avant la phase de publication, et l'archivage des articles après leur publication dans des répertoires ouverts. Pourtant, le *Green Open Access* permettrait de sortir du lock-in dans lequel se trouve la communauté scientifique face à l'accès à la connaissance antérieure, à travers la création d'un outil d'archivage alternatif à celui proposé par les maisons d'édition. Si un tel système permet la discussion, et donc l'amélioration du contenu scientifique avant même son approbation par les rapporteurs, ainsi que la possibilité de créer de nouvelles coopérations entre chercheurs, il rend difficile la citation des œuvres, à cause des métadonnées volatiles (May, 2019). Puisque ce modèle pourrait impacter les profits des maisons d'édition, celles-ci ont adopté des politiques spécifiques pour garder leur contrôle exclusif sur la diffusion des articles scientifiques. On parle de politiques de publication *Green*, blue, yellow et white selon que la maison d'édition autorise le téléchargement de la version publiée de l'article scientifique dans une archive ouverte, ou seulement la version finale, ou une version antérieure à celle finale ou qu'elle interdit tout téléchargement [4]. Même si dans la plupart des cas les politiques adoptées par les maisons d'édition ne sont pas très restrictives (Swan et Brown, 2005), l'accès à 80% des

4. Voir les politiques à la publication classées par Sherpa : https ://www.sherpa.ac.uk/romeo/search.php. Dernier accès 29/02/2020.

articles publiés reste encore soumis à un prix d'accès (Green, 2018 ; Fortson, 2019).

Acteurs	Intérêts	Closed Access	Open Access				
			Gold			*Green*	*Black*
			APC	Hybrid	Platinum	Green, Blue, Yellow, White	
Maison d'édition	Profits > 0	Contrôle de la dissémination de la connaissance scientifique à travers le Copyright	Payement pour la publication des articles scientifiques	Double payement pour certains articles, tout en gardant le modèle Closed Access	Financement par les institutions publiques	Services complémentaires pour la dissémination (Veille, discussion de papier pre-print, archives pour les articles publiés)	Risque de réduction des abonements souscrits individuellement.
Consommateurs	Accès à la connaissance scientifique + réutilisation	Contrainte budgétaire indirecte	Accès *libre* à la connaissance scientifique	Accès libre à une partie de la littérature + contrainte budgétaire indirecte	Accès *libre* à la connaissance scientifique	Accès aux prépublications Accès aux publications en CA après un certain délai	Accès illimité, mais probabilité dedétection de comportement illégal
Producteurs	Réputation = Visibilité + Qualité	Publication sur revue avec IF élevé	a) Majeure visbilité b) Recherche de financements pour les frais d'inscriptions c) Effets de réputation	Choix entre payement pour la publication et majeur visibilité vs. pas de frais mais visibilité reduite	Plus grande visibilité	Avoir des retours avant la publication Bénéficier d'une plus grande visibilité après la publication en CA	Majeure visibilité et moins d'incitation pour contribuer au développment du Green OA
Referee	Réputation et contrôle de la qualité de la nouvelle connaissance produite	Assignation en fonction de l'expertise	Aucun impact sur leur rôle	Aucun impact sur leur rôle	Peer-review ouvert à tous les chercheurs	Aucun impact	Aucun impact sur leur rôle
Bibliothèques	Assurer l'accès et la conservation de la connaissance scientifique	Payement des abonnements a) pour avoir l'accès au savoir scientifique b) pour maintenir l'accès sur plusieures années c) Coût du maintien des copies en format papier	Pas de financement pour l'accès à la connaissance car décentralisation de l'accès	Maintient des abonnements publiés en Closed Access	Pas de frais d'accès ni de publication	Contribution au développment des archives ouvertes et centralisation de l'accès (création d'archives interfacultés)	Aucun impact sur le financement des abonnements.
Société	Accès et production du savoir scientifique	a) Financement de la recherche b) Financement des bibliothèques c) Financement privé	Accès libre au savoir Financement des frais de publication Financement de la recherche	Payement des frais de publication et des abonnements Financement de la recherche	Financement de la production et de la publication de la recherche , limitations à la valorisation économique accès libre à la connaissance finale	Coût de maintient et de mise en service des archives, payement des services complémentaires pour faire face à l'abondance d'information	Aucun impact sur le financement de la recherche mais augmentation des coûts pour le contrôle des comportements déviants

FIGURE 5.3 – Modèles de dissémination du savoir scientifique

Le véritable modèle alternatif au CA est le Gold *Open Access*, un système de publication qui n'impose aucuns frais à l'accès à la connaissance scientifique et qui élimine toute barrière légale à l'utilisation de la connaissance. En suivant les recommandations de la conférence de Budapest (2002), les articles publiés en *Open Access* devraient être accompagnés par la licence CC-BY, qui permet de modifier, partager et commercialiser le bien, à condition d'attribuer la paternité de l'œuvre à son auteur. Selon le business model adopté, le Gold *Open Access* devient *Article Processing Charges Open Access* (APC), Hybrid *Open Access* ou Platinum *Open Access*. L'APC OA place les coûts de publication à la charge des chercheurs-producteurs. Très utilisé par les maisons d'édition établies, ce modèle mène néanmoins à une baisse de leurs profits (Annual Report on Form 20-F 2018). Même si la présence d'un coût à la publication n'empêche pas la ressource d'être définie en tant que open, des frais de publication trop élevés peuvent représenter des barrières au développement du commun. À partir des données de INTACT Open APC Initiative [5], la figure 5.4 montre comme les APCs peuvent varier beaucoup entre maisons d'édition. On constate que les prix à la publication sont plus élevés pour les maisons d'édition établies que pour les nouveaux entrants, tels que PLoS ou Frontiers Media, dont les frais de publication sont en partie couverts par des subventions publiques. En revanche, si on considère la période 2013 [6] – 2018, l'augmentation des frais à la publication a affecté presque toutes les maisons d'édition, (+40% PLoS, +25% Elsevier, +56% Frontiers), permettant d'augmenter les profits dérivant de la publication en *Open Access*. Malgré cette augmentation de prix, ce modèle permet un accès élargi et une réutilisation du savoir majeure. Cependant, les maisons d'édition établies adoptent rarement la licence CC-BY, car cela empêche le maintien d'un contrôle exclusif sur la connaissance [7]. L'utilisation de licences plus restrictives,

5. INTACT OA (Transparent Infrastructure for free-based *Open Access* Publishing) est un projet fondé en 2015 par le German Research Foundation, qui vise à créer une trace de l'évolution des coûts des publications en OA. La base de données Open APC Initiative est développée sur base volontaire par 234 institutions et fournie des informations sur les coûts et les modalités de publication pour plus de 77 000 articles scientifiques.

6. Le 2013 a été choisie comme année de départ pour voir l'évolution des frais à la publication, puisqu'elle est l'année pendant laquelle les maisons d'édition établies ont commencé à tourner leur attention vers le marché de la publication en *Open Access*.

7. À titre d'exemple, on trouve chez Springer et Wiley que plus de 50% des articles publiés n'ont pas une licence CC-BY et même si Elsevier publie 2/3 des articles en adoptant cette licence, moins de 10% de ses journaux font parties des journaux indexés sur DOAJ *Open Access* journals. Le DOAJ est un service

telles que CC-BY-ND ou CC-BY-NC limite la réutilisation de la connaissance, notamment en termes d'analyse de texte et données, ou encore en empêchant l'auteur de télécharger son article sur d'autres plateformes commerciales.

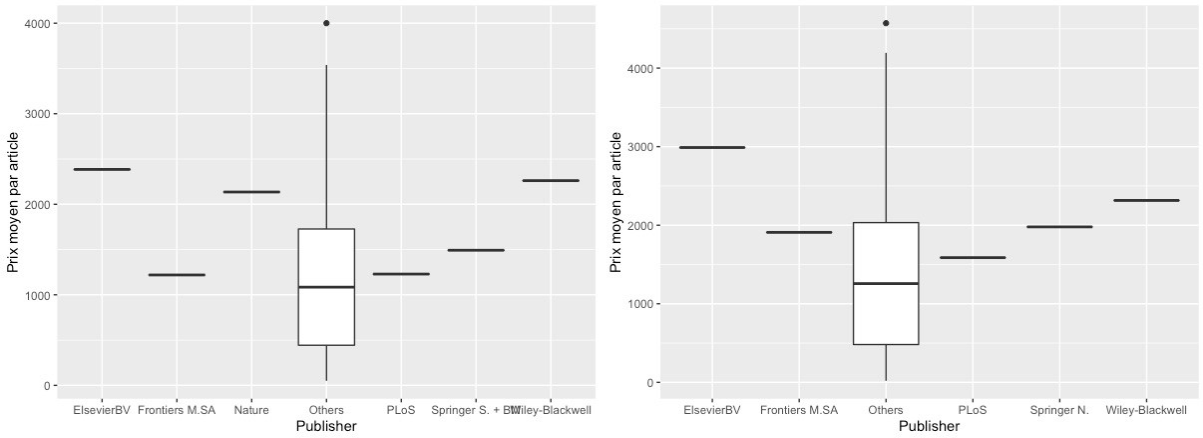

FIGURE 5.4 – Moyenne APC par maison d'édition – année 2013 et année 2018

Le modèle *Open Access* est de facto moins rentable que le *Closed Access* pour les maisons d'édition. Si avec ce dernier les profits augmentent avec l'augmentation des consommateurs, dans le modèle APC les profits ne dépendent que des articles publiés, qui sont en nombre fixe. Pour cette raison, ce sont surtout les maisons d'édition avec une réputation établie qui utilisent un modèle dit hybride, comme le montre la Figure 5.5.

qui garantit la transparence du management des revues OA, le respect des principes imposés par l'*Open Access* Initiative, et la qualité des journaux.

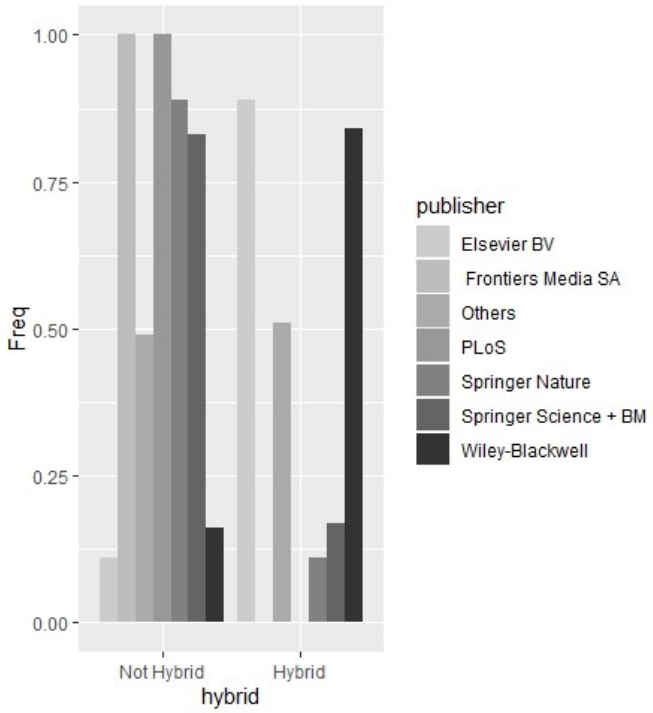

FIGURE 5.5 – Modèle Hybride par maison d'édition

Dans le cas du modèle hybride, l'OA reste une option dans un système de publication qui est en CA. Si la revue continue à être vendue sous un modèle CA, certains articles sont mis en libre accès, suite au paiement des frais de publication. Ce modèle s'avère profitable pour les maisons d'édition et bénéfique pour les chercheurs. Les premières peuvent ajouter aux profits dérivant du modèle en CA les frais pour la publication d'articles contenus dans les mêmes journaux et les scientifiques acquièrent une visibilité majeure et un accès élargi au savoir. Cependant, ce modèle est coûteux pour la société, qui est soumise à un double payement pour avoir accès au même bien.

La dernière forme de Gold OA est l'OA Platinum, qui n'impose aucuns frais de publication. Ce modèle, disruptif pour les maisons d'édition commerciales, se propose comme un modèle totalement non-profit. Pourtant, au lieu d'adopter les licences CC-BY, il interdit toute forme d'exploitation commerciale du bien sous-jacent (Fuchs et Sandoval, 2013). Si ce modèle semble être particulièrement adapté pour le libre accès au savoir scientifique, il limite l'émergence potentielle d'activités commerciales pour la valorisation économique de la connaissance en créant une perte pour la société. Il existe quand même des exceptions à ce type de modèle, comme la revue economics e-journal, qui tout en préservant la gratuité

de la publication, n'impose pas de restrictions à la réutilisation de l'article.

5.4 Entre innovation et adaptation : l'évolution de l'*Open Access*

De nouveaux modèles d'affaire ont été adoptés par les maisons d'édition établies, afin de préserver leurs intérêts, et d'autres ont vu le jour, à partir de la communauté scientifique même, afin de mieux servir l'intérêt commun de l'avancement de la science. Plus qu'un changement de modèles de business, la numérisation de l'information a changé la chaîne de valeur de la connaissance scientifique. S'il est vrai que la pénétration du marché a été faible en ce qui concerne les nouveaux concurrents des maisons d'édition (Björk, 2017), des nouveaux acteurs sont devenus parties prenantes dans la chaîne de valeur de la connaissance, en offrant des services complémentaires à celui de la diffusion des résultats scientifiques. Au lieu d'avoir un marché où l'existence des droits de la propriété intellectuelle est justifiée par une pénurie d'innovation et créativité, la numérisation a mené à une abondance d'information. C'est à partir de cette abondance que toute une série de services a émergé, afin d'assurer une meilleure gestion du savoir scientifique.

5.4.1 La chaîne de valeur de la connaissance scientifique : complémentarité de modèles pour la *Grande Conversation scientifique*

La chaîne de valeur est un ensemble d'activités interconnectées au sein d'une organisation (Porter, 1985), et donc un instrument d'analyse qui peut être adapté aux flux de l'information pour en comprendre la création de valeur (Curry, 2016). En se focalisant sur les seules activités primaires de la chaîne de valeur de la connaissance scientifique au sein de l'organisation responsable de sa circulation, on peut identifier quatre catégories primaires, à savoir la production, la discussion et donc la certification, la diffusion et l'accès au savoir.

Si la production et le financement de la connaissance restent à la charge des institutions publiques et privées, le processus de discussion et de certification des résultats scientifiques a subi des transformations remarquables. Avant tout, la création d'archives ouvertes permet la circulation et la discussion des résultats avant même leur publication, ce qui peut avoir

un impact positif sur la qualité de la connaissance et entraîner une véritable conversation scientifique. Un tel service est offert, par exemple, par ResearchGate, plateforme en ligne qui permet entre autres de formuler des questions techniques et de recherche, dans un esprit de production collective de la connaissance.

Le système de peer-reviewing a également changé. Aujourd'hui, il existe 22 types d'évaluation du savoir scientifique (Ross-Hellauer, 2017), résultant d'une multitude de combinaisons entre l'identité des rapporteurs et des auteurs, mais aussi d'une participation de la communauté scientifique préalable à l'évaluation par le système d'évaluation standard. Cette open discussion, à laquelle chaque membre de la communauté peut participer et où chacun peut améliorer la qualité du manuscrit, pourrait contribuer à la diminution du nombre d'articles rétractés après la publication[8]. Erreurs, fautes et plagiat, qui constituent la plupart des motifs de rétractation (Fang et al., 2012), pourraient diminuer si les articles étaient soumis à l'examen d'une communauté plutôt que d'un comité très restreint de rapporteurs.

L'un des changements majeurs concerne la diffusion des résultats scientifiques à travers les différents modèles qui ont été analysés auparavant. Ces modèles s'inscrivent dans une complémentarité de supports pour la circulation du savoir scientifique, comme le démontre le cas des archives ouvertes. Si ces répertoires sont nés sous l'impulsion d'institutions publiques (*e.g.* HAL, Repec, arXive) pour permettre l'indépendance face à l'accès à la connaissance antérieure détenue par les maisons d'édition, d'autres acteurs ont trouvé leur place dans ce marché. Tel est le cas de SSRN, archive fondée par deux économistes, qui est devenue en 2013 la plus grande archive de contenu scientifique au monde. En analysant les modèles de publication, on trouve l'entrée de nouveaux concurrents comme PLoS et BioMed Central, fondés en 2000, et Frontiers Media créée en 2007. Ces nouveaux acteurs proposent des APC plus faibles par rapport aux maisons d'édition établies et Biomed n'adopte ce modèle que deux ans après sa naissance. Cependant, les maisons d'édition ont elles aussi développé des modèles à partir de ceux des nouveaux concurrents, comme le montre le cas de Scientific Report, fondé par Springer en 2011, en concurrence directe avec Plos One.

8. La rétractation d'articles scientifiques après la publication a augmenté de dix fois dans les dix dernières années (Brainard, 2018).

Avec l'abondance d'information qui caractérise l'âge du numérique et l'élimination d'un coût à l'accès au savoir, les consommateurs et producteurs se trouvent à devoir gérer une quantité toujours croissante de données. Articles prépubliés, rapports de colloques et publications coexistent dans ce panorama hétérogène de la connaissance scientifique, où chaque jour 4 000 nouveaux articles sont publiés (STM Report, 2015). La réputation devient alors une valeur fondamentale pour les chercheurs et pour les maisons d'édition. L'indicateur le plus utilisé pour mesurer la qualité des revues scientifiques est l'*impact factor* (IF), qui mesure le nombre de citations reçues par la revue au cours des deux dernières années (Amin et Mabe, 2000) [9]. Des services de classification de l'IF des différentes revues ont émergé, tout au long des années, afin de centraliser l'information de la qualité des revues (*e.g.* Scopus, Scimagojr, DOAJ) et de permettre aux chercheurs-producteurs de planifier leurs choix stratégiques en termes de publication, et aux chercheurs-consommateurs de maximiser la quantité de connaissance consultée, sous contraintes budgétaire et temporelle.

D'autres acteurs vont enrichir le marché de la publication scientifique. Ce sont les services de personnalisation de la veille scientifique, qui permettent d'établir des critères de recherche pour rester à jour sur les nouveaux articles publiés (*e.g.* Academia.edu, ResearchGate, Researcher), mais aussi pour gérer de façon personnalisée de l'archivage de l'information (*e.g.* Mendeley, Zotero). Souvent, ces acteurs fournissent des services typiques des réseaux sociaux aux chercheurs, en gérant et facilitant les liens entre les membres de la communauté scientifique et en fournissant des signaux de qualité à travers la mesure de citation et de consultation des articles (*e.g.* Scopus, Academai.edu, ResearchGate).

La plupart de ces activités se fonde sur la libre circulation du savoir et donc sur la licence CC-BY. Dans une gestion de type commun, l'accès à la ressource implique aussi sa libre réutilisation, en ce sens que personne ne détient un droit exclusif sur l'exploitation ou sur l'interdiction de son utilisation. Le manque de telles licences dans la publication (stratégie très utilisée par les maisons d'édition établies) limite la liberté de réutilisation du savoir et crée une situation d'incertitude, où l'utilisation du bien sous-jacent est soumise à des asymétries d'information. Cela amène à une réduction de l'utilisation de la connaissance, par exemple à travers le text et data mining, mais aussi à la création de nouvelles barrières

9. En revenant aux stratégies de publication des maisons d'édition établies, on note que de Scientific Report a un IF plus élevé que Plos One (Davis, 2017).

pour l'émergence de services complémentaires dans le marché de la publication scientifique, comme le montre le cas de violation du droit d'auteur soulevé par ACS contre ResearchGate (ACS v. ResearchGate, United States District Court District Of Maryland, 10-02-2018).

5.4.2 Adaptation, innovation et réaction : quelle gestion commune du savoir scientifique ?

La chaîne de valeur de la connaissance scientifique s'est complexifiée avec la numérisation de l'information, comme le montre la figure 5.6. La présence d'un marché en oligopole, fondé sur des droits de propriété intellectuelle, n'a pas pour autant empêché l'entrée sur le marché de nouveaux acteurs capables de satisfaire une demande nouvelle, à savoir la gestion de l'abondance d'information.

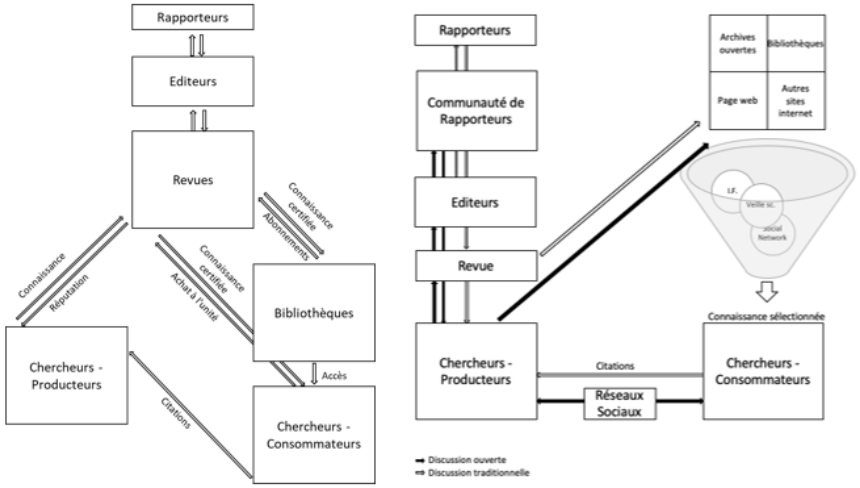

FIGURE 5.6 – Chaîne de valeur pré et post numérisation de l'information

Il est intéressant de noter l'évolution de l'analyse de la connaissance en tant que commun en deux trajectoires différentes.

D'un côté, les entreprises établies ont fait résistance au passage d'un régime de propriété privée à une gestion commune du savoir scientifique, les DPI étant l'instrument le plus profitable. En revenant aux données fournies par INTACT *Open Access*, on observe que sur la période 2005-2007, IOP Publishing, maison d'édition à but non-lucratif, et Springer Science + BM garantissent plus de 97% des publications en OA et jusqu'en 2012, les seuls deux nouveaux entrants avec un poids non-négligeable sont Plos et Frontiers Media,

responsables du 25% des articles OA publiés. Elsevier ne fait son apparition qu'en 2013, suivi en 2014 par Wiley-Blackwekll et en 2015 par Springer Nature, suite à l'acquisition de Springer Science + BM. Nous l'avons vu, souvent l'OA ne reste qu'une option pour les maisons d'édition établies et leurs APC sont au-dessus de la moyenne, ce qui décourage les universités à adopter le modèle *Open Access* (Muller et Watt, 2010).

D'un autre côté, les nouveaux concurrents du marché de la publication scientifique ont apporté des modifications importantes à la chaîne de valeur de la connaissance scientifique. En adoptant des modèles de business non-profit, ces nouvelles plateformes d'intermédiation du savoir (comme Plos) proposent des APC plus faibles et souvent elles couvrent les coûts à travers des financements publics. De plus, ce sont ces revues qui ont proposé un changement dans le système de peer-reviewing en élargissant la discussion à tout membre de la communauté scientifique (*e.g.* economics e-journal). Cependant, les innovations ne s'arrêtent pas au niveau des modèles de publication. L'avantage compétitif des nouveaux acteurs a résidé dans leur capacité d'exploiter une nouvelle demande dérivant de l'abondance d'information. C'est ainsi que, comme nous l'avons constaté, les répertoires ouverts et les bibliothèques numériques personnalisées, les services pour la veille scientifique et les nombreux services pour l'analyse de la qualité des revues, mais aussi pour l'impact scientifique de chaque chercheur et de chaque article scientifique sont nés.

Ainsi, il est possible d'enrichir les phases du développement de l'OA proposées par Laakso et al. (2011), avec ce qu'on appelle une phase de réaction. Pendant les années pionnières (1993-1999) le nombre d'articles publiés en OA était relativement modeste et les coûts de production étaient couverts par le travail bénévole de certains chercheurs. C'est pendant les années de l'innovation (2000-2004) que l'OA est devenu un modèle de publication compétitif par rapport au modèle standard en CA, avec l'émergence de nouvelles maisons d'édition telles que PLoS et BioMedCentral. Les auteurs proposent une dernière phase de développement de l'OA, dite de consolidation (2005-2009), où l'*Open Access* devient un système plus structuré, avec la création du Directory of *Open Access* Journal (DOAJ) comme principal indice pour le classement des revues en OA. À cette phase, il est possible d'ajouter une période dite de réaction où les maisons d'édition établies, conscientes des opportunités créées par la numérisation de l'information à côté du système traditionnel

de circulation du savoir, ont commencé à répondre aux stratégies d'innovation des nouveaux concurrents, avec une approche de type wait for scale qui caractérise nombreuses startups dans le domaine du numérique (Fortson, 2019). En 2013, Elsevier achète Mendeley (Lunden, 2013 ; Bonasio, 2013) qui, à travers une forme freemium, offre des services pour la gestion des références bibliographiques et des citations. Elsevier toujours achète en 2016 SSRN (Gordon, 2016), devenue en 2015 l'archive la plus importante au monde, et en 2017 Plums Analytics, société pour le classement des revues scientifiques. De son côté, Springer achète en 2008 Biomed Central et à partir de 2015 Springer Science + BM font désormais partie du portefeuille de Springer Nature. En 2013, Holtzbtinck Publishing, dont Springer Nature fait partie, devient la société-mère de Frontiers Media, fondée en 2007 comme revue non-profit. Les actions entreprises contre les nouvelles plateformes comme ResearchGate et Academia.edu (Van Noorden, 2017) montrent par ailleurs la volonté des maisons d'édition de type traditionnel de défendre le monopole – garanti par les droits de la propriété intellectuelle – qui protège leur activité principale, à savoir le contrôle de la diffusion du savoir scientifique.

Conclusion

La numérisation de l'information n'a pas modifié que la structure des coûts du marché de la publication scientifique. L'application de l'IAD à ce marché nous a permis de faire émerger les changements majeurs qui ont modifié la chaîne de valeur de la circulation de la connaissance scientifique. De nouveaux business models ont émergé, chacun satisfaisant de façon différente les intérêts des parties prenantes du marché de la publication du savoir scientifique. Si les maisons d'édition établies ont essayé d'adapter leurs business models pour maintenir des profits élevés, des modèles de publication non-profits semblent mieux satisfaire aux exigences des membres de la communauté scientifique. Cependant, une gestion entièrement assurée par les membres de la communauté scientifique reste une utopie difficilement réalisable, étant donné le rôle des chercheurs comme producteurs du savoir et non comme entrepreneurs responsables du management de la connaissance.

Une gestion commune de la circulation du savoir scientifique devrait pourtant être en-

tendue comme un renouvellement de la collaboration existant entre les acteurs principaux du marché de la publication scientifique, collaboration élargie aux nouvelles parties prenantes. Cette complémentarité entre différents acteurs et services est possible seulement dans un système où des contrats privés posent des exceptions aux limites imposées par l'exclusivité du droit d'auteur et qui permettent d'éviter une deuxième enclosure de la connaissance scientifique. C'est ainsi que les licences *Creative Commons* permettent de continuer à assurer les intérêts des chercheurs et des acteurs impliqués dans la gestion du savoir. D'un côté, elles permettent un accès élargi à la connaissance et une plus large réutilisation des données de la recherche scientifique, ce qui pourrait avoir un impact positif non seulement pour la société (Tennant et al., 2006 ; ElSabry et ElHassan, 2017), mais pour l'économie aussi (Houghton et Sheehan, 2009). De l'autre, elles ont créé des nouveaux services capables de satisfaire la demande des consommateurs en termes de gestion du savoir à l'ère de l'abondance, tout en préservant les objectifs de maximisation des profits des acteurs commerciaux. En effet, plus qu'une forme d'uberisation de la publication scientifique (Hall, 2016), la gestion commune de la circulation du savoir devrait être conçue comme un ensemble de services qui permettent de satisfaire les différents intérêts des parties prenantes, dans le cadre d'une chaîne de valeur ainsi renouvelée.

Néanmoins, le renouvellement de cette collaboration devrait tenir en considération un certain nombre d'obstacles – résultant du désalignement des différents intérêts entre les acteurs – auxquels les parties prenantes doivent faire face. Avant tout, le rôle central des répertoires pour l'archivage des articles scientifiques doit être promu parmi les membres de la communauté scientifique, en adoptant des politiques d'obligation au dépôt ou à travers la mise en place d'un système d'incitations. Cela permettrait d'assurer une indépendance de long terme face à l'accès à la connaissance antérieure. En deuxième lieu, le passage d'un prix à l'accès à un prix à la publication de connaissance scientifique pourrait constituer une nouvelle barrière pour la circulation du savoir scientifique. Une régulation des prix à la publication serait souhaitable pour éviter une transposition de la *Serial Crisis*, qui a affecté les budgets des universités à partir des années 1980, à la nouvelle chaîne de valeur à l'ère du numérique. De plus, la reconnaissance, même à travers des signaux de réputation, qui constituent un actif incorporel important pour la communauté scientifique, devrait

être promue pour la publication en *Open Access*, et surtout dans sa forme platinum. Une valorisation au niveau de la reconnaissance par les pairs est demandée à la fois pour les chercheurs qui choisissent de publier dans ces revues, ayant souvent une réputation faible, et au niveau des membres de la communauté scientifiques qui décident de s'engager dans la mise en place de ces revues. Puisque la connaissance libre ne demande pas une gratuité du système, mais l'élimination des barrières à son accès, une reconfiguration des financements consacrés à la publication est demandée, en essayant de soutenir les revues offrant un système platinum, plutôt que les revues dites hybrides, imposant un double payement à la société.

C'est ainsi que la connaissance libre peut être atteinte, à travers un renouvellement de la collaboration entre la communauté scientifique et les différents acteurs du marché de la circulation du savoir scientifique.

Bibliographie du chapitre 5

Beagle, D. (1999). Conceptualizing an Information Commons. *Journal of Academic Librarianship 25*(2), 82–89.

Benkler, Y. (2002). Coase's Penguin, or, Linux and" The Nature of the Firm". *Yale law journal*, 369–446.

Benkler, Y. (2004). Commons-Based Strategies and the Problems of Patents. *Science 305*(5687), 1110–1111.

Benkler, Y. (2006). *The wealth of networks : How social production transforms markets and freedom*. Yale University Press.

Benkler, Y. and H. Nissenbaum (2006). Commons-based peer production and virtue. *Journal of political philosophy 14*(4), 394–419.

Björk, B. (2017). Gold, green, and black open access. *Learned publishing 30*(2).

Bollier, D. (2001). *Public assets, private profits : Reclaiming the American commons in an age of market enclosure*. New American Foundation.

Boyle, J. (1996). *Shamans, Software, and Spleens : Law and the Construction of the Information Society*. Cambridge, MA : Harvard University Press.

Boyle, J. (2002). Fencing off ideas : Enclosure & the disappearance of the public domain. *Daedalus 131*(2), 13–25.

Boyle, J. (2003a). The second enclosure movement and the construction of the public domain. *Law and contemporary problems 66*(1/2), 33–74.

Boyle, J. (2003b). The second enclosure movement and the construction of the public domain. *Law and contemporary problems 66*(1/2), 33–74.

Brainard, J. (2018). Rethinking retractions. *Science 362*, 390–393.

Buranyi, S. (2017). Is the staggeringly profitable business of scientific publishing bad for science ? The Guardian (Long Read.

Callon, M. (1994). Is Science a Public Good ? Science, Technology. *& Human Values 19*, 395–424.

Carroll, M. (2006). Creative commons and the new intermediaries. *Mich. St. L. Rev 45*.

Cole, D. (2017). Laws, norms, and the Institutional Analysis and Development framework. *Journal of Institutional Economics 13*(4), 829–847.

Cole, D., G. Epstein, and M. Mcginnis (2019). The Utility of Combining the IAD and SES Frameworks. *International Journal of the Commons 13*(1).

Cole, D. and M. Mcginnis (2017). Elinor Ostrom and the Bloomington School of Political Economy : A Framework for Policy Analysis. Place : Lexington Books Volume : 3.

Cook-Deegan, R. (2007). The Science Commons in Health Research : Structure, Function, and Value. *The Journal of Technology Transfer 32*(3), 133–156.

Cooke, P. (2001). Regional innovation systems, clusters, and the knowledge economy. *Industrial and corporate change 10*(4), 945–974.

Cooper, M. (2006). Access to the Knowledge Commons in the Digital Age. *Consumer Policy Review 16*(3), 105–116.

Coriat, B. (2015). Le retour des communs : la crise de l'idéologie propriétaire. *Éditions Les Liens qui libèrent*.

Curry, E. (2016). The big data value chain : definitions, concepts, and theoretical approaches. In *New horizons for a data-driven economy*, pp. 29–37. Cham : Springer.

Duncan, J. (1998). The Information Commons : A Model for (Physical) Digital Resources Centers. *Bulletin of the Medical Library Association 86*(4), 576–582.

Edlin, A. and D. Rubinfeld (2004). Exclusion or Efficient Pricing-The Big Deal Bundling of Academic Journals. *Antitrust LJ 72*, 119.

Fang, F., R. Steen, and A. Casadevall (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences 109*(42), 17028–17033.

Ferrucci, E., O. G., and R. E (2018). Free Access to Scientific Knowledge : Sci-hub as a Natural Experiment, Working paper.

Fortson, D. (2019). A Napster moment for ivory towers. *Sunday Times*.

Fuchs, C. and M. Sandoval (2013). The diamond model of open access publishing : Why policy makers, scholars, universities, libraries, labour unions and the publishing world need to take non-commercial, non-profit open access serious. *TripleC : Communication, capitalism & critique 11*(2), 428–443.

Gibbons, M. (Ed.) (1994). *The new production of knowledge : The dynamics of science and research in contemporary societies*. Sage.

Green, T. (2018). We're still failing to deliver open access and solve the serials crisis : To succeed we need a digital transformation of scholarly communication using internet-era principles. *Working paper*.

Hall, G. (2016). *The uberfication of the university*. University of Minnesota Press.

Hardin, G. (1968). The tragedy of the commons. *Science 162*(3859), 1243–1248.

Heckman, J. and S. Moktan (2018). *Publishing and promotion in economics : the tyranny of the top five*. National Bureau of Economic Research. Issue : w25093).

Heller, M. (1998). The tragedy of the anticommons : property in the transition from Marx to markets. Pages : 621–688 Place : Harvard law review.

Hess, C. (2012). The unfolding of the knowledge commons. *St Antony's International Review 8*(1), 13–24.

Hess, C. and E. Ostrom (2003). Ideas, artifacts, and facilities : information as a common-pool resource. *Law and contemporary problems 66*(1/2), 111–145.

Hess, C. and E. Ostrom (2007). *Understanding knowledge as a commons*. The MIT Press.

Himmelstein, D., A. Romero, J. Levernier, T. Munro, S. Mclaughlin, B. Tzovaras, and C. Greene (2018). Sci-Hub provides access to nearly all scholarly literature. *ELife 7*, 32822.

Houghton, J. and P. Sheehan (2009). Estimating the potential impacts of open access to research findings. *Economic Analysis and Policy 39*(1), 127–142.

Kiser, L. and O. E (1982). The Three Worlds of Action : A Metatheoretical Synthesis of Institutional Approaches. In *Strategies of Political Inquiry*, pp. 179–222. Sage.

Laakso, M., P. Welling, H. Bukvova, L. Nyman, B. Björk, and T. Hedlund (2011). The development of open access journal publishing from 1993 to 2009. *PloS one 6*(6), 20961.

Larivière, V., S. Haustein, and P. Mongeon (2015). The oligopoly of academic publishers in the digital era. *PloS one 10*(6), 0127502.

Lessig, L. (2001). *The Future of Ideas : The Fate of the Commons in a Connected World*. New York : Random House.

Lessig, L. (2002). *The future of ideas*. New York : Vintage.

Lessig, L. (2006). .commons. In ed. (Ed.), *Norms and the Law. J.N. Drobak*. Cambridge, MA : Cambridge University Press.

Lukasik, S. (2000). Protecting the Global Information Commons. *Telecommunications Policy 24*(6-7), 519–531.

May, C. (2020). Academic publishing and open access : Costs, benefits and options for publishing research. *Politics 40*(1), 120–135.

Mccabe, M. (2000). Academic journal pricing and market power : A portfolio approach. Type : Working paper.

Mckean, C. and E. Ostrom (2000). *People and forests : Communities, institutions, and governance*. Mit Press.

Merrett, C. (2006). The expropriation of intellectual capital and the political economy of international academic publishing. *Critical Arts 20*(1), 196–111.

Merton, R. (1973). *The sociology of science : Theoretical and empirical investigations*. University of Chicago press.

Moran, E. and E. Ostrom (2005). *Seeing the forest and the trees : human-environment interactions in forest ecosystems*. Mit Press.

Mueller-Langer, F. and R. Watt (2010). Copyright and Open Access for Academic Works. *Review of Economic Research on Copyright Issues 7*(1), 45–65.

Musgrave, R. (1959). *The Theory of Public Finance : A Study in Public Economy*. New York : McGraw Hill.

Nelson, R. (1992). What Is Commercial and What Is Public About Technology, and What Should Be Done ? In N. N. Rosenberg, R. Landau, and D. Mowery (Eds.), *Technology and the Wealth of Nations*, pp. 57–71. Stanford, CA : Stanford University Press.

Nelson, R. (2004). The Market Economy, and the Scientific Commons. *Research Policy 33*(3), 455–471.

Oakerson, R. (1978). The Erosion of Public Highways : A Policy Analysis of the Eastern Kentucky Coal-Haul Road Problem.

Osterloh, M. and B. Frey (2020). How to avoid borrowed plumes in academia. *Research Policy 49*(1), 103831.

Ostrom, E. (2009). The institutional analysis and development framework and the commons. *Cornell L. Rev 95*, 807.

Ostrom, E. (2011). Background on the institutional analysis and development framework. *Policy Studies Journal 39*(1), 7–27.

Ostrom, V. and E. Ostrom (1977). *Public goods and public choices, in Alternatives for Delivering Public Services : Toward Improved Performance*. Boulder, CO : Westview Press.

Pavitt, K. (1987). *On the Nature of Technology. Reprinted in Pavitt, Technology, Management, and Systems of Innovation*. Cheltenham : Edward Elgar.

Phillips, A. (2009). Business models in journal publishing. *The Future of the Academic Journal*, 139–158. Place : Witney Publisher : Chandos Publishing.

Pirie, I. (2009). The political economy of academic publishing. *Historical Materialism 17*(3), 31–60.

Pénin, J. (2013). Are you open ? *Revue économique 64*(1), 133–148.

Ross-Hellauer, T. (2017). What is open peer review ? A systematic review. Pages : 6.

Samuelson, P. (1954). The pure theory of public expenditure. *The review of economics and statistics*, 387–389.

Stiglitz, J. (1999). Knowledge as a global public good. *Global public goods 1*(9), 308–326.

Swan, A. and S. Brown (2005). *Open access self-archiving : An author study*. UK FE and HE funding councils.

Tenopir, C. and D. King (2000). *Towards electronic journals : realities for scientists, librarians, and publishers*. Washington, DC : Special Libraries Association.

Thiel, A., M. Adamseged, and C. Baake (2015). Evaluating an instrument for institutional crafting : How Ostrom's social–ecological systems framework is applied. *Environmental Science & Policy 53*, 152–164.

Van Noorden, R. (2017). Publishers threaten to remove millions of papers from ResearchGate. *Nature 10*.

Weible, C. and P. Sabatier (2017). *Theories of the policy process*. Hachette UK.

Wiley, D. (2016). The access compromise and the 5th R. Available at : http ://opencontent.org/blog/archives/3221.

Conclusion générale

L'*Open Access* est un phénomène qui croise toute discipline de la communauté scientifique, et qui a été largement étudié. Cependant, peu d'études ont essayé de comprendre les causes de sa sous-utilisation et les raisons qui poussent, ou pas, les chercheurs à adopter ce modèle et donc d'envisager une réforme institutionnelle du marché de la publication scientifique. Notre thèse met au centre de l'analyse l'impact de la numérisation de l'information sur le marché de la publication scientifique. Plus précisément, nous nous intéressons aux différents modes d'organisation de la diffusion du savoir scientifique, aux institutions qui en permettent le fonctionnement et à la régulation adoptée, afin de promouvoir une libre circulation de la connaissance scientifique.

La revue de la littérature, proposée dans le premier chapitre, nous a permis de souligner l'existence de différents instruments pour la diffusion et la gestion de la connaissance scientifique, autres que ceux fondés à titre exclusif sur le droit d'auteur. La numérisation de la connaissance, et donc l'élimination du support papier et la conséquente baisse des coûts de la diffusion du savoir scientifique, ont remis en cause le modèle de publication en *Closed Access* et l'efficacité d'un système d'ordre public, dont le droit d'auteur est l'expression principale, pour la gestion de la diffusion des résultats de la recherche. Défini comme modèle de gestion de l'accès, du partage et de la réutilisation de la connaissance, alternatif aux mécanismes de marché typiques du droit d'auteur, l'*Open Access* est aussi un modèle de partage du savoir, dûment analysé par la littérature, qui reste encore aujourd'hui faiblement utilisé dans toutes ses formes, à savoir le *Green* (*i.e.* auto-archivage) et le Gold (*i.e.* publication dans une revue). Notre revue de la littérature a souligné la façon dont la remise en cause du modèle propriétaire doit s'inscrire dans une analyse institutionnelle plus large, en étudiant l'interdépendance existant entre les différentes normes, légales et

sociales, qui règlent et coordonnent les comportements individuels dans le but de faire avancer la science.

L'interdépendance entre ce double système de propriété intellectuelle, fondé à la fois sur les normes légales et sociales, est le cœur de l'analyse du deuxième chapitre. Nous avons montré comment la marchandisation de la connaissance scientifique, mais aussi l'évolution du fonctionnement de la communauté scientifique, ont amené à une forme de saillance du droit d'auteur, en changeant le signifié de la norme sociale de la publication. L'impératif publish or perish transforme cette norme sociale. Il entraîne le fait que la publication n'est plus un moyen pour permettre l'avancement de la science, mais elle devient une fin en soi. Le désalignement qui résulte des deux systèmes de normes, poussant d'un côté à l'ouverture des résultats de la recherche et de l'autre à la recherche d'une rémunération en termes de réputation personnelle, ne permet plus une véritable complémentarité, nécessaire pour instaurer un nouvel équilibre autour de l'*Open Access*.

Les limites d'une analyse portant exclusivement sur le cadre juridique ont été montrées dans le troisième chapitre, où nous avons testé les effets d'une réforme du cadre juridique sur les comportements des chercheurs, notamment en ce qui concerne les pratiques d'auto-archivage. En effet, le faible engagement de la communauté scientifique vers le *Green Open Access* a été attribué par la littérature aux barrières légales imposées par le droit d'auteur. C'est pour ces raisons qu'en France, la loi Pour une république numérique a été introduite. Nos résultats ont montré que le fait de se conformer volontairement à la loi n'est pas la meilleure solution à l'heure actuelle, étant donné le faible impact de cette réforme. De plus, cette analyse fait émerger des différences importantes entre les disciplines scientifiques, en suggérant que différentes pratiques de dissémination des résultats scientifiques caractérisent les sub-communautés scientifiques et qu'une réforme plus ciblée devrait être proposée. Ce travail n'est qu'une première enquête sur l'impact de la réforme du cadre juridique sur le comportement des chercheurs et des futures pistes de recherche sont proposées et encouragées. À partir de ces résultats, une future contribution à cette littérature consisterait à comparer la situation française avec celle d'autres pays qui n'ont pas adopté ce type de réforme juridique.

Si une réforme du droit d'auteur est l'étape préalable pour favoriser la circulation des

résultats de la recherche, d'autres barrières limitent l'adoption de l'*Open Access* comme modèle de diffusion du savoir scientifique. Notre quatrième chapitre s'est alors intéressé à comprendre si la qualité de la revue dans laquelle un article a été publié peut influencer le choix d'adopter la stratégie *Green* par les chercheurs. Ce modèle permet d'éliminer la barrière de prix qui limite l'accès à l'article et donc d'augmenter la probabilité pour un chercheur d'être cité. D'un point de vue stratégique, et surtout dans une optique de contribution au développement de l'*Open Science*, ce sont les producteurs d'articles scientifiques de qualité plus faible qui devraient avoir le plus d'intérêt à adopter ce modèle. En effet, à cause de la contrainte budgétaire des centres de recherche et des bibliothèques universitaires, l'accès à ce type de littérature est fortement limité par la barrière de prix imposée par les maisons d'édition. Nos résultats ont montré que ce modèle est plus utilisé comme moyen de marketing pour souligner l'importance des meilleurs résultats scientifiques que pour garantir un accès, autrement fortement limité, à la littérature scientifique. Pourquoi alors produire un bien, si aucun effort n'est pas fait pour augmenter sa circulation sur le marché ? Cette question de recherche mérite d'être enquêtée dans des travaux futurs, car elle permet de projeter de la lumière sur les incitations à la production de connaissance en fonction de l'avancement de carrière des chercheurs. La finalité ultime de ces incitations devrait être la production de connaissance de qualité et l'avancement collectif de la science, et non pas une submersion du marché de la connaissance scientifique d'articles, dont les auteurs sont les premiers à ne pas les considérer comme de véritables contributions scientifiques.

L'impact du numérique sur le marché de la publication scientifique ne s'arrête pas au modèle *Open Access*. Le cinquième et dernier chapitre de cette thèse en a étudié l'impact sur la chaîne de valeur de la circulation de la connaissance scientifique. De nouveaux modèles d'affaires ont émergé, chacun satisfaisant de façon différente les intérêts des parties prenantes du marché de la publication du savoir scientifique. C'est grâce à la dématérialisation de la circulation du savoir scientifique qu'une gestion commune de la connaissance est devenue possible. Cette gestion commune doit être entendue comme un renouvellement de la collaboration existante entre les acteurs principaux du marché de la publication scientifique, collaboration élargie aux nouvelles parties prenantes. La complémentarité entre

différents acteurs et services est possible seulement dans un système au cœur duquel des contrats privés posent des exceptions aux limites imposées par l'exclusivité du droit d'auteur et qui permettent d'éviter une deuxième enclosure de la connaissance scientifique. C'est ainsi que les licences *Creative Commons* permettent de continuer à assurer les intérêts des chercheurs et des acteurs impliqués dans la gestion du savoir.

Néanmoins, le renouvellement de cette collaboration devrait tenir en considération un certain nombre d'obstacles – résultant du désalignement des différents intérêts entre les acteurs – auxquels les parties prenantes doivent faire face. Avant tout, le rôle central des répertoires pour l'archivage des articles scientifiques doit être promu parmi les membres de la communauté scientifique, en adoptant des politiques d'obligation au dépôt ou à travers la mise en place d'un système renouvelé d'incitations. Ensuite, le passage d'un prix à l'accès à un prix à la publication de connaissance pourrait constituer une nouvelle barrière pour la circulation du savoir scientifique. C'est pour cela que des études portant sur l'évolution des prix à la publication sont fortement recommandées, afin de comprendre si une régulation des prix de ce marché en oligopole est nécessaire. Enfin, la connaissance libre ne demande pas la gratuité du système, mais l'élimination des barrières à son accès. Une reconfiguration des financements consacrés à la publication est recommandée, en essayant de soutenir les revues offrant un système platinum, plutôt que les revues dites hybrides, imposant un double payement à la société.

Cette thèse contribue ainsi à éclaircir le fonctionnement du marché de la publication scientifique à l'heure actuelle et met en exergue la nécessité d'une réforme des normes qui en sont à la base. Si une libre circulation du savoir scientifique est l'objectif des institutions publiques, une protection intellectuelle assouplie par les licences *Creative Commons* et une réforme des normes sociales de la communauté scientifique sont plus que jamais nécessaires.

Résumé

Malgré les possibilités offertes par la numérisation de l'information et les nouvelles méthodes de circulation du savoir scientifique conçues pour élargir l'accès et la réutilisation de la connaissance, le modèle *Open Access* (OA) reste à l'heure actuelle sous-utilisé par rapport au modèle standard, dit de *Closed Access* (CA).

L'objectif de la thèse est d'étudier les barrières et obstacles qui limitent la transition du modèle de CA au modèle d'OA au sein des communautés scientifiques. La question de recherche est traitée en cinq chapitres. Le premier offre une analyse du fonctionnement du système traditionnel de la régulation de l'accès au savoir scientifique, fondé sur le droit d'auteur, et sa remise en cause suite la numérisation de l'information. Dans le deuxième chapitre, nous nous intéressons aux normes sociales et légales régulant le marché de la publication scientifique et à l'influence réciproque qu'elles exercent l'une sur l'autre, de manière à en caractériser l'évolution. En exploitant une base de données entièrement originale construite par nos soins, nous analysons dans le troisième chapitre le rôle du droit d'auteur, traditionnellement identifié par la littérature comme l'une des barrières principales au développement de l'Open Access, dans la transition vers ce modèle. Dans le quatrième chapitre, nous nous intéressons au rôle des classements des revues scientifiques dans le choix du *Green Open Access* comme instrument à la fois alternatif et complémentaire pour la diffusion des résultats de la recherche scientifique. Enfin, dans le cinquième chapitre, une analyse de la connaissance scientifique en tant que commun est proposée, afin de comprendre l'impact de la numérisation sur la chaîne de valeur du marché de la circulation du savoir.

Abstract

Despite the digitization of information and the emergence of new knowledge dissemination models, allowing a free access to the scientific knowledge, the Open Access (OA) is still not the first scholars' publishing choice.

This thesis aims to understand the barriers to the transition from the Open Access system to the traditional one, namely the Closed Access (CA). The research question is developed throughout five chapters. Through a literature review, the first one explains the functioning of the scientific publishing market and the consequences of the digitization of information. In particular, the analysis focuses on three main institutions, namely copyright, social norms and the new legal tools that regulate knowledge dissemination in the digital age. In the second chapter, we study the social and legal norms regulating the scientific publishing market and their reciprocal influence, characterizing the evolution of the institutional framework. Using data from an original dataset we built, in the third chapter we question the role of copyright, defined by the literature as one of the main barriers to the OA publishing. In the fourth chapter, we question the influence of journal rank on scholars' attitudes to adopting the green model. We focus our attention on the economics discipline and use data from HAL, the largest archive in the French scientific landscape. The ranks of both national and international journals were used to enrich our dataset. Finally, in the fifth chapter, we propose an analysis of the scientific knowledge as a commons. We apply the Institutional Analysis Development framework to the scientific publishing market to understand the impact of the digitalization of information over the different stakeholders' interests and the scientific knowledge value chain.