



Livre blanc

**Une science ouverte  
Dans une  
République numérique**

Mars 2016



**cnrs**

dépasser les frontières

Direction de l'information scientifique et technique



DIST - Direction de l'information scientifique et technique



**ISTEX**

Livre blanc – Une Science ouverte dans une République numérique



Livre blanc

**Une Science ouverte  
dans une  
République numérique**

Direction de l'Information Scientifique et Technique  
CNRS

Accompagnement, Expertise, Conseil: Cabinet ALAIN BENSOUSSAN

## SOMMAIRE

<b>Préface</b> .....	<b>6</b>
<b>Synthèse et propositions</b> .....	<b>12</b>
<b>I/ Synthèse</b> .....	<b>12</b>
<b>II/ Constats</b> .....	<b>17</b>
<b>III/ Recommandations principales</b> .....	<b>18</b>
<b>Plan commenté</b> .....	<b>19</b>
<b>1. L'ETAT DES LIEUX : "LA SCIENCE EN TRANSITION"</b> .....	<b>22</b>
<b>1.1 Photographie des usages de la Science</b> .....	<b>22</b>
1.1.1 La Science au cœur de la transition numérique .....	22
1.1.2 Les nouveaux usages des chercheurs .....	24
1.1.3 Une forte demande des laboratoires, des instituts, des organismes .....	33
<b>1.2 Les vides juridiques</b> .....	<b>36</b>
1.2.1 Un droit de la science qui fait défaut .....	36
1.2.2 Droit des plateformes : des évolutions en cours.....	36
1.2.3 Un droit au TDM : une absence aux conséquences graves .....	37
1.2.4 Le besoin d'un droit rénové de l'édition numérique scientifique .....	41
<b>1.3 Les risques de captation abusive</b> .....	<b>49</b>
1.3.1 Captation par les usages économiques .....	49
1.3.2 Captation par les usages de l'édition scientifique.....	49
1.3.3 Captation par les contrats .....	50
<b>1.4 Une validation massive de ces constats</b> .....	<b>54</b>
1.4.1 La synthèse de la consultation nationale.....	54
1.4.2 L'avis du Conseil National du Numérique.....	58
1.4.3 L'étude d'impact du projet de loi.....	60
1.4.4 Le texte adopté par l'Assemblée nationale.....	62
<b>2. L'AVENIR : LA SCIENCE NUMERIQUE OUVERTE</b> .....	<b>64</b>
<b>2.1 Les témoignages du Livre blanc : des principes convergents vers une Science ouverte</b> .....	<b>65</b>
2.1.1 Une valeur commune : La Science, un « bien commun » de l'humanité .....	65

2.1.2	La Science, moteur de l'économie .....	67
2.1.3	Priorité au libre accès et au partage des données de la Science .....	69
2.1.4	L'évaluation numérique des résultats revus par les pairs et métrique des publications .....	75
2.1.5	Publication et embargo .....	78
2.1.6	Analyse et exploration des corpus de données .....	81
2.1.7	Propriété intellectuelle numérique et reconnaissance de paternité .....	84
2.1.8	Les frontières de la valorisation et de la science ouverte .....	84
2.1.9	Pour une éthique de l'IST numérique .....	85
2.1.10	Vers un changement radical de paradigme ? .....	86
	Synthèse globale .....	88
<b>2.2</b>	<b>L'Open Science à l'international .....</b>	<b>89</b>
2.2.1	Au niveau de l'Union européenne : une affirmation de l'Open Science .....	89
2.2.2	A l'étranger : une légalisation de l'Open Science .....	91
2.2.3	Au niveau des instances internationales : une tendance à l'Open Science .....	94
<b>2.3</b>	<b>Concepts clés de la Science ouverte .....</b>	<b>96</b>
2.3.1	Un droit de l'Open : un mouvement international inéluctable d'ouverture et de partage des données .....	96
2.3.2	Le Code de la recherche : fondement du droit de l'IST numérique .....	99
2.3.3	Une nécessaire cohérence avec le droit des données publiques .....	100
2.3.4	Un principe juridique incontournable : la propriété littéraire et artistique .....	102
2.3.5	Les exceptions à aménager : les motifs d'intérêt public et les secrets légaux .....	104
2.3.6	La préservation de la vie privée et des données à caractère personnel .....	105
2.3.7	La valorisation : un intérêt légitime à préserver .....	105
<b>2.4</b>	<b>L'écart entre les usages et le droit .....</b>	<b>108</b>
<b>2.5</b>	<b>La modification de l'article 17 du projet de loi pour une République numérique .....</b>	<b>111</b>
2.5.1	Principes directeurs .....	111
2.5.2	Proposition de modifications de texte .....	113
<b>2.6</b>	<b>Des propositions à affirmer, des pistes pour l'avenir .....</b>	<b>115</b>
2.6.1	Un droit positif de <i>text and data mining</i> .....	115
2.6.2	La définition d'un droit de la Science ouverte : affirmation des valeurs .....	119
2.6.3	Un référentiel des usages de l'IST numérique .....	121
2.6.4	Une norme Afnor ou Iso .....	123

Livre blanc – Une Science ouverte dans une République numérique

2.6.5	Un contrat type de cession de droits d'auteur : pour la protection des chercheurs .....	124
2.6.6	Une charte éthique de la Science numérique.....	125
2.6.7	Une Agence de la Science ouverte .....	125
2.6.8	Une Convention internationale de la Science ouverte universelle.....	126
<b>2.7</b>	<b>Synthèse des propositions pour une Science ouverte .....</b>	<b>126</b>
2.7.1	Avis stratégiques.....	126
2.7.2	Propositions .....	130
	<b>Remerciements .....</b>	<b>131</b>
	<b>Glossaire.....</b>	<b>132</b>
	<b>Annexes.....</b>	<b>138</b>
<b>1.</b>	<b>Présentation du Livre blanc.....</b>	<b>138</b>
<b>2.</b>	<b>Contribution du Conseil Scientifique.....</b>	<b>158</b>
<b>3.</b>	<b>Recommandation du Conseil Scientifique .....</b>	<b>163</b>
<b>4.</b>	<b>Avis du Comité d'éthique du 7 mai 2015 « Les enjeux éthiques du partage des données scientifiques » .....</b>	<b>164</b>
<b>5.</b>	<b>Guide d'entretien des auditions - contribution de la Recherche aux thèmes du projet de loi numérique.....</b>	<b>180</b>
<b>6.</b>	<b>Compte-rendu d'auditions .....</b>	<b>182</b>
a.	UPMC, Jean CHAMBAZ et Paul INDELICATO, 9 juin 2015 .....	182
b.	Conseil Scientifique du CNRS, Bruno CHAUDRET, Claire LEMERCIER, François BONNAREL, François TRONCHE, 30 juin 2015.....	188
c.	Université de Strasbourg, Paul-Antoine HERVIEUX et Françoise CURTIT, 10 juillet 2015 201	
d.	Conseil National du Numérique, Benoit Thieulin et Yann Bonnet, 1 septembre 2015.....	209
e.	Comité ISTEX, Laurent Schmitt, Jean-Marie Pierrel et Grégory Colcanap, 24 septembre 2015	216
f.	ABES, Jérôme KALFON, 5 octobre 2015.....	221
g.	CCSD, Claude Kirchner, 15 octobre 2015.....	227
<b>7.</b>	<b>Tableau récapitulatif des législations étrangères : Open access et TDM</b>	<b>234</b>

## Préface

Lancée par le Premier ministre, la Consultation nationale sur la loi numérique a nourri un débat inédit sur la préparation d'un texte gouvernemental majeur. Parmi les nombreux résultats originaux de cette Consultation, on peut remarquer l'importance accordée par les débats (chercheurs, éditeurs, établissements ou groupements scientifiques) au thème que couvre précisément ce Livre blanc :

### **Comment caractériser, valoriser et partager l'information scientifique à l'heure numérique ?**

Certes, la France n'est pas le premier pays à se poser la question : les solutions abondent hors et dans l'espace européen. Le présent Livre blanc entend innover en fournissant une étude globale, à partir d'éléments de solution construits dans plusieurs grands champs de pratiques scientifiques. Ce travail prend place parmi les efforts récents, stimulés par l'Académie des sciences autour de Jean François BACH, Secrétaire perpétuel, pour revoir en profondeur les bases de l'accès à l'évaluation et à la publication scientifique. Ce regard global a pour but d'inspirer des choix qui, à l'échelle des nouveaux usages numériques de la science, sont des choix internationaux, confrontés comme ceux concernant tous les autres choix numériques - mais plus encore - au poids et aux enjeux du partage planétaire des systèmes d'information et d'Internet.

### **Les enjeux de la Science ouverte : une nécessaire convergence internationale**

L'enjeu de la Science ouverte, précisément, n'est ni une vision lointaine ni un slogan : la Science ouverte, dans l'acception qu'en a aujourd'hui Carlos MOEDAS, Commissaire européen à la Science et à l'innovation, qui vient d'en faire en 2015 l'une des trois priorités de la Commission européenne, est une démarche qui doit concilier impérativement, à la fois, de nouveaux usages numériques, de nouvelles valorisations de ces usages (propriété intellectuelle notamment), et les démarches et droits nouveaux garantissant l'accès et le partage des résultats scientifiques... Sur ces thèmes, l'Amérique du Nord, notamment, poursuit activement la construction de nouvelles bases qui déterminent un champ de compétition internationale particulièrement active.

Chaque pays aujourd'hui cherche donc ses propres solutions et c'est logique qu'il en soit ainsi, même s'il s'avère indispensable de viser la convergence : les pratiques d'édition scientifique-numérique ou "papier" dépendent d'écosystèmes éditoriaux nationaux, historiquement datés et construits, avec leurs édifices juridiques, contractuels, leurs droits acquis. Ces écosystèmes sont eux-mêmes dépendant du poids de l'édition

Livre blanc – Une Science ouverte dans une République numérique

scientifique internationale dans chacun des contextes : les grands éditeurs de science<sup>1</sup> sont inégalement partie-prenantes des systèmes nationaux de publication scientifique, à travers un large éventail de pratiques contractuelles.

Dans le Projet de loi "Pour une République numérique », notre pays va donc chercher à définir le positionnement qu'appelle son histoire, son économie de l'édition scientifique, au regard de ce que sont les besoins de ses communautés de recherche.

Ce Livre blanc se bornera à traiter du seul besoin exprimé par les communautés de la recherche. Ces voix, faut-il le souligner, se sont exprimées sur plusieurs grands thèmes de convergence lors de la Consultation sur le Projet de loi<sup>2</sup>. Axelle LEMAIRE, Secrétaire d'Etat en charge du projet de Loi en a pris acte<sup>3</sup>.

### **Une démarche de la recherche publique pour un dialogue national**

En adoptant ainsi délibérément ce seul point de vue restreint aux besoins des "producteurs finaux" et des "utilisateurs directs" de l'information scientifique, ce Livre blanc encourt évidemment des critiques sur la restriction délibérée de son horizon de préoccupation : n'eut-t-il en effet pas été logique d'y associer largement toutes les parties prenantes à la production et au partage de l'information scientifique ? S'ils sont avertis des limitations imposées par ce choix, les Auteurs de ce Livre blanc en ont également pesé les avantages dans le contexte actuel : le débat avec les Editeurs et tous les fournisseurs de services de l'information scientifique numérique est, à l'évidence, encore mal stabilisé à toutes les échelles, nationale, européenne et internationale, auxquelles ce débat est aujourd'hui mené, et les options mêmes de ce débat sont évolutives.

En adoptant un champ d'analyse restreint aux seuls usages internes à la recherche, les Auteurs de ce Livre blanc estiment pouvoir proposer une "valeur ajoutée" suffisante pour apporter leur pierre à un débat qui excède le champ du projet de loi lui-même. Rappelons ici que ce Livre blanc a été mis en chantier avant que ne s'ouvre le débat sur la Loi numérique, et que bon nombre de ses thèmes excèdent le champ du débat législatif, pour fournir des jalons dans un cadre global de réflexion sur les besoins d'information numérique de la science. Cette démarche s'est largement nourrie dans la réflexion initiée à et par l'INIST, autour de son directeur, Raymond BERARD : à l'heure numérique, cette démarche vient actualiser au XXI<sup>e</sup> siècle, l'une des idées-clefs,

---

<sup>1</sup> Une étude récente de Livre Hebdo (26 juin 2015) nous montre que sur les 12 premiers éditeurs du monde en 2014, les 4 premiers, qui dépassent tous un chiffre d'affaires de 3,5 milliards d'euros, sont des Groupes d'Edition professionnelle, largement enracinés dans la science et dans sa valorisation.

<sup>2</sup> Une synthèse précise de toutes les contributions est accessible par: [http://www.cnrs.fr/dist/z-outils/documents/Projet%20de%20loi%20-%20analyseCNRS\\_DIST.pdf](http://www.cnrs.fr/dist/z-outils/documents/Projet%20de%20loi%20-%20analyseCNRS_DIST.pdf)

<sup>3</sup> Axelle LEMAIRE, synthèse de la Consultation sur le projet de Loi.

Livre blanc – Une Science ouverte dans une République numérique

énoncée par Diderot et d'Alembert dès le Prospectus de 1750, le célèbre programme de travail de l'Encyclopédie, qui était, littéralement, de trouver les représentations et les moyens de "faire tourner les sciences".

### **Deux constats-clefs servent à cet égard de fil conducteur aux deux temps de ce Livre blanc.**

Tout d'abord, l'état des lieux des usages dénote **un fort besoin de rattrapage**, là où aujourd'hui, les usages numériques de la science sont aujourd'hui en décalage avec les grandes pratiques émergentes et/ou installées dans les grands pays de science. Ensuite, une autre idée clef concerne la direction des changements en cours : **aller vers un "droit des ressources partagées et des usages protégés"**, peut résumer pour l'essentiel les propositions et recommandations du Livre blanc. On peut clairement y voir une mise en perspective numérique des propositions émises dans le Rapport "Pour une société apprenante"<sup>4</sup> remis en septembre 2015 au Président de la République.

Qu'il soit également permis aux Auteurs de cette démarche collégiale, d'exprimer leur intention de partage et de dialogue avec tous les acteurs qui apportent leur services et leur savoir-faire au travail de la science : les besoins et contraintes des chercheurs ont été traités en vue de laisser ouverte la question des modèles économiques et juridiques par lesquels ces besoins pourront être satisfaits à l'avenir. Plutôt en effet que d'entrer dans le champ clos d'un débat qui est à l'évidence mal stabilisé, à tous les niveaux auxquels se pose aujourd'hui la question des modèles de convergence entre les besoins et les usages de l'information scientifique, et les moyens de les satisfaire, les Auteurs de ce Livre blanc entendent verser leurs apports au dossier d'un débat à construire avec tous les fournisseurs présents et futurs de la science.

En prenant cette option, les Auteurs de ce Livre blanc affirment leur volonté d'un dialogue national, européen et international, notamment avec tous les représentants de l'Édition scientifique et des fournisseurs de service à la science. Seul ce dialogue donne sens aux présentes propositions et réflexions.

Qu'il soit également acquis que ce travail d'analyse et de réflexion n'entend pas s'approprier et refléter l'opinion de la communauté scientifique nationale dans son ensemble, mais vise à apporter une première contribution construite à un débat national

---

<sup>4</sup>Rapport à François HOLLANDE, Président de la République, en présence de Najat VALLAUD-BELKACEM, Ministre de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, et de Thierry MANDON, secrétaire d'État chargé de l'Enseignement supérieur et de la Recherche Par Sophie BEJEAN, Présidente du comité StraNES et Bertrand MONTHUBERT, Rapporteur général.

Livre blanc – Une Science ouverte dans une République numérique

qui ne fait que s'amorcer, et qui manque aujourd'hui de positions définies comme de débat. Le manque d'études a été souligné de plusieurs côtés, ce que synthétise l'Avis du Conseil d'Etat sur le projet de loi<sup>5</sup>. Sous ces aspects, les enquêtes du CNRS auprès des UMR constituent des exemples isolés, et doivent être approfondies, si systématiques soient-elles<sup>6</sup>.

### **Un Livre blanc : témoignages, réflexions, expertise**

Pour finir, les Auteurs de ce Livre blanc entendent assumer ensemble les choix de construction et de présentation de la démarche qui les a rassemblés : ce travail est fondé à la fois sur le témoignage et sur l'expertise. Une présentation détaillée de cette démarche est proposée en annexe 1 du Livre blanc.

Le témoignage est d'abord celui qu'a construit le Projet ISTEX, premier Investissement d'Avenir numérique national de documentation scientifique, dont les acteurs ont pu explorer ensemble les nouveaux usages de ressources numériques dédiées à la documentation scientifique.

ISTEX propose une validation spectaculaire de la démarche proposée par la Bibliothèque Scientifique Numérique (BSN) : le MENESR, l'ABES, le Consortium Couperin, l'Université de Lorraine, et le CNRS (DIST et INIST) sont les parties prenantes de ce grand projet. Ce Livre blanc a pris naissance dans le contexte du Séminaire ISTEX, présidé en 2014 par Jean Pierre FINANCE, représentant de la Conférence des Présidents d'Universités (CPU) et initiateur d'ISTEX, pour jeter les bases d'une réflexion juridique sur les inflexions fondamentales requises pour la science à l'heure des usages numériques.

Ce Livre blanc s'est par ailleurs construit grâce aux réflexions et témoignages de deux grandes Universités membres de la Ligue Européenne des Universités de Recherche Intensive (LERU) : l'Université Pierre et Marie Curie (UPMC) et l'Université de Strasbourg, représentées par leur Présidents et Vice-Présidents respectifs, Jean CHAMBAZ et Paul INDELICATO, pour l'une, Alain BERETZ et Paul-Antoine HERVIEUX pour l'autre.

Le dispositif national d'Open Access, le CCSD (Centre pour la Communication Scientifique Directe) qui ouvre la voie aux formes nouvelles de partage des connaissances, s'est exprimé par la voix de son Président en exercice, Claude KIRCHNER.

---

<sup>5</sup> Par les Editeurs eux-mêmes à travers le SNE et, plus récemment par l'Avis du Conseil d'Etat sur le projet de loi

<sup>6</sup> <http://www.cnrs.fr/dist/z-outils/documents/Enqu%C3%A0te%20DU%20-%20DIST%20mars%202015.pdf>

Livre blanc – Une Science ouverte dans une République numérique

Le CNRS, qui a impulsé à l'amont en 2013, une stratégie nationale mutualisée sur les besoins et ressources d'information scientifique<sup>7</sup>, s'est ici exprimé par la voix du Président du Conseil scientifique, Bruno CHAUDRET, au terme de près d'un an de réflexion interne du Conseil scientifique. Enfin, Benoit THIEULIN, Président du Conseil National du Numérique, a apporté un précieux concours à cette réflexion globale, en mettant en cohérence les thèmes de la science avec tous ceux qu'il porte dans l'organisation d'une transition nationale vers les usages numériques à l'échelle de notre pays.

Apportée par le Cabinet international ALAIN BENSOUSSAN, l'expertise a donné un indispensable soubassement juridique à ce travail, à travers l'analyse des textes et des options offertes par le droit émergent des technologies du numérique appliquées à la science, au niveau national, européen et international. Le Cabinet ALAIN BENSOUSSAN a donné vie à ce projet en assurant l'ensemble des auditions et la construction de ce Livre blanc, en phase avec les attentes de tous les acteurs de cette démarche, en y apportant ouverture et esprit d'innovation.

Plusieurs Grands Témoins, pour finir, ont bien voulu faire partager leur réflexion sur deux grands domaines où se jouent des enjeux décisifs pour l'information scientifique à l'heure numérique : Bruno DAVID, Président du Muséum d'Histoire Naturelle traite de la diversité des usages de l'information scientifique et des formes nouvelles de son partage, et enfin Daniel EGRET, ancien président de l'Observatoire astronomique de Paris, décrit les nouveaux usages de la métrique des publications scientifiques.

Le CNRS a assuré l'orchestration de cette démarche par les interventions conjointes particulièrement riches, de sa direction de l'information scientifique et technique (DIST) et de sa direction des affaires juridiques (DAJ).

A travers ces nombreux témoignages de chercheurs et de responsables de la recherche ainsi qu'à travers l'expertise juridique internationale présentée par ce Livre blanc, ont été rassemblés les moyens d'une information sur le droit et les usages de la publication numérique scientifique : ceci répond aux besoins, exprimés par tous les Acteurs, professionnels et non professionnels, de disposer d'éléments objectifs sur les changements en cours introduits par le numérique.

Puisse cet ensemble de témoignages, de réflexions et d'expertise, contribuer à la réflexion commune et aux choix dont a besoin la France pour qu'elle soit présente à sa

---

<sup>7</sup> <http://www.cnrs.fr/dist/strategie-ist.htm>

Livre blanc – Une Science ouverte dans une République numérique

place dans le grand projet de la création scientifique numérique du XXI<sup>ème</sup> siècle, qui est désormais engagé.

*Les signataires :*

Pour les membres du Comité exécutif de l'Investissement d'Avenir ISTEX :

Grégory COLCANAP, Coordonnateur du Consortium Couperin

Renaud FABRE, Directeur de la DIST du CNRS

Jérôme KALFON, Directeur de l'ABES

Jean-Marie PIERREL, Professeur à l'Université de Lorraine

Laurent SCHMITT, Responsable du département Projets et Innovation, Inist CNRS

Pour les Grands Témoins :

Alain BERETZ, Président de l'Université de Strasbourg

Jean CHAMBAZ, Président de l'UPMC

Bruno CHAUDRET, Président du Conseil Scientifique du CNRS

Bruno DAVID, Président du Muséum National d'Histoire Naturelle

Daniel EGRET, Astronome (PSL) ancien Président de l'Observatoire de Paris

Claude KIRCHNER, Conseiller du Président d'Inria, Directeur de Recherche

Benoit THIEULIN, Président du Conseil National du Numérique

## Synthèse et propositions

### I/ Synthèse

#### LA "SCIENCE OUVERTE" (OPEN SCIENCE), UN DROIT NOUVEAU POUR DES USAGES NUMERIQUES

##### Directions principales :

- **Créer** : Créer un droit de l'Open Science garantissant le libre accès et la libre réutilisation des données de la recherche publique
- **Equilibrer** : Redéfinir les équilibres économiques de l'écosystème numérique de la science
- **Sécuriser** : Adopter l'article 18 bis (nouveau) du projet de loi pour une République numérique créant une exception au droit d'auteur et au droit du producteur de base de données en faveur du *text and data mining* sur les données de la recherche publique (articles et données de la recherche) afin de sécuriser les pratiques de traitement automatisé de données et réduire les risques de captation abusive
- **Rivaliser** : Permettre à la recherche publique de disposer des moyens légaux et techniques au moins équivalents à ceux de ses homologues européens et américains et en phase avec le mouvement international de l'Open Science
- **Protéger** : Protéger les intérêts légitimes : valorisation, secret, brevet, droit d'auteur, vie privée et données personnelles

##### Qu'est-ce que la Science ouverte ?

L'Open Science est une nouvelle approche transversale de l'accès au travail scientifique, des visées et du partage des résultats de la science mais aussi une nouvelle façon de FAIRE de la science, en ouvrant les processus, les codes et les méthodes.

Les vues globales "encyclopédiques", comme le décloisonnement et le partage à grande échelle des connaissances, renaissent dans le projet de l'Open Science : à l'heure numérique, cette notion souligne la "démultiplication" des connaissances telle qu'elle

Livre blanc – Une Science ouverte dans une République numérique

intervient, par exemple, en explorant en profondeur les bases numériques contenant les revues scientifiques<sup>8</sup>.

L'Open Science, c'est donc un **changement d'optique**, qui veut se comparer à d'autres grands stades antérieurs, comme l'avènement du télescope ou du microscope.

L'Open Science veut ainsi rendre compte des mutations qu'apportent les grandes plateformes « ouvertes » internationales : elles donnent accès à de nouveaux agencements d'acteurs de la recherche (innovation numérique, société civile), facilitées par de nouvelles démarches de partage (lois numériques), et aboutissant à des types originaux de résultats et de régulations (plateformes de données et d'analyses, réseaux sociaux scientifiques, nouvelles formes de collaboration....).

L'Open Science, champ beaucoup plus profond que l'Open access limité à la seule publication, désigne tout l'éventail des voies et moyens d'enrichissement du travail scientifique apportés par le numérique.

### **Un mouvement international inéluctable**

L'Open Science s'inscrit dans un mouvement international d'ouverture. De nombreux pays ont déjà légiféré en faveur de l'Open access et du *text and data mining*. Les instances internationales et européennes prônent cette avancée.

En juin 2015, le Commissaire pour la Recherche, la Science et l'Innovation, Carlos MOEDAS, a défini trois priorités à son action: "Open Innovation, **Open Science**, Openness to the World". L'Open Science est définie à Bruxelles par toutes les "transitions" qui accompagnent les changements numériques de la science<sup>9</sup>. Dès 2015, l'OCDE développe une approche similaire en insistant sur la possible rupture globale que pourrait constituer l'Open Science<sup>10</sup>, sous la condition d'une action concertée.

### **En France: éléments pour une stratégie nationale d'Open Science**

---

<sup>8</sup> Ces explorations fertiles sont aussi bien synchroniques comme sur Pub Med Central, <http://www.ncbi.nlm.nih.gov/pmc/> que diachroniques: comme sur l'exemple cité par la NSF, [http://www.nsf.gov/news/news\\_summ.jsp?cntn\\_id=135258](http://www.nsf.gov/news/news_summ.jsp?cntn_id=135258)

<sup>9</sup> Open Science "describes the on-going transitions in the way research is performed, researchers collaborate, knowledge is shared, and science is organized. It is enabled by digital technologies"

<sup>10</sup> OCDE 2015: "Open science is more than open access to publications or data; it includes many aspects and stages of research processes".

Livre blanc – Une Science ouverte dans une République numérique

La stratégie du CNRS « Mieux partager les connaissances »<sup>11</sup> a fait émerger la nécessité d'un rattrapage en ce qui concerne les pratiques numériques de la publication scientifique sur les plateformes.

Ces thèmes ont également été portés par le Conseil Scientifique du CNRS dans sa recommandation unanime ainsi que par le Comité d'éthique.

De nombreuses voix s'élèvent en faveur de l'Open Science. Le gouvernement lui-même a pris des positions ambitieuses notamment dans la « Stratégie numérique du Gouvernement » du 18 juin 2015, ainsi que dans l'exposé des motifs du projet de loi pour une République numérique. La consultation nationale sur l'avant-projet de loi a fait se démarquer la Science comme un thème prioritaire du consensus national sur le numérique et comme un thème de cohérence globale entre les points de vue des chercheurs et des institutions autour de principes simples :

- la Science est un bien commun de l'humanité ;
- les intérêts légitimes de préservation des secrets et de valorisation doivent être préservés ;
- le *text and data mining* est un droit d'observation numérique naturel nécessaire aux chercheurs dans leur démarche scientifique ;
- les clauses de cession exclusive de droit d'auteur prévues dans les contrats d'édition doivent être déclarées nulles et non avenues ;
- les connaissances doivent librement pouvoir faire l'objet d'exploitation industrielle ou commerciale dans un cadre éthique consolidé.

Ce Livre blanc propose de prendre du recul sur les usages des résultats de la recherche grâce notamment à de grands témoins qui ont bien voulu s'exprimer et dont les opinions sont, de manière générale, convergentes. Ces usages numériques contemporains sont confrontés au cadre juridique existant et mis en perspective.

### **L'absence d'antagonisme juridique**

L'analyse du cadre juridique fait apparaître que les droits français et européen aujourd'hui ne sont pas des obstacles à l'introduction de droits positifs. Une mise à niveau des règles de droit applicables aux pratiques numériques est nécessaire afin de sécuriser l'accès et l'utilisation de l'information scientifique et technique à partir de plateforme en ligne.

---

<sup>11</sup> <http://www.cnrs.fr/dist/strategie-ist.htm>



Livre blanc – Une Science ouverte dans une République numérique

Le projet de loi pour une République numérique dans ses actuels articles 17 et 18 bis nouveau tels qu'adoptés en première lecture par l'Assemblée nationale le 26 janvier 2016 propose des avancées importantes. Le texte français peut aller plus loin en affirmant les valeurs communes du monde de la recherche et en consolidant le droit de procéder au *text and data mining*.

## La nécessité d'un droit nouveau pour la Science

Plusieurs constats entourent la présentation de ce Livre blanc :

1. **La nécessité d'une mise à niveau du droit des usages de la Science numérique en France** : il est nécessaire que ces usages soient de plain-pied avec tous ceux qui ont adoptés, notamment par les grands pays voisins et l'Europe. Ce n'est pas le cas aujourd'hui et cette mise à niveau s'impose.
2. **La diversité des écosystèmes de la publication scientifique et des pratiques numériques qui l'accompagnent** : notre pays à cet égard est spécifique et des adaptations doivent être trouvées, en France comme ailleurs, pour permettre aux éditeurs scientifiques d'être aux côtés de toutes les communautés scientifiques.
3. **La voie dominante, c'est le développement de la Science ouverte numérique** : cette voie se développe dans tous les grands pays qui proposent aujourd'hui des modalités d'Open access, d'Open process dans des processus et selon des modèles qui sont encore loin d'être stabilisés.
4. Les directions de travail d'un nouveau droit des usages font l'objet de **recommandations** précisées dans ce Livre blanc et en particulier par la modification de l'article 17 du projet de loi pour une République numérique dans sa rédaction issue de l'adoption du texte en première lecture par l'Assemblée nationale (26 janvier 2016).

Puisse la représentation nationale et notre société en général tirer le meilleur parti de notre travail prospectif et d'expertise sur les usages numériques de la science : ce travail est désormais à disposition de tous ceux qui sont partie prenante du futur partage des connaissances dont la France va définir l'avenir.

## II/ Constats

Ci-dessous figurent un ensemble de constats apparus dans les entretiens avec les représentants des communautés de recherche et les chercheurs.

<b>CONSTATS</b>
La multiplication des plateformes et la faiblesse de leur encadrement contractuel engendrent un besoin de gouvernance nouvelle.
La multiplicité des objets de l'IST appelle une clarification du droit et des équilibres entre accès à la connaissance scientifique et préservation du potentiel de valorisation de l'IST dans toutes ses composantes.
L'absence de statut légal de l'exploration de données est source d'une insécurité juridique à laquelle la loi doit apporter une réponse.
L'absence de statut légal de l'exploration de données et l'inadaptabilité du droit du producteur de bases de données au traitement dynamique de la connaissance sont source d'une insécurité juridique à laquelle la loi doit apporter une réponse.
Les contrats d'édition signés par les chercheurs sur leurs articles rédigés dans le cadre de la recherche financée sur fonds public prévoient majoritairement des cessions exclusives de droits au profit de l'éditeur. Ces contrats constituent de véritables contrats d'adhésion.
Les chercheurs ont besoin pour réaliser leurs travaux de recherche d'un accès libre et gratuit sous forme numérique à l'ensemble des données de la Science composées : <ul style="list-style-type: none"><li>○ des résultats scientifiques en ce compris les résultats publiés par un éditeur scientifique ;</li><li>○ des données de la recherche entendues comme les données ayant servi à établir ces résultats.</li></ul>
Les chercheurs ont exprimé le besoin de partager les données de Science.
La pratique d'un dépôt en archive ou en plateforme en fonction du domaine des articles, doit être généralisée.
Les chercheurs ont exprimé le besoin : <ul style="list-style-type: none"><li>○ d'une plateforme guichet unique de la connaissance scientifique ;</li><li>○ de régulation des plateformes par le droit.</li></ul>
Les chercheurs ont besoin d'accéder au dernier état des connaissances. Si dans le cadre d'un compromis avec les éditeurs une période d'embargo peut être définie, celle-ci ne doit pas dépasser les délais maximaux prévus dans la recommandation de la Commission européenne (C(2012) 4890) (6 et 12 mois selon les catégories de disciplines) et les délais pratiqués à l'étranger au risque de marginaliser et discriminer la recherche française. La distinction entre sciences exactes et sciences humaines et sociales est contestée dans son principe.
Les dispositions issues de la loi n°2015-1779 ne permettent pas d'adapter la mise à disposition publique des données produites par les établissements et institutions d'enseignement et de recherche. Ces dispositions ne vont pas dans le sens des besoins des chercheurs et des usages

des communautés scientifiques, et ne prennent pas en compte la nature des données (données issues de recherche en cours, de savoir-faire, de ZRR...).

Les écrits et publications scientifiques sont protégeables par le droit d'auteur s'ils sont originaux dans leur forme d'expression. Le chercheur est titulaire des droits d'auteur sur ses articles et écrits scientifiques.

L'Open Science doit préserver les secrets ainsi que la sécurité publique.

Le Code de la recherche énonce déjà dans ses principes les fondements d'un droit numérique de la Science ouverte.

### III/ Recommandations principales

Ci-dessous figurent un ensemble de propositions à débattre avec tous les acteurs, en vue d'optimiser les usages de l'information scientifique numérique.

1	<b>Adoption de l'article 17 du projet de loi pour une République numérique (texte adopté n°663)</b>
2	<b>Adoption de l'article 18 bis (nouveau) du projet de loi pour une République numérique créant une exception au droit d'auteur et au droit du producteur de base de données en faveur du <i>text and data mining</i></b>
3	<b>Participation à la création d'une démarche européenne de « Science ouverte » (priorité du Commissaire à la recherche et à l'innovation pour 2016)</b>
4	<b>Création d'un référentiel des usages de l'IST numérique</b>
5	<b>Définition d'un ensemble de normes Iso sur les usages de l'IST</b>
6	<b>Définition de contrats types de cession de droits d'auteur</b>
7	<b>Rédaction d'une charte éthique de la Science numérique</b>
8	<b>Création d'une Agence pour le développement de la Science ouverte</b>
9	<b>Création d'une convention internationale de la Science ouverte</b>

## Plan commenté

1. Les données de la Science sont, au même titre que nombre de données, au cœur de l'évolution d'un système dont les équilibres sont bouleversés par l'avènement des nouveaux outils numériques et l'affirmation de valeurs d'ouverture, de partage et de collaboration. Des interrogations, des craintes, des jeux et enjeux de pouvoir jaillissent nécessairement de ces évolutions et nécessitent la clarification des dispositions légales et la sécurisation des pratiques.

2. Face au constat d'une « profonde transformation de la production et de la diffusion de la science (caractérisée par une croissance rapide du nombre des publications scientifiques et, dans le même temps, par des coûts toujours plus lourds pour accéder à ces ressources dans le cadre d'un budget au mieux constant) », l'objectif affirmé de la « Stratégie numérique du Gouvernement » telle que présentée par le Premier ministre le 18 juin 2015 est de « Favoriser une science ouverte par la libre diffusion des publications et des données de la recherche ».

3. Le présent Livre blanc propose un raisonnement en deux temps.

4. **Le premier temps permet d'établir un état des lieux de la Science** dans l'environnement numérique en constatant :

- les usages des chercheurs et de leurs équipes. Ces usages ont notamment été recensés dans le cadre d'une enquête sur les usages et besoins d'IST des Unités de recherche, réalisée auprès des Directrices et Directeurs d'Unités du CNRS ; cette enquête a été réalisée à la mi-2014 par la Direction de l'Information Scientifique et Technique du CNRS auprès de 1250 unités publiantes ;
- des carences juridiques pénalisantes et le besoin d'un droit rénové ;
- des risques de captation abusive.

5. Appréhendé comme un véritable outil de travail des communautés scientifiques, les données ou informations scientifiques accessibles sur des plateformes numériques sont l'objet d'usages, échanges, manipulations, traitements multiples, qui bouleversent les conceptions traditionnelles du droit de la propriété matérielle et intellectuelle.

6. Le cadre juridique tel qu'existant et mis en perspective des législations étrangères, ne permet pas aux chercheurs :

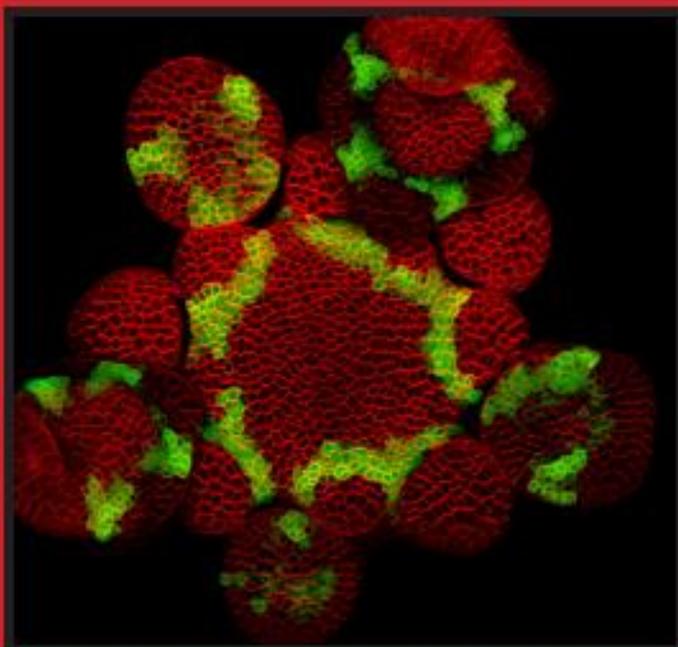
- d'accéder librement aux résultats scientifiques ;
- d'utiliser pleinement les fonctionnalités d'analyse ou de traitement offertes par les plateformes numériques.

7. Ces constats ont par ailleurs été massivement validés dans le cadre de la consultation nationale sur le projet de loi pour une République numérique.

**8. Le second temps formule des propositions juridiques en faveur d'une Science ouverte** dont les pouvoirs publics sont invités à s'emparer. Ces propositions résultent :

- des valeurs et des besoins formulés par les grands témoins de ce Livre Blanc. L'approche privilégiée est en effet celle d'une démarche de consensus qui concourt à l'émergence et au partage de valeurs communes à la recherche publique ;
- d'une analyse des prises de positions des instances communautaires et internationales ainsi que des législations étrangères.

9. Le projet de loi pour une République numérique pourrait incarner l'ambition de la France de promouvoir les valeurs de l'Open Science. Ce cadre est une opportunité d'inscrire la recherche publique française dans le mouvement global de l'Open, tout en préservant les intérêts de l'ensemble des acteurs de la Science, du laboratoire ou de l'organisme public qui finance le projet de recherche à l'éditeur scientifique, au partenaire privé, en passant par le chercheur. Des axes de développement d'un droit nouveau de l'information scientifique et technique numérique sont proposés, dans la perspective d'un rééquilibrage de l'écosystème de la science numérique.



1. L'Etat des lieux :  
« La science en transition »

## **1. L'ETAT DES LIEUX : "LA SCIENCE EN TRANSITION"**

10. La Science est en transition vers un droit nouveau et des transformations majeures de ses usages.

11. Les nouveaux usages numériques de l'IST pratiqués par les chercheurs et les laboratoires (1.1) sont en décalage avec le cadre juridique existant (1.2), absent dans certains domaines, comportant des lacunes dans d'autres.

12. La Science est également aux prises « avec un risque nouveau de captation de ses données, notamment par des éditeurs scientifiques qui demandent des cessions de licence sur les jeux de données intégrés ou associés aux publications de recherche qu'ils éditent »<sup>12</sup> (1.3). L'exposé des motifs du projet de loi pour une République numérique affirme ces constats, ceux-ci sont confirmés notamment par la consultation nationale, par le Conseil national du numérique ainsi que par l'étude d'impact du projet de loi (1.4).

### **1.1 Photographie des usages de la Science**

13. La Science par ses pratiques est au cœur de la transition numérique.

14. Ce paragraphe propose une photographie des pratiques en vigueur en France et à l'étranger concernant :

- l'utilisation de l'IST :
  - o par les chercheurs selon la communauté scientifique à laquelle ils appartiennent, notamment l'utilisation de l'IST comme outil de recherche ;
  - o par les laboratoires, instituts, organismes de l'IST quant à l'organisation de l'accès à l'IST;
- les contrats de l'édition scientifique quant aux conditions de publication, notamment dans les partenariats public / privé.

#### **1.1.1 La Science au cœur de la transition numérique**

15. La transition numérique est marquée par la démultiplication des données disponibles et accessibles à tout instant et partout dans le monde. Cette ère du big data se caractérise par :

- l'absence de frontières et la globalisation de l'information ;
- un esprit collaboratif et de partage ;

---

<sup>12</sup>Projet de loi pour une République numérique, Exposé des motifs, page 6

- l'automatisation des activités notamment professionnelles ;
- la création de valeur.

#### **1.1.1.1 L'absence de frontière**

16. L'internet est l'outil phare de la transition numérique. Ce réseau mondial est marqué par l'absence de frontière physique, de visa d'entrée, et au contraire, par l'existence des flux transfrontaliers et d'échanges à travers le monde des informations.

17. Grâce au protocole IP et à l'interconnexion entre les différents opérateurs, l'internet constitue un réseau dit sans couture ou de bout en bout.

18. Ce réseau mondial, en dépit de failles et d'inégalités comme la fracture numérique ou encore l'équipement en débit différencié selon les politiques publiques des pays, permet à chacun d'avoir accès à une information globalisée.

#### **1.1.1.2 Le droit de la collaboration et du partage**

19. La transition numérique est marquée par le passage du papier au numérique, autrement dit par le passage d'une propriété matérielle unique à une démultiplication de la propriété.

20. La génération du web 2.0 se caractérise notamment par la simplification de la mise en ligne des contenus et par la capacité d'interaction des usagers ; elle repose sur l'autoproduction, la diffusion et la consultation directe par les usagers eux-mêmes des contenus multimédia.

21. Le principe de collectivisation des contenus a transformé fondamentalement les usages. Le partage et l'esprit collaboratif est le modèle du web dont Youtube, Wikipedia et les réseaux sociaux sont les archétypes.

#### **1.1.1.3 Le droit de l'automatisation**

22. Grâce aux nombreuses applications technologiques, les administrations, les organismes de recherche et les entreprises disposent de nombreuses données et métadonnées.

23. Le développement d'outils d'analyse sémantique et lexicographique est nécessaire au traitement de ces données. Les processus de travail dématérialisés voire automatisés permettent la création de contenus générés par l'utilisateur, les user-generated contents.

### 1.1.1.4 Le droit de la création de valeur

24. La transition numérique est génératrice d'une nouvelle économie dite numérique. Elle est basée sur le commerce en ligne mais également sur les échanges non commerciaux tels que le partage de la connaissance et les user-generated contents (données d'usage, wiki, forum de discussion, blog, tweet...)

25. Le rapport de l'OCDE du 12 avril 2007<sup>13</sup> avait déjà mis en exergue l'impact économique très important des user-generated content, et la création de valeur et d'innovation en résultant.

### 1.1.2 Les nouveaux usages des chercheurs

26. L'enquête du CNRS ainsi que les auditions ont permis de mettre en évidence les pratiques des chercheurs et des instituts quant à l'utilisation de l'IST numérique.

#### **Audition Université de Strasbourg Paul-Antoine HERVIEUX, 10 juillet 2015**

« Les usages se modifient et les données de la recherche (BSN10 – données de la recherche<sup>14</sup>) sont en train de changer le paradigme. Les chercheurs, et enseignants-chercheurs, commencent à se rendre compte de la valeur de leurs données. Dès que nous parlons de valeur, il faut réfléchir au droit. Nous sommes aujourd'hui dans un monde où les entreprises privées sont prédominantes et s'intéressent de plus en plus aux données pour en privatiser les usages. »

27. L'utilisation de l'IST par les chercheurs dans le cadre de leur activité de recherche est marquée par les constats suivants :

- l'IST est un outil de travail que l'on peut utiliser, partager, exploiter librement dans le cadre des travaux de recherche ;
- les bases de données sous forme d'archives ouvertes sont utilisées de manière hétéroclite selon les communautés ;
- les communautés scientifiques ne disposent pas d'une connaissance suffisante du droit d'auteur et du droit sui generis des bases de données pour un usage maîtrisé de l'IST.

<sup>13</sup> OCDE, Working Party on the Information Economy "Participative web: user-created content" 12-4-2007 <http://www.oecd.org/sti/38393115.pdf>

<sup>14</sup> <http://www.bibliothequescientifiquenumerique.fr/bsn-10-donnees-de-la-recherche/>

### 1.1.2.1 L'IST numérique : un outil d'exploration et d'analyse

28. L'IST est avant tout vue par les chercheurs comme un instrument de travail auquel ils doivent avoir accès, qu'ils peuvent partager, échanger, réutiliser ou retraiter pour les besoins de leur sujet de recherche. L'IST fait partie de la démarche scientifique en tant que matière première.

29. Les communautés scientifiques distinguent au sein de l'IST deux grandes catégories de données :

- les données de la recherche ;
- les publications.

30. D'autres catégories de données sont spécifiques à certaines communautés :

- le matériel : notamment en biologie ;
- les données de tiers : notamment en Sciences humaines et sociales (SHS).

31. **Accès à l'IST.** La contribution du Conseil Scientifique permet de dégager les grandes pratiques des chercheurs en matière d'accès à l'IST : accès aux publications, accès aux données de la recherche, et pour certaines communautés, accès au matériel de la recherche et accès aux données de tiers utilisées dans le cadre de la recherche.

Communauté	Accès à l'IST
Sciences humaines et sociales	<p><b>Accès aux publications</b> : « En sciences humaines et sociales, en ce qui concerne les publications scientifiques récentes, alors que de plus en plus de revues francophones sont en libre accès immédiat ou après quelques années (notamment grâce à HumaNum, BSN, OpenEdition), les revues anglophones sont souvent cantonnées à des plateformes assez chères. »</p> <p><b>Accès aux données chiffrées / produites par la recherche</b> : « Des plateformes d'échange de données chiffrées (...) ont également été mises en place (Réseau Quetelet, DIMESHS, etc.) : elles assurent à la fois une meilleure circulation des données, le respect des contraintes nécessaires, d'anonymisation par exemple, et la documentation ("métadonnées") sans laquelle les chiffres seraient inutilisables. »</p>

Communauté	Accès à l'IST
	<p><b>Accès aux données de tiers</b> : « Le problème est ici que nombre de données utilisées par les scientifiques en SHS ne sont pas produites par eux (qu'il s'agisse d'une chanson, d'un rapport d'activité d'entreprise ou de l'architecture d'un monument) : d'autres personnes physiques ou morales disposent de droits sur elles. »</p>
Science de l'univers	<p><b>Accès aux données de la recherche</b> : « Les données sont libres d'accès en astronomie pour toute la communauté après la fin d'une période propriétaire. »</p>
Biologie	<p><b>Accès aux publications</b> : « L'édition numérique est généralisée. La recherche d'articles se fait via des plateformes développées par des institutions académiques, permettant d'accéder librement aux résumés d'articles. La plus importante (PubMed) est proposée par le National Institute of Health américain (NIH). »</p> <p>Accès également via PubMed à une version non formatée de tout article publié par un éditeur décrivant des travaux financés par le NIH.</p> <p><b>Accès aux données de la recherche</b> : « De nombreux éditeurs dont Nature, conditionnent d'ailleurs la publication d'un article au dépôt des données en masse, associées à une publication, sur une plateforme accessible gratuitement par tous. »</p> <p><b>Accès au matériel</b> : « Il est à noter que cette exigence va au-delà des données numériques et concerne également le matériel produit dans le cadre d'une publication. »</p>
Physique	<p><b>Accès aux données de la recherche</b> : « l'accès libre aux données brutes n'est pas encore très répandu. »</p> <p>« De nombreuses bibliothèques numériques ont été constituées et mises en libre accès par des groupes de chercheurs. »</p>

Communauté	Accès à l'IST
Chimie	<p><b>Accès aux publications</b> : « La règle, ce sont des publications dans des journaux payants de sociétés savantes (American Chemical Society, Royal Society) ou sociétés commerciales (Wiley, Elsevier, ...) et un développement timide de l'"Open access" type "gold" payé par les auteurs. »</p> <p>« Des bases de données libres d'accès se développent, notamment la base de Cambridge qui contient toutes les structures moléculaires publiées. »</p>
Informatique et Mathématique	<p><b>Accès aux publications</b> : « Les bases de données relatives aux publications sont très importantes pour le travail individuel et communautaire. Une particularité de cette discipline est l'importance de l'accès facile aux publications tant récentes qu'"anciennes" (plusieurs années, décennies voire siècles). La pérennité quant à l'accès à ces publications est donc capitale pour la recherche. »</p>

32. L'accès à ces catégories de données se fait par l'intermédiaire de plusieurs outils techniques :

- les portails des éditeurs ;
- les archives numérisées ;
- les plateformes de dépôt volontaire ;
- le papier (la question du PAP 1, 6 de l'enquête du CNRS<sup>15</sup> note « l'attachement historique des mathématiques et des SHS à la documentation papier porté par des besoins / usages de documents anciens »).

33. **Traitement de l'IST.** Les communautés scientifiques pratiquent de manière différenciée le partage des données ainsi que les techniques de fouilles de textes et de données.

<sup>15</sup> <http://www.cnrs.fr/dist/z-outils/documents/Enqu%C3%A0te%20DU%20-%20DIST%20mars%202015.pdf>

Communauté	Traitement de l'IST
Sciences humaines et sociales	« Partage de données et techniques de "fouille de données et de textes" sont ainsi inégalement répandus selon les types de données, pour des raisons principalement d'obstacles juridiques [données de tiers], et de manque de moyens en personnel pour la production et le maintien de métadonnées de qualité. »
Sciences de l'univers	« Il est nécessaire d'harmoniser et de standardiser les formats, les descriptions et les modes d'accès aux données d'archives, aux métadonnées et aux applications susceptibles de les traiter, de manière à réaliser l'interopérabilité. »
Biologie	« Si les techniques de "fouille de données" sur les textes ne sont pas prioritaires pour la plupart des domaines de la biologie en termes d'outil de découverte, (mais plutôt en termes de collecte documentaire) la "fouille de données" proprement dite prend une place de plus en plus importante ».
Informatique et Mathématique	« Par ailleurs les mathématiques et l'informatique jouent un rôle important dans l'analyse, la gestion et l'exploitation des masses des données (questions autour du Big Data). Il est certainement très important que les données soient accessibles, mais quand elles deviennent de plus en plus massives, il faut aussi pouvoir les exploiter de manière efficace. »

### 1.1.2.2 Le partage des connaissances : dépôt des résultats scientifiques dans des archives ouvertes numériques

34. **Pratique contrastée.** L'enquête du CNRS et la contribution de son Conseil Scientifique ont mis en évidence une utilisation différente des archives ouvertes selon les communautés scientifiques.

35. L'archive ouverte pluridisciplinaire HAL (Hyper Articles en Ligne) est celle dont l'utilisation est la plus répandue dans la communauté nationale. Elle est destinée au dépôt et à la diffusion de documents scientifiques de tout domaine de la recherche, ces documents n'ayant pas nécessairement été validé par un comité éditorial ou un comité

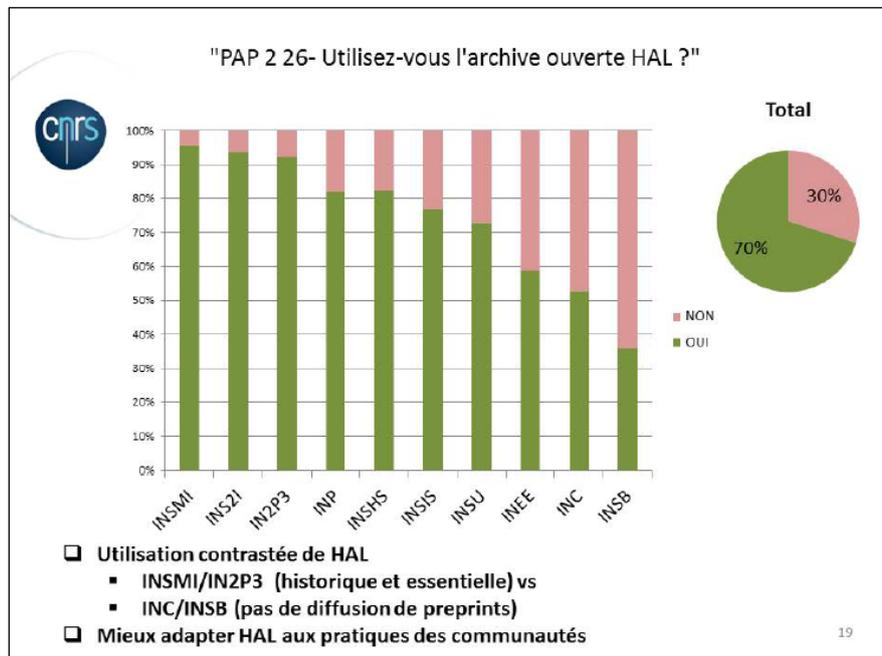
Livre blanc – Une Science ouverte dans une République numérique

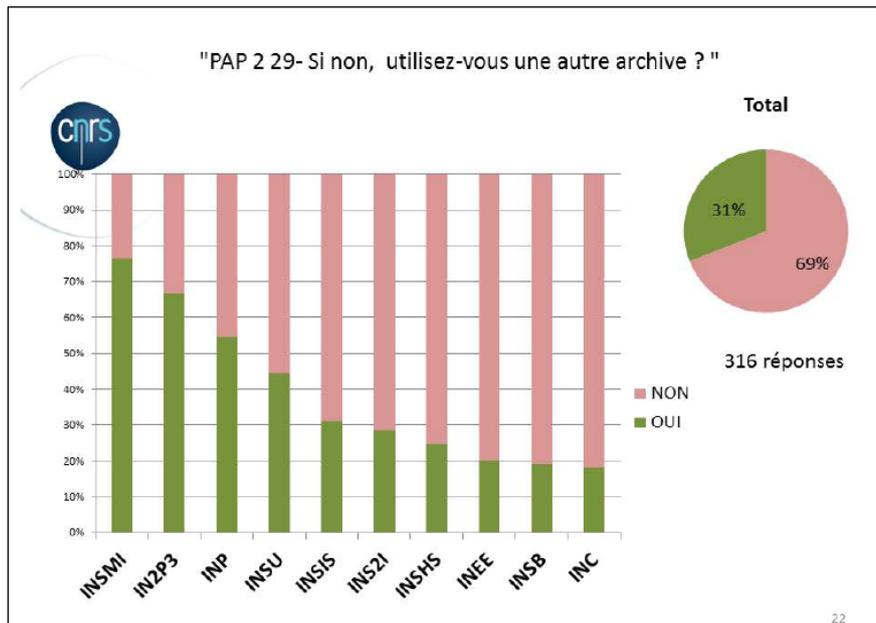
de programme. Ces documents émanent de chercheurs d'établissements d'enseignement et de recherche français ou étrangers, publics ou privés.

36. La question du PAP 2, 26 permet de constater une utilisation contrastée de HAL selon les communautés scientifiques. HAL est largement utilisée par les chercheurs de l'Institut national des sciences mathématiques et de leurs interactions et ceux de l'Institut national de physique nucléaire et de physique des particules ; ces instituts l'utilisent de manière historique ; alors que les chercheurs de l'Institut national de Chimie ou encore ceux de l'Institut des sciences biologiques ne l'utilisent que très peu et n'utilisent pas ou peu d'autres archives.

37. Dans l'hypothèse où les communautés n'utilisent pas HAL comme archive ouverte, elles n'utilisent pas d'autres archives à 69 % (PAP 2, 29). Par conséquent, de nombreuses publications sont uniquement archivées par les éditeurs et ne sont pas librement mises à disposition des communautés scientifiques après une période de réservation. L'accès aux publications n'est possible que par la plateforme de l'éditeur, par un abonnement ou par l'achat de l'article de manière isolée.

Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015





38. Les contributions du Conseil Scientifique apportent certaines précisions sur le dépôt dans des archives ouvertes :

Communauté	Pratique du dépôt
Sciences humaines et sociales	« Le dépôt parallèle d'articles dans des archives ouvertes est peu développé. »
Sciences de l'univers	Pratique très répandue
Physique	<b>Serveur de prépublication (preprint)</b> : « Le numérique joue également un rôle essentiel dans le domaine de la diffusion des résultats, avec l'utilisation quasi systématique des serveurs de prépublication. Le dépôt sur ces serveurs se fait de manière simultanée avec l'envoi à une revue scientifique avec comité de lecture : il permet de « prendre date » immédiatement. »
Chimie	« Il n'y a pas de pré-publication comme ArXiv. »
Informatique et Mathématique	<b>Plateforme d'archivage</b> : « Les plateformes d'archivage des publications comme HAL ou ArXiv répondent donc en partie à cette problématique et sont à soutenir, ainsi que les plateformes de métadonnées ( <a href="#">MathSciNet</a> , <a href="#">Zentralblatt</a> ,...). »

39. **Données déposées.** Les données déposées par les chercheurs sur les archives ouvertes diffèrent d'une communauté à l'autre :

- publications en preprint<sup>16</sup> (article, ouvrage, chapitre d'ouvrage) ;
- publications en postprint<sup>17</sup> (article, ouvrage, chapitre d'ouvrage) ;
- mémoire de thèses (doctorat, master,...) ;
- cours ;
- images, vidéos, sons, cartes ;
- notices bibliographiques seules ;
- notices bibliographiques accompagnées de la publication.
- données de la recherche ;
- métadonnées associées à ces données.

40. L'enquête du CNRS met en exergue que les communautés qui déposent dans HAL :

- déposent les notices bibliographiques à 63 % (PAP 2, 27) ;
- déposent le texte intégral à 56 % (PAP 2, 28) ;

41. Les communautés qui ne déposent pas dans HAL n'utilisent pas d'autre archive à 69 %.

### 1.1.2.3 Un droit des usages à créer entièrement

42. Les opérations de publication, de dépôt dans des archives ouvertes, de *text and data mining* ou relevant d'autres techniques de traitement ou de fouille de données ou des textes, sont pratiquées par les chercheurs en méconnaissance quasi-complète du droit de la propriété intellectuelle.

#### **Audition Université de Strasbourg Paul-Antoine HERVIEUX, 10 juillet 2015**

« Dans les communautés universitaires il y a une méconnaissance quasi générale du droit d'auteur et de ses modalités d'application, quel que soit le type de documents ou de données. »

43. Dans le cadre de l'enquête du CNRS, deux questions permettent d'évaluer le niveau de connaissance des chercheurs sur leurs droits.

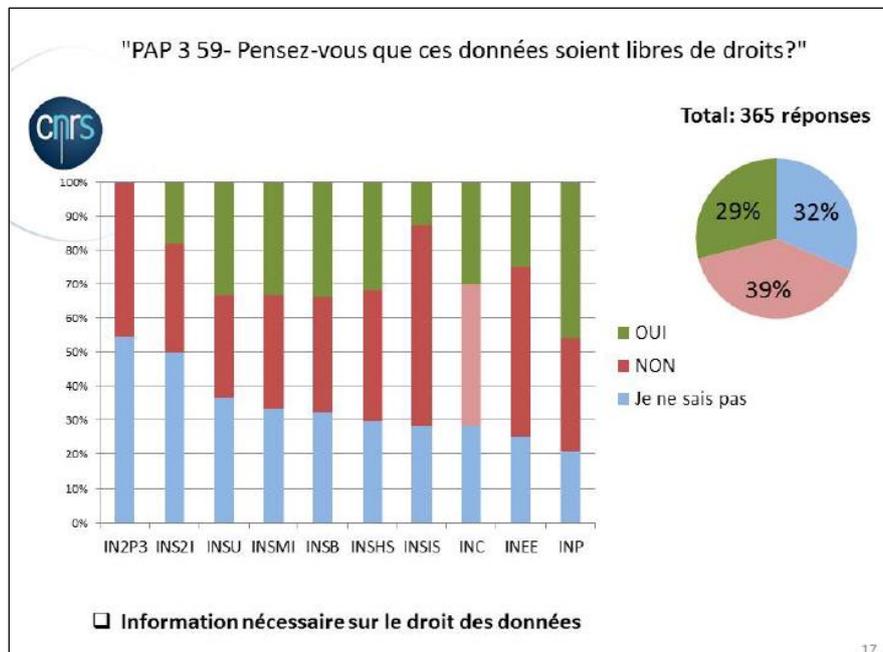
<sup>16</sup> **Version soumise** ("Preprint") : version soumise par le ou les auteurs à la revue avant tout processus de révision ("peer reviewing" ) par le comité de lecture par les pairs\* ("editors and referees").

<sup>17</sup> **Version acceptée** ("Postprint") : version après que le ou les auteurs aient tenu compte des remarques des évaluateurs et que l'article soit accepté par le comité éditorial.

Livre blanc – Une Science ouverte dans une République numérique

44. A la question « Pensez-vous que ces données [les données brutes] soient libres de droits ? », les réponses se partagent entre « Oui » à 29 %, « Non » à 39 % et « Je ne sais pas » à 32 %.

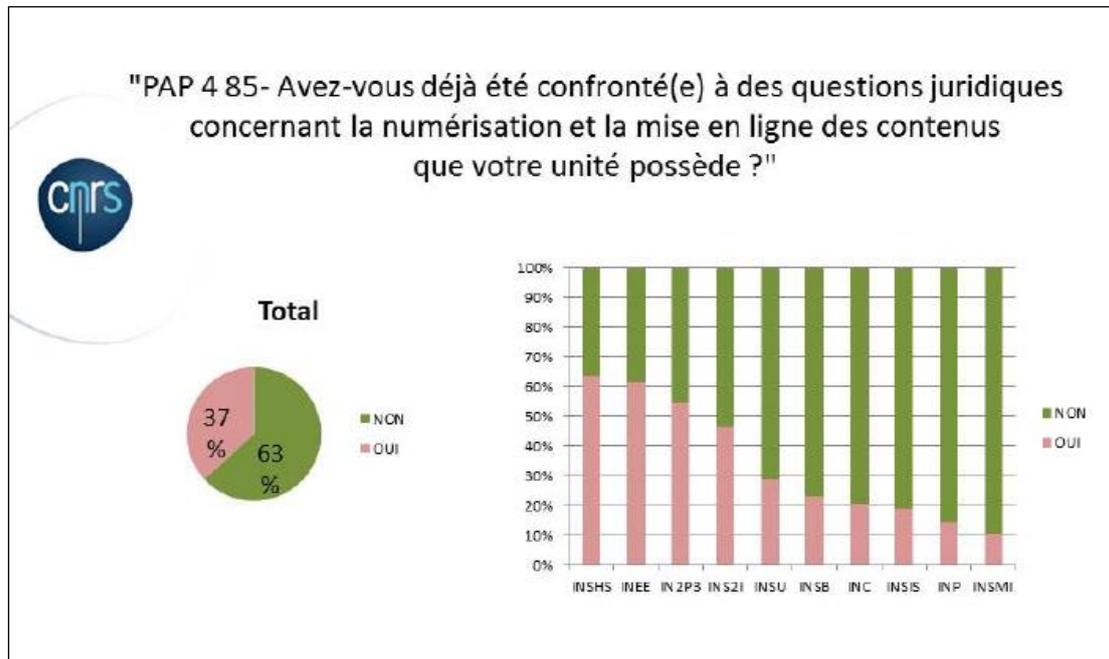
Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015



45. La répartition des réponses à cette question en trois groupes équivalents montre que les chercheurs :

- ne savent pas différencier ce qui est libre de droit de ce qui est réservé par un droit de propriété intellectuelle ;
- ne connaissent pas leur droit sur les données brutes et par extension sur les données protégées (les publications).

46. La deuxième question révélatrice de leur méconnaissance est la suivante :



47. On constate que 63 % des sondés n'ont jamais été confrontés à des questions juridiques concernant la numérisation et la mise en ligne des contenus. Pour être « confronté à des questions juridiques », encore faut-il avoir connaissance de l'existence des problématiques soulevées par la numérisation et la mise en ligne au regard du droit d'auteur.

48. Cette méconnaissance du droit sur les données de recherche révèle que des pratiques illégales peuvent être perpétrées par ignorance et appelle à une clarification voire une affirmation des règles de droit au regard de ces pratiques et des besoins des chercheurs. Des opérations de formation devront accompagner ces évolutions législatives souhaitables.

### 1.1.3 Une forte demande des laboratoires, des instituts, des organismes

49. La multiplication des plateformes et la multiplicité des objets de l'IST engendrent un besoin grandissant de gouvernance et de définition d'un droit numérique de la Science.

### 1.1.3.1 Multiplication des plateformes : un besoin de gouvernance

50. **Multiplication des plateformes.** Il est constaté une multiplication des plateformes d'IST au cours de ces dernières années :

- les plateformes institutionnelles ;
- les plateformes thématiques ;
- les plateformes de dépôt ;
- les plateformes bibliographiques ;
- les plateformes d'archivage...

51. Cette multiplication entraîne une dilution de l'information et un risque de perte de sa valeur. Elle empêche par ailleurs de procéder à des recherches croisées d'informations ainsi qu'à des recherches pluridisciplinaires. Certaines plateformes proposent des services à valeur ajoutée alors que d'autres proposent une simple consultation des textes, voire un simple accès aux données bibliographiques. Cette multiplication a un coût particulièrement élevé tant dans la qualité des services qui en résulte qu'humain ou financier.

52. **Faiblesse du cadre juridique des plateformes.** Ces différentes plateformes sont dotées de conditions générales d'utilisation plus ou moins restrictives quant à l'utilisation de l'IST, pas toujours en phase avec le droit de la propriété intellectuelle et les contrats d'édition conclus entre les chercheurs et les éditeurs.

53. Le projet de loi pour une République numérique propose d'introduire une définition de la notion de plateforme et d'y associer une obligation de loyauté. Ces dispositions sont introduites dans le Code de la consommation mais pourraient être étendues comme « bonnes pratiques » aux plateformes de Science.



La multiplication des plateformes et la faiblesse de leur encadrement contractuel engendrent un besoin de gouvernance nouvelle.

### 1.1.3.2 Multiplicité des objets d'IST (données, analyses, articles) : un besoin de cohérence pour la valorisation

54. La valorisation des fruits de la recherche publique est une préoccupation essentielle et le dépôt de brevets est au cœur du processus de valorisation.

55. Le CNRS est, selon le classement SIR (Scimago Institutions Rankings) 2014 le premier déposant de brevets au monde<sup>18</sup> à partir des publications scientifiques d'établissements. La plupart de ces brevets sont en copropriété avec des universités. Ce palmarès est le résultat d'une recherche de pointe mais également d'une politique de valorisation incitative. La recherche française de manière générale est active et produit de nombreuses innovations brevetables.

56. La recherche publique a mis en place une politique d'intéressement des unités de recherche dont sont issues les inventions et de leurs inventeurs. Les agents reçoivent ainsi une prime forfaitaire au brevet d'invention, dont le montant est fixé par arrêté conjoint des ministres chargés du budget, de la fonction publique et de la recherche. Elle est pour chaque agent affectée du coefficient représentant sa contribution à l'invention<sup>19</sup>.

57. Le brevet ne constitue cependant pas le moyen exclusif de valorisation, et peut s'avérer inadapté à certaines innovations, pour lesquelles le dépôt d'un brevet est impossible ou inopportun.

58. L'IST recouvre des éléments multiples, données, analyses, résultats, résultats de traitements, articles, requêtes, contenus générés par les utilisateurs, etc... qui sont autant de sources potentielles de valorisation dont la recherche française ne doit pas se priver.

59. L'accès ouvert aux plateformes scientifiques ne doit pas être de nature à compromettre la valorisation de l'IST dans ses différentes composantes quelles qu'elles soient.

60. La recherche d'un équilibre entre valorisation et open access passe par la clarification des droits et du régime juridique applicable aux objets de l'IST et impose de tracer une ligne de partage entre bien commun et innovation protégée, entre liberté d'accès et réservation privative.

**+** La multiplicité des objets de l'IST appelle une clarification du droit et des équilibres entre accès à la connaissance scientifique et préservation du potentiel de valorisation de l'IST dans toutes ses composantes.

---

<sup>18</sup> <http://www.gfii.fr/fr/document/le-cnrs-conforte-sa-premiere-place-mondiale-en-nombre-de-publications-scientifiques>

<sup>19</sup> [http://www.cnrs.fr/dire/termes\\_cles/interressement.htm](http://www.cnrs.fr/dire/termes_cles/interressement.htm)

## 1.2 Les vides juridiques

61. L'émergence du numérique dans les pratiques d'IST provoquent un décalage entre le droit et la pratique. Si le Code de la recherche organise la recherche publique et définit ses objectifs, il n'affirme à aucun moment les valeurs communes de la Science.

62. L'IST se manipule à partir de plateformes multiples dont le modèle juridique n'existe pas et proposant des outils de traitement de données dont la légalité est remise en question.

63. Par ailleurs, les pratiques contractuelles sont asymétriques aux pratiques des communautés de la recherche et en défaveur de la Science.

### 1.2.1 Un droit de la science qui fait défaut

64. **Absence de droit de la Science.** Aucune disposition légale, aucun texte ne reflète les valeurs des communautés scientifiques et n'affirme l'intérêt supérieur de la Science.

65. Ce vide juridique se fait de plus en plus sentir dans le cadre de la transition numérique et le développement des services à valeur ajoutée sur les données de Science.

### 1.2.2 Droit des plateformes : des évolutions en cours

66. La notion de plateforme ne connaît ni de statut ni de régime légal. Ce vide juridique entraîne une certaine insécurité juridique déjà soulignée par le Conseil National du Numérique dans son avis du 13 juin 2014 ainsi que par le Conseil d'Etat dans son rapport 2014 « Le numérique et les droits fondamentaux ».

67. L'article 22 du projet de loi pour une République numérique prévoit l'introduction de la définition de la notion de plateforme :

- « activités consistant à classer ou référencer des contenus, biens ou services proposés ou mis en ligne par des tiers, ou à mettre en relation, par voie électronique, plusieurs parties en vue de la vente d'un bien, de la fourniture d'un service, y compris à titre non rémunéré, ou de l'échange ou du partage d'un bien ou d'un service. »

68. Une obligation de loyauté est également imposée à l'opérateur de la plateforme :

- « tout opérateur de plateforme en ligne est tenu de délivrer au consommateur une information loyale, claire et transparente sur les conditions générales d'utilisation du service d'intermédiation qu'il propose et sur les modalités de référencement,

de classement et de déréférencement des contenus, biens ou services auxquels ce service permet d'accéder. »

69. Même si le texte du projet de loi semble régir les relations avec des consommateurs, les plateformes d'Open Science devront appliquer ces principes et fournir aux utilisateurs-chercheurs une information loyale quant à leurs conditions d'utilisation.

### **1.2.3 Un droit au TDM : une absence aux conséquences graves**

#### **1.2.3.1 Enjeux du TDM**

70. **Enjeu majeur.** La pratique du TDM est un enjeu majeur pour la science, la recherche et l'innovation dans la mesure où le TDM permet de dégager de nouveaux sujets de recherche, de nouvelles connaissances et de répondre à des problématiques économiques, sociales et sociétales<sup>20</sup>.

71. En découlent des opportunités en termes de valorisation de ces nouvelles connaissances avec des enjeux d'innovation, de croissance et d'emploi.

72. Les enjeux scientifiques et économiques sont d'autant plus importants que la pratique du TDM est mondiale et fait l'objet d'une normalisation différente d'un pays à l'autre, y compris à l'intérieur de l'Europe. Les Etats-Unis et le Royaume-Uni ont affirmé le droit des chercheurs à procéder à des opérations de TDM ; l'Allemagne a introduit un droit d'exploitation secondaire des publications scientifiques.

73. Un projet de texte révisé de la directive DADVSI (Directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information) était attendu pour fin 2015 mais n'est toujours pas diffusé au jour de la parution du présent Livre blanc. Le processus d'acceptation d'un projet de Directive est long : il faut compter au moins 2 à 3 ans avant que cette directive soit acceptée.

74. La recherche française ne peut pas se permettre d'être discriminée face à ses voisins européens et de prendre un retard qui risque d'être irrémédiable et dont les conséquences préjudiciables sont multiples : retard dans le développement des techniques numériques de recherches, retard dans les sujets de recherche émergents,

---

<sup>20</sup> L'exploration de données a été présentée en 2001 par le Massachusetts Institute of Technology comme l'une des dix technologies émergentes qui « changeront le monde au 21ème siècle ». (Data mining et statistique décisionnelle – l'intelligence des données Stéphane Tuffery Editions Technip 2012)

perte de contrats de partenariat au niveau européen, baisse du nombre de brevets déposés, risque de privatisation des techniques d'exploration de données, etc.

**75. Avant-projet V.1.** En insérant dans la première version de l'avant-projet de loi pour une République numérique des dispositions permettant aux chercheurs de procéder à des opérations de fouilles de données, le gouvernement semble avoir pris la mesure de l'enjeu du TDM pour la recherche et notamment la recherche publique. Or ces dispositions ont été retirées et ne figurent plus dans le projet de loi tel que transmis au Conseil d'Etat.

76. Le retrait de ces dispositions du projet de loi a été relevé dans de nombreuses contributions par la communauté scientifique dans le cadre de la contribution publique sur le projet de loi.

77. La justification de ce retrait avancée par le gouvernement est double :

- le texte prévu organisait une exception au droit d'auteur, or le texte de la directive DADVSI qui règlemente le droit d'auteur entre les pays membres de l'Union européenne empêche les Etats membres de créer des exceptions non prévues par les textes ;
- la directive DADVSI va certainement être révisée et les rapports rendus dans ce cadre prônent de manière quasi-unanime l'introduction de cette exception TDM aux dispositions du droit d'auteur.

**78. Projet de loi TA n°663.** Dans le cadre du débat parlementaire sur le projet de loi pour une République numérique, des députés d'appartenance politique différente ont soutenu l'introduction d'un amendement créant une exception de text and data mining. Une exception au droit d'auteur et au droit du producteur de bases de données a été introduite dans le texte du projet de loi (article 18 bis nouveau du texte adopté n°663).

79. Sous la double impulsion de Fleur Pellerin, ministre de la Culture et de la Communication, et de Thierry Mandon, secrétaire d'Etat chargé de l'Enseignement supérieur et de la Recherche, il a également été confié à Charles Huot, président du Groupement français des industriels de l'information, une « mission de concertation et de proposition pour faciliter le développement du recours en France aux technologies de fouille et d'exploration de textes et de données » (TDM).

### **1.2.3.2 Le *Text and data mining* au regard du droit d'auteur**

80. Le droit d'auteur inclut le monopole de la reproduction y compris l'adaptation des œuvres. L'éditeur, titulaire des droits patrimoniaux sur les écrits scientifiques qu'il édite,

peut par conséquent interdire aux tiers, ainsi qu'aux auteurs, toute reproduction intégrale ou partielle ainsi que toute traduction, adaptation ou transformation, arrangement ou reproduction par un art ou un procédé quelconque (article L.122-4 CPI).

81. Le TDM ne dispose pas de statut légal propre et cette absence de statut est source d'insécurité juridique. En effet, les services d'exploration de données sont la traduction d'opérations techniques multiples parmi lesquelles :

- des opérations d'analyse ou de traitement de la seule connaissance : il s'agit dans ce cas d'actes libres ;
- des opérations techniques impliquant la reproduction en texte plein des données protégeables par le droit d'auteur et leur modification sous des formes telles que coupes, extraits, fusion, compilation, etc: certains de ces actes pourraient concerner les monopoles d'auteur et par voie de conséquence nécessiter l'autorisation préalable du titulaire de droits.

82. L'absence de statut légal du TDM et le débat doctrinal vif sur l'incompatibilité de ces techniques d'exploration avec les dispositions du droit d'auteur sont source d'insécurité juridique et appellent une clarification législative rapide.

**+** L'absence de statut légal de l'exploration de données est source d'une insécurité juridique à laquelle la loi doit apporter une réponse.

### **1.2.3.3 Le *text and data mining* au regard du droit du producteur de base de données**

83. **Principes du droit sui generis.** Si, par principe, les données ne sont individuellement pas protégeables (sauf dans l'hypothèse où la donnée est protégée par un droit privatif, droit de propriété intellectuelle, droit des données à caractère personnel, droit à la vie privée), l'agrégation d'un nombre significatif de données peut, le cas échéant, être protégée au titre du droit sui generis du producteur de la base de données.

84. Les bases de données sont définies dans le Code de la propriété intellectuelle comme étant tout « recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique, et individuellement accessibles par des moyens électroniques ou par tout autre moyen »<sup>21</sup>.

---

<sup>21</sup> [CPI art. L. 112-3](#). -

85. Le cadre légal de protection des données est défini par les dispositions de la directive du 11 mars 1996 sur les bases de données<sup>22</sup> (transposée en France par la loi concernant la protection juridique des bases de données<sup>23</sup>) qui créent un droit « sui generis » au profit du producteur de la base.

86. Le producteur de la base de données est défini comme la personne qui prend l'initiative et le risque de l'investissement. Il peut interdire :

- l'extraction de la totalité ou d'une partie substantielle du contenu de la base de données ;
- la réutilisation de la totalité ou d'une partie qualitativement ou quantitativement substantielle du contenu de la base de données ;
- et/ou l'extraction ou la réutilisation répétée et systématique de parties qualitativement ou quantitativement non substantielles du contenu de la base de données lorsque ces opérations excèdent manifestement les conditions d'utilisation normale de la base<sup>24</sup>.

87. **Base de données de recherche.** L'IST numérique est accessible à partir des bases de données des éditeurs scientifiques qui ont majoritairement petit à petit remplacé leur édition papier par un accès en ligne à leur revue par l'intermédiaire de leur plateforme. L'éditeur, à ce titre, est producteur d'une base de données et peut par conséquent interdire toute extraction qualitativement ou quantitativement substantielle de sa base.

88. L'IST numérique est également disponible à partir de bases de données institutionnelles, ou encore d'épi-revues, de base de données en Open Access. Les producteurs de chacune de ces bases de données sont également titulaires du droit sui generis et peuvent en interdire toute extraction qualitativement ou quantitativement substantielle.

89. **Droit sui generis vs. TDM.** Pour procéder à des opérations de TDM sur des corpus de données, il est nécessaire de procéder à :

- des opérations d'extraction des bases de données relevant du droit privatif du producteur de la base de données ;
- des opérations techniques ne relevant pas du régime du producteur de bases de données.

---

<sup>22</sup> [Dir. 96/9/CE du 11-3-1996](#) concernant la protection juridique des bases de données.

<sup>23</sup> [Loi 98-536 du 1-7-1998](#), portant transposition dans le Code de la propriété intellectuelle de la directive 96/9/CE du Parlement européen et du Conseil, du 11 mars 1996, concernant la protection juridique des bases de données.

<sup>24</sup> [CPI art. L. 342-3.](#)

90. Certains militent pour une révision du droit du producteur de base de données, le texte d'origine correspondant « à une vision statique des traitements de données qui tend à laisser place à des traitements dynamiques ».

« Sur la base d'un état des lieux de la législation, de la jurisprudence existante et des problématiques auxquelles sont confrontés les acteurs du secteur de la production, du traitement et de l'analyse des données, des propositions d'évolutions peuvent être formulées pour adapter le droit des producteurs de bases de données à son nouvel environnement technologique et commercial, en le transformant en un droit des producteurs et des exploitants de données et de bases de données »<sup>25</sup>.



L'absence de statut légal de l'exploration de données et l'inadaptabilité du droit du producteur de bases de données au traitement dynamique de la connaissance sont source d'une insécurité juridique à laquelle la loi doit apporter une réponse.

#### **1.2.4 Le besoin d'un droit rénové de l'édition numérique scientifique**

91. Les pratiques de l'édition scientifique sont fondamentalement différentes de celles de l'édition littéraire, en effet :

- l'objet de la publication est différent : à la différence de l'œuvre littéraire, dans la publication scientifique, le contenu informatif prévaut sur la forme d'expression, qui peut être accessoire ;
- le contenu de la publication est différent : la création littéraire est celle propre à son auteur alors que le scientifique, dont la « matière première » est la science, exploite majoritairement le travail de ses prédécesseurs ;
- le travail fourni par l'éditeur est différent : l'éditeur littéraire accompagne son auteur dans la rédaction de l'ouvrage, l'encourage y compris financièrement par le versement d'avoir, l'intéresse sur les ventes, procède à la mise en forme, à la mise en valeur, organise l'édition, la diffusion et la promotion, suit l'auteur dans ces activités de promotion, etc. A l'inverse, l'éditeur scientifique ne reçoit que des articles terminés (après passage au comité de lecture), n'intéresse pas l'auteur sur la vente des revues ou des abonnements, procède à la mise en page et à la diffusion en ligne ;

---

<sup>25</sup> « La nécessaire évolution du droit du producteur de base de données pour permettre son adaptation à l'émergence du Big data » par Nicolas COURTIER pour « [La propriété intellectuelle & la transformation numérique de l'économie](#) » Inpi, 10-9-2015

92. L'édition numérique scientifique entraîne une mutation des pratiques et impose nécessairement une révision des règles du jeu contractuel :

- dans la relation entre éditeur et chercheur ;
- dans les contrats d'abonnement ;
- dans les contrats de partenariat public / privé et l'organisation de la publication des articles issus du partenariat.

#### 1.2.4.1 Contrat éditeur / chercheur : contrat de cession de droits d'auteur

93. La publication d'articles par les chercheurs est assurée par des éditeurs dans le cadre d'un contrat d'édition. Par ailleurs, la publication est particulièrement importante pour le chercheur, son évaluation étant réalisée notamment à partir de cet indicateur.

94. Les éditeurs exercent plusieurs missions importantes :

- la datation de l'article ;
- l'assurance d'une relecture et d'une validation de l'article par un comité éditorial constitué de chercheurs spécialistes (peer-review) le plus souvent à titre gracieux ;
- la mise en page éventuelle de l'article ;
- la correction éventuelle de la langue
- la diffusion de l'article par les canaux qui lui sont propres ;
- l'inscription dans une base de données ou attribution d'un DOI<sup>26</sup>
- l'archivage de l'article
- la gestion des droits d'auteur
- la publicité du journal et de l'article

95. Le contrat d'édition entre un chercheur et un éditeur prend le plus souvent la forme d'un contrat d'adhésion. Il prévoit une cession des droits d'auteur du chercheur au profit de l'éditeur, le plus généralement **à titre exclusif et gracieux, pour une exploitation dans le monde entier et pour toute la durée légale des droits d'auteur**. De nombreux témoignages ont permis de mettre en évidence la pratique des éditeurs de faire signer un « Copyright Transfer Form ». Ce contrat est « rédigé d'une façon que seul un juriste spécialiste du droit du « copyright » puisse comprendre ».

96. Le chercheur le signe souvent sans même le lire car il n'a souvent pas l'appui de son établissement pour lui donner un avis motivé et un moyen de défendre ses intérêts de créateur. Enfin la rapidité de publication est souvent un point important dans le cadre de

---

<sup>26</sup> Digital Object Identifier - <https://www.doi.org/>

la compétition internationale et le chercheur n'a pas souvent le temps de mettre en œuvre une procédure de validation du contrat approprié<sup>27</sup>.

**97. Modèle hybride.** L'auteur publiant en accès gratuit dans des journaux électroniques scientifiques hybrides (accès gratuit et accès abonnement) doit généralement payer un droit (Article Processing Charges) pour couvrir les frais de la revue.

98. Le Comité d'éthique du CNRS, dans un avis « sur les relations entre chercheurs et maisons d'édition scientifique » du 31 janvier 2011<sup>28</sup>, décrit cette situation de la manière suivante :

- « La cession du droit d'auteur pour un article accepté par le comité éditorial d'une revue, qui peut être basée dans un pays ou un autre, sur recommandation d'un ou plusieurs relecteurs est le plus souvent demandée par l'éditeur à titre gracieux. Si un auteur refuse de signer le formulaire de cession de son droit d'auteur, son article, bien qu'accepté par le comité éditorial, ne sera en général pas publié. Si en revanche il signe ce formulaire, il renonce a priori à diffuser lui-même son article, et à utiliser les figures et tableaux de données qu'il contient, puisque la maison d'édition en est devenue propriétaire, ceci presque toujours sans avoir dédommagé l'auteur pour la perte de ces droits. »

99. En effet, la conclusion d'un contrat d'édition avec cession exclusive empêche notamment le chercheur :

- de mettre en ligne son article sur la plateforme institutionnelle de son organisme employeur qui a financé les recherches ayant abouti à l'article ;
- de partager l'article avec d'autres chercheurs intéressés par ses travaux ;
- de diffuser l'article sur son site web ;
- de réutiliser les graphiques, les supports, dans d'autres publications ou conférences orales ;
- de déposer en archive ouverte l'article.

100. Certains éditeurs, conscients de l'importance pour la recherche de la mise à disposition des articles et de la tendance à l'Open Science, autorisent après le respect d'une période d'embargo la mise en ligne sur une archive ouverte de l'article (postprint). Le site HELOISE <https://heloise.ccsd.cnrs.fr/> est un service d'information sur les politiques des éditeurs en matière de dépôt des articles. Ce service concerne

---

<sup>27</sup> Contribution de Marie FARGE, Directrice de Recherche CNRS

<sup>28</sup> [http://www.cnrs.fr/comets/IMG/pdf/03-avis\\_relations-chercheurs-maisons-edition-2.pdf](http://www.cnrs.fr/comets/IMG/pdf/03-avis_relations-chercheurs-maisons-edition-2.pdf)

uniquement les dépôts sur les sites des scientifiques eux-mêmes et des institutions scientifiques.

#### **1.2.4.2 Contrat éditeur / institut : contrat de licence**

101. Afin de disposer de l'accès aux revues et ouvrages scientifiques, les instituts concluent avec les éditeurs des contrats d'abonnement ouvrant l'accès aux revues et aux services en ligne de l'éditeur.

102. Cet abonnement relatif à l'accès aux collections courantes et/ou aux archives de l'éditeur, est :

- soit conclu directement entre l'institut et l'éditeur;
- soit conclu par l'intermédiaire d'un groupement de commande (dans le cadre de « licence nationale »).

103. Les contrats de licence contiennent des dispositions différentes selon les éditeurs mais prévoient de manière générale un socle commun de droits au profit des instituts <sup>29</sup> :

- un droit d'accès, de consultation, de visualisation des collections ;
- un droit d'impression ou de copie électronique pour l'usage propre de l'abonné ;
- parfois un droit de pratiquer du TDM via l'API (Application Programming Interface) de l'éditeur qui récupère alors les données d'usage; la diffusion de la production issue du TDM peut inclure des extraits du texte intégral dans la limite d'un certain nombre de mots ou d'un pourcentage du texte, sous licence Creative Commons CC BY-NC et avec un lien vers le texte complet de l'article sur le site de l'éditeur.

104. Les utilisateurs n'ont pas l'autorisation d'extraire modifier, traduire ou créer une quelconque œuvre dérivée à partir des données mises à disposition par l'éditeur dans le cadre de l'abonnement.

#### **1.2.4.3 La publication dans les contrats de valorisation industrielle**

105. La mise à disposition des données de recherche doit être organisée et doit prendre en compte la nature des données. Les données issues d'une Zone à Régime Restrictif

---

<sup>29</sup> Analyse des marchés :

- Elsevier - Abes du 31/01/2014 abonnement à la Freedom Collection
- Elsevier - Abes pour ISTEX marché 2013-20
- Elsevier - CNRS marché 2010-09

(ZRR)<sup>30</sup>, de recherches en cours, les données liées à des savoir-faire ou à des secrets, à des titres de propriété industrielle ne doivent pas être mises à disposition de manière systématique.

106. Par ailleurs, dans le cadre des contrats de collaboration de recherche et les contrats de partenariat public / privé, les modalités de publication des résultats de recherche font l'objet de dispositions spécifiques notamment lorsque le contrat aboutit à un dépôt de brevet ou est soumis à une obligation de confidentialité.

107. **Clause Brevet.** Le contrat peut prévoir que dans l'hypothèse où les résultats issus de la recherche sont brevetables, les partenaires procéderont au dépôt d'un brevet.

108. Le contrat doit alors prévoir les modalités de dépôt du brevet (propriétaire unique ou copropriété) et les modalités d'exploitation du brevet (licence exclusive d'exploitation, licence d'exploitation par secteur d'activité ou par zone géographique, conditions du partage des redevances, cession).

109. Une invention est brevetable si elle constitue une nouveauté au regard de l'état de la technique. Celle-ci est entendue comme tout ce qui a été rendu accessible au public avant la date de dépôt de la demande de brevet, par une description écrite ou orale, un usage ou tout autre moyen, y compris le contenu des demandes de brevet français, européen ou international, à condition qu'elles désignent la France, déposées antérieurement, et qui n'ont pas encore été publiées à la date du dépôt de la demande de brevet considérée<sup>31</sup>.

110. Ainsi, si l'invention a été rendue publique en quelque endroit du monde que ce soit, non seulement par un brevet antérieur mais aussi par une publication, une exposition publique (à un salon par exemple) ou une même une simple divulgation orale, elle n'est plus nouvelle.

111. Toutefois, lorsque la divulgation de l'invention résulte d'un abus<sup>32</sup> elle n'affecte pas la nouveauté de l'invention si elle a lieu dans les six mois précédant le dépôt de la demande de brevet.

---

<sup>30</sup> Régime issu de l'Arrêté du 3 juillet 2012 relatif à la protection du potentiel scientifique et technique de la nation et de la Circulaire interministérielle de mise en œuvre du dispositif de protection du potentiel scientifique et technique de la nation du 7 novembre 2012

<sup>31</sup> <http://www.entreprises.gouv.fr/propriete-intellectuelle/inventions>

<sup>32</sup> [Article L. 611-13 du CPI.](#)

112. La divulgation pourra être considérée comme abusive lorsqu'elle intervient:

- sans l'accord de l'inventeur (vol de l'invention, espionnage industriel) ;
- en violation d'un secret (professionnel par exemple) ;
- en violation d'une obligation contractuelle de confidentialité.

113. Lorsque le contrat prévoit le dépôt d'un brevet à l'issue de la coopération, aucune publication ne peut être réalisée par le ou les chercheurs en charge de la recherche et développement au risque d'anéantir le caractère de nouveauté de l'innovation et empêcher tout dépôt de brevet.

114. **Clause Confidentialité.** Un nombre croissant de contrats de recherche prévoit des clauses encadrant la confidentialité des connaissances propres de chacune des parties et les conditions de publication des résultats communs issus de la recherche.

115. Les clauses partielles suivantes peuvent par exemple figurer dans les contrats de recherche passés par le CNRS avec des industriels :

- l'exemple 1 est tiré d'un contrat type de collaboration de recherche entre le CNRS et un partenaire industriel ;
- les exemples 2 et 3 sont issus de contrats cadres entre le CNRS et des partenaires industriels.

	Exemple 1	Exemple 2	Exemple 3
<b>Définitions structurantes</b>	Informations Confidentielles Connaissances propres Connaissances nouvelles	Informations Confidentielles Résultats Résultats communs	Informations Confidentielles (liste) Données Métiers de X Résultats Résultats Valorisables
<b>Confidentialité</b>	Obligation de confidentialité sur les Informations Confidentielles pendant la durée du contrat et pendant les 5 ans qui suivent la résiliation ou le terme du contrat	Obligation de confidentialité sur les Informations Confidentielles pendant la durée du contrat et pendant les 5 ans qui suivent la résiliation ou le terme du contrat	Sont considérées comme des Informations Confidentielles : <ul style="list-style-type: none"> <li>- Les Connaissances Propres d'une autre Partie et les Données métiers de X pour une durée de confidentialité de 10 ans à compter de la fin du Contrat quelle qu'en soit la cause</li> <li>- Les Résultats Valorisables</li> </ul>
<b>Publication</b>	<p>Accord écrit de l'autre partie pour toute publication sur les Connaissances nouvelles pendant la durée du présent contrat et les 6 mois qui suivent son expiration.</p> <p>Tout projet de publication ou communication sera soumis à l'avis de l'autre partie qui pourra supprimer ou modifier certaines précisions dont la divulgation serait de nature à porter préjudice à l'exploitation industrielle et commerciale, dans de bonnes conditions, des connaissances nouvelles. De telles suppressions ou modifications ne porteront pas atteinte à la valeur scientifique de la publication.</p> <p>Si des informations contenues dans la</p>	<p>Les publications ou communications relatives aux Résultats communs qui en sont issus font référence à la collaboration des Parties.</p> <p>Tout projet de publication ou communication, relatif aux Résultats communs, devra recevoir, pendant la durée de l'accord spécifique et les vingt-quatre (24) mois qui suivent son expiration ou sa résiliation, l'accord préalable et écrit de l'autre Partie.</p> <p>Cette décision pouvant consister :</p> <ul style="list-style-type: none"> <li>- à accepter sans réserve le projet de publication ou communication ; ou</li> <li>- à demander que les Informations confidentielles lui appartenant soient retirées du projet ; ou</li> <li>-</li> <li>- à demander que des Résultats</li> </ul>	<p>Tout projet de publication ou communication par une Partie, quels qu'en soient la forme ou le support, relatif au programme de recherche et aux Résultats, doit faire l'objet d'une demande écrite préalable par lettre recommandée avec accusé de réception, auprès de l'autre Partie, pendant la durée de confidentialité des Informations Confidentielles.</p> <p>L'autre Partie peut demander la suppression ou la modification de certains éléments de la publication dont elle estime que la divulgation serait de nature à lui porter préjudice, ou à porter préjudice à l'utilisation industrielle ou commerciale des Résultats du Programme de Recherche ou à la protection d'un titre de propriété intellectuelle.</p>

	Exemple 1	Exemple 2	Exemple 3
	<p>publication ou communication doivent faire l'objet d'une protection au titre de la propriété industrielle, une des parties pourra retarder la publication ou la communication d'une période maximale de 18 mois à compter de la demande.</p>	<p>communs qui ont fait l'objet d'un dossier technique secret soit retirés du projet ; ou</p> <ul style="list-style-type: none"> <li>- à demander des modifications, en particulier si certaines informations contenues dans le projet de communication sont de nature à porter préjudice à l'exploitation industrielle et commerciale des Résultats communs ; ou</li> <li>- à demander que la communication soit différée si des causes réelles et sérieuses lui paraissent l'exiger, en particulier si des informations contenues dans le projet ou de communication doivent faire l'objet d'une demande de protection au titre de la Propriété Industrielle.</li> </ul>	<p>En particulier, une Partie peut demander à retarder la publication ou la communication, pour une période maximale de dix-huit (18) mois, si certaines informations doivent faire l'objet d'une protection au titre d'un droit de propriété intellectuelle.</p>
<b>Exception</b>	Rapports d'activité du chercheur et soutenance de thèse libres	Rapports d'activité du chercheur et soutenance de thèse libres	Rapports d'activité du chercheur et soutenance de thèse libres. Communication et dépôt de demande de brevet sur les résultats propres libres.

116. La pratique contractuelle grandissante, et de plus en plus négociée, visant à encadrer la publication par des clauses de brevet et / ou des clauses de confidentialité doit être prise en compte pour la mise à disposition des données de la recherche, afin de préserver les équilibres nécessaires à la valorisation des innovations.

### 1.3 Les risques de captation abusive

117. La captation des données et résultats scientifiques peut être légitime si elle répond aux intérêts légitimes de valorisation, de préservation des secrets ou de respect de la vie privée. En revanche, lorsqu'elle répond à des intérêts privés au détriment de la Science, a fortiori s'agissant de résultats issus de la recherche publique, la captation s'avère abusive.

118. Cet abus peut se manifester dans :

- le choix du modèle économique de l'édition scientifique ;
- le détournement du droit de la propriété intellectuelle ;
- les pratiques contractuelles.

#### 1.3.1 Captation par les usages économiques

119. Le modèle Gold hybride, que certains considèrent comme transitoire, permet un libre accès aux publications scientifiques sous réserve du paiement d'« article processing charges » et représente ainsi une première source de captation.

120. Le marché de l'édition scientifique est divisé en deux groupes entre lesquels un écart se creuse de manière irréversible : les « majors » (Elsevier, Springer, Nature), qui se développent plus vite que le marché en captant les activités d'édition des sociétés savantes, ont permis de faire croître leur catalogue beaucoup plus rapidement que celui des éditeurs académiques et commerciaux de plus faible taille.

121. Par ailleurs, le secteur de l'édition scientifique est un secteur de concurrence imparfaite où la concurrence par les prix n'existe pas puisque les revues ne sont pas substituables les unes aux autres. Cela explique les taux de marge très élevés des grands éditeurs et la fixation de manière unilatérale et non concurrentielle des montants des abonnements et des « article processing charges »<sup>33</sup>.

#### 1.3.2 Captation par les usages de l'édition scientifique

122. Le droit de la propriété intellectuelle appliqué aux données de la science aboutit de fait à une privatisation de la connaissance au profit des éditeurs, ce qui a pour conséquence :

---

<sup>33</sup> Informations extraites de l'étude de la DIST-CNRS « L'Édition de sciences à l'heure numérique : dynamiques en cours (2015) »

- de poser des barrières juridiques et financières à l'accès aux informations scientifiques et techniques ;
- de ralentir la recherche et les avancées scientifiques ;
- de favoriser la concentration des informations scientifiques et des thématiques de recherche financées au moins pour moitié par des fonds publics entre les mains d'éditeurs privés.

123. Les éditeurs exploitent et tirent des bénéfices importants de l'édition des articles scientifiques alors :

- que les coûts de l'édition numérique sont plus faibles que les coûts de l'édition papier ;
- que ni les auteurs, ni le comité éditorial, ni les experts ne sont rémunérés par les éditeurs ; souvent même les auteurs doivent payer des « article processing charges » et les organismes doivent payer leurs abonnements.

124. Les bénéfices des éditeurs n'ont d'ailleurs jamais été aussi importants qu'aujourd'hui. Le marché mondial de l'édition scientifique de recherche est estimé à 12,8 milliards d'euros. Les services numériques représentent en moyenne 60% des chiffres d'affaires. Chez les grands éditeurs qui ont investi de façon précoce dans le numérique et dans une logique de plateformes ce ratio est de 75%.

125. Les conditions de réutilisation des leurs propres articles par les chercheurs sont très restrictives voire inexistantes, les auteurs cédant dans la grande majorité des cas leurs droits à titre exclusif par de véritables contrats d'adhésion.

### 1.3.3 Captation par les contrats

#### 1.3.3.1 Le contrat d'édition scientifique

126. **Contrat nommé.** L'auteur d'un écrit scientifique peut par un contrat d'édition céder « à des conditions déterminées à une personne appelée éditeur le droit de fabriquer ou de faire fabriquer en nombre des exemplaires de l'œuvre, à charge pour elle d'en assurer la publication et la diffusion »<sup>34</sup>.

127. Il s'agit d'un contrat nommé régi par les articles L.132-1 à L.132-17 du Code de la propriété intellectuelle, qui pose des obligations importantes à la charge de l'éditeur et notamment :

---

<sup>34</sup> [CPI art L132-1](#)

- une obligation de publication : à défaut de publication, le contrat sera résilié (article L. 132-17 CPI) ;
- une exploitation permanente et suivie : l'article L.132-12 du CPI met à la charge de l'éditeur l'obligation d'assurer à l'ouvrage une disponibilité permanente, et donc de procéder à des réimpressions d'office, de procéder ou faire procéder à des éditions populaires ;
- une reddition des comptes : l'article L 132-13 précise que « l'éditeur est tenu de rendre compte ».

128. L'édition implique une cession des droits de l'auteur à son éditeur, et cette cession de droits, avec l'objectif qui lui est assigné, est l'élément qui caractérise les contrats d'édition par rapport aux autres contrats de cession de droits de propriété intellectuelle.

129. L'article L.132-8 du Code de la propriété intellectuelle dispose que « l'auteur doit garantir à l'éditeur l'exercice paisible et, sauf convention contraire, exclusif du droit cédé. »

130. **Rémunération.** La rémunération proportionnelle est la contrepartie de la cession des droits au profit de l'éditeur. La loi prévoit donc, par principe, la rémunération de l'auteur proportionnellement « aux recettes » ou « aux produits d'exploitation ». En d'autres termes, afin de protéger l'auteur et de lui permettre de participer au succès de son œuvre, l'article L. 132-5 du code prévoit une rémunération de l'auteur proportionnelle à l'exploitation de son œuvre, ce qui inclut non seulement l'édition, mais aussi le droit de représentation, la traduction ou l'adaptation de son œuvre.

131. Par la règle de la participation proportionnelle, le législateur a voulu protéger l'auteur contre toute cession des droits d'exploitation qu'il pourrait consentir pour une somme dérisoire, au regard des profits dégagés par l'éditeur. Par dérogation, une rémunération forfaitaire peut être prévue dans les cas limitativement énumérés aux articles L. 131-4 et L. 132-6 du CPI.

132. L'auteur peut toutefois renoncer contractuellement à une rémunération proportionnelle à l'exploitation de son œuvre. En effet, l'article L. 122-7 du CPI relatif à la cession du droit de reproduction d'une œuvre prévoit qu'une telle cession peut être gratuite. Pour cela, une clause de cession du droit de reproduction à titre gratuit doit être prévue formellement dans le contrat d'édition.

133. La violation des dispositions relatives à la rémunération est sanctionnée par une nullité relative dont l'action se prescrit par cinq ans, délai qui court à compter de la signature du contrat<sup>35</sup>.

134. **Contrat d'adhésion.** En cédant à titre exclusif à l'éditeur ses droits patrimoniaux sur son article, l'auteur-chercheur ne peut plus exploiter ses articles, les partager ou les auto-archiver même à titre gratuit et au bénéfice de la communauté scientifique. Dans la majorité des cas, le contrat signé est un véritable contrat d'adhésion prévoyant une cession exclusive des droits et n'est pas assorti d'une rémunération. Ces contrats pourraient être révisés au regard des dispositions du Code de la propriété intellectuelle.

135. Il serait intéressant de revenir à un régime plus protecteur des auteurs, autorisant l'auteur-chercheur à partager ses articles sans limite, y compris contractuelle, et ce au nom :

- de l'intérêt supérieur de la Science ;
- du financement par des fonds publics des recherches ayant abouti à la rédaction de l'article.

136. Ce système autorisant l'auteur-chercheur à partager librement ses publications ne doit pas pour autant pénaliser le monde de l'édition scientifique ; des aménagements peuvent être trouvés notamment par la définition de période(s) d'embargo.

 Les contrats d'édition signés par les chercheurs sur leurs articles rédigés dans le cadre de la recherche financée sur fonds public prévoient majoritairement des cessions exclusives de droits au profit de l'éditeur. Ces contrats constituent de véritables contrats d'adhésion. Les contrats d'édition qui n'indiquent pas formellement que les droits sont cédés à titre gratuit encourent la nullité.

### 1.3.3.2 Le contrat d'abonnement à la plateforme de l'éditeur

137. **CGU ou abonnement.** Les conditions générales d'utilisation ainsi que les contrats d'abonnement auprès des éditeurs organisent les conditions d'utilisation des articles et des services accessibles à partir des plateformes. Ces contrats prévoient généralement des conditions limitées d'utilisation des articles :

---

<sup>35</sup> Dalloz Action Droit d'auteur Chapitre 112 – Transmission, cession et contrats relatifs aux droits d'auteur – André R. Bertrand – 2010

- l'accès au texte intégral, l'impression, le téléchargement pour le seul usage de l'abonné ;
- l'impossibilité de procéder à l'exploration des articles à l'extérieur de la plateforme.

138. A titre d'exemple les conditions générales d'utilisation d'Elsevier Masson disponibles sur son site web stipulent à l'article Propriété intellectuelle :

- « 6.1 Sauf disposition écrite contraire, les droits d'auteur et autres droits de propriété intellectuelle sur l'ensemble des offres, publications et autres Produits et/ou Services d'Elsevier demeureront sa propriété. Les droits concédés par Elsevier sont destinés au seul usage par le Client et ne peuvent être cédés, transférés ou concédés par voie de sous licence sans l'accord préalable écrit d'Elsevier. Les droits concédés par Elsevier le sont à titre non exclusif et aux fins dont les parties sont expressément convenues. Toute autre utilisation requerra l'accord préalable écrit d'Elsevier. Le Client n'acquerra aucun droit de propriété intellectuelle sur les Produits.
- 6.2 Aucune partie des offres, publications ou Produits d'Elsevier ne peut être stockée dans un fichier de données automatisé et/ou reproduite, par un procédé mécanique, électronique, par télécopie, photocopie, enregistrement ou par tout autre moyen ou sous toute autre forme sans l'accord préalable écrit et exprès d'Elsevier. »

139. **Contenus générés par les utilisateurs.** Conscient de la nécessité de proposer à leurs abonnés des services de traitement de l'information (croisement, analyse sémantique et lexicographique, synthèse automatique, traduction...), les éditeurs ont développé des services à valeur ajoutée disponibles depuis leur plateforme.

140. A titre d'exemple, Elsevier propose via son API un contrat de licence TDM :

- limitant l'utilisation de l'API à des fins non commerciales ;
- limitant l'exploration des résultats de l'utilisation de l'API sur des jeux de données, les « user generated content », à :
  - o la mise en ligne des données bibliographiques accompagnée d'un lien DOI (Digital Object Identifier) qui pointe vers l'article en texte plein ;
  - o l'ajout d'une mention de propriété ;
- interdisant :
  - o l'utilisation d'un extrait du texte plein de plus de 200 caractères ;
  - o de modifier, traduire ou créer des travaux dérivés sur la base des jeux de données ;

- de reproduire, conserver ou redistribuer les jeux de données ;
- d'extraire ou d'utiliser les jeux de données à titre commercial ;
- d'utiliser des robots ou d'autres programmes automatisés, des algorithmes de recherche ;
- d'utiliser les résultats de l'utilisation de l'API pour améliorer des dépôts institutionnels ou d'une manière qui concurrencerait l'article tel que validé par les pairs.

141. La propriété des contenus générés par les utilisateurs n'est pas clairement affirmée par l'éditeur mais en concédant des droits sur les « TDM Output », l'éditeur présume en être titulaire.

142. Cette captation de la Science financée sur fonds publics par les éditeurs privés appellent à légiférer en faveur du droit des chercheurs :

- sur le libre accès aux données et aux résultats de la recherche ;
- sur la libre exploration des données et des résultats de la recherche.

143. Afin de répondre aux besoins des chercheurs au regard de l'environnement macro de la Science, il est proposé en partie 2 une démarche et des solutions juridiques en faveur de l'Open Science.

## 1.4 Une validation massive de ces constats

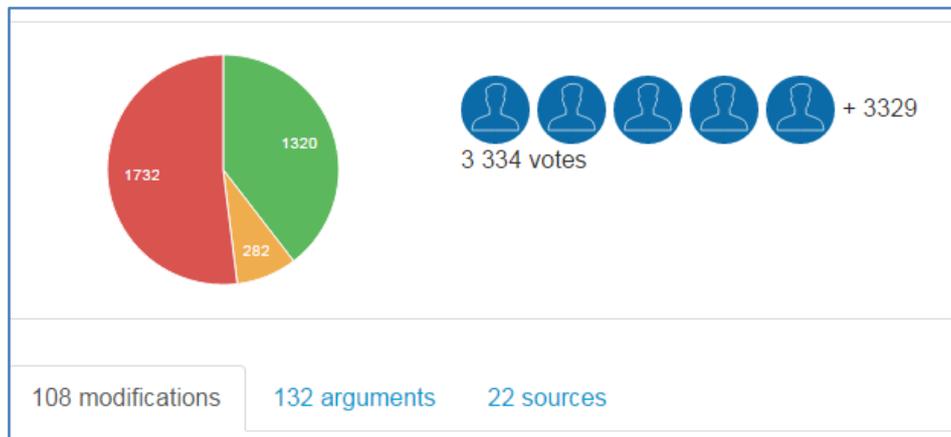
### 1.4.1 La synthèse de la consultation nationale

144. Pour la première fois de l'histoire législative, le projet de loi pour une République numérique a été mis en ligne pour faire l'objet d'une discussion publique et recevoir des contributions des citoyens entre le samedi 26 septembre et le dimanche 18 octobre 2015.

145. L'initiative a mobilisé les internautes : les compteurs affichant à la clôture de la consultation :

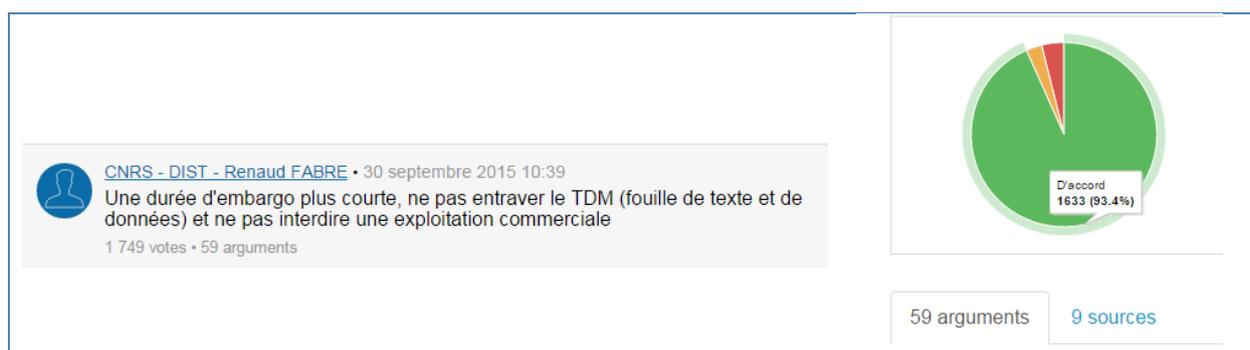


146. Le bilan de la consultation sur l'article « Libre accès aux publications scientifiques de la recherche publique » est le suivant :



147. Avec 3 334 votes et 108 propositions de modifications, l'ex article 9 de l'avant-projet de loi consacré au « libre accès aux publications scientifiques de la recherche publique » aura suscité le plus de réactions chez les internautes, au terme de la consultation publique sur le projet de loi numérique le 18 octobre, devant des articles dédiés à l'Open data ou encore à la libre réutilisation des données des services publics industriels et commerciaux.

148. La proposition qui a reçu le plus de vote est celle de la Direction de l'Information scientifique et technique du CNRS intitulée « Une durée d'embargo plus courte, ne pas entraver le TDM (fouille de texte et de données) et ne pas interdire une exploitation commerciale »<sup>36</sup> avec 1633 votes « pour » sur 1749.



149. Roberto DI COSMO, professeur en informatique à l'Université Paris-Diderot, a posté le deuxième amendement ayant reçu recueilli le plus de voix avec 1 511 votes pour sa proposition visant à « protéger le droit des auteurs d'articles scientifiques, pour

<sup>36</sup> <https://www.republique-numerique.fr/consultations/projet-de-loi-numerique/consultation/consultation/opinions/section-2-travaux-de-recherche-et-de-statistique/article-9-acces-aux-travaux-de-la-recherche-finanee-par-des-fonds-publics>

permettre le libre accès à la recherche scientifique », déposée 10 jours avant la fin de la consultation.

150. Des institutions et organismes de recherche ont aussi pris officiellement position, comme l'Inria, l'INRA, Couperin avec l'ADBU, le SNESUP, l'UPMC, l'INP Toulouse, l'association des archivistes français, Cairn.info, etc. Leurs propositions étaient majoritairement les suivantes :

- permettre le libre accès aux résultats de la recherche scientifique ;
- réduire les périodes d'embargo à 6 et 12 mois, voire pour certains à néant, et pour d'autres sans distinction entre les domaines scientifiques ;
- permettre le dépôt en archive ouverte ;
- autoriser les opérations de *text and data mining* ou de fouilles de données ;
- garantir la possibilité de valoriser les résultats issus du traitement des connaissances.

151. Le Syndicat National de l'Édition (SNE) et la Fédération Nationale de la Presse d'Information Spécialisée (FNPS) se montrent favorables aux périodes d'embargo de 12 et 24 mois prévues par le texte ; le Groupement Français de l'Industrie de l'Information (GFII) propose qu'elles soient décidées en fonction d'études d'impact à mettre en place.

152. Le 18 octobre 2015 un Gouv'camp a été organisé pour la clôture de la plateforme de contribution nationale et a réuni dans des groupes de travail les différents contributeurs au projet de loi. Un compte rendu présentant une position consensuelle a été mise en ligne par Alain Bensoussan, co-rapporteur avec Grégory COLCANAP du groupe « article 9 » (devenu article 17 dans la dernière version du projet de loi). Ce compte-rendu insiste sur le fait que :

- « les écrits scientifiques doivent devenir des biens communs » ;
- la question du TDM doit être abordée ;
- les périodes d'embargo pour la publication en libre accès doivent être ramenées à six et douze mois, au maximum.
- le dépôt « en archive ouverte publique et pérenne » doit être encouragé ;
- la valorisation commerciale des contenus d'un article scientifique doit rester ouverte.

153. Les besoins tels qu'exprimés par les chercheurs et les grands témoins auditionnés dans le cadre de ce Livre blanc recoupent les points cruciaux mentionnés ci-dessus par les contributeurs à la plateforme.

**154. Synthèse officielle.** Une synthèse officielle de la consultation publique préfacée par Madame Axelle LEMAIRE énumère les contributions qui ont conduit à la modification de l'avant-projet de loi, au rang desquelles la contribution de la DIST. Le commentaire du Gouvernement explique le périmètre des modifications apportées au regard de la proposition formulée :

De manière quasiment consensuelle, il émerge de la consultation une demande claire pour renforcer les droits des chercheurs à diffuser librement leurs travaux, lorsque ces travaux ont été financés par des fonds publics. A la recherche d'un nouvel équilibre entre les positions des différents acteurs en présence à l'heure du numérique et de la société de la connaissance, le Gouvernement a fait évoluer la mesure de la manière suivante :

- Les délais « d'embargo », au terme desquels l'auteur d'une publication financée sur fonds publics peut, au plus tard, mettre librement à disposition son écrit, ont été réduits de moitié.
- Si l'article est mis à disposition gratuitement par l'éditeur en ligne, l'auteur pourra immédiatement faire usage de son droit.
- Le texte précise désormais également que les données de la recherche liées à cet écrit peuvent être immédiatement réutilisées, et que leur circulation ne peut être freinée à l'occasion de l'édition de l'écrit.

Parmi les demandes exprimées auxquelles il n'a pas été donné suite à ce stade, il faut citer la création d'une exception au droit d'auteur pour l'analyse de textes et de données à des fins de recherche (« *text and data mining* », TDM), qui est nettement soutenue par la communauté scientifique. Le droit européen ne permet pas actuellement de créer de nouvelles exceptions, et le Gouvernement souhaite que cette question soit abordée dans le cadre des travaux européens en cours.

155. Des modifications importantes ont pu émerger de cette contribution nationale ; il reste toutefois des points importants non pris en compte dans cet article 17 (notion de publication dans sa version auteur / dans sa version éditeur) ou renvoyés à un prochain texte. Des propositions d'amendements du texte du projet de loi figurent en partie 2 du présent Livre (2.5 « La modification de l'article 17 du projet de loi pour une République numérique »).

## 1.4.2 L'avis du Conseil National du Numérique

156. Le 30 novembre 2015, le Conseil National du Numérique a émis un avis sur le projet de loi pour une République numérique tel qu'issu de la consultation publique.

157. Le Livre blanc reprend ici l'avis du Conseil National du Numérique sur les dispositions relatives à l'article 17 du projet de loi et le partage dans son intégralité.

### **Sur le libre accès aux publications scientifiques et aux données de la recherche publique (Open Access) : compléter les dispositions prévues par le projet de loi**

Le CNNum salue la limitation des périodes de cession exclusive des publications scientifiques de la recherche publique à une période de 6 mois pour les sciences, la technique et la médecine et de 12 mois pour les sciences humaines et sociales, par la reconnaissance d'un droit d'exploitation secondaire pour les chercheurs.

Le Conseil recommandait dans son rapport Ambition numérique de compléter cette disposition par une obligation de rendre accessible gratuitement ces publications sur un site institutionnel, dans une revue ouverte ou sur un site d'archive ouverte. Elle pourrait s'appliquer aux organismes de recherche. Les Etats-Unis ont emprunté cette voie en prévoyant une durée de l'embargo égale à 1 an. Le Royaume-Uni a quant à lui décidé de créer des incitations au libre accès par la prise en compte des publications scientifiques ouvertes dans l'évaluation – et le financement – de la recherche.

Enfin, le CNNum se félicite de l'inscription des données issues d'activités de recherche publique et rendues publiques légalement dans un régime de chose commune, au sens de l'article 714 du code civil.

158. Le Conseil National du Numérique relève également que le projet de loi accuse certains manques notamment en matière de circulation des données et du savoir et plus particulièrement sur la fouille de textes et de données pour la recherche (*text and data mining*).

159. Le Livre blanc est également en adéquation avec l'avis du Conseil National du Numérique sur ce point et préconise la voie de l'exception au droit d'auteur. .

### **Autoriser la fouille de textes et de données pour la recherche (*text and data mining*)**

La fouille de textes et de données désigne un ensemble de traitements informatiques consistant à extraire de connaissances selon un critère de nouveauté ou de similarité dans des textes ou des bases de données. Elle permet par exemple la recherche de « signaux faibles » difficiles à appréhender par la lecture cursive, le repérage ou l'analyse de comptes rendus d'expérimentations ratées.

Elle a été considérée comme porteuse de nombreux potentiels pour la découverte scientifique et le développement de nouvelles connaissances. Elle doit permettre à la recherche de profiter des avancées en matière de l'analyse des méga-données ("big data"), qui a vocation à devenir un enjeu majeur de la compétitivité internationale. L'Irlande, le Royaume Uni mais aussi les Etats-Unis ou encore le Japon le permettent aujourd'hui.

#### **Considérant que :**

- La fouille automatisée de textes et de données, en tant qu'activité de lecture et d'extraction d'informations, est une pratique qui ne se distingue pas fondamentalement du relevé manuel des informations qui a toujours été effectué par la recherche ;
- Le droit d'auteur, qui protège la forme d'expression et non les idées, permet aujourd'hui de lire et de réutiliser des informations ou données incluses dans un texte sur lequel on a obtenu un droit d'accès.

#### **Le CNNum estime qu'il n'y a pas de raison légitime à restreindre ce droit dans le cadre d'un traitement automatisé.**

Les grands éditeurs qui détiennent la majeure partie des publications scientifiques, peuvent aujourd'hui proscrire, par des solutions contractuelles, la fouille de textes et de données – notamment la copie provisoire, techniquement nécessaire afin de la réaliser – aux chercheurs, même lorsque ces derniers disposent d'un accès légal à l'ensemble des publications scientifiques comprises dans les bases de données fouillées. Cette interdiction s'appuie notamment sur le droit sui generis des bases de

données. Cette pratique nécessite donc la création d'une exception au droit d'auteur, sur la base d'une réinterprétation de l'exception pour la recherche, à l'image de l'interprétation du Royaume-Uni.

Compte tenu des limites et contraintes imposées par les solutions contractuelles, le CNNum recommande donc **d'instaurer une véritable exception au droit d'auteur autorisant la fouille de textes et de données.**

### 1.4.3 L'étude d'impact du projet de loi

160. Alors que le Conseil d'Etat dans son avis du 9 décembre 2015 déplorait le manque d'étude d'impact notamment sur l'article 17 du projet de loi<sup>37</sup>, le Gouvernement a proposé le même jour cette étude d'impact dont les termes confortent la présente analyse.

161. L'étude insiste en effet que les objectifs d'une telle nouvelle législation :

- « Il s'agit d'ouvrir la possibilité d'une diffusion en accès libre des travaux scientifiques financés sur fonds publics, au terme d'une durée dite « d'embargo » préservant les droits exclusifs des éditeurs. Il s'agit en même temps de sécuriser juridiquement des pratiques existantes dans la communauté scientifique et bien tolérées par les éditeurs. La création de ce nouveau droit pour les auteurs des travaux nécessite l'intervention du législateur, afin qu'il s'impose dans l'ensemble des contrats d'édition à venir. Cette disposition crée un nouveau droit pour l'auteur de la publication et promeut un nouvel équilibre dans la relation entre un chercheur et son éditeur. »

---

<sup>37</sup> Conseil d'Etat 3-12-2015 Avis sur un projet de loi pour une République numérique N° 390741 page 5  
« 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'Etat 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 faire obstacle à l'adoption de cette mesure. En outre, le Conseil d'Etat n'a pu retenir cette disposition ainsi que celle qui qualifie de « choses communes » au sens de l'article 714 du code civil les données issues de recherches financées sur fonds publics et qui permet la libre réutilisation de ces données une fois publiées, au motif qu'il n'a pas disposé d'une véritable étude des impacts positifs ou négatifs, juridiques aussi bien qu'économiques qu'on peut en attendre. »

- « La mesure vise également à favoriser et à protéger la libre réutilisation des données de la recherche, à partir du moment où elles sont rendues publiques. »

## 162. L'impact de la loi a été analysé aux niveaux :

- de la puissance publique : elle « favorise une meilleure régulation des coûts de l'information scientifique et technique, aujourd'hui largement supportés par la puissance publique » ;
- économique et social : « Le partage des données de la recherche concourt au développement économique et social »
- sur la recherche : « La mise en accès ouvert de publications et la libre réutilisation des données de la recherche favorise le partage des connaissances et des découvertes, anciennes et récentes, au sein de la communauté scientifique. Elle encourage les collaborations et l'interdisciplinarité, limite la duplication des efforts de recherche, contribue à l'amélioration générale de la qualité des travaux. Elle ouvre également la voie à une meilleure prise en compte des attentes de la société civile, favorisant une recherche et une innovation responsables. Elle profite enfin aux entreprises qui cherchent à innover, en particulier aux petites et moyennes entreprises qui n'ont pas les capacités d'investir dans la recherche et développement. »
- sur l'économie de l'édition scientifique : l'impact est très faible :
  - des expériences étrangères ont montré « un recul limité des accès via les sites des éditeurs » ;
  - le risque de désabonnement est très faible ;
  - au niveau de l'édition scientifique française l'impact est à relativiser dans la mesure où la majorité de leur chiffre d'affaires est aujourd'hui constituée de subventions apportées par des établissements ou des laboratoires. Toutefois et compte tenu du tissu éditorial français, le Gouvernement a décidé de prendre en compte la crainte exprimée par de nombreux éditeurs et directeurs de revues de sciences humaines et sociales en définissant un plan de transition au libre accès des revues de SHS ;
  - au niveau de l'édition scientifique mondiale, elle « se caractérise aujourd'hui par une forte concentration, de nature oligopolistique, autour

de quelques groupes internationaux » et « l'information scientifique et technique constitue une activité exceptionnellement rentable ».

#### 1.4.4 Le texte adopté par l'Assemblée nationale

163. Les députés ont pris la mesure des enjeux pour la recherche publique non seulement en validant le principe de libre accès aux données scientifiques mais en plus en introduisant un cadre juridique aux pratiques de text and data mining.

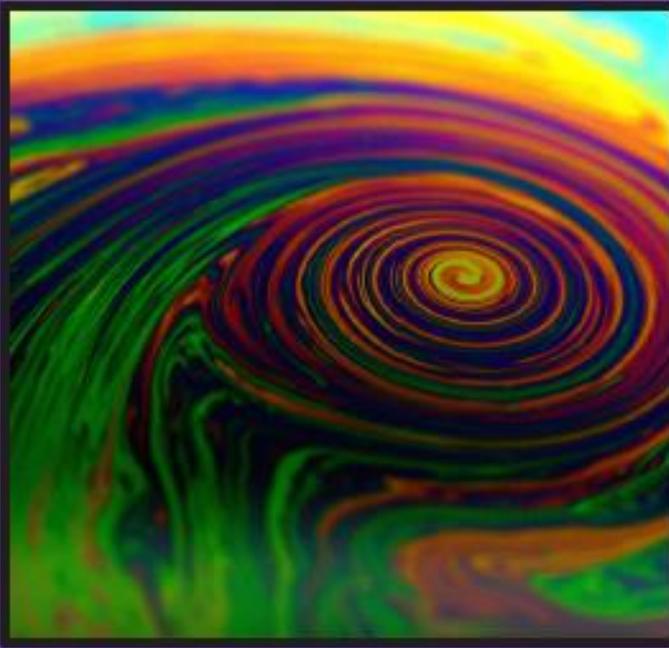
164. Un rapport sur l'impact du principe de libre accès aux données scientifiques sur le marché de l'édition scientifique et sur la circulation des idées et des données scientifiques devra être remis au Parlement par le Gouvernement « au plus tard deux ans après la promulgation » de la loi pour une République numérique.

165. Un nouvel article 18 bis a été introduit et crée une exception au droit d'auteur et une exception au droit du producteur de base de données en faveur de l'exploration de textes et de données<sup>38</sup>.

166. Un tableau en annexe propose de comparer les différentes versions du texte du projet de loi (annexe 1) de la consultation publique à l'adoption du projet de loi par l'Assemblée nationale.

---

<sup>38</sup> Projet de loi pour une République numérique texte adopté par l'Assemblée nationale n°663 le 26-1-2016  
<http://www.assemblee-nationale.fr/14/ta/ta0663.asp>



## 2. L'Avenir : « La science numérique ouverte »

## 2. L'AVENIR : LA SCIENCE NUMERIQUE OUVERTE

167. Le présent chapitre présente un ensemble de propositions (2.5 et 2.6) visant à faire évoluer le droit de l'information scientifique et technique pour prendre en compte les besoins et pratiques de la communauté scientifique, sécuriser les usages de la recherche publique, rééquilibrer la protection des intérêts en présence, et ce dans l'intérêt supérieur de la Science.

168. Ces propositions sont formulées à partir :

- des axes de réflexions engagés par les grands témoins de ce Livre blanc et par l'expression des valeurs communes et universelles de la Science telles qu'elles ont découlé de l'audition de ces témoins (2.1) ;
- de l'analyse des règles émergentes au niveau international (2.2) ;
- de l'étude de l'écart entre le droit constant et les pratiques et besoins (2.4).

169. La France doit se conformer aux valeurs et règles émergentes dans les instances et législations supranationales et étrangères, faute de quoi la Science française risque d'être marginalisée.

170. L'émergence nécessaire de ce droit nouveau de la Science ouverte repose sur des concepts clés existants et des intérêts légitimes qu'il convient de préserver (2.3).

171. L'adoption de l'article 17 du projet de loi pour une République numérique issue d'une position consensuelle qui a émané au Gouv Camp et la confirmation de l'article 18 bis (nouveau) en faveur du *text and data mining* permettraient au monde de la recherche publique de sécuriser ses pratiques. Il est également formulé une proposition de création d'un véritable droit positif et global de la Science ouverte.

172. Afin de garantir une certaine souplesse d'application des dispositions légales au regard notamment des pratiques différenciées des communautés scientifiques, il est proposé la création d'un référentiel des usages, et la création d'une norme.

173. Par ailleurs les contrats de cession des droits entre les auteurs et les éditeurs étant de véritables contrats d'adhésion, un contrat type de cession de droits d'auteur plus protecteur des intérêts légitimes des chercheurs et de la communauté scientifique pourrait être rendu obligatoire.

174. La Science ouverte nécessite également la définition de règles éthiques communes dont les valeurs pourront s'étendre à l'international et dont l'application pourra être

garantie par une Agence de la Science ouverte. Enfin, le rayonnement de la Science française à travers le monde légitime la France à prendre l'initiative d'une convention internationale de la Science ouverte universelle.

175. Des opérations de formation devront accompagner les évolutions législatives souhaitables.

## **2.1 Les témoignages du Livre blanc : des principes convergents vers une Science ouverte**

### **2.1.1 Une valeur commune : La Science, un « bien commun » de l'humanité**

176. Historiquement, la science a toujours été appréhendée comme un bien commun ; la méthode scientifique elle-même implique une construction collective de la connaissance (travaux en partenariat, échange d'informations, peer review...). La place grandissante des technologies de l'information dans la recherche scientifique en tant qu'outil de recherche, de partage et de collaboration a réactivé la notion de « bien commun » associée à la Science.

177. **Position internationale partagée.** « Bien universel, par essence, l'article scientifique, n'est pas un bien comme les autres ». Cette position a été affirmée dans un communiqué de presse signée par la Conférence des présidents d'universités, la Conférence des grandes écoles, la Conférence des directeurs des écoles françaises d'ingénieurs et Couperin.

178. Cette dimension universelle de la connaissance scientifique avait déjà été soutenue dans la « Déclaration de Berlin sur le libre accès à la connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales » de 2003. Le libre accès avait pour caractéristique d'être « une source universelle de la connaissance humaine et du patrimoine culturel ayant recueilli l'approbation de la communauté scientifique »<sup>39</sup>.

179. Dans la tribune « Favorisons la libre diffusion de la culture et des savoirs » parue dans Le Monde le 10 septembre 2015<sup>40</sup>, tribune signée par près de 1820 personnes, il est réaffirmé que « les biens communs - ou communs - nourrissent depuis toujours les

---

<sup>39</sup> Déclaration de Berlin sur le libre accès à la connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales, 22 octobre 2003, <http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre>

<sup>40</sup> Tribune « Favorisons la libre diffusion de la culture et des savoirs » [http://www.lemonde.fr/idees/article/2015/09/10/favorisons-la-libre-diffusion-de-la-culture-et-des-savoirs\\_4751847\\_3232.html](http://www.lemonde.fr/idees/article/2015/09/10/favorisons-la-libre-diffusion-de-la-culture-et-des-savoirs_4751847_3232.html)

pratiques d'échange et de partage qui structurent la production scientifique et la création culturelle. »

180. La Conférence des présidents d'universités dans une motion d'octobre 2015 mentionne « que la connaissance est un bien commun de l'humanité et que les données de la science doivent être érigées en données d'intérêt général »<sup>41</sup>.

181. L'Open Science doit être érigée comme un « principe universel » permettant un accès aux communs que sont les données scientifiques, favorisant ainsi le bien de l'humanité et les avancées scientifiques.

### **Recommandation du Conseil Scientifique du CNRS**

« La science est un bien commun de l'humanité qui ne saurait souffrir de captation abusive par des intérêts privés. »

182. En effet, l'Open access a un impact bien réel sur l'avancée de la recherche, voire dans certains cas sur la préservation de la santé publique :

- l'équipe en charge au Liberia du virus Ebola n'a pas pu accéder à certains articles du fait de leurs coûts importants, alors qu'ils auraient été nécessaires afin d'identifier le virus plus tôt et ainsi adapter plus rapidement les mesures de prévention et de soin. La privatisation de la connaissance a ici entraîné la mort d'un certain nombre de personnes ; un accès libre et immédiat à la connaissance est ici une nécessité vitale ;
- le projet de séquençage complet de l'ADN du génome humain a pu être réalisé grâce à une collaboration à grande échelle des chercheurs, et ce partout dans le monde, et par une diffusion publique et immédiate des résultats de recherche. L'Internet en tant que catalyseur permet également de réduire davantage le temps imparti à une étude de cette envergure. L'accès libre, gratuit et immédiat aux résultats scientifiques a permis ici une avancée scientifique majeure.

183. Par ailleurs, cette évolution du système économique et technologique existant est nécessaire afin d'éviter :

- que ce soit le seul droit du commerce qui régisse l'usage des plateformes numériques centralisant les résultats et les données de la recherche ;

<sup>41</sup> <http://www.cpu.fr/actualite/les-donnees-de-la-science-un-bien-commun/>

- un gaspillage de l'argent public en réalisant des recherches qui ont déjà été réalisées par d'autres instituts ;
- un surcoût pour les instituts de recherche, les laboratoires, les universités.

184. Partout dans le monde, des prises de position en faveur de la Science comme bien commun sont affirmées :

- au Québec, l'association « Science et bien commun » fondée en juillet 2011 a pour objectif de défendre et promouvoir une vision des sciences au service du bien commun ;
- l'Open Science Federation qui milite en faveur de l'Open Science ;
- sur le plan international, l'organisation Creative Commons a lancé en 2005 un projet nommé Science Commons qui proposait d'introduire dans le milieu scientifique les principes d'ouverture et de partage en érigeant la science en bien commun et en élargissant l'utilisation des licences Creative Commons à la recherche scientifique et technique<sup>42</sup>.

### 2.1.2 La Science, moteur de l'économie

185. L'Open Science favorise la recherche et l'innovation dans la mesure où elle permet de partager des connaissances, de dégager de nouveaux sujets de recherche, de nouvelles connaissances et de répondre à des problématiques économiques, sociales et sociétales. En découlent également des opportunités en termes de valorisation des nouvelles connaissances avec des enjeux d'innovation, de croissance et d'emploi.

186. Ce moteur de l'innovation est reconnu a de nombreuses reprises :

- par l'Unesco : le libre accès « permet de favoriser la circulation du savoir à l'échelle planétaire pour contribuer aux découvertes scientifiques, à l'innovation et au développement socio-économique »<sup>43</sup> ;
- pour l'OCDE, l'Open Science permet d'améliorer l'efficacité et la qualité de la recherche en réduisant les coûts de la collecte de données, en facilitant l'exploitation des données « dormantes » ou inaccessibles à faible coût et en augmentant les possibilités de collaboration dans la recherche ainsi que l'innovation. L'Open Science permet également de réduire la fracture de l'accès à la Science et de renforcer les capacités dans les pays en développement ;

---

<sup>42</sup> <http://creativecommons.org/science>

<sup>43</sup> <http://www.unesco.org/new/fr/communication-and-information/access-to-knowledge/open-access-to-scientific-information/>

- dans la stratégie numérique du Gouvernement du 18 juin 2015 : « La libre circulation de la science et sa libre exploitation contribue en outre à l'innovation, encourage la collaboration, améliore la qualité des publications, évite la duplication des efforts, permet l'exploitation des résultats de recherche antérieurs et favorise la participation des citoyens et de la société civile ».

187. Dans sa contribution, le Conseil scientifique souligne l'enjeu pour les chercheurs de disposer des données et publications scientifiques sous format numérique afin de les partager et de les explorer :

### **Contribution du Conseil Scientifique du CNRS**

« La numérisation des données utilisées par les scientifiques et de leurs publications permet un traitement automatisé, un transfert rapide, une harmonisation des méthodes d'accès et des descriptions ; tout cela est susceptible de mettre à la portée du chercheur une immense matière, riche et diversifiée, en des temps singulièrement raccourcis »

188. Le Comité d'éthique<sup>44</sup> du CNRS a également souligné que :

- « Faciliter l'accès et la réutilisation de ces données est ainsi devenu un enjeu crucial pour partager et faire circuler plus rapidement les résultats de la recherche. »

189. Au cours des auditions réalisées dans le cadre de la rédaction de ce Livre blanc, les besoins suivants ont systématiquement été soulevés par les témoins du Livre blanc :

- le libre accès aux données de la Science
- le nécessaire peer-review et la nécessité de disposer de nouveaux indicateurs d'évaluation ;
- la publication et le temps de réservation de l'éditeur ;
- l'exploration de données ;
- la reconnaissance de paternité et la visibilité ;
- le respect d'intérêt légitime (brevet, secret) ;
- une éthique de l'IST.

190. Des propositions de changement total des paradigmes ont également émergé.

---

<sup>44</sup> Auto saisine du COMETS « Les enjeux éthiques du partage des données scientifiques » Groupe « Data sharing » 12-12-2014

### 2.1.3 Priorité au libre accès et au partage des données de la Science

191. Tous s'accordent à affirmer la nécessité d'avoir un accès libre et massif aux données de la Science, au nom de l'intérêt supérieur de la recherche dont les enjeux sont humains, sociaux, économiques.

192. Cette notion de données de la Science inclut non seulement les données de la recherche mais également les résultats de la recherche, qu'ils soient publiés par un éditeur scientifique ou non publiés.

#### **Audition Inria, Claude Kirchner 15 octobre 2015**

« L'ensemble des données scientifiques doit rester sous la maîtrise des scientifiques. »

#### **Recommandation du Conseil Scientifique**

« Le libre accès aux résultats de l'activité scientifique (publications, données de la recherche, métadonnées, services à valeur ajoutée) ne saurait être entravé sans remettre en cause le développement même de la science. »

193. Ce besoin d'Open data, d'ouverture des données de la recherche, et d'Open Access, ouverture de l'accès aux publications scientifiques protégées par le droit d'auteur, forme un consensus au sein des communautés scientifiques.

194. Par ailleurs, les chercheurs accèdent majoritairement aux données par des plateformes dont le modèle technique et juridique n'est pas sécurisé.

195. **Périmètre des données.** Afin de déterminer leurs besoins en matière de libre accès, il convient de définir le périmètre des données nécessaires aux chercheurs dans le cadre de la démarche scientifique et dans l'exercice de leurs travaux de recherche.

196. Le projet de loi pour une République numérique apparaît insuffisant à cet égard.

197. La pré-version du projet de loi pour le numérique de juillet 2015 employait la notion de « contributions scientifiques ». La notion de « contribution » ne faisant l'objet d'aucune définition légale, elle était imprécise et sujette à interprétation.

198. Le projet de loi pour une République numérique lui a préféré la notion d'« écrits scientifiques », terminologie reprise de l'article L.112-2 du Code de la propriété intellectuelle qualifiant d'œuvres de l'esprit protégeables par le droit d'auteur « les livres, brochures et autres écrits littéraires, artistiques et scientifiques ».

199. Le texte ne s'intéresse qu'aux « écrits scientifiques », appréhendés en tant qu'œuvres protégeables par le droit d'auteur. Cette interprétation est confirmée :

- par le rappel du principe de protection par le droit d'auteur des écrits scientifiques ;
- par la disposition limitant le droit du chercheur de mettre à disposition son écrit dans sa « dernière version acceptée de son manuscrit par son éditeur et à l'exclusion du travail de mise en forme qui incombe à ce dernier ».

200. Le projet de loi limite la possibilité de libre accès aux travaux scientifiques en « postprints ».

#### **Audition Inria, Claude Kirchner le 15 octobre 2015**

« La ``version auteur acceptée''<sup>45</sup> d'un article scientifique, intégralement réalisée par cet auteur jusqu'à sa transmission à l'éditeur pour publication, doit rester libre de toute contrainte et pouvoir être diffusée en ligne sous la forme souhaitée par l'auteur (ou son institution), en particulier dans une archive ouverte.

Les éventuelles contraintes d'embargo ne peuvent porter que sur la ``version éditeur'' dans sa mise en forme finale et ce afin de respecter l'exploitation commerciale éventuelle. Elles ne sont acceptables que si la "version auteur" est effectivement libre de diffusion, et la durée de l'embargo devrait alors être fixée en cohérence avec les pratiques internationales. »

201. **Les besoins.** Or, les chercheurs ont exprimé le besoin de disposer d'un accès à l'ensemble des données et résultats scientifiques issus d'une activité de recherche financée au moins pour moitié par des fonds publics.

202. Ce besoin a été réaffirmé par le Conseil Scientifique dans sa recommandation jointe au Livre blanc.

---

<sup>45</sup> Le rappel des définitions des versions ``version auteur initiale'' ou preprint, ``version auteur acceptée'' ou postprint et ``version éditeur'' est disponible dans le glossaire et de modifier dans le glossaire "Postprint / Post publication / Version auteur acceptée " et "Preprint / Prépublication / version auteur initiale".

203. La communication de la Commission européenne du 17 juillet 2012<sup>46</sup> affirme également que « nombreux sont les résultats de travaux de recherche financés par des fonds publics existant sous forme de données qui ne sont pas rendus largement disponibles à des fins de vérification ou d'exploitation, ce qui compromet sérieusement l'efficacité des investissements dans la recherche. »

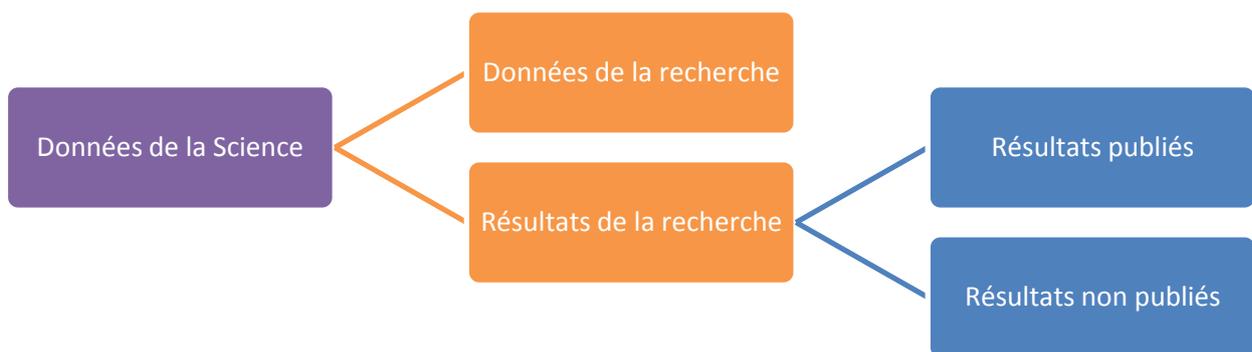
204. Cette « lost Science », valeur économique indéniable, financée par des fonds publics, doit pouvoir être valorisée et exploitée par la recherche publique.

205. Par ailleurs, le Code de la recherche définit parmi les missions de la recherche publique (article L.112-1 du Code de la recherche) :

- « le partage et la diffusion des connaissances scientifiques » ;
- « l'accès libre aux données scientifiques ».

**+** Les chercheurs ont besoin pour réaliser leurs travaux de recherche d'un accès libre et gratuit sous forme numérique à l'ensemble des données de la Science composées :

- des résultats scientifiques en ce compris les résultats publiés par un éditeur scientifique ;
- des données de la recherche entendues comme les données ayant servi à établir ces résultats.

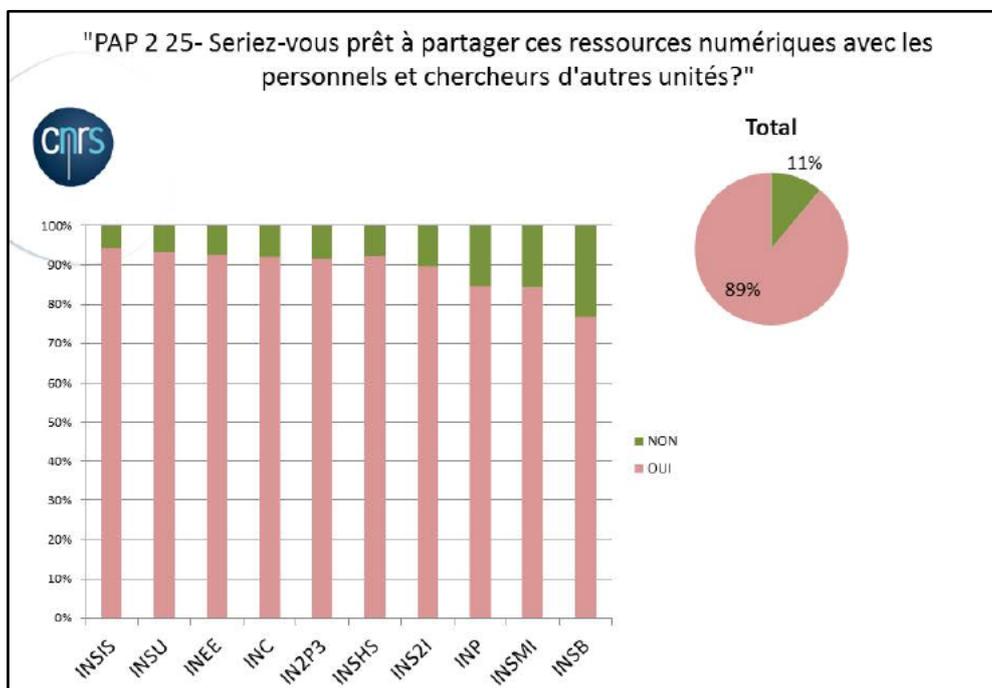


<sup>46</sup> Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions « Pour un meilleur accès aux informations scientifiques : dynamiser les avantages des investissements publics dans le domaine de la recherche » C(2012) 401 du 17-7-2012

206. **Partage des données.** Valeur importante et fondement historique de la démarche scientifique, le partage des connaissances est le moteur même de la recherche. La transition numérique a bouleversé la pratique par un accès à une masse de données grandissante et globale, de manière instantanée et ce partout dans le monde.

207. Les chercheurs sont prêts à plus de 89% à partager des ressources numériques avec les personnels et chercheurs d'autres unités.

Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015



208. Le partage des données de Science permet en effet d'enrichir la connaissance.



Les chercheurs ont exprimé le besoin de partager les données de Science.

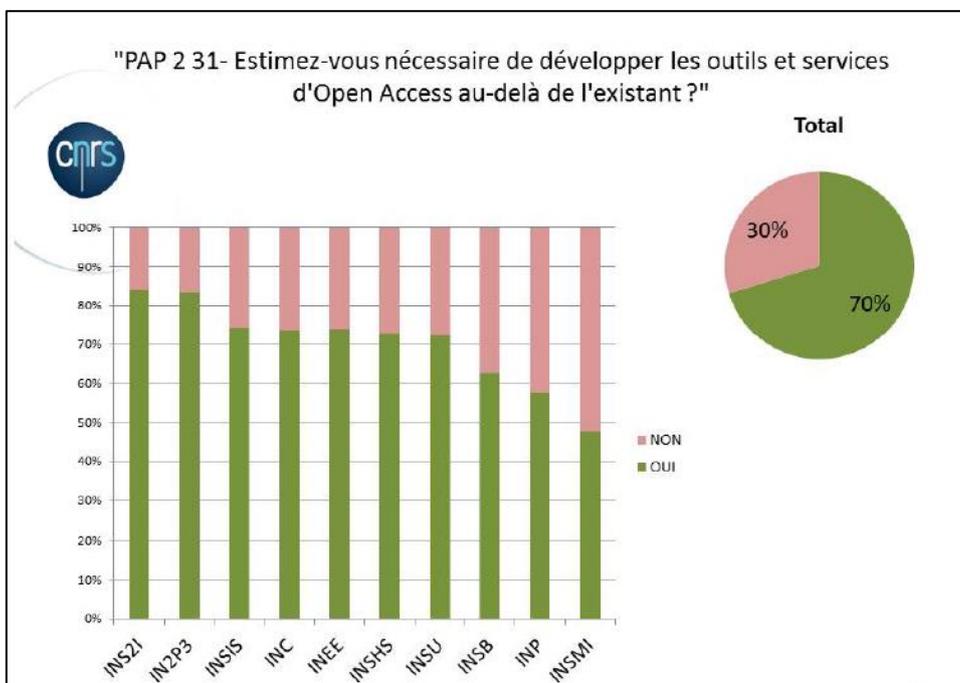
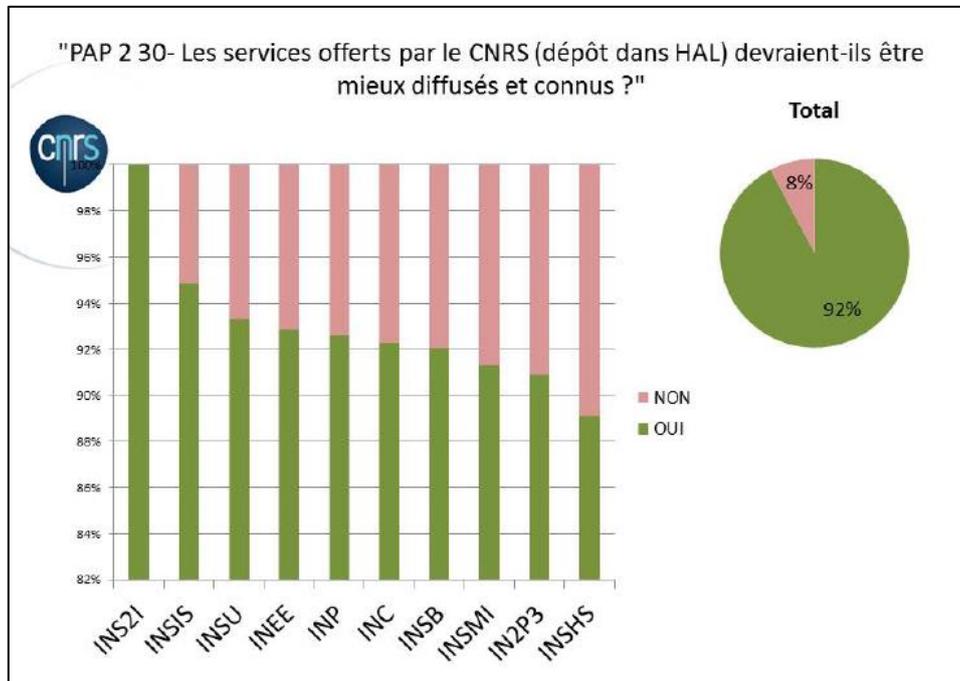
209. **Les plateformes multiusages.** L'accès et le partage des données de Science peuvent s'opérer à partir « d'outils innovants et conviviaux »<sup>47</sup>, simples d'utilisation, et permettant un accès large à la connaissance.

210. Les plateformes en ligne de type HAL ou ArXiv se sont développées. Les communautés scientifiques souhaiteraient disposer d'une meilleure information sur les

<sup>47</sup> Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015 page 5

services de type plateforme de dépôt comme HAL et estiment à 70% que d'autres outils et services d'Open access doivent être développés.

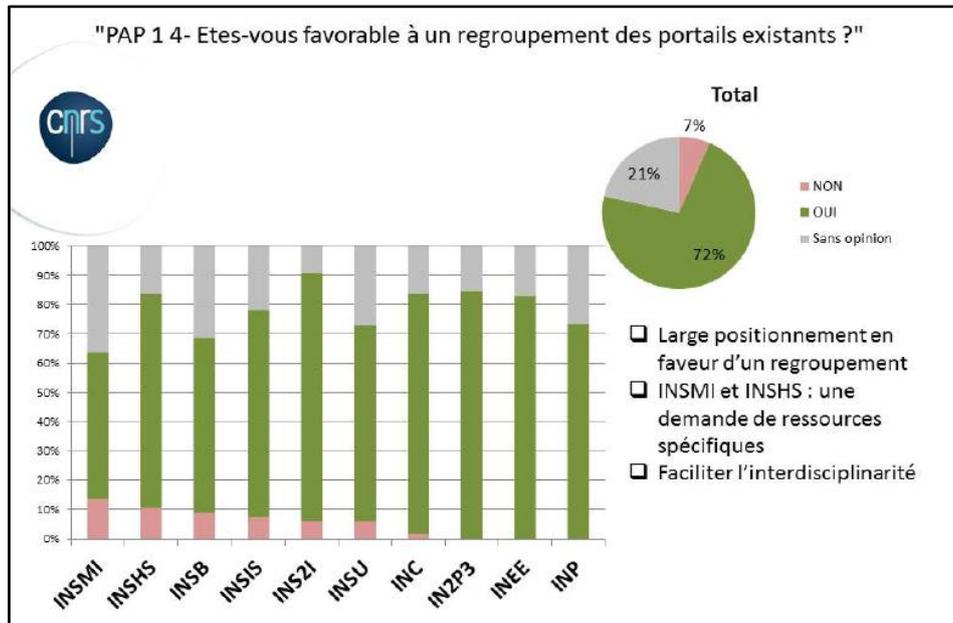
Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015



211. Les chercheurs se positionnent largement en faveur d'un regroupement des portails existants (72%). Le regroupement de la connaissance sur un portail unique permet de limiter la déperdition des connaissances et de travailler à partir de documents provenant

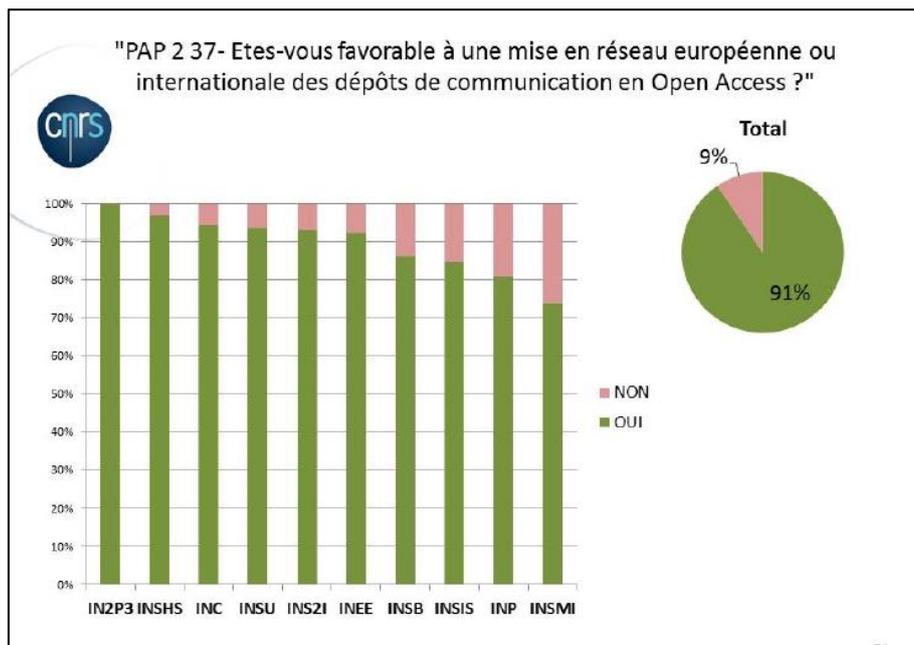
de différentes disciplines, facilitant ainsi les recherches transdisciplinaires et à partir de corpus provenant d'éditeurs différents.

Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015



212. Les chercheurs sont de même favorables à 91 % à une mise en réseau à l'échelle européenne ou internationale des dépôts de communication en Open Access.

Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015



213. Les chercheurs expriment largement un besoin technique d'outil numérique de type plateforme(s) en ligne, permettant l'accès et le partage des données et des résultats à minima au niveau national.

+ La pratique d'un dépôt en archive ou en plateforme en fonction du domaine des articles, doit être généralisée.

214. Des inquiétudes quant aux règles du jeu régissant la plateforme se sont également exprimées :

- besoin d'une définition de la notion de plateforme ;
- définition de l'étendue des droits des chercheurs à partir de ces plateformes ;
- format des données et interopérabilité des données ;
- précision des métadonnées.

+ Les chercheurs ont exprimé le besoin :

- d'une plateforme guichet unique de la connaissance scientifique ;
- de régulation des plateformes par le droit.

#### 2.1.4 L'évaluation numérique des résultats revus par les pairs et métrique des publications

215. La notion d'évaluation désigne deux procédures distinctes :

- le peer review : la relecture et l'évaluation par un comité de lecture des travaux du chercheur avant publication dans une revue scientifique ;
- la procédure d'évaluation du chercheur par une unité de recherche dans le cadre de la promotion interne ou lors d'une procédure de recrutement.

216. **Comité de lecture.** Les chercheurs sont attachés à ce système d'évaluation des travaux scientifiques avant publication dans des revues scientifiques. Le Comité de lecture est généralement composé de chercheurs travaillant dans le même domaine que celui de l'article proposé. Les pairs ont pour mission de juger la qualité scientifique de l'article, et la validité méthodologique de la démonstration exposée. En fonction de leur avis, l'article sera alors accepté ou rejeté, le comité de rédaction conservant la décision définitive.

217. L'organisation de cette évaluation fait toutefois l'objet de critiques dont l'article « Peer review »: déontologie et fraudes chez les chercheurs scientifiques »<sup>48</sup>, paru le 2 février 2014, fait état :

- un premier tri est généralement réalisé par le comité de rédaction avant de soumettre un article au comité de lecture ;
- le comité de lecture est bien souvent incarné par une personne seule qui réalise l'évaluation bénévolement pour le compte de l'éditeur ;
- ce pair est bien souvent submergé de nombreuses demandes d'évaluation et doit également procéder à un tri des articles à évaluer ;
- en résulte une déperdition importante d'articles et de connaissances scientifiques ;
- les risques liés au développement des « article processing charges » pour la publication de son article notamment en termes de qualité.

218. **Evaluation.** Les chercheurs du CNRS doivent présenter un rapport d'activité et une liste exhaustive de leurs productions scientifiques en vue de leur évaluation par leur(s) section(s) et/ou une commission interdisciplinaire. C'est une obligation statutaire posée par les articles 10, 29 et 49 modifiés du décret n°83-1260 du 30 décembre 1983 fixant les dispositions statutaires communes aux corps de fonctionnaires des établissements publics scientifiques et technologiques.

219. La publication est un critère de recrutement des chercheurs, d'avancement et de financement des projets de recherche.

220. Toutefois ce critère quantitatif du nombre de publication a été critiqué et le Comité d'éthique du CNRS a formulé la recommandation suivante : « L'évaluation qualitative de la recherche par les pairs doit rester la règle. »<sup>49</sup> Impliquant une lecture des publications qui restent centrales dans l'évaluation.

221. La mise en ligne des données et leur partage devront être pris en compte dans le cadre de l'évaluation des chercheurs. Le Comité d'éthique préconise sur ce point que :

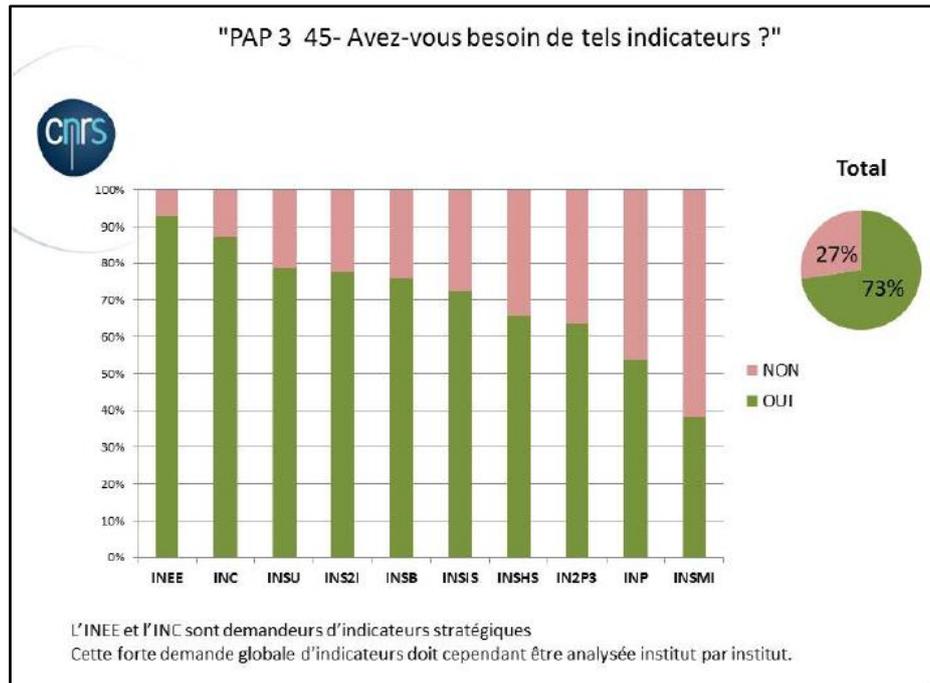
- « La contribution aux travaux de partage des données doit être reconnue dans l'évaluation et les décisions de promotion des chercheurs. Pour faciliter cette reconnaissance, le COMETS préconise que des indicateurs pertinents soient

---

<sup>48</sup> <http://www.contrepoints.org/2014/02/02/155325-peer-review-deontologie-et-fraude-chez-les-chercheurs-scientifiques>

<sup>49</sup> Avis du COMETS « Problèmes éthiques pour les métiers de la recherche publique en mutation » CNRS, 12 février 2014





223. Il est par conséquent nécessaire de « construire des outils de métrique des publications, des données, des analyses sur les Plateformes et développer des démarches de mise en réseau des plateformes et de e-infrastructures »<sup>52</sup>.

### 2.1.5 Publication et embargo

224. L'Open Science n'est pas incompatible avec la publication par un éditeur des articles scientifiques, bien au contraire, ces deux modes de diffusion sont complémentaires, l'éditeur proposant couramment des services sur les données.

225. **Publication.** L'éditeur, en plus du peer review, fait un travail de mise en page, d'insertion dans une revue, de mise en ligne et de diffusion de l'article. Le travail de l'éditeur est donc complémentaire à celui de la mise en ligne de l'article et des traitements de données transdisciplinaires et multi-corpus.

226. Par ailleurs, dans les pays et disciplines où l'Open Access est entré dans les pratiques des éditeurs et des chercheurs, on ne constate pas de baisse de chiffre d'affaires des éditeurs scientifiques, qui bénéficient au contraire d'une certaine publicité par la mise à disposition des articles aux communautés scientifiques.

<sup>52</sup> Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015 page 39

227. **Délais d'embargo.** Dans le cadre de la démarche scientifique, le chercheur lit des articles, recueille de la connaissance, croise des informations anciennes et récentes. Pour ce faire, il a besoin d'avoir un accès immédiat aux articles publiés par les éditeurs scientifiques.

228. Lors des auditions ainsi que dans les recommandations du Conseil Scientifique et du Comité d'éthique, les chercheurs ont unanimement affirmé que, la Science étant un bien commun de l'humanité, aucune période d'embargo ne devrait s'appliquer<sup>53</sup>.

229. Par ailleurs, la distinction opérée dans le projet de loi entre « sciences, technique et médecine » et « sciences humaines et sociales » pour la durée d'embargo est contestée dans son principe. Une durée d'embargo plus longue en SHS ne s'accorde pas avec les besoins des chercheurs tels qu'exprimés notamment par Maya BACACHE-BAUVALLET, Françoise BENHAMOU et Marc BOURREAU dans un rapport de l'Institut des études politiques n°11 de juillet 2015.

230. Ce rapport intitulé « Les revues de sciences humaines et sociales en France : libre accès et audience » conclut en page 5 que les résultats de l'étude menée « objectivent donc la mise en place d'une durée d'embargo relativement courte relativement aux durées évoquées dans le débat public pour les SHS ».

**Audition Université de Strasbourg Paul-Antoine HERVIEUX, 10 juillet 2015**

« La clé de la science est l'instantanéité de la recherche. Une donnée à un cycle de vie plus ou moins long, mais plus elle est fraîche plus elle est intéressante. Pour que la Science avance, l'embargo ne doit pas faire obstacle à la diffusion des résultats. Cependant, en fonction de la communauté scientifique concernée et pour des raisons diverses (e.g. compétition, processus de valorisation en cours...) une période d'embargo pourra être instaurée. »

231. Toutefois, afin d'assurer la transition vers l'Open Science et de préserver les intérêts économiques des éditeurs notamment français, les communautés scientifiques se sont accordées pour reconnaître un délai d'embargo aux éditeurs.

232. Cette durée d'exclusivité doit être « suffisamment longue pour maintenir l'équilibre économique de revues numériques, et suffisamment courte pour élargir de manière significative l'audience ayant accès à l'article dans sa version en accès libre. »<sup>54</sup>

---

<sup>53</sup> «Recommandation du CS du 25-9-2015 « Les scientifiques doivent pouvoir mettre à disposition gratuitement, sous une forme numérique, ces données et ces résultats, a priori sans période d'embargo imposée par les éditeurs »

**Audition ABES, Jérôme KALFON, 5 octobre 2015****« On quitte le principe du désir pour revenir au principe de réalité ».**

233. Le Conseil National du Numérique souligne également dans sa contribution la nécessité d'un « court délai d'embargo permettant l'activité commerciale de l'éditeur ».

234. Les délais d'embargo acceptés par les chercheurs sont les délais maximaux prévus par la recommandation de la Commission européenne (C(2012) 4890) :

- « Les publications issues de la recherche financée par des fonds publics soient librement accessibles dans les meilleurs délais, de préférence immédiatement et, dans tous les cas, au plus tard six mois après leur date de publication, et au plus tard douze mois pour les publications dans les domaines des sciences sociales et humaines ».

235. En tout état de cause, les délais ne peuvent pas être plus longs que ceux prévus par les législations nationales de nos voisins européens :

- en Allemagne : délai d'embargo de 12 mois sans distinction entre les disciplines ;
- en Espagne : dépôt dans une archive institutionnelle le plus rapidement possible, sans dépasser 12 mois, sans distinction entre les disciplines.

236. La « Commission de réflexion sur le droit et les libertés à l'âge du numérique » de l'Assemblée nationale a émis dans un rapport remis le 8 octobre 2015 à Claude Bartolone des recommandations en faveur de l'Open access parmi lesquelles :

- « Rendre librement accessibles les publications scientifiques financées sur fonds publics" après "une durée d'exclusivité de 6 à 12 mois ».

237. Le projet de loi pour une République numérique dans sa version issue de la contribution a pris en compte ces arguments et est revenu aux délais d'embargo raisonnables de 6 et 12 mois.

---

<sup>54</sup> [Rapport d'information](#) de la Commission de réflexion sur le droit et les libertés à l'âge du numérique de l'Assemblée nationale remis le 8-10-2015 page 241



Les chercheurs ont besoin d'accéder au dernier état des connaissances. Si dans le cadre d'un compromis avec les éditeurs une période d'embargo peut être définie, celle-ci ne doit pas dépasser les délais maximaux prévus dans la recommandation de la Commission européenne (C(2012) 4890) et les délais pratiqués à l'étranger au risque de marginaliser et discriminer la recherche française.

La distinction entre sciences exactes et sciences humaines et sociales est contestée dans son principe.

### 2.1.6 Analyse et exploration des corpus de données

238. **Droit historique.** L'analyse et l'exploration de données sont pratiquées depuis toujours par les chercheurs et sont au cœur même de la démarche scientifique. Seuls les outils d'observation ont changé, les outils numériques permettant de balayer un plus grand volume d'informations en moins de temps.

239. **Droit d'observation.** Les chercheurs s'accordent à soutenir que le droit d'exploration de données est un droit d'observation naturel qui ne doit souffrir d'aucune limitation. Le TDM permet au même titre qu'un microscope l'observation d'objets scientifiques.

#### **Audition de Jean-Marie PIERREL et Gregory COLCANAP, 24 septembre 2015**

« Le TDM est un droit d'observation des objets scientifiques, nécessaire à la Science ».

« L'informatique n'est rien de plus qu'un appareillage spécifique pour observer des données ».

#### **Contribution du Conseil National du Numérique**

« Le TDM n'est pas en soi une activité nouvelle.

Il revient à lire et dégager des informations, du sens, à partir de documents. Il ne se distingue pas fondamentalement du relevé manuel d'informations, pratiqué par la recherche depuis la nuit des temps. »

240. Dans ce cadre, les chercheurs ont besoin d'outils numériques de traitement des données, libre d'accès et gratuit :

- « Demande de consolidation et d'un partage de nouvelles pratiques de gestion des données, des publications, mais aussi des outils d'analyse des résultats, des outils innovants de métrique des publications »<sup>55</sup> ;

<sup>55</sup> Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015 page 5

- « Mettre à disposition des outils et services facilitant l'exploitation de données de la recherche »<sup>56</sup> ;

### **Contribution du Conseil Scientifique**

« Cette exigence de la mise à disposition des données s'étend aux services à valeur ajoutée (traitement massif de type "Big data", fouilles de données, lien aux métadonnées, interopérabilité) qui doivent également être publics et libres d'accès pour éviter toute captation abusive »

### **Audition Université Pierre et Marie Curie Jean CHAMBAZ et Paul INDELICATO, 9 juin 2015**

« La donnée scientifique évolue, et bien que la science des données ne se substituera pas à la Science en tant que démarche, la donnée et plus encore sa réutilisation se place au cœur de cette nouvelle démarche »

**241. Risques.** Si ce droit d'observation semble un acquis théorique, celui-ci est remis en cause :

- par les éditeurs scientifiques qui privatisent son accès ;
- par les nombreuses interrogations et divergences d'opinion émises quant à la compatibilité des pratiques de *text and data mining* avec le droit d'auteur et le droit du producteur de base de données.

**242. Risque de captation.** La production de valeur par l'utilisation de ces techniques de traitement de données, notamment aujourd'hui par les techniques de TDM, ne doit pas être préemptée par les éditeurs privés. Or cette privatisation est déjà en marche par :

- l'obligation des chercheurs de procéder au traitement des données d'un éditeur à partir de l'application programming interface (API) de l'éditeur en question ;
- la réservation par les conditions générales d'utilisation de l'API ou les contrats d'abonnement de tous les résultats issus de l'utilisation des techniques de TDM, les « TDM output » ou les « user generated contents » ; l'éditeur se réservant par conséquent la possibilité de les distribuer à des tiers.

---

<sup>56</sup> Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015 page 35

**Audition Université de Strasbourg Paul-Antoine HERVIEUX, 10 juillet 2015**

« Les éditeurs cherchent à se protéger en contrôlant la possibilité de TDM via les API (Application Programming Interface), des clauses contractuelles (licences Creative Commons, limitation du nombre de mots « explorables »). »

243. Cette captation empêche les chercheurs de procéder à des explorations transdisciplinaires, à partir de corpus de plusieurs éditeurs et prive la Science de connaissances fondamentales et de valeurs scientifiques et économiques issues notamment des user generated contents.

**Audition Université Pierre et Marie Curie Jean CHAMBAZ et Paul INDELICATO, 9 juin 2015**

« La possibilité de faire du TDM sur les données scientifiques est essentielle. Si cette possibilité dépend des éditeurs, les chercheurs seront soumis à un contrôle important qui ne pourra pas être maîtrisé ».

244. **Enjeux.** Le droit d'exploration et de traitement des données représente un enjeu fort pour la science, la recherche et l'innovation dans la mesure où il permet de dégager de nouveaux sujets de recherche, de nouvelles connaissances et de répondre à des problématiques économiques, sociales et sociétales. Le « *text and data mining* » (TDM) est l'une des applications techniques de ce droit d'exploration des données sur laquelle l'attention est aujourd'hui focalisée<sup>57</sup>.

**Audition Université de Strasbourg Paul-Antoine Hervieux, 10 juillet 2015**

« La relation avec les éditeurs est au centre des préoccupations des chercheurs, notamment au regard du TDM. Le TDM sera l'outil par excellence dans les années à venir. »

245. Le droit d'exploration ouvre ainsi des opportunités considérables en termes de valorisation des nouvelles connaissances avec les enjeux qui y sont attachés d'innovation, de croissance et d'emploi, sans perdre de vue pour autant l'intérêt des acteurs économiques du secteur.. « Les services de "fouille de données" et assimilés

---

<sup>57</sup> L'exploration de données a été présentée en 2001 par le Massachusetts Institute of Technology comme l'une des dix technologies émergentes qui « changeront le monde au 21ème siècle ». (Data mining et statistique décisionnelle – l'intelligence des données Stéphane Tuffery Editions Technip 2012)

jouent un rôle considérable dans la valorisation scientifique des données et écrits en libre accès »<sup>58</sup>.

246. Les enjeux scientifiques et économiques sont d'autant plus sensibles que la pratique du TDM est mondiale et fait l'objet d'une normalisation différente d'un pays à l'autre, y compris à l'intérieur de l'Europe (voir 2.2 « L'Open Science à l'international »). A défaut d'adoption de dispositions autorisant les pratiques de TDM, il existe un risque important d'assister à l'émergence d'une recherche à deux vitesses au sein de l'Union européenne et de mettre en péril certains partenariats de recherche entre la France et l'Angleterre par exemple.

+ L'affirmation dans la loi, comme principe universel, du droit d'observation des données de la Science par l'utilisation de techniques numériques d'exploration est nécessaire pour ne pas pénaliser la recherche française.

### 2.1.7 Propriété intellectuelle numérique et reconnaissance de paternité

247. Open Science ne signifie pas renonciation au droit moral de l'auteur. Les chercheurs souhaitent le respect de leur droit à la paternité, droit d'autant plus important que le nombre de citations fait partie des critères d'évaluation du chercheur.

248. La mention du nom de l'auteur et de l'article dans les citations ou dans les bibliographies des rapports d'étude doit faire partie des « bonnes pratiques » des chercheurs dans l'utilisation de l'IST. Des règles éthiques pourront rappeler ces bonnes pratiques.

+ Le droit à la paternité est une composante intangible du droit moral de l'auteur. Il doit être respecté strictement et renforcé par des règles éthiques.

### 2.1.8 Les frontières de la valorisation et de la science ouverte

249. L'Open Science ne doit pas faire obstacle aux enjeux économiques de la recherche.

250. La mise à disposition des données scientifiques sur des plateformes Open Science ne doit pas aller à l'encontre :

- de la valorisation des données notamment par brevet ;
- du respect des secrets et des dispositions spécifiques telles que les Zones à Régime Restrictif ;

---

<sup>58</sup> Recommandation du Conseil Scientifique du CNRS

- du respect des règles contractuelles de confidentialité.

251. La mise à disposition des données de la recherche doit être encadrée également au regard des pratiques non homogènes des communautés scientifiques.

**+** L'Open Science doit préserver les intérêts légitimes, notamment ceux liés à la valorisation des innovations et à la protection des secrets, et respecter les pratiques des communautés.

### 2.1.9 Pour une éthique de l'IST numérique

252. Les chercheurs ont exprimé un besoin de régulation à plusieurs niveaux :

- au niveau légal afin de sécuriser les pratiques ;
- au niveau éthique par la mise en place d'une « charte d'éthique de l'IST » qui pose des « Principes éthiques qui soient de nature à transcender les catégories instrumentales et à affirmer les buts de la recherche publique dans un contexte global de Science Ouverte »<sup>59</sup>.

253. Le Comité d'éthique a également fait valoir que « face à cette dynamique de circulation des données relayée par leurs autorités de tutelle et par leur communauté, les chercheurs doivent :

- prendre conscience de leur responsabilité individuelle, déontologique<sup>60</sup> et éthique, vis-à-vis de la communauté à laquelle ils appartiennent ;
- avoir connaissance des engagements internationaux des institutions dont ils dépendent ;
- participer à la définition de principes éthiques propres à leur discipline dans le domaine du *data sharing* et du *big data* en général. »<sup>61</sup>

---

<sup>59</sup> Résultats de l'enquête sur les usages et besoins d'IST - CNRS - mars 2015 page 59

<sup>60</sup> "Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences", *Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council. National Academy of Sciences.*

<sup>61</sup> Autosaisine du COMETS « Les enjeux éthiques du partage des données scientifiques » Groupe Data Sharing 12-12-2014

### 2.1.10 Vers un changement radical de paradigme ?

254. Des propositions plus radicales ont également émergé. Face au constat de la dérive du modèle économique existant, certains proposent de tendre vers un changement structurel complet.

255. **Modèle 1.** La première proposition consiste à considérer la publication scientifique comme ouverte par essence, dépourvue de droits patrimoniaux, immédiatement accessible, et ce sous réserve du seul droit à la paternité des auteurs. Dans ce modèle, les éditeurs scientifiques sont des prestataires de service de « labellisation » (peer review), de diffusion de la connaissance et de développement de service, rémunéré pour ces services.

#### **Audition Jérôme KALFON, 5 octobre 2015**

« L'éditeur doit être rémunéré sur le travail qu'il effectue, le "travail à façon". »

« La Science est un objet particulier, facteur d'enrichissement collectif et de développement, qui doit être de libre parcours et dépourvu de droits patrimoniaux ».

256. **Modèle 2.** La seconde proposition est basée sur deux principes fondamentaux :

- l'ensemble des données scientifiques doit rester sous la maîtrise des chercheurs scientifiques ;
- les services autour de ces données sont ouverts à la concurrence.

257. Dans ce modèle, la notion de données est très large et couvre tout article publié ou non, page web, communication, blog, vidéo, photo, données de la recherche y compris acquise par des capteurs, simulation de calcul mécanique ou humain, cahier de laboratoire, code source, requête...

258. Ces données ne donnent pas lieu à un droit de propriété et doivent être librement accessibles et librement réutilisables dans les seules limites de la déontologie scientifique.

259. Les services développés autour de la donnée et notamment les techniques d'exploration peuvent être développés par des éditeurs privés et soumis à la concurrence puisque les données issues des traitements seront de libre accès et permettront aux scientifiques de vérifier les résultats à partir de ces données.

260. Or, aujourd'hui les requêtes formulées auprès d'un éditeur sont biaisées par l'algorithme de recherche. La vérification du H Index<sup>62</sup> est par exemple impossible aujourd'hui sur Google Scholar. Il est nécessaire qu'en plus du principe de neutralité du net, un droit à la transparence des données soit affirmé.

**Audition Inria, Claude KIRCHNER le 15 octobre 2015**

« Un droit à la transparence doit être affirmé et ce droit n'existe que si le chercheur a la possibilité de vérifier les données ; cette vérification est essentiel dans la méthode scientifique. »

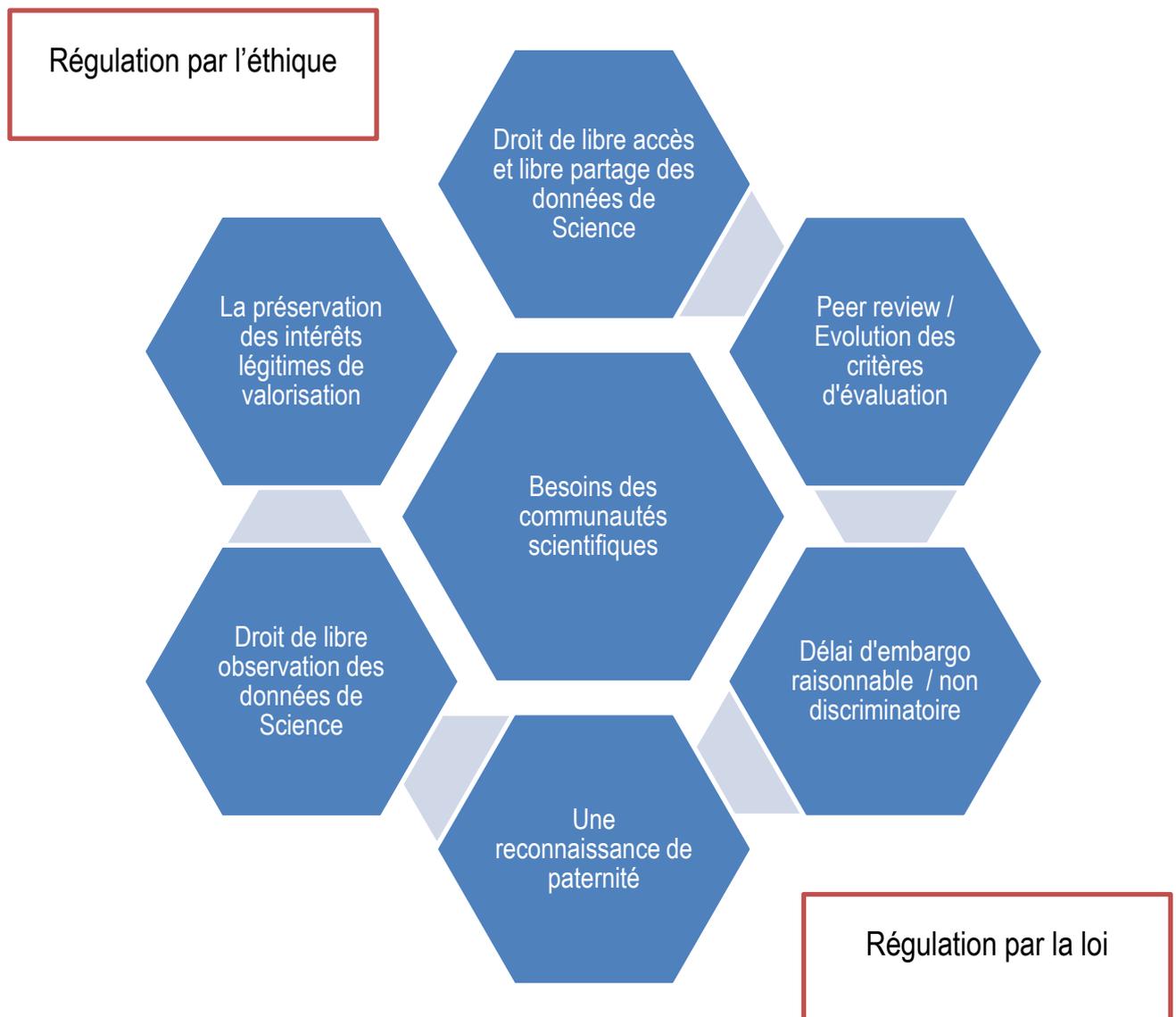
« Ce droit à la transparence sur les requêtes s'étend à toutes les informations qui traduisent toutes les interrogations des chercheurs (requêtes, discussions...) »

---

<sup>62</sup> Indice essayant de quantifier la productivité scientifique et l'impact d'un scientifique en fonction du niveau de citation de ses publications. Source Wikipedia.

## Synthèse globale

261. Les besoins des chercheurs dans l'utilisation de l'IST comme outil d'analyse au regard des contraintes techniques, contractuelles, légales existantes sont synthétisés dans le schéma ci-après.



## 2.2 L'Open Science à l'international

262. Le principe d'ouverture des données de Science a été affirmé à plusieurs reprises par les instances européennes ainsi que par les instances internationales.

263. De nombreux pays et notamment des pays dont l'activité de recherche scientifique est vue comme un moteur économique ont adopté des législations en faveur de l'Open Science et / ou du TDM.

264. La marche vers l'Open Science s'inscrit dans le mouvement de l'histoire.

### 2.2.1 Au niveau de l'Union européenne : une affirmation de l'Open Science

265. **Recommandation Commission européenne 2012.** Dans une communication du 17 juillet 2012 (C(2012)4890) relative à l'accès aux informations scientifiques et à leur conservation, la Commission européenne recommande aux Etats de :

- « définir des politiques claires en matière de diffusion des publications scientifiques issues de la recherche financée par des fonds publics et du libre accès à ces dernières » et de veiller « à ce que les publications issues de la recherche financée par des fonds publics soient librement accessibles dans les meilleurs délais, de préférence immédiatement et, dans tous les cas, au plus tard six mois après leur date de publication, et au plus tard douze mois pour les publications dans les domaines des sciences sociales et humaines » ;
- « de définir des politiques claires en matière de diffusion des données de la recherche financée par des fonds publics et de libre accès à ces dernières » et de veiller « à ce que les données de la recherche financée par des fonds publics deviennent accessibles, utilisables et réutilisables par le public au moyen d'infrastructures électroniques ».

266. Par ailleurs, soulignant le rôle fondamental de la publication dans l'évaluation du chercheur, la recommandation prévoit une révision des politiques d'avancement :

- en récompensant « les chercheurs qui adhèrent à une culture de partage de leurs résultats de recherche, notamment en garantissant le libre accès à leurs publications » ;
- en élaborant, encourageant et utilisant « de nouveaux modèles, critères et indicateurs alternatifs pour l'évaluation des carrières ».

267. **Horizon 2020.** Dans le cadre du programme Horizon 2020 (instrument financier de développement de l'Union de l'innovation pour le financement de la recherche et de

l'innovation pour la période 2014-2020), la Commission européenne a fait du libre accès aux publications scientifiques un principe général.

268. Les lignes directrices de l'Open access aux publications scientifiques ont été rédigées dans le cadre de ce programme et publiées une première fois le 16 décembre 2013<sup>63</sup>. Elles prévoient qu'à partir de 2014, toutes les publications scientifiques rédigées au cours de projets financés ou cofinancés dans le cadre du programme Horizon 2020 devront être mises à disposition en libre accès :

- soit immédiatement par l'éditeur, qui les publiera en ligne (approche dite de la «voie dorée»); les coûts de publication engagés pourront être remboursés par la Commission européenne ;
- soit par les chercheurs, 6 mois au plus tard après la publication (12 mois pour les sciences sociales et humaines), via des archives libres d'accès (approche dite de la «voie verte»). Les publications et les données scientifiques issues de la recherche financée par des fonds publics seront accessibles plus rapidement à un plus large public, ce qui permettra aux chercheurs et aux entreprises de les exploiter plus facilement.

269. **Rapport Reda.** Le Parlement européen a adopté le 9 juillet 2015 la version finalisée du rapport sur la mise en œuvre de la directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information dont la rapporteuse est Julia REDA. Ce rapport préconise dans le cadre de la révision de la directive DADVSI l'adoption des dispositions suivantes :

- souligne la nécessité de permettre des techniques analytiques automatisées des textes et des données (par exemple la "fouille de textes et de données") à toutes les fins, pour autant que la permission de lire l'œuvre ait été acquise (point 48 du rapport) ;
- demande une large exception à des fins de recherche et d'éducation, qui devrait couvrir non seulement les établissements d'enseignement mais également tout type d'activité éducative ou de recherche, y compris l'enseignement non formel (point 51 du rapport).

270. **Révision du droit d'auteur.** La Commission européenne a également annoncé le 9 décembre 2015 « ses premières mesures pour améliorer l'accès aux contenus en ligne

---

<sup>63</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

et présente sa vision d'un droit d'auteur modernisé ». Dans ce cadre, la Commission annonce avoir « l'intention de travailler sur les exceptions au droit d'auteur » et notamment de réviser « les règles de l'Union afin de permettre aux chercheurs d'utiliser plus facilement les techniques de fouille data mining et de text mining pour analyser de grandes séries de données. »

### 2.2.2 A l'étranger : une légalisation de l'Open Science

271. De nombreuses initiatives et législations en faveur de l'ouverture des données scientifiques et des opérations de *text and data mining* sur les données scientifiques sont présentes sur la scène internationale.

272. **Etats-Unis.** Les Etats-Unis ont été l'un des premiers pays à introduire des dispositions légales (Consolidated Appropriations Act, 2008) relatives à la mise à disposition du public des travaux de recherches financés par le National Institutes of Health (NIH). Cette loi prévoit que tous les articles publiés dans des revues faisant suite à des travaux financés par le NIH, doivent être versés dans l'archive ouverte en ligne propre au NIH, le « National Library of Medicine's PubMed Central ». Les contrats avec les éditeurs doivent explicitement le permettre. Il existe même une liste d'éditeurs pour lesquels ce versement est automatique, le chercheur n'ayant aucune démarche à accomplir. Cette loi précise que ce versement doit avoir été effectué au plus tard 12 mois après la publication effective dans une revue.

273. En février 2013, le "Fair Access to Science and Technology Research Act" (loi sur l'accès équitable à la recherche scientifique et technologique) (FASTR) a été présenté au Congrès.

274. En mai 2013, « The networking and information technology development program » a déposé un budget complémentaire au budget fédéral 2014. Ce budget complémentaire présente plusieurs programmes de recherche et développement dans le domaine du développement des technologies et notamment du partage de la connaissance dans le secteur public.

275. Des initiatives telles que celle de l'Université de Berkeley, qui a mis en place un fonds dédié au financement des articles en libre accès, voient également le jour.

276. Les juges américains ont par ailleurs explicitement reconnu que les opérations de TDM peuvent être couvertes par l'exception de « fair use » dans l'affaire Authors Guild v. Google (14 novembre 2013), dans le cadre de la mise en œuvre d'un vaste programme

de numérisation des livres dans le but de constituer une bibliothèque numérique accessible au plus grand nombre.

277. Enfin, la très récente décision Google Books du 16 octobre 2015 vient élargir le champ du TDM aux chercheurs américains. Google avait lancé en 2004 un grand projet de numérisation de livres ; en 2005, un ensemble d'auteurs et d'éditeurs ont attaqué le moteur de recherche pour violation du droit d'auteur. La Cour d'appel américaine a reconnu à Google le bénéfice du *fair use* en considérant que le service offert par Google aux utilisateurs était susceptible d'apporter un bénéfice à la société en termes d'accès à la connaissance et justifiant que soit écartée l'application des droits exclusifs d'auteur. Cette décision ouvre par conséquent à Google, à ses concurrents mais également aux institutions publiques la possibilité de numériser des données et de fournir des services d'exploration. Avec cette jurisprudence, les Etats-Unis donnent un avantage significatif à ses chercheurs en leur ouvrant la possibilité de numériser des ensembles très large de données accessibles licitement, de mutualiser les corpus et de développer des fonctionnalités de recherches et de traitements algorithmiques des données<sup>64</sup>.

278. **Royaume-Uni.** Le Royaume-Uni est un leader dans le développement de l'Open Access. Le Groupe Finch, groupement indépendant d'acteurs mis en place par le ministère des Entreprises, de l'Innovation et des Compétences du Royaume-Uni, a publié un rapport sur la façon dont les résultats de la recherche pourraient être rendus plus accessibles en juin 2012.

279. Le Parlement britannique en septembre 2013 a publié un rapport sur l'Open Access et semble adopter une solution duale de « green » et de « gold » Access aux travaux scientifiques.

280. Enfin, les chercheurs dont les travaux sont financés en tout ou partie par l'agence de financement non gouvernementale britannique *Wellcom Trust*, qui s'est positionnée clairement en faveur du libre accès, doivent déposer une copie électronique de tout article, accepté pour publication dans une revue à comité de lecture, dans PlubMed Central et dans PlubMed Central UK. Ce dépôt doit se faire le plus rapidement possible avec un délai maximum de 6 mois après la date de publication.

281. Le Royaume-Uni a également introduit en 2014 une exception pour l'exploration de données (droit de réaliser une copie dédiée exclusivement à l'opération de TDM dans le

---

<sup>64</sup> Article « Comment l'affaire Google Books se termine en victoire pour le Text mining » 21-10-2015 <http://scinfolex.com/2015/10/21/comment-laffaire-google-books-se-terme-en-victoire-pour-le-text-mining/>

but de conduire une recherche à des fins non commerciales sans accord des titulaires des droits ni compensation financière au bénéfice de ces derniers)<sup>65</sup> sur le fondement du « fair dealing »<sup>66</sup>.

282. **US-UK.** Par ailleurs, les Etats-Unis et le Royaume-Uni ont lancé ensemble une initiative intitulée « UK-US Global Innovation Initiative » qui a pour but de faciliter la collaboration universitaire entre les deux pays ainsi qu'avec les pays émergents au cours des cinq prochaines années.

283. **Allemagne.** L'Allemagne a créé une plateforme de dépôt des contributions scientifiques. L'organisme fédéral participe à l'acquisition de revues, par le biais de groupements de commande, et pose comme condition l'acceptation par les éditeurs d'une diffusion en accès libre des contributions à l'expiration d'un délai de 12 mois à compter de la publication. L'acquisition reste, à ce jour, limitée à des bouquets de niche, moins coûteux que les principaux abonnements.

284. Par ailleurs a été proposé en février 2013 et adopté en novembre 2013 un projet de loi<sup>67</sup> amendant la loi allemande sur le droit d'auteur (« Entwurf eines Gesetzes zur Nutzung verwaister Werke und zu weiteren Änderungen des Urheberrechtsgesetzes und des Urheberrechtswahrnehmungsgesetzes »). Ce texte aménage le droit d'auteur dans le domaine de la communication scientifique en introduisant « un droit d'exploitation secondaire » (Zweitverwertungsrecht).

285. Le texte adopté est le suivant :

- « L'auteur d'une contribution savante, née d'une activité de recherche financée au moins pour moitié par des ressources publiques et publiée dans une collection périodique paraissant au moins deux fois par an, est en droit, même lorsqu'il a cédé un droit d'exploitation exclusif à l'éditeur, de rendre publiquement accessible cette contribution dans la version acceptée du manuscrit, après un délai de douze mois suivant sa première publication, toute fin commerciale étant exclue. La

---

<sup>65</sup> Copyright, Designs and Patents Act, Article 29A "Copies for text and data analysis for non-commercial research", Octobre 2014

<sup>66</sup> Copyright, Designs and Patents Act 1988 Art 29 "Fair dealing with a literary, dramatic, musical or artistic work for the purposes of research or private study does not infringe any copyright in the work or, in the case of a published edition, in the typographical arrangement."

<sup>67</sup> « Entwurf eines Gesetzes zur Nutzung verwaister Werke und zu weiteren Änderungen des Urheberrechtsgesetzes und des Urheberrechtswahrnehmungsgesetzes »

source de la première publication doit être indiquée. Un accord dérogatoire au détriment de l'auteur est sans effet ».<sup>68</sup>

286. Ce texte organise un droit d'exploitation secondaire légal des contributions scientifiques nées d'une activité d'enseignement ou de recherche financée au moins pour moitié par des ressources publiques et ce même si un droit d'exploitation exclusif a été cédé à un éditeur. Le texte organise les modalités pratiques de ce droit d'exploitation secondaire afin de tenir compte des intérêts des éditeurs (délai de 12 mois, publication dans des collections périodiques, une mise à disposition à des fins non commerciales) et des auteurs (droits d'exploitation secondaire inaliénable, renforcement de la position de l'auteur par une seconde publication et meilleure diffusion des résultats).

### 2.2.3 Au niveau des instances internationales : une tendance à l'Open Science

287. Enfin, des prises de positions en faveur de l'ouverture des données scientifiques ont été adoptées dans des instances internationales telles que l'UNESCO, le G8 et l'OCDE.

288. **UNESCO.** Dans le cadre de la conférence mondiale sur la Science du 1<sup>er</sup> juillet 1999 organisée par l'UNESCO, intitulée « La science pour le XXI<sup>ème</sup> siècle, Un nouvel engagement », deux documents ont été adoptés :

- « La Déclaration sur la science et l'utilisation du savoir scientifique » : dans ce texte il est affirmé la nécessité de partager les données et la connaissance scientifique et de favoriser, faciliter la coopération, « indispensable au travail scientifique et à la transformation des résultats de la recherche scientifique en avantages tangibles pour la société »<sup>69</sup>. Le texte déclare que « les parlements et les gouvernements devraient être invités à fournir une assise juridique, institutionnelle et économique au développement des capacités scientifiques et technologiques dans les secteurs publics et privés et à faciliter l'interaction de ces deux secteurs »<sup>70</sup> ;
- un « Agenda pour la science - cadre d'action » : ce document définit les principes directeurs pour « faire face aux problèmes et aux défis de la recherche scientifique ». « La mise en commun de l'information et des connaissances scientifiques » est l'un de ces principes. Le document incite les différents acteurs de la recherche à collaborer à l'échelon international<sup>71</sup> et « à faciliter la publication

---

<sup>68</sup> Traduction du texte adopté : <http://openaccess.inist.fr/?Point-sur-le-Libre-Acces-en>

<sup>69</sup> « Déclaration sur la science et l'utilisation du savoir scientifique », UNESCO, point 38

<sup>70</sup> « Déclaration sur la science et l'utilisation du savoir scientifique », UNESCO, point 37

<sup>71</sup> « Agenda pour la science-cadre d'action », UNESCO, point 17

et la plus large diffusion des résultats de la recherche scientifique (...) et ce grâce à la formation, à l'échange d'information et au développement de services bibliographiques et de systèmes d'information satisfaisant mieux les besoins de communautés scientifiques à travers le monde »<sup>72</sup>. Le texte suggère également aux établissements de recherche à favoriser l'utilisation des nouvelles technologies de l'information « en développant l'édition électronique et en créant des environnements virtuels de recherche et d'enseignement ou des bibliothèques numériques »<sup>73</sup>. Enfin, le texte incite les gouvernements à octroyer des budgets couvrant les infrastructures nécessaires de diffusion et de partage des connaissances scientifiques et à prévoir des cadres juridiques appropriés<sup>74</sup>.

**289. G8 Londres 2013.** Dans une déclaration commune du 12 juin 2013, les ministres des sciences du G8 affirment qu'une collaboration scientifique internationale est un nouveau défi mondial nécessitant une modification et amélioration des infrastructures de recherche afin de rendre les données scientifiques publiées et validées par les pairs accessibles dans le cadre d'un accès global. Les acteurs du G8 proposent le cadre :

- d'une infrastructure de recherche mondiale ;
- d'un Open data de la recherche scientifique ;
- d'un élargissement de l'accès aux résultats de la recherche scientifique.

**290. Rapport OCDE.** L'OCDE a publié un rapport « *Enquiries Into Intellectual Property's Economic Impact* » (août 2015) sur les impacts économiques des droits de propriété intellectuelle. Le chapitre 7 « *Legal Aspects of Open access to Publicly Funded Research* » dresse un état des lieux des législations nationales sur l'accès, la diffusion et l'utilisation des résultats des recherches financées par des fonds publics dans le cadre du libre accès.

**291.** Le rapport attire également l'attention sur deux autres questions :

- le problème du libre accès dans le cadre de partenariats entre le secteur public et le secteur privé ;
- le régime du *text and data mining*.

**292.** Ces prises de positions institutionnelles et étrangères en faveur de l'Open Access montrent l'évolution naturelle de l'édition scientifique vers une mise à disposition gratuite

---

<sup>72</sup> « Agenda pour la science-cadre d'action », UNESCO, point 19

<sup>73</sup> « Agenda pour la science-cadre d'action », UNESCO, point 20

<sup>74</sup> « Agenda pour la science-cadre d'action », UNESCO, point 21

et large des données scientifiques. La France a également affirmé à plusieurs reprises son engagement dans l'Open Access.



L'Open Science procède d'un mouvement international marqué par des engagements et des prises de positions fortes des instances européennes, supranationales ainsi qu'étrangères.

La recherche française doit s'inscrire dans ce mouvement au risque d'être discriminée, marginalisée et non compétitive.

## 2.3 Concepts clés de la Science ouverte

293. Le droit nouveau de la Science ouverte s'inscrit dans un corpus de concepts existants et d'intérêts légitimes qui doivent être préservés voire renforcés.

### 2.3.1 Un droit de l'Open : un mouvement international inéluctable d'ouverture et de partage des données

294. Open data, open format, open source, open access, open process sont autant de sphères d'ouverture des données dont la philosophie commune est le partage et la libre réutilisation des données.

#### 2.3.1.1 Open data

295. Le mouvement d'ouverture des données, autrement appelé « Open data », est apparu au cours des années 1957-1958 aux Etats Unis avec la création du *World Data Center System*. Le mouvement s'est légalisé par l'adoption en 1966 de la loi sur le libre accès à l'information, la *Freedom of Information Act*<sup>75</sup>. Puis, en 2007, un amendement de cette loi par la loi *OPEN Government Act*, a mis au centre de l'action gouvernementale américaine les notions de transparence, de gouvernance et d'ouverture des données publiques.

296. Ce mouvement a trouvé un large écho en Europe, notamment dans le secteur public. Au Royaume-Uni, un projet analogue à celui mené aux Etats-Unis est officiellement lancé en janvier 2010<sup>76</sup>. En France, la libération des données publiques connaît un intérêt particulier dans le secteur public depuis 2009. La mission Etalab est

---

<sup>75</sup> <http://www.foia.gov/>

<sup>76</sup> <http://data.gov.uk/project>

créée en 2011 sous l'autorité du Premier ministre. Elle a pour mission de piloter la politique d'ouverture et de partage des données publiques.

297. Dans ce cadre, Etalab administre le portail unique interministériel data.gouv.fr destiné à rassembler et à mettre à disposition librement l'ensemble des informations publiques de l'Etat, de ses établissements publics et, si elles le souhaitent, des collectivités territoriales et des personnes de droit public ou de droit privé chargées d'une mission de service public.

298. De nombreuses villes se dotent également de plateformes Open data par lesquelles elles mettent à disposition des administrés leurs informations publiques notamment dans les domaines de la culture, de la citoyenneté, des déplacements, de l'urbanisme, de l'environnement, des finances, de l'administration, des services et du stationnement.

### **2.3.1.2 Open format**

299. La notion de « standard ouvert » ou « format ouvert » est définie par l'article 4 de la loi n° 2004-575 du 21 juin 2004 comme « Tout protocole de communication, d'interconnexion ou d'échange et tout format de données interopérable et dont les spécifications techniques sont publiques et sans restriction d'accès ni de mise en œuvre ».

300. La loi 2015-1779 « relative à la gratuité et aux modalités de la réutilisation des informations du secteur public » ainsi que le projet de loi « pour une République numérique » réaffirment la nécessité de mettre à disposition les données dans un « standard ouvert aisément réutilisable ».

### **2.3.1.3 Open source**

301. Les licences ouvertes permettent de mettre à disposition des tiers des données, des bases de données, des créations numériques, des logiciels sous des licences plus ou moins permissives.

302. Les licences les plus utilisées notamment en Open data sont les suivantes :

- la licence Etalab ;
- la licence ODbL ;
- les licences Creative Commons ;
- la licence PDDL.

303. En matière de logiciel, les licences utilisées sont la licence CeCILL, la licence GNU-GPL, la licence MIT, etc.

#### 2.3.1.4 Open access et Open process

304. **Définition.** « Le mouvement de l'Open Access est une prise de position de la communauté scientifique internationale qui demande une mise à disposition, un accès libre et gratuit aux résultats de la recherche scientifique »<sup>77</sup>.

305. L'Open process est le droit de procéder librement à l'observation de données par l'utilisation d'outils numériques de traitement, d'analyse ou d'exploration, comme le *text and data mining* (TDM).

306. **Engagement international.** Cette prise de position en faveur de l'accès ouvert a vu le jour il y a près de 15 ans :

- dans « la lettre ouverte de Public Library of Science » de 2001 poussant les éditeurs à autoriser la mise à disposition en libre accès, dans des bibliothèques publiques en ligne, des documents issus de la recherche qui sont publiés dans leurs revues ;
- dans le cadre de « l'Initiative de Budapest en faveur de l'accès libre » (14 février 2002)<sup>78</sup>. Cette campagne mondiale prônait l'accès libre pour toute nouvelle recherche évaluée par les pairs<sup>79</sup>.

307. Aujourd'hui, ce mouvement prend de l'ampleur et se manifeste notamment :

- par des publications :
  - Tribune : « Favorisons la libre diffusion de la culture et des savoirs » (Le Monde du 10/9/2015) et pétition associée ;
  - Conseil national du numérique : Avis et rapport Ambition Numérique, « Pour une politique française et européenne de la transition numérique » (remis le 18 juin 2015) ;
  - SavoirCom1 : <http://www.savoirscom1.info/2014/10/savoirscom1-soutient-le-projet-dune-charte-universelle-de-lopen-science> ;
  - Inist et le site <http://openaccess.inist.fr/> ;
  - Sciences Communes <http://scoms.hypotheses.org/458> ;

---

<sup>77</sup> [http://corist-shs.cnrs.fr/gold\\_open\\_access](http://corist-shs.cnrs.fr/gold_open_access)

<sup>78</sup> <http://openaccess.inist.fr/?Initiative-de-Budapest-pour-l>

<sup>79</sup> <http://www.budapestopenaccessinitiative.org/>

- dans des conférences ou colloques :
  - Colloque CNRS « Dynamiques de l'édition scientifique, de l'industrie de l'information, de la documentation. Un Agenda 2015 pour la science publique ouverte » des 4 et 5 novembre 2014 ;
  - Workshop CERN / CNRS (DIST) Multi-core Platforms for science ;
  - Colloque « Innovation et gouvernance de l'IST dans l'ESR » 18 et 19 mars 2014 ;
  - GFII (Groupement français de l'Industrie de l'Information) Groupe Open access le 15 octobre 2014 ;
  - Congrès de l'ADBU (Association des Directeurs et personnels de direction des Bibliothèques Universitaires et de la Documentation) des 2 et 4 septembre 2014 « Le droit à l'épreuve des évolutions de l'IST et des besoins de la science » ;
  - Salon du Livre de Paris, Espace « Savoir & Connaissances », 21 mars 2015, conférence « Vers une science ouverte : quels impacts sur la publication scientifique ? » ;
- par la création de plateformes de dépôt parmi lesquelles :
  - HAL : archive ouverte qui a reçu 9 millions de visiteurs uniques en 2014 ;
  - HAL-SHS : déclinaison en sciences sociales de HAL ;
  - ArXiv ;
- par la création de plateformes de dépôt en Open Access parmi lesquelles :
  - Openedition.org et revues.org
  - Persée ;
- par des pratiques d'entraide :
  - <http://rue89.nouvelobs.com/2015/09/08/hashtag-clandestin-partager-science-inaccessible-261102> : la création d'un hashtag [#IcanhazPDF](#) permettant aux scientifiques d'avoir accès à un article scientifique.

### 2.3.2 Le Code de la recherche : fondement du droit de l'IST numérique

308. Le Code de la recherche organise institutionnellement les organismes qui participent en France à la recherche scientifique mais ne définit pas les principes ou les valeurs de la communauté scientifique.

309. L'article L.112-1 du Code de la recherche précise les objectifs de la recherche publique parmi lesquels :

- « le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès » ;
- « l'organisation de l'accès libre aux données scientifiques ».

**Article L112-1** [En savoir plus sur cet article...](#)

Modifié par [LOI n°2013-660 du 22 juillet 2013 - art. 16](#)

La recherche publique a pour objectifs :

- a) Le développement et le progrès de la recherche dans tous les domaines de la connaissance ;
- b) La valorisation des résultats de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie ;
- c) Le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès ;
- c bis) Le développement d'une capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique, et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;
- d) La formation à la recherche et par la recherche ;
- e) L'organisation de l'accès libre aux données scientifiques.

310. En affirmant comme objectifs de la recherche publique « le partage et la diffusion des connaissances en donnant priorité aux formats libres d'accès » et « l'organisation de l'accès libre aux données scientifiques », les fondements d'un droit numérique de la Science ouverte sont posés.

### 2.3.3 Une nécessaire cohérence avec le droit des données publiques

311. **Principes généraux.** L'ordonnance n°2005-650 du 6 juin 2005 a érigé en principe général, le droit du public de réutiliser les données détenues par les personnes publiques à quelque fin que ce soit, et notamment à des fins commerciales et privées<sup>80</sup>.

312. Les personnes publiques soumises à cette loi sont listées à l'article 1 de la loi CADA de 1978<sup>81</sup> : l'Etat, les collectivités territoriales, les autres personnes de droit public et les personnes de droit privé chargées d'une mission de service public.

313. L'article 11 de la loi CADA prévoyait que « les conditions dans lesquelles les informations peuvent être réutilisées sont fixées, le cas échéant, par les administrations

<sup>80</sup> Article 10 de la loi n° 78-753 du 17-7-1978 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques modifiée par l'ordonnance n°2005-650 du 6 juin 2005

<sup>81</sup> Loi n° 78-753 du 17 juillet 1978 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques

mentionnées aux a et b du présent article lorsqu'elles figurent dans des documents produits ou reçus par :

- Des établissements et institutions d'enseignement et de recherche ;
- Des établissements, organismes ou services culturels ».

314. Cette exception permettait de :

- ne pas faire obstacle à la politique de mise à disposition des données de recherche, puisque le principe est bien la communication et la réutilisation, mais dans des conditions déterminées par les établissements eux-mêmes ;
- permettre à cette mise à disposition de se faire dans des conditions respectueuses du rythme des travaux des chercheurs, des usages de chaque communauté et de la nature des données ;
- éviter que la réutilisation puisse être un droit invocable par tout tiers, sans que la nature des données ne puisse être prise en compte (données issues d'une Zones à Régime Restrictif, recherches en cours, données liées à des savoir-faire ou des titres de propriété industrielle, etc.). De fait, l'absence d'automaticité du droit à réutilisation permet de limiter l'exercice du droit à communication des supports de données.

315. **Evolution et loi Valter.** Cette exclusion du principe de libre réutilisation des informations publiques des données produites par les établissements et institutions d'enseignement et de recherche a été purement et simplement supprimée de la loi CADA par la loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public (loi Valter).

316. Les termes adoptés ne permettent pas de concilier le principe de mise à disposition des données publiques et la protection raisonnée des intérêts des chercheurs et du patrimoine immatériel public.

317. Le régime adoptée est par ailleurs incohérent avec les discussions sur l'avant-projet de loi pour une République numérique et les besoins des communautés scientifiques sur une circulation régulée des publications et des données tel qu'exprimés dans le cadre de la consultation publique.

+ Les dispositions issues de la loi n°2015-1779 ne permettent pas d'adapter la mise à disposition publique des données produites par les établissements et institutions d'enseignement et de recherche.

Ces dispositions ne vont pas dans le sens des besoins des chercheurs et des usages des communautés scientifiques, et ne prennent pas en compte la nature des données (données issues de recherche en cours, de savoir-faire, de ZRR...).

### 2.3.4 Un principe juridique incontournable : la propriété littéraire et artistique

318. **Origine.** Historiquement, la reconnaissance légale du droit d'auteur initiée par BEAUMARCHAIS a été proclamée par l'Assemblée constituante le 13 janvier 1791 (loi ratifiée le 19 janvier 1791 par Louis XVI). C'est la première loi édictée dans le monde pour protéger les auteurs et leurs droits : elle accorde aux auteurs le droit exclusif d'autoriser la reproduction de leurs œuvres pendant toute leur vie puis aux héritiers pendant une durée de cinq années. À l'issue de ce délai, l'œuvre entre dans le domaine public.

319. Ce droit d'auteur avait été pensé en réaction :

- aux imprimeurs-libraires-éditeurs qui disposaient de fait de la propriété des textes qu'ils avaient achetés, souvent pour un prix dérisoire, à leurs auteurs, et les exploitaient sans autre contrepartie pour les auteurs ;
- aux artistes interprètes qui gardaient les rémunérations devant revenir aux auteurs.

320. Le droit d'auteur est donc historiquement protecteur de l'auteur de l'œuvre et non du titulaire de droit.

321. **Droit d'auteur et protection des écrits scientifiques.** Les écrits scientifiques, composant de l'IST (articles, ouvrages, ...) sont éligibles à la protection par le droit d'auteur en vertu de l'article L. 112-2 du Code de la propriété intellectuelle (CPI) qui dresse une liste non limitative d'œuvres considérées comme des œuvres de l'esprit et parmi lesquelles :

- 1° Les livres, brochures et autres écrits littéraires, artistiques et scientifiques.

322. En revanche, la connaissance, le savoir, mais également les écrits purement techniques ou descriptifs, ne relèvent pas du droit d'auteur. Ils appartiennent au domaine public, sont librement réutilisables par le public sans autorisation préalable et ne peuvent

donc faire l'objet d'une protection privative ; seule la forme originale sous laquelle elles sont exprimées et formalisées pourra être protégeable.

323. Dès que l'œuvre formalise l'idée ou la connaissance, elle est protégeable par le droit d'auteur<sup>82</sup>. L'article L. 111-2 du CPI dispose à cet égard que :

- « l'œuvre est réputée créée, indépendamment de toute divulgation publique, du seul fait de la réalisation, même inachevée, de la conception de l'auteur ».

324. Seules les créations originales sont éligibles à la protection par le droit d'auteur.



Les écrits et publications scientifiques sont protégeables par le droit d'auteur s'ils sont originaux dans leur forme d'expression.

En revanche, la connaissance et les informations scientifiques contenues dans ces écrits sont par principe de « libre parcours ».

325. **Chercheur : titulaire de droit.** Aux termes de l'article L. 111.1 du CPI, c'est l'auteur, entendu comme personne physique qui a réalisé la création, qui est investi des droits de propriété intellectuelle sur son œuvre. Ce même article précise que ni le contrat de travail, ni le contrat de commande n'emporte dérogation à ce principe.

326. La règle est invariable quel que soit le statut, public ou privé, du donneur d'ordre. Toutefois, les droits d'auteur des agents publics sont aménagés dans l'intérêt du service public. Tout en reconnaissant aux auteurs agents publics un droit de propriété sur leurs œuvres, la loi l'aménage dans l'intérêt du service public que remplit la personne publique employeur :

- d'une part, en restreignant la portée des droits moraux de leurs agents :
  - o le droit de divulgation est limité ;
  - o l'agent public-auteur ne peut s'opposer à la modification de son œuvre lorsqu'elle est décidée dans l'intérêt du service public ;
  - o l'agent public-auteur ne peut exercer son droit de repentir et de retrait, sauf accord de l'autorité hiérarchique ;
- d'autre part, en lui octroyant certaines prérogatives.

327. Ce régime issu de la loi prévoit également, dans certains cas, un intéressement au profit de l'agent public.

---

<sup>82</sup> Cass. Civ. 1<sup>ère</sup>, 17-10-2000, RG n°97-20820 : « la protection de l'idée comme œuvre de l'esprit suppose la création de l'œuvre par la réalisation, même inachevée, de la conception de l'auteur ».

328. Cependant, certaines catégories d'agents publics ne sont pas soumises à ce régime spécifique. Les « agents auteurs d'œuvre dont la divulgation n'est soumise, en vertu de leur statut ou des règles qui régissent leurs fonctions, à aucun contrôle préalable de l'autorité hiérarchique » sont soumis au principe général de titularité. Les professeurs d'universités, enseignants-chercheurs, et plus généralement, selon les débats parlementaires « les agents qui disposent dans leurs fonctions d'une grande autonomie intellectuelle, voire une indépendance de jugement, même si celle-ci s'inscrit dans une hiérarchie » sont soumis au régime général applicable à tout auteur et disposent de la plénitude de leurs droits d'auteur.



Le chercheur est titulaire des droits d'auteur sur ses articles et écrits scientifiques.

### 2.3.5 Les exceptions à aménager : les motifs d'intérêt public et les secrets légaux

329. La libre mise à disposition des données et des résultats doit avoir pour limite la préservation de l'intérêt public tel que :

- la sûreté de l'Etat, la sécurité publique ou la sécurité des personnes ;
- les secrets protégés par la loi.

330. Cette limite existe déjà dans la loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal (dite loi Cada). L'article 6 prévoit une liste de restrictions au principe de communication des documents administratifs dont la consultation ou la communication porterait atteinte :

- « Au secret des délibérations du Gouvernement et des autorités responsables relevant du pouvoir exécutif ;
- Au secret de la défense nationale ;
- A la conduite de la politique extérieure de la France ;
- A la sûreté de l'Etat, à la sécurité publique ou à la sécurité des personnes ;
- A la monnaie et au crédit public ;
- Au déroulement des procédures engagées devant les juridictions ou d'opérations préliminaires à de telles procédures, sauf autorisation donnée par l'autorité compétente ;

- A la recherche, par les services compétents, des infractions fiscales et douanières ;
- Ou, sous réserve de l'article L. 124-4 du Code de l'environnement, aux autres secrets protégés par la loi (...) »



L'Open Science doit préserver les secrets ainsi que la sécurité publique.

### 2.3.6 La préservation de la vie privée et des données à caractère personnel

331. La libre mise à disposition des données et des résultats dans le cadre de l'Open Science doit avoir pour limite la préservation de la vie privée et des données personnelles des personnes en application également des limites posées par la loi Informatique et libertés.

### 2.3.7 La valorisation : un intérêt légitime à préserver

332. La libre mise à disposition des données scientifiques doit également avoir pour limite la valorisation des résultats.

333. **Cadre légal.** En effet, l'article L. 112-1 du Code de la recherche liste parmi les objectifs de la recherche publique :

- « b) **La valorisation des résultats** de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie. »

334. Le chapitre III du Code de la recherche prévoit les modalités de la valorisation des résultats de la recherche par les établissements et organismes de recherche (articles L.533-1 à L.533-3). Ainsi, le Code de la recherche encourage « les agents de l'Etat et des personnes publiques investies d'une mission de recherche, auteurs d'une invention brevetable (dans le cadre de recherches financées par des dotations de l'Etat et des collectivités territoriales ou par des subventions d'agences de financement nationales) à déclarer leur invention afin que leur employeur valorise l'invention brevetée ».

335. **Objectif.** La valorisation de la recherche consiste à augmenter la valeur des résultats de recherche et développement.

336. Le Conseil National d'Évaluation de l'enseignement supérieur (CNE, France) définit la valorisation comme le moyen de « rendre utilisables ou commercialisables les résultats, les connaissances et les compétences de la recherche »<sup>83</sup>.

337. Par conséquent, la valorisation de la recherche consiste à :

- mettre en relation le monde de la recherche avec le monde socio-économique ;
- donner de la valeur aux résultats de la recherche ;
- mettre à disposition de la société les résultats de la recherche qu'elle a contribué à financer<sup>84</sup>.

338. La loi n°99-587 du 12 juillet 1999 sur l'innovation et la recherche favorise le transfert de technologies de la recherche publique vers le secteur économique et la création d'entreprises innovantes. Les grandes voies de la valorisation sont les suivantes :

- le contrat de collaboration de recherche avec un partenaire socio-économique ;
- la mise à disposition d'équipements ;
- l'expertise ou la consultance ;
- la protection des résultats et le transfert à un partenaire via les contrats de licence ou de cession de droits de propriété intellectuelle ;
- la création d'entreprise ainsi que la mobilité des chercheurs vers les entreprises.

339. **Moyens.** La valorisation est de trois ordres :

- le brevet ;
- la confidentialité ;
- le secret.

340. **Brevet.** Sont brevetables, les inventions nouvelles impliquant une activité inventive et susceptibles d'application industrielle<sup>85</sup>. Ainsi, pour être brevetable, l'invention doit constituer une nouveauté au regard de l'état de la technique. L'invention n'est pas nouvelle lorsqu'elle se trouve déjà de toutes pièces dans l'état de la technique, soit parce qu'il existe une antériorité, soit parce que l'inventeur a lui-même rendu son invention publique avant de déposer une demande de brevet.

---

<sup>83</sup> <http://www.senat.fr/rap/r05-341/r05-3411.html>

<sup>84</sup> <https://www.univ-lille3.fr/recherche/valorisation/valorisation/>

<sup>85</sup> Art L.611-10.1.CPI

341. Si l'invention a été rendue publique en quelque endroit du monde que ce soit par une publication, une exposition publique (à un salon par exemple) ou même une simple divulgation orale, elle n'est plus nouvelle (sauf hypothèse de divulgation abusive).

342. **Confidentialité.** Les contrats de recherche ou de partenariat public / privé dans le domaine de la recherche organisent la confidentialité des résultats de la recherche issus du contrat. Cette confidentialité « négociée » est bien souvent limitée dans le temps et les conditions de publication sur les résultats de recherche sont encadrées contractuellement.

343. La mise à disposition des données et résultats scientifiques doit prendre en compte ces impératifs de confidentialité et définir un régime aux résultats de recherche financés au moins pour moitié par des fonds publics. Deux types de régime sont envisageables :

- l'ineffectivité des clauses de confidentialité sur résultats de recherche financés au moins pour moitié par des fonds publics et par conséquent le caractère d'ordre public des dispositions ;
- la possibilité de déroger au principe de libre accès et réutilisation des résultats de recherche financés au moins pour moitié par des fonds publics dans l'hypothèse où des clauses de confidentialité sont contractuellement prévues et qu'elles sont encadrées dans le temps.

344. **Secret.** Il existe deux types de secrets :

- les secrets légaux qui doivent constituer une limitation au principe d'Open Science comme développé au paragraphe 2.3.5 « les exceptions à aménager : les motifs d'intérêt public et les secrets légaux » ;
- les secrets contractuels, autrement dit la confidentialité comme mentionné ci-dessus.

345. La protection des actifs immatériels de la recherche publique s'inscrit dans le cadre de cet objectif de valorisation. L'Open Science doit également s'inscrire dans ce cadre et être au service de cet enjeu de valorisation.



Le Code de la recherche énonce déjà dans ses principes :

- les fondements d'un droit numérique de la Science ouverte ;
- l'équilibre nécessaire entre Science ouverte et valorisation.

## 2.4 L'écart entre les usages et le droit

346. Le tableau ci-après propose une analyse des écarts entre :

- le cadre juridique existant, les vides et lacunes constatés notamment par les grands témoins ;
- les usages des chercheurs.

347. Pour chaque pratique ou besoin identifié, un écart avec l'état du droit existant est défini entre 1 et 5. L'échelle utilisée est la suivante :

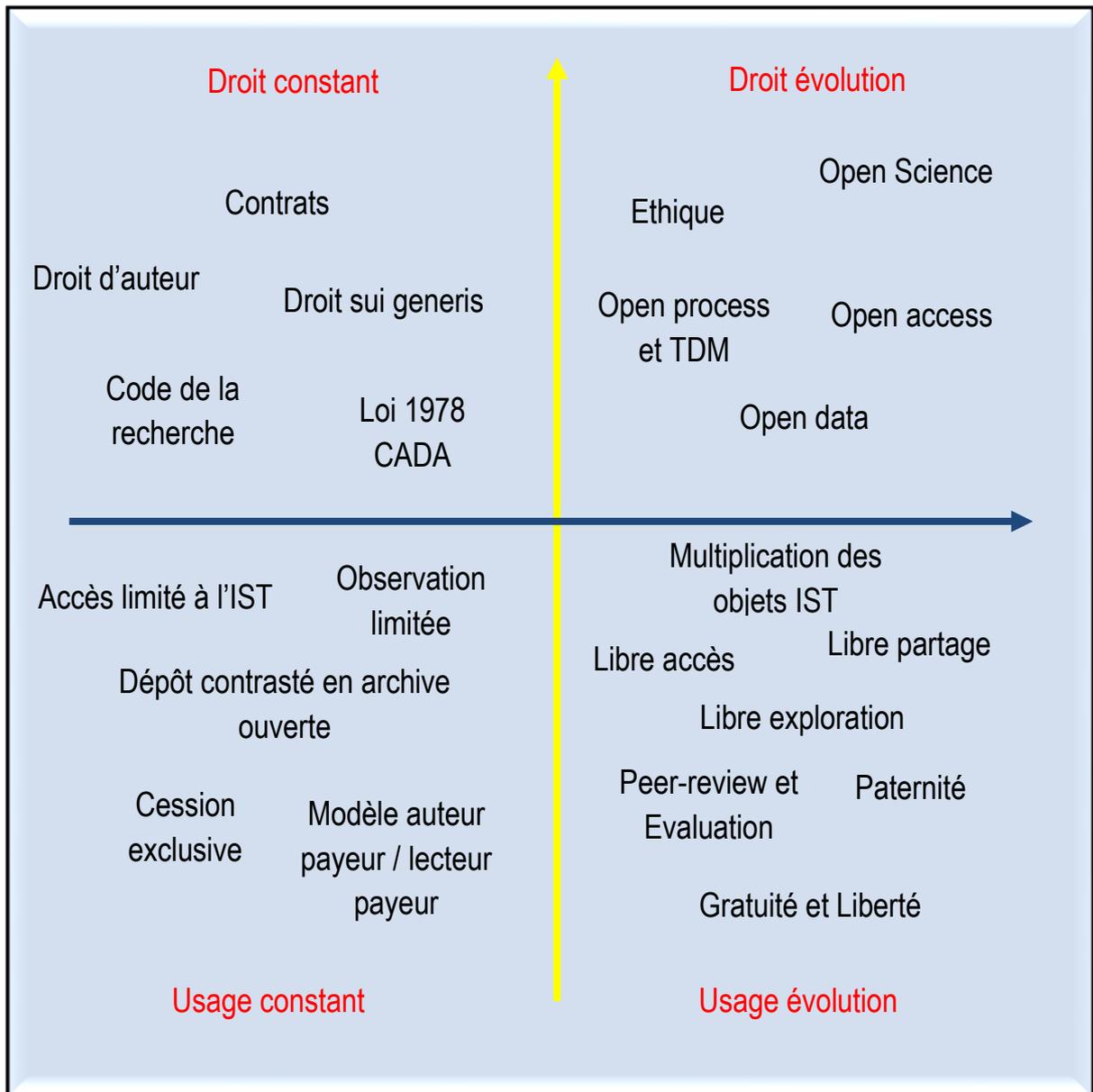
1/5 : écart inexistant	4/5 : écart fort
2/5 : écart faible	5/ 5 : incompatibilité complète
3/5 : écart moyen	

348. Un commentaire rapide justifiant cet écart au regard des développements présentés dans ce Livre blanc est proposé en dernière colonne.

Usage numérique	Cadre juridique français	Ecart	Commentaires
Libre accès et libre exploration des données	Code de la recherche Loi de 1978 modifiée par la loi Valter 2015-1779	3/5	A droit constant : incompatibilité
Libre accès aux écrits scientifiques publiés	Protection par le droit d'auteur Contrat d'édition et clause d'exclusivité	5/5	Incompatibilité complète surtout dans les contrats d'édition avec cession exclusive
Libre exploration des écrits scientifiques publiés	Absence de cadre légal Débat sur l'incompatibilité avec le droit d'auteur et le droit sui generis Conditions générales d'utilisation des plateformes éditeurs / contrat d'abonnement	5/5	Insécurité juridique Privatisation par contrat et par les API des éditeurs
Dépôt des données en archive ouverte et pérenne	Loi de 1978 modifiée par la loi Valter 2015-1779	3/5	Dépôt déjà organisé dans certaine communauté
Dépôt des publications en archive ouverte et	Protection par le droit d'auteur Contrat d'édition et clause d'exclusivité	3,5/5	Incompatibilité complète surtout dans les contrats d'édition avec cession

Usage numérique	Cadre juridique français	Ecart	Commentaires
pérenne			exclusive  En pratique, certains éditeurs l'autorisent après une période d'embargo
Peer review	Absence de cadre légal	<b>NA</b>	Aucun encadrement légal Règles éthiques à définir
Evaluation des chercheurs et prise en compte de la publication en Open	Décret n°83-1260 du 30 décembre 1983 fixant les dispositions statutaires communes aux corps de fonctionnaires des établissements publics scientifiques et technologiques	<b>4/5</b>	Obligation d'évaluation posée dans le décret  Evolution nécessaire des critères d'évaluation  Règles éthiques à définir
Reconnaissance de paternité	Droit d'auteur	<b>2/5</b>	Application du droit moral de l'auteur  Règles éthiques à définir
Valorisation	Code de la recherche	<b>2/5</b>	Dispositions légales existantes  Enjeu à prendre en compte dans l'Open Science numérique (multiplication des objets IST)
Ethique de l'IST	Absence de cadre légal	<b>NA</b>	Règles éthiques à définir notamment en matière de Peer review / Evaluation des chercheurs / Reconnaissance de paternité

349. Le carré juridique ci-après formalise ces écarts :



## 2.5 La modification de l'article 17 du projet de loi pour une République numérique

350. L'article 17 du projet de loi pour une République numérique dans sa version finale produite par le Gouvernement a fait sien un grand nombre de principes directeurs dégagés dans le cadre de la consultation nationale et du GouvCamp. Des améliorations prenant en compte les idées formulées par les grands témoins sont proposées.

### 2.5.1 Principes directeurs

351. La proposition de texte est basée sur les 8 principes directeurs suivants. Ces principes directeurs ont été présentés par Alain BENSOUSSAN et Grégory COLCANAP, co-rapporteurs dans le cadre du compte-rendu consensuel du groupe GouvCamp de l'article Open Access (atelier de travail dans le cadre du GouvCamp sur l'avant-projet de loi pour une République numérique 16-10-2015) et ont fait l'objet d'un rapport remis à Axelle LEMAIRE, Secrétaire d'État chargée du Numérique, vendredi 16 octobre 2015.

- 1. Les écrits scientifiques doivent devenir des biens communs.** En effet, les connaissances scientifiques et résultats de la Science sont des biens communs destinés à un usage universel dans l'intérêt de l'humanité. Les écrits scientifiques ne peuvent être un moyen d'interdire ou restreindre l'accès à la connaissance scientifique.
- 2. Le texte "Data mining" doit être réintroduit de manière absolue.** Il s'agit d'un enjeu économique (innovation et moteur de la recherche) et de positionnement concurrentiel (se doter de dispositions comparables à d'autres pays comme la Grande Bretagne). La demande est que la recherche française puisse disposer *a minima* des réglementations favorables aux recherches anglaises (TDM) afin d'être compétitive. Le Data mining est un "télescope" consacrant un droit à l'observation numérique en toute liberté.
- 3. Les données scientifiques (à financement majoritairement public) doivent devenir un bien commun informationnel.** L'objectif est d'autoriser le dépôt en même temps que des articles, des données de base de la recherche. Un tel dépôt faciliterait la reproductibilité de la recherche en même temps qu'il favoriserait l'innovation dans la société civile.
- 4. La valorisation des contenus d'un article scientifique doit pouvoir faire l'objet d'une exploitation notamment commerciale.** L'écrit scientifique en tant que tel ne peut faire l'objet d'une exploitation commerciale sans autorisation des titulaires de droit. En revanche, le contenu d'un écrit scientifique et l'information

scientifique et technique qu'il recèle est la source potentielle d'innovations dont les retombées commerciales peuvent être considérables. La valorisation des avancées scientifiques, et par conséquent, des écrits qui les décrivent, est l'une des missions fondamentales attribuées aux organismes de recherche et aux universités. Interdire l'exploitation commerciale du contenu d'un article scientifique par ses auteurs et leurs employeurs serait contraire aux missions fondamentales des écoles, universités et instituts d'enseignement et de recherche, et constituerait un lourd handicap à l'innovation en France. La limitation de l'exploitation commerciale des publications mises en ligne en accès ouvert peut porter sur l'article en lui-même par l'auteur ou un tiers, mais pas sur le contenu et conclusions de l'article.

**5. Les clauses de cession exclusive doivent être déclarées nulles et non avenues.** Il est souhaitable de rééquilibrer les intérêts en présence par la prise en compte des risques d'asymétrie contractuelle.

**6. La dernière version acceptée du manuscrit doit être disponible immédiatement ou dans un délai de de 6 à 12 mois.** La durée du délai proposé constituerait un handicap pour la recherche française et sa diffusion aux autres pays ; elle n'est pas non plus en concordance avec les recommandations européennes, ce qui créerait des incohérences dans le cas de contrats européens.

**7. Le dépôt doit être en archive ouverte publique et pérenne.** Il apparaît essentiel de mentionner la conservation du droit de dépôt dans les archives ouvertes. Ces infrastructures ont pour vocation de recueillir, préserver et mettre en libre accès la production scientifique et répondent à des standards internationaux. Ne pas les mentionner expose à ne pas leur donner une juste reconnaissance en tant qu'outil stratégique, à une mise en ligne désordonnée et à un risque de refus de dépôt en archive ouverte sous prétexte de mise en ligne sous une autre forme numérique.

**8. La loi s'applique aux contrats selon les règles d'application de la loi dans le temps.**

352. Certains de ces principes ont été pris en compte dans la rédaction de l'article 17 tel qu'issu du Conseil des ministres et notamment :

- la clarification de la disposition relative à l'exploitation commerciale ;
- la déclaration de certaines clauses de cession nulles et non avenues ;
- les délais d'embargo ramenés au maximum à 6 et 12 mois ;

- l'application de la loi selon les règles d'application de la loi dans le temps.

## 2.5.2 Proposition de modifications de texte

**353. Résultats scientifiques publiés.** Les modifications complémentaires proposées sur l'article 17 dans sa version adoptée par l'Assemblée nationale apparaissent en vert ci-dessous.

### Article 17

A la fin du chapitre III du Titre III du Livre V du code de la recherche, il est ajouté un article L. 533-4 ainsi rédigé :

« Art. L. 533-4. - I. - Lorsque un écrit **résultat** scientifique, issu d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne, est publié ~~dans un périodique paraissant au moins une fois par an,~~ son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique **notamment dans une archive ouverte publique et pérenne**, sous réserve des droits des éventuels coauteurs, toutes les versions successives du manuscrit jusqu'à la version finale acceptée pour publication, dès lors que l'éditeur met lui-même celle-ci gratuitement à disposition par voie numérique et, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication. Ce délai est de six mois pour une publication dans le domaine des sciences, de la technique et de la médecine et de douze mois dans celui des sciences humaines et sociales. Un délai inférieur peut être prévu pour certaines disciplines, par arrêté du ministre chargé de la recherche.

La version mise à disposition en application du premier alinéa ne peut faire l'objet d'une exploitation dans le cadre d'une activité d'édition à caractère commercial.

« II. – Dès lors que les données issues d'une activité de recherche, financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne, ne sont pas protégées par un droit spécifique, ou une réglementation particulière, et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre.

« III. - L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.

« IV. - Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite.»

**354. Notion d'écrit scientifique.** La notion d'écrit scientifique utilisée par le texte du projet de loi ne correspond pas à une typologie de données telle qu'utilisée par les chercheurs. Il serait préférable d'utiliser le vocabulaire issu de la pratique dans la loi (résultat scientifique publié).

355. La notion d'écrits scientifiques fait référence à l'article L.112-2 du Code de la propriété intellectuelle. Les écrits scientifiques sont abordés dans leur dimension d'œuvre de l'esprit, protégée par le droit d'auteur dans sa forme d'expression originale. Cette approche ne rend pas compte de la valeur scientifique ou informationnelle de l'écrit scientifique, ni des données scientifiques qu'il renferme.

**356. Suppression de la notion de bien commun.** La référence à la notion de bien commun et à l'article 714 du Code civil a été supprimée de la version du projet de loi ; le texte préférant affirmer un principe de libre réutilisation. La justification de cette reformulation est la suivante ;

- « Le Gouvernement a tenu compte de l'avis du Conseil d'Etat, qui estime que les effets d'une référence à l'article 714 du Code civil seraient incertains, faute d'une jurisprudence suffisante. La formulation choisie, issue des échanges avec le Conseil d'Etat, vise les mêmes finalités, à savoir la libre réutilisation des données. »

**357. TDM.** Les dispositions de l'avant-projet de loi numérique dans sa version de juillet intégraient dans le Code de la propriété intellectuelle une exception au droit d'auteur et au droit du producteur de base de données en faveur du *text and data mining*. Ce texte a été supprimé puis a été réintroduit par amendements devant l'Assemblée nationale. Des propositions de modification et de renforcement de ce texte ouvrant la voie au TDM sont proposées ci-après.

## 2.6 Des propositions à affirmer, des pistes pour l'avenir

### 2.6.1 Un droit positif de *text and data mining*

#### 2.6.1.1 Principes directeurs

358. L'enjeu scientifique, économique, humain du TDM est crucial pour la recherche française.

359. Les chercheurs français ne doivent pas être discriminés face à leurs homologues étrangers au risque de faire naître une recherche à deux vitesses et de mettre en péril des partenariats avec des instituts de recherche étrangers.

360. La définition d'un cadre légal au TDM est nécessaire et peut se réaliser de deux manières :

- par la voie de l'exception au droit d'auteur ; voie choisie par les députés ;
- par la voie du droit positif : voie alternative à l'exception.

#### 2.6.1.2 L'introduction d'une exception au droit d'auteur et au droit du producteur de base de données

361. Dans le cadre de la révision de la directive 2001/29/CE DADVSI, l'Union européenne a conscience de la nécessité de réglementer la pratique du TDM. Dans ce cadre sera certainement proposée l'insertion dans la directive DADVSI révisée d'une exception au droit d'auteur et au droit du producteur de bases de données en faveur du TDM.

362. Plusieurs rapports ont été réalisés et tous vont dans le sens de l'introduction d'une régulation du TDM :

- le rapport Sirinelli pour le CSPLA « Rapport de la mission sur la révision de la directive 2001/29/CE sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information » de décembre 2014 demande la « création de nouvelles exceptions au droit d'auteur concernent notamment les activités dites de *text and data mining* (TDM) »<sup>86</sup> ;
- l'étude du cabinet Wolf & Partners de mars 2014, intitulée « Study on the legal framework of *text and data mining* »<sup>87</sup> pour la Commission européenne ;

---

<sup>86</sup> Rapport CSPLA page 8

<sup>87</sup> [http://ec.europa.eu/internal\\_market/copyright/docs/studies/1403\\_study2\\_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf)

- un groupe d'experts de la Commission européenne a également publié en avril 2014 un rapport intitulé « Standardisation in the area of innovation and technological development, notably in the field of *Text and data mining* »<sup>88</sup> ;
- le rapport Reda : ce rapport adopté par le Parlement européen le 9 juillet 2015 affirme « l'impératif d'évaluer avec soin la mise à disposition des techniques analytiques automatisées des textes et des données (par exemple la "fouille de textes et de données") à des fins de recherche. » ;
- le communiqué de presse de la Commission européenne du 9 décembre 2015 présentant les mesures pour améliorer l'accès aux contenus en ligne et présentant sa vision d'un droit d'auteur modernisé.

363. Dans la synthèse officielle de la consultation publique, il est précisé que « le droit européen ne permet pas actuellement de créer de nouvelles exceptions [au droit d'auteur], et le Gouvernement souhaite que cette question [du TDM] soit abordée dans le cadre des travaux européens en cours. »

364. En effet, un projet de directive révisée est attendu pour début 2016. Toutefois, il faut compter au moins 2 à 3 ans avant que cette directive soit acceptée et 2 ans supplémentaires pour qu'elle soit transposée en droit français. La recherche française ne peut pas se permettre de prendre un tel retard et de risquer de réduire les partenariats avec les universités ou unités de recherche étrangères, de ralentir les chercheurs dans leur démarche de recherche, d'être irrémédiable et de reléguer la recherche française à des pratiques d'un autre âge.

365. **Projet de loi TA n°663.** Dans le cadre de l'examen du projet de loi pour une République numérique devant l'Assemblée nationale, des députés d'appartenance politique différente ont soutenu l'introduction d'un amendement en faveur du text and data mining. Le texte suivant a été adopté :

Article 18 bis (nouveau)

Le code de la propriété intellectuelle est ainsi modifié :

1° Après le second alinéa du 9° de l'article L. 122-5, il est inséré un 10° ainsi rédigé :

« 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données pour les besoins de la

---

<sup>88</sup> [http://ec.europa.eu/research/innovation-union/pdf/TDM-report\\_from\\_the\\_expert\\_group-042014.pdf](http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf)

recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche ; »

2° Après le 4° de l'article L. 342-3, il est inséré un 5° ainsi rédigé :

« 5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données dans un cadre de recherche, à l'exclusion de toute finalité commerciale. La conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche pour lesquelles elles ont été produites, sont assurées par des organismes désignés par décret. Les autres copies ou reproductions sont détruites. »

366. Les députés ont fait le choix de définir un cadre légal au text and data mining par la voie de l'exception au droit d'auteur et au droit du producteur de base de données. Ce choix ne peut être que salué et les Auteurs du Livre blanc encourage à l'adoption définitive de cet article. Cette voie de l'exception a pour alternative éventuelle, la création d'un droit positif sectorisé au domaine de la recherche.

### **2.6.1.3 Une alternative avec la création éventuelle d'un droit positif dans le Code de la recherche**

367. Compte tenu des enjeux fondamentaux du TDM pour la recherche et de ces contraintes de calendrier, il est recommandé de réfléchir à une alternative éventuelle, dans le Code de la recherche, avec l'introduction d'un droit positif pour réaliser des opérations de traitement et de fouilles sur les données scientifiques, les données de la recherche et les publications scientifiques.

368. **Intérêt supérieur de la recherche.** Cette proposition ne constitue apparemment pas une exception au droit d'auteur, même si elle peut paraître constitutionnellement fragile, et par conséquent elle ne se heurte pas à l'interdiction posée par la directive DADVSI de créer des exceptions autres que celles expressément prévues par la directive.

369. Elle ne traite pas des œuvres en général, mais s'attache exclusivement aux données et résultats de la recherche scientifique publique, qui comprennent entre autres des publications scientifiques.

370. Ce droit positif introduit au nom de l'intérêt supérieur de la recherche prend forme dans le droit à l'information et le droit d'accès à la connaissance. Cet intérêt général « n'est désormais plus géré exclusivement dans les exceptions au droit d'auteur ; il prend de plus en plus la forme de limitations externes au droit d'auteur, en plus donc des exceptions »<sup>89</sup>.

**371. Droit d'observation préexistant.** L'introduction d'un droit d'exploration numérique constitue l'affirmation d'un principe général d'observation et la transposition d'une pratique existant à l'ère du papier, l'analyse de texte par des outils matériels (stabilo, commentaires, etc.) à une pratique numérique d'analyse par des outils automatiques. L'objectif est le même mais l'outil diffère et permet le traitement simultané d'une grande masse de données.

372. L'affirmation de ce droit positif préexistant permettrait de lever l'incertitude quant aux droits des chercheurs d'analyser, d'explorer les données scientifiques légalement accessibles et réduirait le risque de privatisation des données de la recherche par les éditeurs.

373. La création de ce droit au TDM est proposée pour ajout dans le Code de la recherche sous les termes suivants :

Au chapitre III du titre III du Livre V du Code de la recherche, il est inséré un article L.533-5 ainsi rédigé :

Art. L. 533-5 :

- « Les données de la recherche et les écrits scientifiques dans les conditions mentionnées à l'article L. 533-4 du Code de la recherche, peuvent librement et gratuitement faire l'objet d'une copie technique à des fins d'observation, de traitement et d'exploration numérique, pour les besoins de la recherche publique et dans le respect du droit moral de l'auteur. »

<sup>89</sup> Droit d'auteur et intérêt général, Yves GAUBIAC, Revue Propriétés intellectuelles Juillet 2010 - N°36

## 2.6.2 La définition d'un droit de la Science ouverte : affirmation des valeurs

### 2.6.2.1 Principes directeurs

374. Le Code de la recherche organise institutionnellement les organismes qui participent en France à la recherche scientifique mais aucun texte ne définit les principes et les valeurs de la communauté scientifique.

375. Un droit de la Science, défini par consensus par les chercheurs pour la recherche publique, serait la retranscription des valeurs des communautés scientifiques telles que :

- le partage de la connaissance ;
- le libre accès aux données scientifiques ;
- la liberté de traitement des données scientifiques ;
- les enjeux de valorisation.

376. Un texte posant les principes d'une Science ouverte permettrait à la France d'être pionnière dans ce domaine.

### 2.6.2.2 Proposition de texte

377. Il est proposé d'ajouter les dispositions suivantes dans le Code de la recherche :

Il est ajouté au

Code de la recherche,

LIVRE Ier : L'ORGANISATION GÉNÉRALE DE LA RECHERCHE ET DU DÉVELOPPEMENT TECHNOLOGIQUE  
TITRE Ier : ORIENTATION DE LA RECHERCHE ET DU DÉVELOPPEMENT TECHNOLOGIQUE

Chapitre II : Objectifs et moyens institutionnels de la recherche publique

Les articles L.112-6 à L.112-11 suivants :

#### **Article L.112-6**

1° Sans préjudice des dispositions de l'article L. 112-1, les données scientifiques issues de la recherche publique mises à la disposition de la communauté scientifique et librement réutilisées pour les besoins de la recherche publique dans les conditions prévues par le présent chapitre, sont des biens communs relevant du régime du domaine commun informationnel.

2° Les données scientifiques sont réputées être issues de la recherche publique lorsqu'elles sont issues d'une activité de recherche financée au moins pour moitié par des fonds publics ou lorsqu'elles sont le fruit de travaux intégralement conduits par des unités de recherche créées par des établissements participant à la mise en œuvre du service public de la recherche.

3° On entend par données scientifiques tous les résultats de la recherche ainsi que toutes les données de la recherche ayant servi à établir ces résultats.

### **Article L.112-7**

A moins que des exigences de valorisation ou de transfert ou un motif d'intérêt général s'opposent à leur publicité, la mise à disposition de données scientifiques prévue à l'article L. 122-6 est réalisée par un dépôt sur des plateformes numériques ouvertes à la communauté scientifique. Celles-ci offrent des fonctionnalités d'accès en ligne, de référencement, de partage et de traitement des connaissances et ce sans préjudice du droit d'auteur et du droit du producteur de base de données.

Les établissements publics chargés d'une mission de recherche peuvent mettre en place des plateformes numériques à cet effet.

### **Article L.112-8**

Les établissements et organismes de recherche ayant rendu public une donnée sous quelque forme que ce soit ont le droit de se réserver l'exploitation de la donnée, y compris à des fins de recherche, dans des conditions et pendant une durée définie par lesdits établissements et organismes. Il est alors tenu compte des usages et pratiques dans les domaines scientifiques concernés. En tout état de cause, cette durée ne peut excéder cinq ans.

### **Article L.112-9**

Lorsque des données scientifiques font l'objet d'une publication par un éditeur privé, elles doivent faire l'objet d'un dépôt parallèle, sous une forme appropriée, sur une plateforme numérique ouverte telle que visée à l'article L. 112-7.

Ces données scientifiques sont rendues accessibles et réutilisables immédiatement dans leur version auteur et au plus tard à l'issue d'un délai de six mois pour les sciences et de douze mois pour les sciences humaines et sociales à compter de la première publication dans leur version éditeur.

### **Article L.112-10**

La réutilisation des données scientifiques s'effectue dans le respect du droit de paternité des enseignants-chercheurs, des chercheurs et, plus généralement, de toute personne ayant directement contribué à leur obtention.

La réutilisation inclut notamment la reproduction, la modification et le traitement des données quelle qu'en soit la forme par tous outils informatiques et toutes techniques automatisées d'exploration, d'analyse, d'indexation, d'agrégation, de classement, de traitement des données, telle la fouille de texte et de données.

### **Article L.112-11**

Les contrats portant sur la publication de données scientifiques issus de la recherche publique ne peuvent avoir pour objet ou pour effet de faire obstacle, en tout ou en partie, à l'application des dispositions du présent chapitre, notamment des articles L. 112-9 et L. 112-10.

## **2.6.3 Un référentiel des usages de l'IST numérique**

378. L'objectif de la création d'un répertoire des usages des communautés scientifiques est de disposer d'un référentiel souple et adaptable. Ce répertoire recenserait :

- les pratiques communes de l'ensemble des communautés scientifiques ;
- les pratiques propres à chacune des communautés scientifiques ;
- les définitions structurantes de la Science.

379. Ce référentiel pourrait prendre deux formes :

- un répertoire créé par une instance représentative de la communauté scientifique ; cela pourrait être le rôle de la future Agence pour la Science ouverte ;
- un référentiel général des usages publié par arrêté.

### **2.6.3.1 Un droit souple : un répertoire des usages**

380. **Droit souple.** Le Conseil d'État recommande de doter les pouvoirs publics d'une doctrine de recours et d'emploi du droit souple pour contribuer à la politique de simplification des normes et à la qualité de la réglementation (codes de bonne conduite, recommandations de bonnes pratiques).

381. Le droit souple est défini par le Conseil d'Etat comme répondant à trois conditions cumulatives :

- avoir pour objet de modifier ou d'orienter les comportements de ses destinataires en suscitant, dans la mesure du possible, leur adhésion ;
- ne pas créer par lui-même de droits ou d'obligations pour ses destinataires ;
- présenter, par son contenu et son mode d'élaboration, un degré de formalisation et de structuration qui l'apparente aux règles de droit<sup>90</sup>.

382. **Droit coutumier.** En plus d'être qualifié de droit souple, ce répertoire des usages pourrait constituer une véritable coutume des communautés scientifiques, s'imposant à ce titre comme un élément de droit positif.

383. La coutume est définie par le dictionnaire de « Vocabulaire juridique » de Gérard Cornu comme la « norme de droit objectif fondée sur une tradition populaire qui prête à une pratique constante, un caractère juridiquement contraignant ; véritable règle de droit mais d'origine non étatique que la collectivité a fait sienne par habitude dans la conviction de son caractère obligatoire ».

384. **Agence de la Science ouverte.** Le répertoire doit être rédigé, tenu, modifié, mis à jour par une instance représentative de la communauté scientifique qui pourrait être la future Agence de la Science ouverte.

385. Il pourrait également être mis en ligne pour être commenté, pour que des propositions de modifications émergent des communautés scientifiques, et que ce répertoire soit un véritable instrument de la communauté scientifique aux services de la Science.

386. La plateforme ISTEEX pourrait servir de pré-configurateur à ce répertoire des usages.

### 2.6.3.2 Un droit réglementaire : un référentiel général des usages

387. Ce répertoire des usages pourra prendre la forme d'un « Référentiel général des usages » qui comme le Référentiel général de sécurité serait issu d'un arrêté et aurait pour effet de s'imposer comme un cadre contraignant tout en étant adaptable et adapté aux enjeux et besoins de la recherche publique.

---

<sup>90</sup> <http://www.conseil-etat.fr/Decisions-Avis-Publications/Etudes-Publications/Rapports-Etudes/Etude-annuelle-2013-Le-droit-souple>

388. Le référentiel général constitue également un document de référence pour le secteur privé.

## **2.6.4 Une norme Afnor ou Iso**

### **2.6.4.1 Un référentiel de l'état de l'art et des usages**

389. Une norme désigne un ensemble de spécifications décrivant un objet, une entité ou une manière d'opérer. Il en résulte un principe servant de règle et de référence technique.

390. Une norme n'est pas obligatoire, sauf exception, son adhésion est un acte volontaire.

391. Néanmoins, même si l'application des normes n'est pas impérative, les juges ne manquent pas de se référer à ces normes lorsqu'elles existent, comme constitutives de l'état de l'art ou des usages de la profession.

392. C'est ainsi que pour les tribunaux, le non-respect d'une norme (facultative ou obligatoire) peut être constitutif d'une faute susceptible d'engager la responsabilité du professionnel.

393. Ainsi, bien que non obligatoire, la norme présente une importance non négligeable. Notamment sa reconnaissance comme étant représentative des usages et des règles de l'art est susceptible de lui conférer une présomption de conformité à la réglementation.

### **2.6.4.2 Modalités pratiques**

394. Les normes sont élaborées par des organismes dont les plus connus sont :

- au niveau international : l'ISO (International Organization for Standardization);
- au niveau européen : le CEN (Comité Européen de Normalisation)
- au niveau français : l'AFNOR (Association Française de Normalisation)

395. Il convient d'adresser à l'organisme de certification choisi, la demande de normalisation. Le bureau de certification va procéder à l'évaluation de l'opportunité de lancer des travaux dans le domaine envisagé. Selon le résultat de cette étude d'opportunité, il peut être décidé soit d'inscrire ces travaux dans une commission de normalisation existante, soit de créer un nouveau domaine d'activité de normalisation ou d'étendre un domaine d'activité existant.

396. Un premier document de travail sera rédigé par un groupe de travail ou par un chef de projet désigné à cet effet. Une enquête publique est ensuite lancée. Si la consultation

réalisée en vue d'obtenir l'accord final sur le document est positive, la mise au point définitive du texte peut être effectuée. L'homologation de la norme par l'organisme est alors prononcée.

397. La norme pourra contenir :

- le domaine d'application ;
- les définitions structurantes du domaine ;
- la démarche scientifique et les usages associés ;
- les conditions de publication des résultats ;
- le dépôt en archive ouverte (format du dépôt, métadonnées...) ;
- le respect de la paternité...

### **2.6.5 Un contrat type de cession de droits d'auteur : pour la protection des chercheurs**

398. Afin de garantir les droits des chercheurs sur leur publication et de prendre en compte les risques d'asymétrie contractuelle, un décret pourrait organiser un contrat type de cession de droit d'auteur destiné à la recherche publique.

399. Ce contrat définirait les règles du jeu entre les parties et la protection du chercheur dans sa relation avec l'éditeur. Il permettrait notamment de s'assurer de l'absence de cession à titre exclusif et de garantir les droits des chercheurs :

- d'autoriser le dépôt et la reproduction en archive ouverte de la publication dans la version auteur immédiatement et dans la version éditeur après le respect d'une période d'embargo ;
- de permettre l'exploration immédiate du contenu de l'article à partir d'outils numériques de traitement de données ;
- d'empêcher toutes formes de privatisation ou de réserve de propriété sur le contenu de l'article.

400. Ce contrat pourrait faire l'objet d'un décret et ainsi avoir une valeur réglementaire qui s'imposerait à l'éditeur pour toute publication scientifique constituant un résultat de la recherche publique.

## 2.6.6 Une charte éthique de la Science numérique

401. Le Comité d'éthique du CNRS soutenu par le Conseil Scientifique du CNRS soutient l'introduction d'une charte éthique de la Science numérique. Cette charte définirait les valeurs d'accès et de partage des données scientifiques ainsi que les bonnes pratiques des chercheurs telles que :

- le dépôt des données scientifiques sur des plateformes Open Science ;
- le respect des mentions de paternité.

402. Un comité d'éthique serait garant du respect de cette charte en veillant notamment à :

- la diffusion et à la compréhension de son contenu ;
- la sensibilisation des chercheurs à l'importance de l'éthique: « Les chercheurs et les personnels du monde de la recherche doivent être formés aux dimensions éthiques de la gestion des données, en particulier au respect de la vie privée, de la propriété intellectuelle, de la qualité et de l'intégrité des données. Ils doivent être informés de l'état actuel et de l'évolution des règles juridiques concernant le partage responsable de données utilisées »<sup>91</sup> ;
- formuler des avis assortis de recommandations afin de préciser les lignes de conduite définies dans la charte.

## 2.6.7 Une Agence de la Science ouverte

403. Une Agence de la Science ouverte pourrait être créée.

404. Elle aurait notamment pour rôles :

- le contrôle du respect des règles éthiques posées dans la Charte éthique ;
- le contrôle du respect du principe de libre accès ;
- un rôle consultatif ;
- le suivi des évolutions techniques et des usages ainsi que la gestion du répertoire des usages ;
- des propositions d'adaptation des cadres légaux existants au regard de l'évolution des pratiques et de besoins.

---

<sup>91</sup> Avis COMETS « Les enjeux éthiques du partage des données scientifiques » 7-6-2015

405. L'Agence pourrait également être en charge de la rédaction du rapport sur « l'impact du principe de libre accès aux données scientifiques sur le marché de l'édition scientifique et sur la circulation des idées et des données scientifiques » prévu par l'article 17 ter (nouveau) du projet de loi adopté par l'Assemblée nationale.

### **2.6.8 Une Convention internationale de la Science ouverte universelle**

406. La France pourrait être à l'initiative d'une Convention internationale de la Science ouverte universelle.

407. La notion de « convention internationale » est utilisée en droit international pour décrire des déclarations formelles de principes qui n'ont au départ pas de force obligatoire. Ces conventions doivent généralement être ratifiées par les États signataires de la convention pour obtenir une force obligatoire et ainsi devenir de véritables traités internationaux.

408. Compte tenu des positions déjà prises par l'UNESCO en faveur de l'Open Science, cette institution pourrait être l'organisme porteur de cette convention.

## **2.7 Synthèse des propositions pour une Science ouverte**

### **2.7.1 Avis stratégiques**

**Avis n°1 - Le numérique et la science : faire circuler les connaissances – Bruno DAVID, Président du Muséum national d'Histoire naturelle**

*Multiplicité des usages des communautés scientifiques*

Les [résultats de l'enquête](#) sur les usages et les besoins d'IST des Unités de recherche réalisée auprès des Directeurs des Instituts du CNRS et des Directeurs des Unités publiantes a permis de démontrer un continuum de pratiques entre les disciplines, sur un large éventail d'usages.

Les évolutions de l'Information scientifiques et techniques devront donc se définir à partir des orientations communes, sans opposer les communautés les unes aux autres, tout en esquissant des lignes directrices qui intègrent la diversité des pratiques et des usages.

*La gestion des publications à l'heure numérique*

Les équipes de recherche attendent, outre la gestion courante des publications qui entrent et sortent de leurs laboratoires, des outils d'analyse de contenu qui leur permettent de se projeter sur le front de la science. Tout d'abord par l'analyse des publications extérieures, pour explorer les avancées en cours, mais également par l'analyse des productions de leurs laboratoires pour l'aide à l'évaluation et au pilotage.

Si nous comparons les bases de publications scientifiques numériques à des bibliothèques classiques, là où il fallait des jours pour construire une synthèse, l'accès quasi-automatique et quasi-instantané aux contenus et la possibilité d'en extraire des signaux deviennent un formidable outil stratégique.

### *Exploration et partage des données de la recherche*

L'accès aux grandes bases de connaissance et la possibilité de fouille de texte et de données (Text & Data Mining) est un enjeu majeur à l'heure numérique. Le TDM s'apparente à l'utilisation des télescopes, par l'ampleur des données accessibles, mais aussi des microscopes par la précision atteinte.

La formalisation et la synthétisation des données liées aux articles scientifiques sont indispensables pour permettre d'interroger ces contenus le plus globalement possible, c'est à dire en ayant accès à la fois aux signaux tirés du contenu des publications et à l'ensemble des données les supportant.

C'est ici que se situe la plus grande marge de progrès : aujourd'hui les données sont accessibles au cas par cas, l'évolution numérique représenterait donc un grand pas en avant qui permettrait un accès global conduisant à rebondir vers de nouvelles questions et explorations.

Les données de la recherche sont un formidable capital commun. Elles doivent pouvoir être partagées sans contraintes, entre les communautés scientifiques.

Ces changements sont devant nous, la recherche doit pouvoir s'en saisir.

Bruno DAVID, Président du Muséum national d'Histoire naturelle

## **Avis 2 : De nouveaux usages pour partager les connaissances : la métrique des publications scientifiques – Daniel EGRET, Astronome, conseiller du Président de Paris Sciences et Lettres (PSL) sur la bibliométrie**

### *Mieux partager les connaissances par le numérique*

En peu de décennies, l'irruption du numérique a profondément modifié les pratiques et les enjeux du partage de l'information scientifique. Chaque laboratoire ou unité de recherche se doit maintenant de disposer d'un site Web ou d'un portail numérique ouvrant une fenêtre sur l'activité scientifique de l'unité. Bien souvent, ce portail se borne à fournir une présentation générale, des pointeurs, des contacts, mais la soif d'information de la communauté scientifique –et de la société en général, particulièrement flagrante pour les domaines scientifiques mobilisant le plus l'intérêt du public— est grande et ne se satisfera pas longtemps de ces timides ébauches. Il est devenu impératif de mieux éclairer nos collègues chercheurs et nos concitoyens sur notre activité de recherche, sur ses résultats (et en premier lieu les publications scientifiques, les réalisations instrumentales, les innovations, etc.). Impératif, également, d'organiser un accès plus large aux jeux de données produits par cette activité. Le numérique offre de larges gammes de solutions techniques, mais bien sûr le coût humain et la nécessaire mobilisation de ressources et de compétences restent les obstacles les plus difficiles à franchir.

### *Caractériser la production scientifique par la métrique*

Plus spécifiquement, dans le domaine des publications scientifiques, il s'agit non seulement de se donner les moyens de partager mieux encore la connaissance et les données de la recherche, mais aussi d'organiser la caractérisation de la production scientifique du laboratoire ou de l'institution, et de la rendre bien identifiable, visible et accessible, afin de pouvoir l'analyser, la comparer et la valoriser.

Les enjeux pour la recherche française sont d'importance puisqu'il s'agit là d'augmenter le rayonnement et l'impact de la recherche produite dans nos laboratoires. Fournir aux utilisateurs, quelle que soit leur fonction, des outils d'analyse de l'impact des publications, c'est les aider à formuler, avec des outils et services professionnels, les réponses aux défis, parfois perçus comme menaçants, de l'évaluation et de la comparaison internationale.

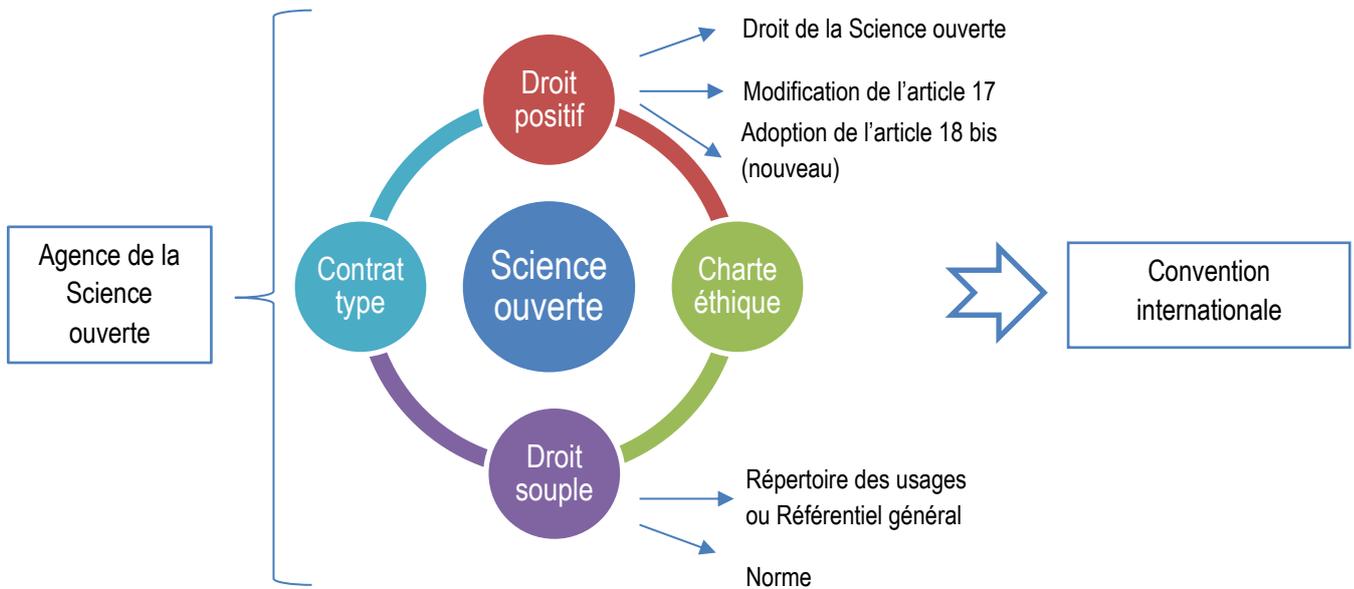
Il s'agit en outre de fournir aux institutions (organismes de recherche, universités de recherche, établissements) et aux instances d'évaluation les outils et services qui permettent de valoriser cette production scientifique, d'en évaluer l'impact sur la connaissance et sur la société, d'effectuer des comparaisons plus équitables.

Enfin, les nouveaux usages de la métrique doivent permettre non seulement de favoriser le partage et la diffusion des informations bibliographiques, mais aussi d'en exploiter les contenus pour mieux fonder notre compréhension des nouveaux processus de recherche et des pratiques en matière de production, utilisation et valorisation des résultats de la recherche. Les outils de fouille des données deviennent ainsi des instruments d'observation de nouvelle génération, au service de l'activité de recherche et d'innovation, donnant accès à de nouveaux mondes de résultats.

Daniel EGRET – Astronome, conseiller du Président de Paris Sciences et Lettres (PSL) sur la bibliométrie

## 2.7.2 Propositions

409. L'ensemble des propositions peut être résumé par le schéma suivant :



## Remerciements

**Les rédacteurs du Livre blanc remercient particulièrement et chaleureusement l'implication de :**

**La Direction de l'Information Scientifique et Technique du CNRS :**

- Renaud FABRE, Directeur DIST
- Laurence EL KHOURI, Directrice adjointe DIST
- Charlotte AUTARD, Chargée de projet ISTEEX

**La Direction des affaires juridiques du CNRS :**

- Nicolas CASTOLDI, Délégué général à la valorisation du CNRS précédemment Directeur des Affaires juridiques du CNRS

**Le Cabinet ALAIN BENSOUSSAN :**

- Alain BENSOUSSAN
- Laurence TELLIER-LONIEWSKI
- Sarah LENOIR

## Glossaire

### Article Processing Charges (APC) / Frais de publication

Les frais de publication se rapportent à la somme d'argent demandée par les éditeurs aux auteurs d'articles scientifiques pour que ceux-ci soient librement accessibles aux lecteurs. Deux types de revue sont concernés : les revues dont l'intégralité des articles est librement accessible et les revues dont une partie des articles est librement accessible (modèle hybride). (Source : Glossaire Inist)

### Archive institutionnelle

Une archive institutionnelle relève d'une institution (université, grande école, organisme de recherche, association professionnelle) et a pour objectif de contenir, valoriser et conserver l'ensemble de la production scientifique de celle-ci. (Source : Glossaire Inist)

### Archive ouverte

Le terme archive ouverte désigne un réservoir où sont déposées des données issues de la recherche scientifique et de l'enseignement et dont l'accès se veut ouvert c'est-à-dire sans barrière. Cette ouverture est rendue possible par l'utilisation de protocoles communs qui facilitent l'accessibilité de contenus provenant de plusieurs entrepôts maintenus par différents fournisseurs de données. (Source : Glossaire Inist)

### Données de la recherche

Ensemble des données ayant servi à l'établissement d'un résultat scientifique.

### Données de la Science

Ensemble des données de la recherche et des résultats scientifiques.

## Ecrit scientifique

Selon le Code de la propriété intellectuelle, Œuvre de l'esprit protégeable par le droit d'auteur (article L.112-2 CPI). Fait principalement référence aux ouvrages, articles scientifiques et actes de colloques ou de congrès, rapports.

## Gold Road / Voie dorée

La voie dorée s'applique à la publication d'articles dans des revues en libre accès, quel que soit leur mode de financement. Elle correspond à la deuxième stratégie recommandée dans l'Initiative de Budapest pour l'Accès Ouvert : « Revues alternatives : en second lieu, les savants ont besoin des moyens pour lancer une nouvelle génération de revues alternatives engagées dans le libre accès et pour aider les revues existantes qui choisissent d'opérer la transition vers l'accès libre. » (Source : Glossaire Inist)

## Green Road / Voie verte

La voie verte qualifie l'auto-archivage par les chercheurs ou l'archivage par une tierce personne des articles dans des archives ouvertes. Elle correspond à la première stratégie préconisée dans l'Initiative de Budapest pour l'Accès Ouvert : « Auto-archivage : en premier lieu, les savants ont besoin d'outils et d'assistance pour déposer leurs articles de revues à comité de lecture dans des archives électroniques ouvertes, une pratique communément appelée auto-archivage. » (Source : Glossaire Inist)

## Information Scientifique et Technique (IST)

L'information scientifique et technique (IST) regroupe l'ensemble des informations produites par la recherche et nécessaires à l'activité scientifique comme à l'industrie. De par sa nature, l'IST couvre tous les secteurs scientifiques et techniques et se présente sous de multiples formes : articles, revues et ouvrages scientifiques, spécifications techniques décrivant des processus de fabrication, documentation technique accompagnant les produits, notices de brevet, bases de données bibliographiques, littérature grise, banques de données brutes,

archives ouvertes et entrepôts de données accessibles sur Internet, portails, etc.»<sup>92</sup>.

## Métadonnées

Les métadonnées sont un ensemble de données structurées décrivant des ressources physiques ou numériques. Elles sont un maillon essentiel pour le partage de l'information et l'interopérabilité des ressources électroniques. Elles sont classiquement divisées en métadonnées descriptives, administratives ou de structure. (Source : Glossaire Inist)

## Modèle auteur-payeur

On parle du modèle auteur-payeur quand l'auteur ou son institution d'appartenance ou le bailleur de fonds donne une contribution à l'éditeur pour rendre l'article librement et gratuitement accessible à tout lecteur. Il est opposé aux modèles lecteur-payeur et sponsor-payeur. (Source : Glossaire Inist)

## Modèle hybride

Dans une revue, il peut coexister deux types d'articles : ceux qui sont librement accessibles — contre une redevance versée à l'éditeur par l'auteur ou son financeur (voir : modèle auteur-payeur) — et ceux qui sont accessibles par un abonnement. Ce système est appelé modèle hybride. (Source : Glossaire Inist)

## Modèle lecteur payeur

Le modèle lecteur-payeur correspond au modèle traditionnel de l'édition, l'abonnement. Le lecteur ne peut avoir accès qu'aux revues et aux ouvrages pour lesquels il a pris, mais le plus souvent il s'agit de son institution, un abonnement auprès d'un ou plusieurs éditeurs. Il est opposé aux modèles auteur-payeur et sponsor-payeur. (Source : Glossaire Inist)

---

<sup>92</sup><http://www.enseignementsup-recherche.gouv.fr/cid20438/les-missions-de-l-information-scientifique-et-technique.html>

## Open access

Désigne l'accès permanent et gratuit pour le lecteur, sur internet, aux données issues de la recherche scientifique et de l'enseignement (Source : Invisu Inha)

## Open data

Désigne Données qu'un organisme met à la disposition de tous sous forme de fichiers numériques afin de permettre leur réutilisation.

### Notes

1. Les données ouvertes n'ont généralement pas de caractère personnel
2. Elles sont accessibles dans un format favorisant leur réutilisation.
3. La réutilisation des données ouvertes peut être soumise à conditions (Source : Vocabulaire de l'informatique et du droit JORF n°0103 du 3 mai 2014 page 7639)

## Open process

Désigne le droit de procéder librement à l'observation de données par l'utilisation d'outils numériques de traitement, d'analyse ou d'exploration.

## Open Science

Désigne l'accès permanent et gratuit sur internet aux données issues de la recherche scientifique et de l'enseignement ainsi que le droit de procéder librement à l'observation de ces données par l'utilisation d'outils numériques de traitement, d'analyse ou d'exploration. (Open Science = Open access + Open process)

## Peer review / Evaluation par les pairs

L'évaluation par les pairs désigne la validation d'un article par un comité de lecture composé de scientifiques, experts dans le même champ disciplinaire que le contenu de l'article. Ce processus est destiné à lui assurer une qualité scientifique. (Source : Glossaire Inist)

## **Postprint / Post publication / Version acceptée**

La post publication est la version définitive d'un manuscrit produit par un (des) auteur (s) après révision par les pairs, comportant les modifications apportées par les pairs mais sans la mise en forme de l'éditeur. (Source : Glossaire Inist).

## **Preprint / Prépublication/ Version soumise**

La prépublication désigne les versions d'un texte produit par un (des) auteur(s) avant acceptation par un comité de rédaction et éventuellement par un comité de lecture (évaluation par les pairs). (Source : Glossaire Inist)

## **Recherche publique**

La recherche publique est organisée dans les services publics, notamment les établissements publics d'enseignement supérieur, les établissements publics de recherche et les établissements de santé, et dans les entreprises publiques. (article L.112-2 Code de la recherche)

## **Résultat de la recherche**

Ecrit scientifique ou données quelconques issus d'une activité de recherche et produits à partir des données de la recherche. Ce résultat de la recherche peut faire l'objet d'une publication (« résultat publié ») ou non (« résultat non publié »).

## **Résultat de la recherche publique**

Résultat issu de la recherche publique ou d'une recherche financée au moins pour moitié par des fonds publics

## ***Text and data mining (TDM)***

Technique de traitement automatique de la connaissance

## Version éditeur

La version éditeur est la version finale, publiée, d'un manuscrit produit par un (des) auteur (s), après évaluation par les pairs et avec la mise en forme de l'éditeur. (Source : Glossaire Inist)

## Annexes

### 1. Présentation du Livre blanc

#### **Le Livre blanc : Une démarche de la recherche publique pour la recherche publique**

410. A l'heure où le projet de loi « pour une République numérique » propose l'insertion dans le Code de la recherche de dispositions relatives à l'Open Access, le Centre national de la recherche scientifique (CNRS), aux côtés de ses partenaires dans le projet ISTEEX ainsi que d'un grand nombre de chercheurs et d'acteurs du domaine de la recherche publique, propose par ce Livre blanc le fruit de ses réflexions et analyses.

411. Depuis plusieurs années, la communauté scientifique de la recherche publique avance la nécessité de créer un cadre juridique et organisationnel pour l'accès aux données et informations scientifiques et techniques dans le monde numérique, en particulier aux données issues de sa propre activité de recherche.

412. Ce Livre blanc rend compte de ces réflexions, des pratiques des chercheurs quant à l'utilisation de l'information scientifique et technique et des outils numériques. L'ensemble des propositions formulées pour la création d'une Science ouverte est le fruit d'un travail commun et de témoignages forts du monde de la recherche (1.3). La genèse (1.1), les objectifs (1.2) et la démarche de réalisation (1.4) de ce Livre blanc sont présentés ci-après.

#### **1.1 Genèse du Livre blanc**

413. Le projet de rédaction de ce Livre blanc est né à l'occasion des réflexions sur la sécurisation du projet de plateforme ISTEEX (1.1.1) et du constat de départ selon lequel le modèle économique de l'édition scientifique, secteur où les prix ont augmenté de manière considérable, n'est plus viable en l'état pour les organismes d'enseignement et de recherche (1.1.2).

414. Par ailleurs, la revendication d'Open Science s'inscrit pleinement dans le mouvement international de l'Open favorable au partage de la connaissance scientifique (1.1.3) et dans une ambition française déjà affirmée à plusieurs reprises (1.1.4).

### 1.1.1 L'investissement d'avenir ISTE<sup>X</sup> : première plateforme d'accès à la Science ouverte

415. **ISTEX : une plateforme numérique multi-usages.** ISTE<sup>X</sup>, Initiative d'excellence de l'Information Scientifique et Technique, est un projet de plateforme numérique multi-usages (base de données de bases de données) aux meilleurs standards internationaux, accessible à distance par l'ensemble des communautés scientifiques et offrant « tous les moyens accessibles de consultation et d'analyse aujourd'hui disponibles dans toutes les communautés de la Science »<sup>93</sup>. Cette base de données de bases de données ambitionne :

- de donner librement et gratuitement accès aux chercheurs à l'ensemble de l'Information Scientifique et Technique (IST) mondiale, contenue dans les archives et les collections courantes ;
- de fournir aux chercheurs des services à haute valeur ajoutée de traitement de la connaissance et des données scientifiques et techniques.

416. Le ministère de l'Enseignement supérieur et de la recherche donne la définition suivante de l'IST :

- « L'information scientifique et technique (IST) regroupe l'ensemble des informations produites par la recherche et nécessaires à l'activité scientifique comme à l'industrie. De par sa nature, l'IST couvre tous les secteurs scientifiques et techniques et se présente sous de multiples formes : articles, revues et ouvrages scientifiques, spécifications techniques décrivant des processus de fabrication, documentation technique accompagnant les produits, notices de brevet, bases de données bibliographiques, littérature grise, banques de données brutes, archives ouvertes et entrepôts de données accessibles sur Internet, portails, etc ... »<sup>94</sup>.

417. Dans le cadre de la sécurisation juridique du projet d'investissement d'avenir ISTE<sup>X</sup> (ANR-10-IDEX-0004-02 - [www.istex.fr](http://www.istex.fr)), il est apparu des lacunes et des vides juridiques.

418. **Les problématiques.** L'analyse du cadre technique, économique et juridique d'ISTEX a permis de mettre en évidence plusieurs problématiques auxquelles le droit positif n'offrait pas de réponses satisfaisantes au regard des besoins de la Science :

---

<sup>93</sup> <http://www.istex.fr/>

<sup>94</sup> <http://www.enseignementsup-recherche.gouv.fr/cid20438/les-missions-de-l-information-scientifique-et-technique.html>

- le modèle économique et juridique de l'édition scientifique ne correspond plus au modèle technique des plateformes numériques ;
- l'accès et le partage des données scientifiques comme outils de travail des communautés scientifiques se heurtent aux clauses de cession exclusive des droits de propriété intellectuelle, ainsi qu'aux principes du droit d'auteur et du droit des bases de données ;
- la plateforme ISTEEX comporte des services à valeur ajoutée parmi lesquels la pratique du « *text and data mining* » (fouille de texte et de données) : cette pratique permet aux chercheurs d'utiliser des outils de recherche intelligente, de croisement de données, d'exploration, de recherche transdisciplinaire. Cette pratique n'est pas encadrée légalement et vient se heurter, sur certains aspects, au droit d'auteur et au droit des bases de données.

419. **Les défis.** ISTEEX s'inscrit plus globalement dans les deux défis de l'IST de l'ère numérique tels que rappelés par le CNRS dans sa stratégie ouverte pour une IST d'avenir <sup>95</sup> :

- « ouvrir les conditions d'accès » à l'IST ;
- « offrir une réponse à tous les besoins », notamment pour tenir compte des pratiques différenciées selon les communautés scientifiques.

420. C'est dans ce cadre, face à ces constats, notamment de l'inadaptation du modèle juridico-économique de l'édition scientifique, et à ces défis, qu'est née l'idée de rédiger un Livre blanc recensant les besoins des acteurs de la science afin de faire évoluer la législation en vigueur.

### 1.1.2 L'impératif : faire évoluer les modèles économiques de la Science numérique

421. **Les modèles existants.** Coexistent dans l'édition scientifique numérique plusieurs modèles économiques dont les caractéristiques sont décrites précisément par le site [www.openaccess.inist.fr](http://www.openaccess.inist.fr) et dont les principes sont repris ci-après :

- le modèle de l'auteur-payeur : « lorsque l'auteur ou son institution d'appartenance ou le bailleur de fonds donne une contribution [Article Processing Charge - APC] à l'éditeur pour rendre l'article librement et gratuitement accessible à tout lecteur » ;
- le modèle du lecteur-payeur : « modèle traditionnel de l'édition, l'abonnement. Le lecteur ne peut avoir accès qu'aux revues et aux ouvrages pour lesquels il a pris,

---

<sup>95</sup> « Une stratégie ouverte pour une information scientifique et technique d'avenir » CNRS page 8

mais le plus souvent il s'agit de son institution, un abonnement auprès d'un ou plusieurs éditeurs » ;

- le modèle sponsor-payeur : « le financement de la revue est pris en charge par une société savante, un organisme de recherche, une fondation... » ;
- le modèle hybride : certains éditeurs rendent librement accessibles les articles publiés dans leurs revues contre une redevance versée par l'auteur ou son financeur (modèle auteur-payeur) ; les lecteurs devant payer un abonnement pour avoir accès aux revues ou ouvrages (lecteur-payeur)<sup>96</sup> ;
- la « Green road » : » : la voie verte concerne l'auto-archivage et les dépôts institutionnels, centralisés (comme HAL en France), thématiques (comme Arxiv en physique), permettant l'accès gratuit et l'usage des articles scientifiques, lors, ou peu après, la publication dans une revue arbitrée par les pairs.

422. **Modèle hybride critiqué.** Aujourd'hui, le modèle le plus répandu en pratique est le modèle hybride engendrant un double paiement, le plus souvent du laboratoire. Il peut être résumé ainsi :



423. Les dérives de ce modèle hybride, auteur-lecteur-payeur, ont largement été relayées par les scientifiques eux-mêmes :

- la naissance d'éditeurs prédateurs qui « ont pollué le système mondial de l'édition scientifique en profitant de l'Open access pour faire la publicité de pseudo-revues scientifiques »<sup>97</sup> ;

<sup>96</sup> Libre accès à l'IST – Inist CNRS : <http://openaccess.inist.fr/>

<sup>97</sup> <http://sciences.blogs.liberation.fr/home/2013/10/open-access-du-r%C3%AAve-au-cauchemar-bis.html>

- le paiement par l'auteur d'un APC élevé : l'étude « Developing an effective market for Open Access APC » montre que les APC les plus élevés sont ceux du modèle hybride, de l'ordre de 2727 \$ (2328 €) par article<sup>98</sup> ;
- les trois sociétés savantes françaises de mathématiques (Société Française de Statistique (SFdS), Société de Mathématiques Appliquées et Industrielles (SMAI), Société Mathématique de France (SMF)) ont publié dès septembre 2012 une déclaration intitulée « Open Access : mise en garde et effets pervers du système auteur-payeur » ;
- l'avis du Comité d'éthique du CNRS du 29 juin 2012 sur « Le libre accès aux publications scientifiques (« Open access ») » met en garde contre les dangers de ce modèle auteur-payeur ;
- le paiement par le lecteur d'un abonnement en continuelle augmentation :
  - o « les établissements d'enseignement supérieur et de recherche dépensent annuellement plus de 80 millions d'euros pour avoir accès aux ressources électroniques. Les prix d'accès ont d'ailleurs continuellement augmenté : de 7% par an depuis 10 ans »<sup>99</sup> ;
  - o des appels au boycott des grands éditeurs ont déjà été relayés par les universités françaises en 2012, mettant en cause le prix des abonnements aux revues scientifiques que payent les bibliothèques des universités. « Elles dépenseraient souvent plus de la moitié de leur budget dans ces achats à trois grands éditeurs commerciaux : Elsevier, Springer et Wiley »<sup>100</sup>.

424. Par conséquent, la marche vers l'Open Science s'accompagne d'une réflexion plus générale :

- « sur le partage des valeurs dans la chaîne éditoriale, sur les marges associées à l'activité des groupes globaux, sur les modèles économiques optimaux de publication »<sup>101</sup>.

---

<sup>98</sup> <http://www.cnrs.fr/dist/z-outils/documents/Distinfo2/Distinfo4.pdf>

<sup>99</sup> [http://www.lemonde.fr/idees/article/2015/09/10/favorisons-la-libre-diffusion-de-la-culture-et-des-savoirs\\_4751847\\_3232.html](http://www.lemonde.fr/idees/article/2015/09/10/favorisons-la-libre-diffusion-de-la-culture-et-des-savoirs_4751847_3232.html)

<sup>100</sup> [http://www.lemonde.fr/sciences/article/2012/04/25/harvard-rejoint-les-universitaires-pour-un-boycott-des-editeurs\\_1691125\\_1650684.html](http://www.lemonde.fr/sciences/article/2012/04/25/harvard-rejoint-les-universitaires-pour-un-boycott-des-editeurs_1691125_1650684.html)

<sup>101</sup> Schéma d'orientation stratégique de l'IST page 13 « Les modèles économiques de l'Édition »

### 1.1.3 Un contexte international largement ouvert au partage de la connaissance

425. Ce Livre blanc s'inscrit dans un contexte international favorable au partage de la connaissance scientifique.

426. L'évolution vers le libre accès est un phénomène mondial. La fédération européenne des académies nationales de sciences et de sciences humaines a adopté en avril 2012 une déclaration intitulée « *Open Science for the 21st century* » qui prône le partage des résultats et des outils de recherche.

427. L'accessibilité des données de la recherche fait également l'objet de débats dans de nombreuses enceintes internationales, dont l'OCDE et l'UNESCO.

428. Les exemples se sont également multipliés au niveau national, les Etats intégrant dans leur législation des dispositions favorisant l'Open Access et / ou le *text and data mining*.

429. Les discours politiques français s'inscrivent également dans cette tendance.

### 1.1.4 Une nouvelle ambition pour la France

430. **Origine.** Le débat sur l'Open access des données scientifiques, né dans les années 2000, a connu un regain de force et d'envergure en France ces derniers mois dans le cadre du projet de loi pour une République numérique autour de deux principaux sujets :

- la nécessité de mettre en ligne les publications scientifiques ainsi que les données sous-jacentes à l'hypothèse scientifique ;
- la nécessité de permettre aux chercheurs de procéder à des traitements des données, à de la fouille de texte : le *text and data mining* (TDM).

431. **Discours janvier 2013.** Lors des discours des 5<sup>èmes</sup> journées Open access sur le thème « Généraliser l'accès ouvert aux résultats de la recherche » (les 24 et 25 janvier 2013), Madame Geneviève FIORASO, alors ministre de l'Enseignement supérieur et de la Recherche, avait déjà présenté le principe et les enjeux de l'Open Science en déclarant que :

- « L'information scientifique est un bien commun qui doit être disponible pour tous ».

432. **Projet de loi de finances pour 2014.** Par ailleurs, l'annexe au projet de loi de finances pour 2014 intitulé « Rapport sur les politiques nationales de recherche et de

formations supérieures » prévoit un point 8 sur l'information scientifique et technique et les réseaux documentaires mentionnant notamment :

- « le développement du libre accès (Open access) aux publications scientifiques ».

**433. Stratégie numérique du Gouvernement.** Dans la stratégie numérique du Gouvernement du 18 juin 2015, il est par ailleurs clairement affirmé comme mesure emblématique du plan numérique l'action de « Favoriser une science ouverte par la libre diffusion des publications et des données de la recherche ».

**434.** Le texte précise :

- « Afin que notre recherche soit toujours plus compétitive dans le concert mondial, la France accentue son engagement dans l'ouverture des publications et des données de la recherche financées sur fonds publics » ;
- « La libre circulation de la science et sa libre exploitation contribue en outre à l'innovation, encourage la collaboration, améliore la qualité des publications, évite la duplication des efforts, permet l'exploitation des résultats de recherche antérieurs et favorise la participation des citoyens et de la société civile » ;
- « L'accès libre aux données de la recherche, dont les modalités font l'objet de travaux suivis, constituera le prolongement de l'accès libre aux publications »<sup>102</sup>.

**435. Les versions de l'avant-projet de loi.** Plusieurs versions de l'avant-projet de loi pour une République numérique ont été dévoilées avant la version officielle objet de la consultation publique.

**436.** Une première version de l'avant-projet de loi relatif à l'ambition numérique de la France, dont le texte était disponible en ligne le 21 juillet 2015, prévoyait :

- une section 3 « Libre accès aux travaux de recherche » créant un droit de mise à disposition publique des contributions scientifiques financées au moins pour moitié par les fonds publics après le respect d'une période d'embargo :

---

<sup>102</sup> « Stratégie Numérique du Gouvernement » Premier Ministre - Gaité Lyrique - Jeudi 18 juin 2015

### Article 39

Il est créé dans le code de la propriété intellectuelle un article L. 132-8-1 ainsi rédigé :

« Art. L. 132-8-1. – L'auteur d'une contribution scientifique, issue d'une activité de recherche financée au moins pour moitié par des fonds publics et publiée dans le cadre d'une collection paraissant au moins une fois par an, dispose du droit, même s'il a cédé un droit d'exploitation exclusif à l'éditeur, de rendre publiquement accessible la version acceptée de son manuscrit, au terme d'un délai de six mois pour les sciences et de douze mois pour les Sciences humaines et sociales à compter de la première publication, toute fin commerciale étant exclue.

- une section 4 « Exceptions de fouille de texte et de données et de panorama » permettant l'exploration de textes et de données pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale :

I. - L'article L. 122-5 du code de la propriété intellectuelle est ainsi modifié :

1° Après le neuvième alinéa, il est inséré un alinéa ainsi rédigé :

« f) Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de destruction des fichiers au terme des activités de recherche pour lesquelles elles ont été produites ; »

2° Après le vingt-et-unième alinéa, il est inséré un alinéa ainsi rédigé :

« 10° les reproductions et représentations, intégrales ou partielles, à l'exclusion de toute finalité commerciale, d'œuvres architecturales ou de sculptures, réalisées pour être placées en permanence dans des lieux publics. »

II. - Après le sixième alinéa de l'article L. 342-3 du même code, il est inséré un alinéa ainsi rédigé :

« 5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données dans un cadre de recherche, à l'exclusion de toute finalité commerciale. Ces copies et reproductions sont assurées par un organisme désigné par décret, qui garantit la destruction des fichiers au terme des activités de recherche pour lesquelles elles ont été produites. »

437. La version 2 de l'avant-projet de loi de septembre 2015 proposait l'insertion dans le Code de la recherche (au chapitre III « La valorisation des résultats de la recherche par les établissements et organismes de recherche de dispositions ») d'un article sur l'Open access.

**Article 11 (39)**  
*Open access*  
*(arbitrage politique nécessaire)*

Au chapitre 3 du titre 3 du livre V du code de la recherche, il est inséré un article L. 533-4 ainsi rédigé :

« I. – Les droits d'exploitation sous une forme numérique d'un écrit scientifique, issu d'une activité de recherche financée au moins pour moitié par des fonds publics, sont cessibles à titre exclusif à un éditeur, dans les conditions mentionnées à la section première du chapitre II du titre III du Livre 1er du code de la propriété intellectuelle.

II. Lorsque un écrit scientifique est publié dans un périodique, un ouvrage paraissant au moins une fois par an, des actes de congrès ou de colloques ou des recueils de mélanges, son auteur, même en cas de cession exclusive à un éditeur, dispose du droit de mettre à disposition gratuitement sous une forme numérique, sous réserve des droits des éventuels coauteurs, la dernière version acceptée de son manuscrit par son éditeur et à l'exclusion du travail de mise en forme qui incombe à ce dernier, au terme d'un délai de douze mois pour les sciences, la technique et la médecine et de vingt-quatre mois pour les sciences humaines et sociales, à compter de date de la première publication. Cette mise à disposition ne peut donner lieu à aucune exploitation commerciale.

« III. – Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. Elles ne s'appliquent pas aux contrats en cours. »

438. Cette version 2 de l'avant-projet de loi :

- propose la création d'un droit de dépôt dans le Code de la recherche plutôt que dans le Code de la propriété intellectuelle ;
- multiplie par deux les délais d'embargo par rapport à la première version ;
- supprime purement et simplement l'article sur le TDM.

439. **Versions officielles.** Le tableau ci-après recense les textes officiels du projet de loi pour une République numérique et leurs évolutions :

- le texte mis en ligne pour consultation publique à partir du 26 septembre 2015 ;
- le texte issu de la contribution publique tel qu'adressé au Conseil d'Etat le 6 novembre 2015 ;
- la version finale du texte adoptée par le Conseil des ministres le 9 décembre 2015 ;
- le texte adopté par l'Assemblée nationale le 26 janvier 2016.

<b>Texte objet de la consultation publique</b> <b>26-9-2015</b>	<b>Texte issu de la consultation publique</b> <b>transmis au Conseil d'Etat</b> <b>6-11-2015</b>	<b>Texte adopté en Conseil des ministres</b> <b>9-12-2015</b>	<b>Texte adopté par l'Assemblée nationale</b> <b>26-1-2016</b>
<p><b>Article 9 - Libre accès aux publications scientifiques de la recherche publique</b></p> <p>Au chapitre 3 du titre 3 du livre V du code de la recherche, il est inséré un article L. 533-4 ainsi rédigé :</p> <p>« Art. L. 533-4 –                      I. Lorsque un écrit scientifique, issu d'une activité de recherche financée au moins pour moitié par des fonds publics, est publié dans un périodique, un ouvrage paraissant au moins une fois par an, des actes de congrès ou de colloques ou des recueils de mélanges, son auteur, même en cas de cession exclusive à un éditeur, dispose du droit de mettre à disposition gratuitement sous une forme numérique, sous réserve des droits des éventuels coauteurs, la dernière version acceptée de son manuscrit par son éditeur et à l'exclusion du travail de mise en forme qui incombe à ce dernier, au terme d'un délai de douze mois pour les sciences, la technique et la médecine et de vingt-quatre mois pour les sciences humaines et sociales, à compter de date de la première publication. Cette mise à disposition ne peut donner lieu à aucune exploitation commerciale.</p> <p>« II. – Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. Elles ne s'appliquent pas aux contrats en cours. »</p>	<p><b>Article 14</b></p> <p>A la fin du chapitre III du Titre III du Livre V du code de la recherche, il est ajouté un article L. 533-4 ainsi rédigé :</p> <p>« Art. L. 533-4. –                      I. - Lorsque un écrit scientifique, issu d'une activité de recherche financée au moins pour moitié par des fonds publics, est publié dans un périodique, un ouvrage paraissant au moins une fois par an, des actes de congrès ou de colloques ou des recueils de mélanges, son auteur, même en cas de cession exclusive à un éditeur, dispose du droit de mettre à disposition gratuitement sous une forme numérique, sous réserve des droits des éventuels coauteurs, la version finale du manuscrit acceptée pour publication, au plus tard six mois pour les sciences, la technique et la médecine et douze mois pour les sciences humaines et sociales à compter de la date de la première publication, ou au plus tard lorsque l'éditeur met lui-même l'écrit gratuitement à disposition sous une forme numérique.</p> <p>Il est interdit d'exploiter la mise à disposition permise au titre du premier alinéa dans le cadre d'une activité d'édition à caractère commercial.</p> <p>« II. – Les données de la recherche rendues publiques légalement issues d'une activité de recherche financée au moins pour moitié par des fonds publics et qui ne sont pas protégées par un droit spécifique sont des choses communes, au sens de l'article 714 du code civil.</p>	<p><b>Article 17</b></p> <p>A la fin du chapitre III du Titre III du Livre V du code de la recherche, il est ajouté un article L. 533-4 ainsi rédigé :</p> <p>« Art. L. 533-4. –                      I. - Lorsqu'un écrit scientifique, issu d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne, est publié dans un périodique paraissant au moins une fois par an, dans des actes de congrès ou de colloques ou des recueils de mélanges, son auteur dispose, même en cas de cession exclusive à un éditeur, du droit de mettre à disposition gratuitement sous une forme numérique, sous réserve des droits des éventuels coauteurs, la version finale du manuscrit acceptée pour publication, dès lors que l'éditeur met lui-même l'écrit gratuitement à disposition sous une forme numérique, et, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication. Ce délai est de six mois pour les sciences, la technique et la médecine, et de douze mois pour les sciences humaines et sociales.</p> <p>« Il est interdit d'exploiter la mise à disposition permise au titre du premier alinéa dans le cadre d'une activité d'édition à caractère commercial.</p> <p>« II. - Dès lors que les données issues d'une activité de recherche, financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de</p>	<p><b>Article 17</b></p> <p>Le chapitre III du titre III du livre V du code de la recherche est complété par un article L. 533-4 ainsi rédigé :</p> <p>« Art. L. 533-4. –                      I. – Lorsqu'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations de l'État, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne est publié dans un périodique paraissant au moins une fois par an, son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique, sous réserve de l'accord des éventuels coauteurs, toutes les versions successives du manuscrit jusqu'à la version finale acceptée pour publication, dès lors que l'éditeur met lui-même celle-ci gratuitement à disposition par voie numérique et, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication. Ce délai est de six mois pour une publication dans le domaine des sciences, de la technique et de la médecine et de douze mois dans celui des sciences humaines et sociales. Un délai inférieur peut être prévu pour certaines disciplines, par arrêté du ministre chargé de la recherche.</p> <p>« La version mise à disposition en application du premier alinéa ne peut faire l'objet d'une exploitation dans le cadre d'une activité d'édition à caractère commercial.</p> <p>« II. – Dès lors que les données issues d'une</p>

Texte objet de la consultation publique 26-9-2015	Texte issu de la consultation publique transmis au Conseil d'Etat 6-11-2015	Texte adopté en Conseil des ministres 9-12-2015	Texte adopté par l'Assemblée nationale 26-1-2016
	<p>« III. - L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.</p> <p>« IV. - Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite.»</p>	<p>financement nationales ou par des fonds de l'Union européenne, ne sont pas protégées par un droit spécifique, ou une réglementation particulière, et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre.</p> <p>« III. - L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.</p> <p>« IV. - Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. »</p>	<p>activité de recherche financée au moins pour moitié par des dotations de l'État, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre.</p> <p>« III. – L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.</p> <p>« IV. – Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. »</p>
			<p><b>Article 17 ter (nouveau)</b></p> <p>Le Gouvernement remet au Parlement, au plus tard deux ans après la promulgation de la présente loi, un rapport qui évalue les effets de l'article L. 533-4 du code de la recherche sur le marché de l'édition scientifique et sur la circulation des idées et des données scientifiques françaises.</p>
			<p><b>Article 18 bis (nouveau)</b></p> <p>Le code de la propriété intellectuelle est ainsi modifié :</p> <p>1° Après le second alinéa du 9° de l'article L. 122-5, il est inséré un 10° ainsi rédigé :</p> <p>« 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données pour les</p>

Texte objet de la consultation publique 26-9-2015	Texte issu de la consultation publique transmis au Conseil d'Etat 6-11-2015	Texte adopté en Conseil des ministres 9-12-2015	Texte adopté par l'Assemblée nationale 26-1-2016
			<p>besoins de la recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche ; »</p> <p>2° Après le 4° de l'article L. 342-3, il est inséré un 5° ainsi rédigé :</p> <p>« 5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données dans un cadre de recherche, à l'exclusion de toute finalité commerciale. La conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche pour lesquelles elles ont été produites, sont assurées par des organismes désignés par décret. Les autres copies ou reproductions sont détruites. »</p>

## 1.2 Objectif du Livre blanc : un objectif spécifique à la recherche publique

440. Ce Livre blanc a pour objectif de présenter les besoins des chercheurs publics dans leur activité de recherche et de proposer un cadre juridique de nature à faire de la recherche publique française une recherche compétitive dotée d'un arsenal législatif ambitieux et précurseur :

- en favorisant l'accès aux données et aux résultats scientifiques et leur réutilisation;
- en encadrant juridiquement de manière à les sécuriser les pratiques et situations de fait existantes et nécessaires aux communautés scientifiques de la recherche publique ;
- en prenant en compte les impératifs de valorisation de l'innovation ;
- en rétablissant une situation d'équilibre avec les éditeurs scientifiques.

441. Les objectifs de ce Livre blanc s'inscrivent dans le respect des droits de propriété intellectuelle des auteurs tels qu'établis par le Code de la propriété intellectuelle.

## 1.3 Les grands témoins

442. Afin d'identifier et de dresser un panorama des pratiques et des besoins des chercheurs dans le cadre d'une Science numérique, des auditions de grands témoins, basées sur un guide d'entretien, ont été menées. Les comptes rendus de ces auditions ainsi que le guide d'entretien sont annexés au présent Livre blanc.

443. Les grands témoins auditionnés ont apporté une contribution essentielle au présent Livre blanc.

444. Ce Livre blanc est le fruit de partage, de réflexion commune, d'audition et de travail collaboratif, depuis plus d'un an, sur l'Open access et l'Open process par ces grands témoins de la recherche et pour la recherche scientifique.

### 1.3.1 Universités et LERU

445. Ont été auditionnées des personnalités du monde universitaire, présidents et représentants d'universités et membres de la *League of European Research Universities* (Leru) :

- Alain BERETZ, Président de l'Université de Strasbourg et Président de la LERU
- Jean CHAMBAZ, Président de l'UPMC et président de la CURIF (Coordination des Universités de recherche intensive françaises)
- Françoise CURTIT, CNRS, Chargée de mission « Science ouverte » à l'Université de Strasbourg ;
- Jean-Pierre FINANCE, Président du Consortium Couperin, Délégué permanent pour la CPU à Bruxelles, ex-président de l'Université Nancy 1 ;
- Paul-Antoine HERVIEUX, Vice-président délégué Partenariats avec les EPST et les collectivités, à l'Université de Strasbourg ;
- Paul INDELICATO, Vice-président Recherche et innovation de l'UPMC.

### 1.3.2 Conseil scientifique du CNRS

446. **Missions.** Le Conseil Scientifique du CNRS veille à la cohérence de la politique scientifique du CNRS en liaison avec l'ensemble des instances scientifiques consultatives du Comité national de la recherche scientifique (CoNRS). Il donne notamment un avis sur :

- les grandes orientations de la politique scientifique du CNRS ;
- les principes communs d'évaluation de la qualité des recherches et de l'activité des chercheurs.

447. Par ailleurs, le décret organique du CNRS, modifié par le décret n°2015-1151 du 16 septembre 2015<sup>103</sup> prévoit que dans le cadre de la politique scientifique définie par le Gouvernement, en relation avec les besoins culturels, économiques et sociaux de la nation et en liaison avec les établissements d'enseignement supérieur et de recherche, le CNRS est en charge des missions suivantes :

- « participer à l'analyse de la conjoncture scientifique nationale et internationale et de ses perspectives d'évolution en vue de l'élaboration de la politique nationale dans ce domaine ;
- assurer l'élaboration et la diffusion de la documentation scientifique et la publication des travaux et données de la recherche, notamment en mettant à disposition de la communauté scientifique et universitaire des plates-formes documentaires et en contribuant à leur enrichissement. »

---

<sup>103</sup> Décret n°82-993 du 24 novembre 1982 portant organisation et fonctionnement du Centre national de la recherche scientifique modifié par le décret n°2015-1151 du 16 septembre 2015

448. Le décret du 16 septembre 2015 a ainsi érigé en mission nationale du CNRS la diffusion et l'enrichissement de la documentation scientifique et technique notamment par l'intermédiaire de l'outil numérique de la plateforme.

449. C'est au regard de cette mission nationale que le CNRS a initié et piloté la rédaction de ce Livre blanc.

450. **Groupe de travail.** Dans le cadre de ses missions et après audition de Renaud FABRE, le Conseil Scientifique du CNRS s'est saisi des sujets présentés et a proposé un groupe de travail composé des personnes suivantes :

- Bruno CHAUDRET, Président du Conseil scientifique du CNRS, Directeur de recherches CNRS, Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique ;
- Pierre BINETRUY, Physicien, Directeur du Laboratoire Astroparticule et Cosmologie, professeur à l'Université Paris 7 ;
- François BONNAREL, Ingénieur CNRS, Centre de données astronomiques de Strasbourg ;
- Claire LEMERCIER, Directrice de recherche au CNRS en histoire, Centre de sociologie des organisations, Paris ;
- Sophie POCHIC, Responsable de l'équipe PRO (Professions, Organisations, Réseaux), Centre Maurice Halbwachs ;
- François TRONCHE, Directeur de recherche CNRS, Institut de Biologie Paris-Seine

451. La réflexion conduite par ce groupe de travail, autour de la sécurisation des pratiques et des besoins des chercheurs, a abouti à deux documents annexés au présent Livre blanc :

- une contribution acceptée par l'ensemble du Conseil Scientifique et à laquelle le présent Livre blanc fait de nombreuses fois référence ;
- une recommandation unanime sur l'Open Science.

### 1.3.3 Comité exécutif ISTEEX

452. Ce Livre blanc est né de l'initiative et d'une réflexion commune des membres du Comité exécutif du projet ISTEEX sous l'impulsion de Renaud FABRE, Directeur de l'Information Scientifique et Technique du CNRS et chef de file du projet ISTEEX.

453. Les membres du comité ISTEEX se sont particulièrement impliqués dans les réflexions et analyses qui ont présidé à la réalisation du présent Livre blanc, et ont exprimé leurs positions respectives.

- Raymond BERARD, Directeur de l'Institut national de l'information scientifique et technique (Inist) et Laurent Schmitt, responsable du département projets et innovation ;
- Grégory COLCANAP, Coordonnateur de Couperin, Consortium universitaire de publications numériques accompagné par Monique JOLY, responsable du département services et prospectives ;
- Jérôme KALFON, Directeur de l'Agence bibliographique de l'enseignement supérieur (ABES) ;
- Jean-Marie PIERREL, Professeur des Universités à l'Université de Lorraine, agissant pour le compte de la Conférence des Présidents d'Université (CPU).

454. Marie-Pascale LIZEE (Département de l'Information Scientifique et Technique et Réseau Documentaire (DISTRD, sous-direction du pilotage stratégique et des territoires, Service de la Coordination des stratégies de l'enseignement supérieur et de la recherche) ainsi qu'Alain ABECASSIS (Chef du service de la coordination des stratégies de l'enseignement supérieur et de la recherche au ministère de l'Education nationale, de l'enseignement supérieur et de la recherche) ont suivi l'avancement des réflexions menées autour du projet ISTEEX et du projet de loi pour une République numérique.

455. Les analyses et réflexions, ainsi que le témoignage des acteurs de la recherche publique quant aux besoins et aux valeurs des communautés scientifiques, ont été traduits juridiquement par la Direction juridique du CNRS et par Nicolas CASTOLDI, délégué général à la valorisation du CNRS, en termes de proposition de textes de lois ou de règlements.

### 1.3.4 Comité d'éthique du CNRS

456. Le Comité d'éthique du CNRS (COMETS) est une instance consultative, indépendante, placée auprès du Conseil d'administration. Il développe la réflexion sur les aspects éthiques que suscite la pratique de la recherche, en tenant compte de ses finalités et de ses conséquences ; il dégager des principes éthiques qui concernent les activités de recherche, les comportements individuels, les attitudes collectives et le fonctionnement des instances de l'organisme.

457. Dans le cadre de ses missions, et après audition de Renaud FABRE, le Comité d'éthique s'est autosaisi de la question relative au lien entre éthique et partage des données scientifiques.

458. Le COMETS a publié un avis du 7 mai 2015 intitulé « Les enjeux éthiques du partage des données scientifiques » (en annexe) dont le présent Livre blanc reprend les conclusions à de nombreuses reprises.

459. Ont été auditionnées dans le cadre du Livre blanc :

- Danièle BOURCIER, directrice de recherche au CNRS, juriste et membre du COMETS ;
- Michèle LEDUC, Directrice de recherche émérite au CNRS au Laboratoire Kastler Brossel à l'École normale supérieure, Présidente du COMETS.

### 1.3.5 Le Président du conseil national du numérique

460. Auteur du rapport « Ambition numérique – Pour une politique française et européenne de la transition numérique » remis en juin 2015 au Premier ministre, le Conseil National du Numérique (CNNum) est un acteur important de la consultation sur le projet de loi pour une République numérique. Le CNNum a par ailleurs publié un avis relatif au projet de loi pour une République numérique le 30 novembre 2015 dont les conclusions intéressant le présent Livre blanc seront reprises<sup>104</sup>.

461. Le Président du Conseil National du Numérique, Benoit THIEULIN, accompagné de Yann BONNET et de Charly BERTHET ont également été auditionnés dans le cadre du Livre blanc afin de connaître la position du CNNum sur l'Open Science.

462. La contribution du CNNum est annexée au Livre blanc.

### 1.3.6 Personnalités du monde de la recherche et de l'Open access

---

<sup>104</sup> <http://www.cnummerique.fr/avis-du-cnum-relatif-au-projet-de-loi-pour-une-republique-numerique/>

463. Des personnalités représentatives et reconnues du monde de la recherche et de l'Open Access ont enfin été auditionnées. Témoins privilégiés des pratiques des chercheurs et de leurs besoins, ils ont fait part de leur engagement et de leur prise de position en faveur du mouvement de l'Open Science :

- Claude KIRCHNER, Président en exercice du CCSD, Directeur de recherche à l'Institut national de recherche en informatique et en automatique (Inria) ;
- Christophe PERALES, Président de l'Association des Directeurs et personnels de direction des bibliothèques universitaires et de la documentation (ADBU) ;
- Christoph SORGER, Directeur de l'Institut des sciences mathématiques et de leurs interactions (Institut national des sciences mathématiques et de leurs interactions - Insmi).

## 1.4 La démarche

464. L'approche privilégiée est une démarche de consensus qui doit concourir à l'émergence et au partage de valeurs communes par l'ensemble des communautés scientifiques.

465. Pour cela, la première étape du Livre blanc a consisté à procéder au recensement et à la photographie :

- des usages de l'information scientifique et technique par les communautés scientifiques ;
- des pratiques de valorisation notamment dans le cadre des partenariats public / privé.

466. Dans une seconde étape, les résultats de ces analyses ont été confrontés aux cadres normatifs français et internationaux existants afin d'en déterminer les écarts et définir les nouveaux besoins émergents de l'usage numérique de l'IST.

467. La troisième et dernière étape a consisté à élaborer avec les communautés scientifiques des propositions afin de réduire au maximum cet écart entre les besoins et le cadre normatif.

### 1.4.1 Pratiques numériques émergentes

468. Deux types de pratiques ont été recensés dans le cadre du présent Livre blanc :

- les pratiques des chercheurs ;
- les pratiques de valorisation.

**469. Pratiques des chercheurs.** Deux principaux outils ont permis de recueillir les pratiques des chercheurs quant à l'utilisation de l'information scientifique et technique par les outils numériques.

**470. Enquête CNRS.** Le premier outil est une enquête menée par le CNRS sur les usages et les besoins d'IST des Unités de recherche. Cette enquête, réalisée auprès des Directrices et Directeurs d'Unités du CNRS à la mi-2014, a fait l'objet d'un questionnaire de 91 questions adressé par les Directrices et Directeurs d'Unités du CNRS aux Directrices et Directeurs des 1250 Unités publiantes. Un tiers d'entre eux ont répondu à la totalité des questions posées : 432 réponses complètes ont été exploitées.

471. Les réponses des Unités étant globalement proportionnelles à la répartition des Unités dans les champs de recherche considérés, l'échantillon est représentatif de l'ensemble de la recherche publique.

**472. Audition.** Le second outil est l'audition. L'association forte avec les universités et les organismes a permis l'audition de grands témoins et une parole plurielle sur les évolutions souhaitées.

473. Ces auditions ont été réalisées sur la base d'un guide d'entretien annexé au présent Livre blanc.

474. Ce guide propose trois questions ouvertes, l'objectif étant de laisser libre cours à la réflexion de la personne auditionnée :

- sur les pratiques et besoins des chercheurs en termes d'accès et d'utilisation des données et des résultats de la recherche publique, notamment au regard de la privatisation des publications par le droit de la propriété intellectuelle et le contrat d'édition ;
- sur les équilibres à trouver entre le partage des données de la science et le marché de l'édition scientifique ; entre le partage des données et les enjeux de valorisation des innovations. Autrement dit, il s'agit de distinguer une captation abusive des données scientifiques d'une captation légale ;
- sur la nécessité de définir des règles de partage et d'exploration des données : le lieu du partage et de l'observation (la plateforme), le périmètre des données partagées (données brutes, données enrichies, résultats, publications...) et les conditions de leur mise à disposition, la qualité des métadonnées associées, le statut des contenus générés par les utilisateurs (*user generated contents*).

475. **Pratique de valorisation.** Conformément à la mission légale de valorisation de la recherche publique (article L. 112-1 du Code de la recherche), et dans un contexte de concurrence internationale, il est nécessaire que les propositions formulées dans le cadre de ce Livre blanc prennent en compte les enjeux de valorisation de la recherche.

476. Une analyse des pratiques de valorisation a été menée à partir d'exemples concrets et de contrats conclus par le CNRS notamment avec des industriels :

- exemple issu d'un contrat type de collaboration de recherche entre le CNRS et un partenaire industriel ;
- exemple issu de contrat cadre entre le CNRS et les partenaires industriels.

### 1.4.2 Evolution des règles et du droit

477. Le recensement de ces pratiques a permis d'identifier un certain nombre de besoins des communautés scientifiques dont les grands témoins se sont fait l'écho.

478. Ces besoins ont été confrontés avec l'ordre existant :

- l'ordre juridique ;
- l'ordre éthique et les valeurs communes des communautés scientifiques ;
- l'ordre économique et le respect des équilibres avec le monde de la publication scientifique, et avec celui des industriels ;
- le mouvement inéluctable et historique vers l'Open Science.

479. L'état des lieux des pratiques, la définition des écarts entre les pratiques et l'ordre existant, ainsi que la confrontation de ces deux éléments notamment avec l'article 17 du projet de loi pour une République numérique a permis la formulation de propositions normatives ou organisationnelles afin de réduire ces écarts tout en préservant les équilibres.

480. Les propositions formulées sont issues des réflexions des groupes de travail et auditions et traduisent un consensus, comme en témoignent les comptes rendus d'audition.

## 2. Contribution du Conseil Scientifique

### Introduction

La place du "numérique" dans l'activité scientifique est devenue capitale aujourd'hui, même si bien entendu, il faut avoir conscience qu'il ne fait en partie que figer ou "photographier" le réel et la recherche dans un état donné. L'activité scientifique comporte bien d'autres facettes que la gestion de données. Cependant, la numérisation des données utilisées par les scientifiques et de leurs publications permet un traitement automatisé, un transfert rapide, une harmonisation des méthodes d'accès et des descriptions ; tout cela est susceptible de mettre à la portée du chercheur une immense matière, riche et diversifiée, en des temps singulièrement raccourcis. Dégageant ainsi les scientifiques d'une partie des tâches répétitives et grandes consommatrices de temps, le numérique peut libérer leurs capacités réflexives et créatrices.

On peut sans doute à cet égard, comme cela a été maintes fois écrit, comparer les possibilités ouvertes à la recherche par le numérique à celles qu'ont connues les savants des 16<sup>e</sup> et 17<sup>e</sup> siècles avec l'invention de l'imprimerie et l'accélération des échanges de savoir qui en a résulté.

### A travers les disciplines

#### SHS

En sciences humaines et sociales, en ce qui concerne les publications scientifiques récentes, alors que de plus en plus de revues francophones sont en libre accès immédiat ou après quelques années (notamment grâce à HumaNum, BSN, OpenEdition), les revues anglophones, sont souvent cantonnées à des plateformes assez chères ; le dépôt parallèle d'articles dans des archives ouvertes est peu développé. La numérisation des sources imprimées utilisées par nombre de disciplines, qu'il s'agisse de publications scientifiques plus anciennes ou de romans, journaux, traités juridiques, etc. est, elle, en bonne voie ; elle se fait souvent en libre accès, même si certaines grandes entreprises (comme Gale) produisent également des bases de données à des prix prohibitifs, pratiquement inaccessibles en Europe.

Des plateformes d'échange de données chiffrées, qu'il s'agisse des échelles les plus détaillées de la statistique publique ou de données produites par la recherche, ont également été mises en place (Réseau Quetelet, DIMESHS, etc.) : elles assurent à la fois une meilleure circulation des données, le respect des contraintes nécessaires, d'anonymisation par exemple, et la documentation ("métadonnées") sans laquelle les chiffres seraient inutilisables.

Cela dit, les données des sciences humaines et sociales, extrêmement variées selon les disciplines (de l'histoire de l'art à l'économie en passant par la linguistique), sont loin de se limiter à des imprimés libres de droits ou des chiffres. Les plateformes de partage restent à inventer pour, par exemple, les photographies d'archives ou d'œuvres d'art prises dans un but scientifique (ce qui se heurte au droit de reproduction) ; elles sont encore peu développées pour les données issues d'enquêtes qualitatives de terrain (qui posent de délicats problèmes d'anonymisation, de mise en forme et de documentation). Le problème est ici que nombre de données utilisées par les scientifiques en SHS ne sont pas produites par eux (qu'il s'agisse d'une chanson, d'un rapport d'activité d'entreprise ou de l'architecture d'un monument) : d'autres personnes physiques ou morales disposent de droits sur elles. Partage de données et techniques de "fouille de données et de textes" sont ainsi inégalement répandus selon les types de données, pour des raisons principalement d'obstacles juridiques, et de manque de moyens en personnel pour la production et le maintien de métadonnées de qualité. Du fait de ces contraintes, pour beaucoup de types de données en SHS, il semble difficile d'imaginer un libre partage qui irait au-delà d'un partage pour usage scientifique, avec toutes les difficultés que présente la définition de ce périmètre. Au-delà, il y aurait en effet des dangers de captation de données qui peuvent être très sensibles. Par ailleurs, pour certains types de données, le temps d'exploitation pouvant mener à des publications est assez long, ce qui plaide pour des délais avant mise en partage ajustés à ces particularités.

## **Sciences de l'univers**

En astronomie, et plus généralement dans une partie des sciences de l'univers ou des sciences d'observation, se généralise le paradigme d'observatoire virtuel. Les données sont libres d'accès en astronomie pour toute la communauté après la fin d'une période propriétaire. C'est l'approche du réemploi maximalisé des données qui est privilégiée. Pour cela, il est nécessaire d'harmoniser et de standardiser les formats, les descriptions et les modes d'accès aux données d'archives, aux métadonnées et aux applications susceptibles de les traiter, de manière à réaliser l'interopérabilité. Cette interopérabilité s'étend à la liaison des données de la recherche avec les publications en ligne. Les dangers de captation à des fins commerciales ont été peu prégnants jusqu'à ce jour, même si les choses pourraient changer à l'avenir (par exemple avec la "météorologie spatiale" et l'observation détaillée des éruptions solaires).

## **Biologie**

En biologie, l'édition numérique est généralisée et la recherche d'articles se fait via des plateformes thématiques développées par des institutions académiques et permettant d'accéder librement aux résumés d'articles. La plus importante (PubMed) est proposée

par le National Institute of Health américain (NIH). L'accès à la totalité des articles est lui généralement payant, le transfert du copyright à l'éditeur étant la pratique la plus courante. Il est à noter que le NIH, s'est opposé à cette pratique et propose en accès libre, via PubMed, une version non formatée de tout article publié par un éditeur décrivant des travaux financés par le NIH. Depuis une dizaine d'années, l'Open Access se développe. Le coût de publication est alors généralement payé par les auteurs lors de la publication.

Si les techniques de "fouille de données" sur les textes ne sont pas prioritaires pour la plupart des domaines de la biologie en termes d'outil de découverte, (mais plutôt en termes de collecte documentaire) la "fouille de données" proprement dite prend une place de plus en plus importante. Le libre accès à ces données est répandu, à l'instar de ce qui s'est passé pour le génome humain. De nombreux éditeurs dont Nature, conditionnent d'ailleurs la publication d'un article au dépôt des données en masse, associées à une publication, sur une plateforme accessible gratuitement par tous. Il est à noter que cette exigence va au-delà des données numériques et concerne également le matériel produit dans le cadre d'une publication.

Lorsqu'un papier décrit un matériel particulier (lignée de cellules, microorganismes ou souris génétiquement modifiées, virus, anticorps, ...), l'éditeur (Nature...) demande un engagement de l'auteur pour la donation de ce matériel aux autres chercheurs académiques. Des plateformes internationales existent pour la conservation et la distribution de ce matériel. Au-delà de la question des données en masse, plusieurs éditeurs dont Nature songent à mettre en place un système permettant d'accéder, via leur site, aux données brutes ayant conduit à l'élaboration des figures d'un article. Si cela permettra au lecteur de s'assurer de la bonne interprétation des résultats, la question se pose de la propriété de ces données et de leur cession éventuelle.

## Physique

En ce qui concerne la physique hors « grands instruments », l'accès libre aux données brutes n'est pas encore très répandu. En revanche, de nombreuses bibliothèques numériques ont été constituées et mises en libre accès par des groupes de chercheurs ; régulièrement mises à jour, elles portent aussi bien sur la modélisation théorique de problèmes génériques (conduction électrique, dynamique moléculaire) que sur la mise au point et le pilotage d'expériences (interfaçage d'appareils, bibliothèques pour le traitement des données). Le numérique joue également un rôle essentiel dans le domaine de la diffusion des résultats, avec l'utilisation quasi-systématique des serveurs de prépublications. Le dépôt sur ces serveurs se fait de manière simultanée avec l'envoi à une revue scientifique avec comité de lecture, et permet de « prendre date » immédiatement.

## Chimie

Le domaine de la chimie est en fait vraiment intermédiaire entre les pratiques des sciences de la vie et celles de la physique. La règle, ce sont des publications dans des journaux payants de sociétés savantes (American Chemical Society, Royal Society) ou sociétés commerciales (Wiley, Elsevier, ...) et un développement timide de l'"open access" type "gold" payé par les auteurs. Il y a en fait peu de différences entre les deux, les négociations avec l'ACS ont été à un moment donné plus dures qu'avec Elsevier. Il n'y a pas de pré-publication comme ArXiv. Des bases de données libres d'accès se développent, notamment la base de Cambridge qui contient toutes les structures moléculaires publiées.

## Mathématiques et informatique

En mathématiques, les bases de données relatives aux publications sont très importantes pour le travail individuel et communautaire. Une particularité de cette discipline est l'importance de l'accès facile aux publications "anciennes" (plusieurs années, décennies voire siècles). La pérennité quant à l'accès à ces publications est donc capitale pour la recherche. Les plateformes d'archivage des publications comme HAL ou Arxiv répondent donc en partie à cette problématique et sont à soutenir, ainsi que les plateformes de métadonnées (MathSciNet, Zentralblatt,...). En ce qui concerne les données numériques, il est nécessaire que pour des questions de reproductibilité, comparaison et interprétation de méthodes de simulation et calcul, elles soient librement accessibles et elles-aussi pérennisées (archivage, catalogues de jeux de données, ...) et ceci concerne aussi bien les logiciels que les codes de calcul. Par ailleurs les mathématiques jouent un rôle important dans l'analyse, la gestion et l'exploitation des masses des données (questions autour du Big Data). Il est certainement très important que les données soient accessibles, mais quand elles deviennent de plus en plus massives, il faut aussi pouvoir les exploiter de manière efficace. Dans ce domaine il y a d'importants défis à relever pour la recherche mathématique.

### La "publication des données" :

Une question importante transversale aux disciplines est celle de la "publication" des données.

L'exigence du libre accès est claire dans le cas de données associées à des publications dûment validées par les revues à comité de lecture. Mais qu'en est-il de données qui seraient mise en ligne avant publication, par exemple pour analyse et interprétation dans le cadre de larges collaborations? C'est une réalité qui monte dans nombre de disciplines. Ce problème est d'autant plus aigu que la définition même de ce qu'est une donnée publiée est parfois floue.

## **Dangers et garde-fous juridiques :**

On constate donc, que comme toute avancée, la numérisation des données et des résultats de la science peut avoir des contre-effets. Les résultats scientifiques jouent un rôle capital dans la concurrence économique mondiale en conférant à leurs détenteurs des avantages compétitifs parfois considérables. En retour, la science moderne a besoin, pour se développer et expérimenter, de technologies que va souvent lui offrir le monde de la production régi par le marché. C'est particulièrement le cas des fonctionnalités qui peuvent être fournies par les éditeurs scientifiques

Pourtant chacun sait que le développement de la connaissance se fait à travers les échanges d'idées, de résultats et de données entre scientifiques. Il est donc impératif de limiter la captation du travail de ces derniers par les intérêts privés et en même temps de fournir un cadre légal pour permettre de libérer autant que faire se peut l'échange des données d'usage scientifique.

## **Trois principes :**

Le conseil scientifique soutient trois principes importants qui permettraient de répondre à ces objectifs :

- La liberté complète de circulation et d'usage des données scientifiques pour réutilisation dans le cadre de la science, sous réserve qu'un temps minimal de réserve permettant aux producteurs de données de les interpréter et de publier soit garanti par la loi. Cette exigence de libre circulation des données s'étend d'une part aux publications et d'autre part aux données et textes non scientifiques à l'origine mais constituant la matière première de beaucoup de recherches, notamment en sciences humaines et sociales.
- Cette exigence de la mise à disposition des données s'étend aux services à valeur ajoutée (traitement massif de type "Big data", fouilles de données, lien aux métadonnées, interopérabilité) qui doivent également être publics et libres d'accès pour éviter toute captation abusive. Ceci supposerait, dans le cas de création de services et de plateformes par les éditeurs et plus généralement le secteur privé, des garanties légales de juste prix non discriminants.
- Elle suppose aussi la clarification des droits à disposer de l'usage de leurs productions scientifiques et de leurs publications par les auteurs face aux éditeurs et aux autres acteurs privés. Les droits de propriété intellectuelle des scientifiques ne doivent en aucun cas être concédés gratuitement aux éditeurs, de manière à favoriser la libre circulation des résultats scientifiques.

Le conseil scientifique tient également à saluer le travail réalisé par le COMETS dans son rapport intitulé “les enjeux éthiques du partage des données scientifiques” et il souscrit aux recommandations contenues dans ce texte.

### **3. Recommandation du Conseil Scientifique**

Le conseil scientifique du CNRS s'est tenu régulièrement informé des discussions entourant la préparation du projet de loi sur le numérique qui doit être soumis au Parlement à l'automne 2015. Il a produit sa propre contribution au Livre blanc de l'organisme sur ces questions d'importance capitale pour les activités de recherche scientifique.

Le conseil réaffirme deux principes essentiels: i) la science est un bien commun de l'humanité qui ne saurait souffrir de captation abusive par des intérêts privés, ii) le libre accès aux résultats de l'activité scientifique (publications, données de la recherche, métadonnées, services à valeur ajoutée) ne saurait être entravé sans remettre en cause le développement même de la science. Ce principe de libre accès est bénéfique autant pour les auteurs et la communauté scientifique, que pour les organismes de financement et plus largement l'enseignement supérieur.

Fort de ces principes, le conseil scientifique s'inquiète de possibles reculs en termes de durée d'embargo et de libre accès aux publications scientifiques qui pourraient intervenir dans le projet de loi. Il rappelle que d'autres pays, comme l'Allemagne, le Canada, les États-Unis, le Royaume-Uni, ont mieux su résister aux demandes des éditeurs privés en faisant adopter dans leurs législations le principe du libre accès.

Il rappelle son exigence de voir consolider par la loi les pratiques actuelles en matière d'accès aux données scientifiques, comme c'est déjà le cas dans ces pays.

- lorsque l'activité de recherche a été financée en partie par des fonds publics, la cession à un éditeur des droits sur les données et les écrits issus de cette recherche ne saurait être exclusive ;
- les scientifiques doivent pouvoir mettre à disposition gratuitement, sous une forme numérique, ces données et ces résultats, a priori sans période d'embargo imposée par les éditeurs ;
- les services de "fouille de données" et assimilés jouent un rôle considérable dans la valorisation scientifique des données et écrits en libre accès. Ils ne doivent pas être entravés par les plate-formes à finalité commerciale de diffusion de ces données et écrits.

#### 4. Avis du Comité d'éthique du 7 mai 2015 « Les enjeux éthiques du partage des données scientifiques »

##### Groupe « Data sharing »

##### Les enjeux éthiques du partage des données scientifiques

##### I- Autosaisine

- 1- Depuis deux décennies, les données ont acquis un rôle central dans la production scientifique quelles que soient les disciplines. Les chercheurs ont de plus en plus besoin de vastes entrepôts de données (*big data*) - mais aussi de jeux de données de dimensions plus modestes (*small data*) - pour explorer, visualiser et comparer des résultats, valider des hypothèses ou en formuler de nouvelles, voire pour construire automatiquement de nouvelles connaissances (*machine learning*). De grandes infrastructures et des plateformes communes continuent d'être créées pour l'archivage, le stockage et le traitement de l'information. Les avancées rapides dans les technologies numériques ont considérablement amélioré la façon dont les données, informations et outils peuvent être diffusés, gérés, utilisés et réutilisés entre les chercheurs et constituer un écosystème autour des publications scientifiques. Les mouvements favorisant l'accès ouvert deviennent donc cruciaux pour exploiter au mieux de vastes gisements de données. Aucune organisation n'a assez de moyens pour mener seule ses travaux.<sup>105</sup> L'effort demandé pour exploiter les connaissances numérisées est immense, d'autant plus qu'il requiert l'intervention humaine à un certain moment du processus (la fouille de textes est nécessaire mais non suffisante). Faciliter l'accès et la réutilisation de ces données est ainsi devenu un enjeu crucial pour partager et faire circuler plus rapidement les résultats de la recherche.
- 2- Cependant les attitudes vis à vis du partage et de l'ouverture sont très différentes suivant les types de données et les disciplines. Dans certaines disciplines (en astrophysique ou en génomique par exemple) les avantages de ce partage de données sont apparus assez forts et les inconvénients assez faibles pour que ce développe un courant dit du *data sharing*. Pour ces communautés les rapprochements et comparaisons de données sont à l'évidence sources de nouvelles découvertes et elles considèrent que toute entrave à la circulation des résultats scientifiques est non seulement inefficace mais contraire aux principes

---

<sup>105</sup> Voir l'article "Dix laboratoires mondiaux partageront données et chercheurs" (*Le Monde* 4 février 2014). Ce projet orchestré par le NIH demande notamment aux laboratoires privés ET publics "ne pas développer leur propre médicament à partir des découvertes obtenues avant qu'elles n'aient été rendues publiques".

fondamentaux de mise en commun généralisée et ouverte des connaissances.

Mais pour d'autres disciplines (particulièrement en Sciences humaines), les données sont souvent collectées individuellement : liées à l'objet de la recherche, les données ne peuvent être partagées qu'avec le même embargo que celui de la publication des résultats.

- 3- Au-delà du cas de ces communautés scientifiques « auto-organisées », les politiques gouvernementales d'*Open data*<sup>106</sup> visent depuis quelques années à diffuser largement les données subventionnées sur fonds publics. Certaines de ces données peuvent intéresser les scientifiques et, réciproquement, les données des scientifiques peuvent concerner la société. Le mouvement de *data sharing* doit donc être ajusté aux politiques gouvernementales d'ouverture des données qui poursuivent des objectifs sensiblement différents et qui sont soumises à des contraintes éthiques et juridiques différentes.
  
- 4- Le présent avis a pour objectif de voir comment les différentes politiques scientifiques peuvent s'articuler dans un champ plus vaste qui serait l'éthique du partage des données de recherche. Si un grand nombre de chercheurs soutiennent le *data sharing*, la plupart se sentent démunis voire réticents devant cette obligation gouvernementale de diffusion (l'*open data*) qui peut paraître paradoxale, voire contreproductive : incités à diffuser largement - ce que confirme le Programme européen HORIZON 2020<sup>107</sup>, ils doivent en même temps appliquer des restrictions légales à cette diffusion publique des données, au nom du respect de la vie privée, du droit d'auteur, de l'obligation de secret, de la confidentialité et de la sécurité. Face à ces injonctions qui peuvent paraître contradictoires, il devient nécessaire d'informer les chercheurs sur leurs diverses obligations et sur la portée éthique de leurs choix relatifs aux données qu'ils collectent, partagent ou réutilisent.

---

<sup>106</sup> Notons au passage que nous considérons dans ce texte essentiellement les données scientifiques, ce qui exclut de notre analyse toutes les données retraçant les activités individuelles qui posent des problèmes éthiques d'une toute autre nature.

<sup>107</sup> Qui consacre le principe du « libre accès aux publications et aux données de la recherche » : voir [www.horizon2020.gouv.fr](http://www.horizon2020.gouv.fr)

## II- Analyse

Les données scientifiques (*research data*) considérées ici concernent toutes les données collectées dans le contexte de la recherche scientifique<sup>108</sup> c'est-à-dire :

- les données primaires (empiriques, observées, mesurées)
- les données secondaires, dérivées des données primaires, annotées, enrichies, interprétées ajoutant de la valeur aux données initiales et pouvant impliquer d'autres acteurs.
- les métadonnées qui structurent, gèrent et facilitent l'accessibilité des données primaires et secondaires.

Ces données peuvent être des documents textuels, graphiques, picturaux, multimedia, ou des représentations numériques. L'écart entre données et publications tend d'ailleurs à se réduire avec le concept de l'Open process, qui consiste à diffuser les données et connaissances utilisées et construites au cours du processus d'élaboration et d'écriture (*thinking and writing*) de la publication scientifique<sup>109</sup>.

### 1- Contexte stratégique

Des accords et chartes successives ont jalonné l'histoire du mouvement du *data sharing*. En 1996 pour la première fois des chercheurs, impliqués dans le séquençage du génome humain, ont signé un ensemble d'accords consacrant les bases du partage ouvert des données prépubliées. Puis la première définition de l'Open data a été donnée par la Déclaration internationale sur le libre accès de Budapest qui s'est tenue le 14 février 2002, connue sous l'acronyme *BOAI*<sup>110</sup> (*Budapest Open Access Initiative*). De là, de nombreuses autres initiatives ont vu le jour le jour, avec par exemple la Déclaration de Berlin de 2003 sur le libre accès à la connaissance dans toutes les sciences y compris les sciences humaines<sup>111</sup>, suivie en mars 2005 d'une nouvelle Déclaration appelée Berlin III visant à renforcer les mesures adoptées dans le cadre de Berlin I. La plupart des organismes scientifiques dont le CNRS ont signé ces déclarations et légitimé cette culture de l'accès ouvert. Actuellement plusieurs recommandations générales sont disponibles (consortiums scientifiques, OCDE...) et des agences de financement s'appuient sur de tels principes pour fixer leurs exigences dans ce domaine qui sont autant de conditions

---

<sup>108</sup> H.Tjalsma & j. Rombouts, *Selection of research data- Guidelines for appraising and selecting research data*, The Hague : Data archiving and Networked services 2011, p.13,14 at <http://www.dans.knaw.nl>

<sup>109</sup> P.Uhlir « Revolution et evolution in scientific Communication : Moving from Restricted dissemination of publicly-funded knowledge to open knowledge environments » : [http://www.communia-project.eu :communi/files/conf202009\\_P\\_UhlirBS.pdf](http://www.communia-project.eu :communi/files/conf202009_P_UhlirBS.pdf), 2008

<sup>110</sup> <http://www.budapestopenaccessinitiative.org/read>

<sup>111</sup> <http://openaccess.inist.fr/>

d'octroi de subventions. Ainsi en 2013, l'initiative Horizon 2020 a défini une politique européenne d'ouverture et de partage de données scientifiques à laquelle les organismes nationaux doivent désormais se conformer lorsque des fonds européens sont en jeu. De même, une des dernières initiatives en date vient de nouveau du domaine de la biologie humaine avec le lancement en 2013 de la « Global alliance for genomics and health »<sup>112</sup>.

La nécessité d'une réflexion est déjà inscrite dans le cadre du plan stratégique du CNRS « *Mieux partager la connaissances, une stratégie ouverte pour une information scientifique et technique d'avenir* »<sup>113</sup>. Cette stratégie mobilisatrice doit être traduite en obligations éthiques pour les chercheurs, principaux producteurs et utilisateurs de données scientifiques. Une partie de ce programme prévoit d'ailleurs de « mettre en place une charte éthique transcendant les catégories instrumentales » et réaffirmant les buts de la recherche publique.

## 2- Contexte scientifique

L'activité scientifique s'appuie de plus en plus sur la création et l'utilisation partagées d'infrastructures de données multi-sources et multi-usages. Ces transformations récentes de la démarche scientifique sont liées à trois types de changements :

- l'évolution technologique des instruments de mesure et de capteurs de données brutes permettant de produire des données en masse ;
- les capacités informatiques de stockage, d'archivage et d'analyses (naissance de la bio-informatique par exemple) ;
- l'internet collaboratif et la mise en réseau, permettant un enrichissement direct en ligne de bases de données et de plateformes par de nombreux intervenants et permettant des économies d'échelle.

Cette évolution aboutit à un bouleversement des principes et des pratiques, passant de recherches de type *hypothesis-driven research* à la généralisation de *data-driven research*, c'est-à-dire à un processus d'élaboration partant de données déjà constituées. Dans ce contexte, les données qui font désormais l'objet d'annotation, de fouille et d'analyse deviennent les composantes essentielles de l'activité de recherche. Les utilisations multiples de données deviennent la règle et les masses de données produites imposent la création d'infrastructures de données à usages multiples. Une étape importante de la science devient la constitution d'infrastructures de données comme appui

---

<sup>112</sup> *Nature* 498, 16–17 (05 June 2013) | doi:10.1038/498017a ; L'initiative <http://genomicsandhealth.org/> est en train de produire un « *International Code of conduct for genomic and health-related data sharing* », actuellement à disposition pour commentaire sur le site web (<http://genomicsandhealth.org/our-work/work-products/international-code-conduct-genomic-and-health-related-data-sharing-draft-6>).

<sup>113</sup> <http://www.cnrs.fr/dist/strategie-ist.htm>

à des opérations de recherche (*research process*) et non plus seulement à l'archivage de résultats. Les capacités d'enrichissement, d'annotation des données rendent indispensables d'une part un suivi des bases de données correspondantes et de leur évolution. D'autre part une organisation des métadonnées, permettant l'utilisation effective des données transformées, agrégées, corrélées. Enfin les données, ainsi produites en masse en dehors d'une hypothèse particulière, nécessitant souvent des budgets importants sur lesquels vont s'appuyer de nombreux travaux de recherches, deviennent ce qu'on appelle des *community resources* dont le maintien et l'accès sont essentiels pour l'action collective du groupe.

Cependant si les publications permettent d'accéder à 10 % des données d'expérience disponibles, 90 % de celles-ci restent sur les disques durs d'ordinateurs. Les données ne circulent pas suffisamment rapidement dans le monde scientifique. Quelles que soient les disciplines, trop de résultats restent non publiés et beaucoup de données sont sous-utilisées ou perdues<sup>114</sup>. Les données issues de résultats négatifs sont oubliées. Comment inciter les chercheurs à participer à l'ouverture et à la diffusion de leurs données ?

### 3- Nouvelles responsabilités face à l'évolution mondiale des pratiques

Le *data sharing* ouvert des données scientifiques croise un autre mouvement mondial qui dépasse le champ scientifique : l'ouverture des données publiques, c'est-à-dire subventionnées sur fonds publics (*open data*). Les politiques d'*open data*, développées au niveau des Etats, obligent les administrations et établissements publics à rendre accessibles et à partager leurs données. Or ces deux tendances se renforcent mais ne relèvent pas de la même logique.

#### **Le *data sharing* et le mouvement des communs scientifiques**

La diffusion du savoir se fit d'abord à travers les cercles de savants puis dans les échanges entre Académies des sciences<sup>115</sup>. Les savants pouvaient débattre directement par l'intermédiaire des comptes rendus de leurs expérimentations et se transformèrent ainsi en communauté <sup>116</sup> Les données brutes étaient partagées et répliquées car elles se situaient en dehors d'une quelconque appropriation exclusive. On avait coutume de dire

---

<sup>114</sup> Au début de 2014, cinq articles ont été publiés dans le *Lancet* sur le thème Research : increasing value, reducing waste. Voir notamment : An-Wen Chan, Fujian Song, Andrew Vickers, Tom Jefferson, Kay Dickersin, Peter C Gøtzsche, Harlan M Krumholz, Davina Ghersi, H Bart van der Worp, "Increasing value and reducing waste: addressing inaccessible research" in *Lancet* 2014; 383: 257–66.

<sup>115</sup> « L'Académie des sciences doit son origine à la fois aux cercles de savants qui dès le début du XVII<sup>e</sup> siècle se réunissent autour d'un mécène ou d'une personnalité érudite, et aux sociétés scientifiques permanentes qui se constituent à la même époque, telles l'*Academia dei lincei* à Rome (1603), la *Royal Society* à Londres (1645)... Par ses travaux et ses publications, l'Académie contribue de façon essentielle à l'expansion de l'activité scientifique. » <sup>1</sup>

<sup>116</sup> Evelyne Barbin (dir.), *Arts et sciences à la Renaissance*, Ellipses, 2007

que le piège de la souris était breveté mais que les données de l'expérience ne l'étaient pas. Le marché des données scientifiques, brutes ou non, n'existait pas encore. Mais depuis quelques années, après l'explosion du *big data* et la naissance d'une science des données, la reconnaissance de leur valeur voire de leur monétisation et les verrous juridiques qui en ont résulté ont incité la communauté scientifique à s'organiser pour réaffirmer les principes de l'ouverture et de la disponibilité des données.

Les données sont confrontées à de nombreux obstacles juridiques ou techniques à leur diffusion : soit les grandes bases de données sont soumises à des droits d'accès, soit les données sont disponibles sous des formats fermés, ou nécessitant des logiciels propriétaires. C'est pourquoi en 2005 une communauté de chercheurs, conscients des résistances rencontrées lors de la mise en place des politiques d'ouverture des données, a lancé une initiative globale visant à créer concrètement des Communs Scientifiques (*Science Commons*) avec des outils et des méthodes (plateformes d'accès, contrats d'auteur type,...) pour accélérer la circulation des résultats et la réutilisation des données sur lesquels ils sont fondés<sup>117</sup>.

De même, le Protocole appelé "Open data" (qui était plutôt du « data sharing ») est né dans le milieu de la recherche : il incite lui aussi concrètement les membres de la communauté scientifique mondiale à mettre en commun leurs ressources quel que soit leur statut juridique. Ces plateformes de recherche partageables<sup>118</sup> facilitent en effet le développement de nouveaux services pour :

- **la réutilisation de la recherche grâce à des politiques et des outils qui aident les individus et les organisations à rendre leur production accessible et à indiquer, sur leurs résultats et leurs données, cette possibilité de réutilisation**
- **l'accès immédiat à des outils (calculs en ligne) grâce à des contrats standard qui apportent des économies d'échelle aux chercheurs, de sorte que ceux-ci peuvent dupliquer vérifier et étendre leur recherche, et aussi participer à toute la chaîne de valeurs jusqu'au *peer review***
- **l'intégration de sources fragmentées d'information en fournissant aux chercheurs les moyens de trouver analyser et utiliser des données de sources disparates en marquant et intégrant l'information à travers un langage commun, standardisé et traduisible dans la machine.**

Ce mouvement d'ouverture et du partage des données a été facilité par des politiques d'archivage ouvert développées au sein des institutions scientifiques (ArXiv, 1991). En France, HAL (Hyper Article en Ligne), créé en 2000 est fondé sur « le modèle de la

---

<sup>117</sup> Désormais Science at Creative Commons. Voir aussi : <http://sciencecommons.org/about/>.

<sup>118</sup> Ce fut le fondement du projet Science Commons voir : D. BOURCIERr, Web, « Science et Communication : l'exemple de Science Commons », *Revue Hermès* 57, 2010, pp. 53-160.

communication directe entre chercheurs »<sup>119</sup> de leur *preprints* : son pilotage et ses missions, en cours de refonte, sont encore à définir vis-à-vis de l'archivage des données scientifiques, pour intégrer dans le *design* le lien sur le délai d'embargo et de la licence ouverte dont le choix dépend uniquement du chercheur lors du dépôt.

En 2013 une initiative issue du domaine de la biologie humaine a lancé la « Global alliance for genomics and health »<sup>120</sup>. Il s'agit d'un mouvement unifié de fédérations membres de 170 pays qui ont décidé de fournir une plateforme d'engagement en direction d'organisations non gouvernementales et de créer un réseau puissant destiné à dresser le profil de maladies non communicables. D'autres disciplines comme la physique et les sciences de la Terre et de l'Espace insistent sur d'autres impératifs. Ainsi, l'observation pérenne des phénomènes naturels met en jeu des processus dont les constantes de temps peuvent être grandes par rapport à la vie humaine (les variations du champ magnétique terrestre, la tectonique, l'évolution du climat, le cycle sismique...). Elles sont par essence non reproductibles et sont à la base de nos connaissances sur le monde qui nous entoure, de ses évolutions et des risques qu'il fait peser sur nos sociétés. La nature étant un bien commun, l'archivage et la diffusion libre de ces données est une exigence vis à vis du public.

La numérisation des observations et l'échange des fichiers numériques offraient de nouvelles opportunités à ceux qui souhaitaient réinterpréter les données ou les confronter. Ces divers mouvements ont eu un effet heuristique certain sur le processus scientifique traditionnel qui était décrit séquentiellement de la conception à la restitution écrite, qu'elle soit digitale ou imprimée. Désormais le discours scientifique n'est plus descriptible de façon linéaire mais ressemble à un processus où des résultats partiels sont interdépendants et évolutifs. Ces interactions cognitives se manifestent dans ce qu'on appelle les « *knowledge hubs* »<sup>121</sup> où coexistent plusieurs couches de connaissance plus ou moins élaborée. Dans ce cadre, c'est l'accès aux données primaires qui devient déterminant en permettant de vérifier leur qualité et de jauger de la méthodologie et de l'interprétation qui s'en est ensuivi. De plus, la structure de la connaissance en accès ouvert a une influence sur la recherche elle-même qui n'est plus une « variable indépendante »<sup>122</sup> du processus d'élaboration, mais le fruit dynamique d'un *brain storming* continu entre chercheurs (*e-mail science*).

---

<sup>119</sup> Voir Rapport Serge BAUIN, *L'Open acces à moyen terme : une feuille de route pour HAL*, DIST, CNRS, septembre 2014

<sup>120</sup> *Nature* 498, 16–17 (05 June 2013) | doi:10.1038/498017a. L'initiative <http://genomicsandhealth.org/> est en train d'élaborer un « International Code of conduct for genomic and health-related data sharing », actuellement à disposition pour commentaire sur le site web (<http://genomicsandhealth.org/our-work/work-products/international-code-conduct-genomic-and-health-related-data-sharing-draft-6>).

<sup>121</sup> [https://mpira.ub.uni-muenchen.de/8778/1/MPRA\\_paper\\_8778.pdf](https://mpira.ub.uni-muenchen.de/8778/1/MPRA_paper_8778.pdf)

<sup>122</sup> La devise de la Royal Society britannique était « *Nullius in verba* » (*Take no man's word for it*)

Conséquence de ce mouvement, les éditeurs ont pris l'habitude de demander aux chercheurs, en plus des résultats scientifiques, de mettre leurs données en ligne<sup>123</sup> pour vérifier la reproductibilité de l'expérience ou du processus. Cela leur a permis d'abord de pouvoir contrôler les résultats publiés en les confrontant aux données et de tenter d'éviter ainsi le plagiat et les fraudes, qui les obligent à la rétractation des articles. A plus long terme, il n'est pas exclu que ces données accumulées finissent par créer un « marché de données » au profit des éditeurs. Ce phénomène très inquiétant doit être dénoncé et refusé par les chercheurs.

Le *data sharing* scientifique renvoie à une pratique communautaire traditionnelle en science: celle de rendre les données utilisées pour la recherche scientifique disponibles pour les autres chercheurs et de créer des *pools* de ressources gérés par la communauté scientifique. Beaucoup d'agences, d'institutions et de revues ont *soutenu* les politiques de partage de données parce que l'ouverture et la transparence étaient considérées comme des principes éthiques inhérents au travail scientifique.

Cependant rendre universelle cette pratique de partage pose des questions du type : à qui ces données doivent-elles être ouvertes ? A quelle communauté scientifique ? Doit-elle être ouverte aux citoyens, au public ?

Les politiques de partage impliquent d'informer les chercheurs sur les limites de ce partage. Les données concernées peuvent être indisponibles en tant que données personnelles non anonymisées, ou peuvent être soumises à des régimes particuliers comme celui de la sécurité nationale et du secret professionnel, à des clauses contractuelles restrictives ou à divers intérêts commerciaux. Ensuite, si le chercheur est titulaire d'un droit sur ces données et qu'il souhaite les partager, il lui est conseillé de mettre ses données protégeables sous une licence libre comme *Creative Commons*<sup>124</sup> pour au moins informer les futurs utilisateurs que la plupart des œuvres sont protégées mais « libérées » aux conditions de l'auteur.. Les chercheurs doivent être vigilants quand ils cèdent à des tiers leurs droits d'exclusivité sur leurs données ou sur des banques de données<sup>125</sup>, ou inversement quand ils utilisent des données générées par d'autres chercheurs ou par des plateformes gouvernementales ouvertes.

---

<sup>123</sup> Toronto International Data Release Workshop Authors (2009), "Prepublication data sharing", *Nature* 461: 168-170)

<sup>124</sup> [www.creativecommons.fr](http://www.creativecommons.fr) qui est un projet de partage de contenus et une plateforme de licences ouvertes dont les options s'étalent de la licence la plus ouverte (mention de l'attribution) à la plus "marchande".

<sup>125</sup> Le statut des banques de données en Europe, défini comme *sui generis*, n'est pas transposable dans la plupart des autres pays. Les Etats Unis par exemple ne reconnaissent pas le droit d'auteur sur les banques de données.

## ***L'open data comme politique publique***

Contrairement au mouvement de partage de données développé par les chercheurs eux-mêmes, les politiques publiques d'*Open data* ont pris naissance hors de la communauté scientifique. Chaque jour, un nombre croissant de données sont produites ou collectées par différents acteurs opérant dans différents secteurs d'activité qui se distinguent par leurs objectifs et finalités. L'Etat, tout d'abord, est un important producteur de données. De plus en plus de données statistiques sont produites, reproduites, collectées, diffusées ou rediffusées par les administrations publiques dans le cadre de leurs missions institutionnelles. Il s'agit notamment de données démographiques, géographiques, météorologiques, économiques, financières, culturelles, touristiques, etc., qui visent à assurer la qualité et la continuité du service public - mais qui peuvent aussi constituer de nouvelles ressources pour les chercheurs. C'est ainsi qu'en Europe, suite à la Directive sur la réutilisation des informations du secteur public,<sup>126</sup> puis à la Directive créant des infrastructures d'information géographique<sup>127</sup>, la majorité des pays ont adopté une politique visant à promouvoir l'ouverture des données publiques. Dans le secteur public, l'*Open data* pourrait donc être défini comme la mise à disposition *libre* et (quasi) *gratuite* de données qui implique la possibilité de les réutiliser de façon la moins contrainte possible. En France, Etalab<sup>128</sup>, service qui gère l'*open data* public sous l'autorité du premier ministre, a pour mission de communiquer les données de la recherche. Il serait souhaitable qu'il inclue à côté de son réseau d'experts des compétences éthiques propres à définir les données intéressant la recherche.

Cependant les politiques qui promeuvent l'ouverture de données publiques, c'est à dire promouvant le partage et la réutilisation, n'ont pas les mêmes objectifs et les mêmes destinataires que le *data sharing*. Un des objectifs de l'*Open data* du secteur public est de permettre la valorisation voire la monétisation de ces données en créant de la richesse pour les entreprises qui les exploitent. En outre, les destinataires de cette ouverture concernent tous les acteurs du secteur public, administrations et collectivités bénéficiant de l'argent public. Ces politiques ont cependant en commun de vouloir favoriser la transparence des modes de production de connaissances et de créer des gisements de données accessibles et partageables.

La différence fondamentale entre le *Data sharing* et l'*Open data* est que dans le monde scientifique le mouvement est parti de la communauté elle-même et que la finalité est éthique car elle concerne les valeurs, c'est à dire la délimitation de ce qui est bien ou mal pour la communauté qui l'applique. En revanche pour l'*Open data* gouvernemental,

---

<sup>126</sup> Directive 2003/98/CE du 17 novembre 2003 sur la réutilisation des informations du secteur public.

<sup>127</sup> Directive 2007/2/CE directive 2007/2/CE du [14 mars 2007](#) établissant une infrastructure d'[information géographique](#) dans la [Communauté européenne](#) (appelée Directive *Inspire*).

<sup>128</sup> Mis en place par la Circulaire du 17 septembre 2013.

l'incitation est d'origine normative et s'applique *légalement* à l'ensemble des agents publics, y compris aux agents travaillant dans le service public de la recherche.

Une des façons d'éclaircir les régimes applicables seraient de différencier données scientifiques et données publiques. Mais les données scientifiques, si elles sont majoritairement produites sur des fonds publics, ont sauf exception toutes vocations à devenir publiques. Pourtant on s'aperçoit que les chercheurs suivant les disciplines ne sont pas tous favorables à l'ouverture de leurs données. L'embargo de six mois à un an suivant les délais de publication serait dans les sciences humaines et sociales le délai nécessaire pour que les données primaires soient rendues disponibles. Cette exigence se décline différemment en sciences exactes et dépend en fait beaucoup des domaines : dans de nombreux cas il n'y a pas d'embargo sur les données ; par exemple en biologie les données sont généralement fournies en même temps que la publication. Pour ce qui est de l'exploitation des données issues des grands instruments de la physique ou de l'astronomie, il y a un délai avant leur mise au service de toute la communauté, car les données brutes doivent en général être traitées avant d'être exploitables ; en outre en astronomie des règles très précises, fixées par avance, donnent une préférence pour un temps limité (en général un an) pour l'exploitation des données aux chercheurs ayant construit un instrument sur le grand équipement (satellite, télescope). Cela signifie pour les chercheurs que les politiques générales d'ouverture des données publiques ne rejoignent pas les habitudes de la communauté à laquelle ils appartiennent et qu'ils ne peuvent y contribuer qu'en définissant les limites de leurs pratiques.

D'un autre côté, les chercheurs doivent bénéficier de l'Open data public promu par l'Etat. Les données publiques, notamment en matière de santé, ont vocation à devenir des données scientifiques qu'ils pourront utiliser. Ainsi la SNIIRAM <sup>129</sup>est définie comme la plus grande base de données sur la santé du monde : elle est alimentée depuis des décennies par les informations générées par la prise en charge de la totalité des consommations de soins et hospitalisations en France. Ces données par définition sont sensibles : elles ont donc fait l'objet de procédures d'anonymisation. Cependant chacun – et particulièrement le chercheur- sait que ces procédures ne sont pas fiables à 100 %<sup>130</sup> Répondant à cette demande d'ouverture de ce gisement de données (sensibles) que les auteurs du rapport qualifient de « bien commun de recherche en santé publique », la Caisse Nationale d'Assurance Maladie a donc mis à la disposition, pour les chercheurs et les organismes publics en charge de la santé publique, un échantillon randomisé de bénéficiaires (un dossier sur 100) et non la totalité des données. Parmi les principes parfois contradictoires qu'elle veut mettre en œuvre, la Caisse a décidé d'ouvrir l'accès aux lots de données anonymes en distinguant la publication (gratuite) et des extractions

---

<sup>129</sup> Schéma général du système d'information de l'Assurance maladie. Voir Sur la proposition d'ouverture et de partage de ces « données publiques » : *Rapport sur la gouvernance et l'utilisation des données de santé* par Louis Bras et André Loth, Septembre 2013.

<sup>130</sup> Ce qui implique que le chercheur doit aussi en informer le patient.

ou des tableaux de bord à façon (payants). Elle entend développer aussi sa politique d'ouverture, avec pour critères l'intérêt public (notamment de la recherche), la qualité du protocole, le besoin d'accéder aux données, la sécurité des procédures et la qualité du demandeur. On comprend ainsi comment à la fois les données publiques même sensibles peuvent être utiles pour les chercheurs et faire l'objet de négociations spécifiques à cette communauté, tout en respectant le droit des personnes.

Veiller à la qualité des données et valider les traitements au regard des méthodes scientifiques fondées sur la preuve représente aussi un autre grand défi. Les Big Data – qui se réfèrent à des volumes d'information massifs, complexes et susceptibles d'être connectés- peuvent améliorer en effet notre compréhension et prédiction (par *machine learning*) des comportements susceptibles d'affecter la santé et accélérer le cycle de la connaissance à la diffusion. Cependant « Big error » peut menacer « Big data »<sup>131</sup> : Dans cet article, les auteurs demandent la réplication systématique des résultats et des études collaboratives épidémiologiques à large échelle pour tester les outils prédictifs et passer des corrélations obtenues aux *vraies* causalités.

### **Les contraintes concernant le traitement de données personnelles**

Le chercheur qui utilise des données à caractère personnel - que la personne soit identifiée ou identifiable indirectement par profilage ou ciblage- se trouve confronté d'abord à des contraintes juridiques fortes. Or dans certains cas le modèle de traitement de données à partir duquel la politique de protection de la vie privée avait été élaborée dans la loi dite *Informatique et Libertés* du 6 janvier 1978 peut paraître trop contraignante dans les nouveaux contextes du *big data* par rapport aux objectifs de la recherche, voire obsolète comme le signalent les chercheurs interrogés<sup>132</sup>. En effet la circulation rapide et ouverte des données entre chercheurs bouleverse l'ordre des procédures et rend les flux de données relativement autonomes par rapport à leurs sources ou leurs auteurs. Il devient souvent impossible de respecter le principe de finalité (les hypothèses ne sont pas élaborées *a priori*), le principe de proportionnalité (les données nécessaires ne sont pas connues avant leur utilisation) et le principe de non conservation (on ne détruit pas les données à la fin d'une recherche à cause de l'accès ouvert et de la réutilisation). Signalons cependant que toutes les données archivées font l'objet d'une exception aux fins de recherche, les délais de la recherche dépassés.

Ce besoin de ressources existe dans beaucoup de disciplines. Prenons l'exemple de la vision par ordinateur, qui a pour objet de reconnaître automatiquement des scènes visuelles. La reconnaissance de visages est un des domaines de la vision par ordinateur, qui peut servir à des applications de biométrie. Le problème se trouve alors rendu plus

---

<sup>131</sup> M.J. Khoury & J.P.A. Ioannidis, « Big data meets public health » in *Science*, 26 November 2014, vol. 346, 6213 p.1054-1055.

<sup>132</sup> Une vingtaine d'auditions ont été organisées au CNRS.

complexe du fait du Droit à l'image, l'identité d'une personne pouvant être retrouvée à partir de son visage. Le cas a été soulevé récemment dans le cadre de l'organisation d'une campagne d'évaluation de systèmes de reconnaissance de visages. Après de longues négociations, la CNIL a donné son accord à la condition que ces données ne soient pas conservées au-delà du terme du projet, sauf demande de prolongation nécessitant de refaire un dossier, et accord de la CNIL. Cela entraîne donc le paradoxe d'interdire d'expérimenter d'autres systèmes sur les mêmes données pour en comparer les performances, alors que cela constitue une démarche scientifique normale. Un même paradoxe existe dans la demande faite à Google par le G29 alliant les différentes « CNIL européennes », de ne pas conserver plus de trois mois les images de Google Street View servant au développement d'algorithmes de « floutage » automatique des visages. Il paraît paradoxal de limiter l'utilisation de ces images, alors que le but est de permettre le développement d'algorithmes de floutage de visage les plus efficaces possibles en mode opérationnel. Il semble donc que l'on mélange dans ce cas les besoins de la recherche et les contraintes d'une utilisation opérationnelle, qui sont de nature différente. Il serait utile de réfléchir à l'introduction en droit français de la notion de *fair use* utilisé en Common law dans le domaine du droit d'auteur. On peut traduire cette expression par « usage loyal » ou « usage acceptable ». Transposé au domaine de la recherche, il s'agirait de laisser le législateur et le juge, définir un ensemble de règles de droit, qui essaieraient de prendre en compte à la fois les intérêts de la recherche et l'intérêt public, en autorisant certains usages qui seraient, autrement, considérés comme illégaux. Une telle loi consacrant un *Fair research use* faciliterait le développement des recherches nécessitant sous certaines conditions l'utilisation de données protégées.

Plus précisément il est parfois difficile d'appliquer les principes de base du traitement des données personnelles comme informer les personnes sur le devenir et l'utilisation des données, ou obtenir leur consentement. Ainsi par exemple la démarche du chercheur peut imposer d'obtenir des informations à l'insu de la personne objet de son enquête. Il serait nécessaire de prévoir des principes à respecter s'il n'y a pas consentement, comme l'engagement à informer *a posteriori* cette personne. Dans d'autres disciplines, les données appartiennent à des jeux de données non identifiants mais leurs combinaisons peuvent provoquer des ré-identifications qui nécessitent des procédures pour le changement éventuel de « statut », identifiant ou non, des données. De même la protection de l'« anonymat » impose d'autres types de garanties (engagement d'institutions et de chercheurs sur la non-utilisation de la caractéristique identifiante des données, en cas de ré-identification). En tout état de cause, le chercheur doit informer les sujets de l'impossibilité de garantir l'anonymat strict des données, et leur donner l'assurance que tout sera fait pour veiller aux mesures de protection de leurs droits.

Enfin, l'enjeu de l'apport de chacun passe par une identification unique, non ambiguë et persistante des chercheurs, qui leur attribuera la reconnaissance de ces contributions donc de ses responsabilités (voir initiative ORCID par exemple).

## **Appel à la vigilance des chercheurs sur les données partagées**

Face à cette dynamique de circulation des données relayée par leurs autorités de tutelle et par leur communauté, les chercheurs doivent :

- prendre conscience de leur responsabilité individuelle, déontologique<sup>133</sup> et éthique, vis à vis de la communauté à laquelle ils appartiennent ;
- avoir connaissance des engagements internationaux des institutions dont ils dépendent ;
- participer à la définition de principes éthiques propres à leur discipline dans le domaine du *data sharing* et du *big data* en général.

Le data sharing avait été lancé par des communautés de chercheurs et relevait de la *soft law*, c'est-à-dire de règles de conduite non contraignantes. Aujourd'hui les engagements institutionnels auxquels le chercheur public est soumis sont devenus contraignants depuis les politiques d'Open data précitées<sup>134</sup>. Il lui convient d'apprécier au cas par cas les implications de ces politiques face aux dimensions éthiques de la recherche et d'appliquer l'ouverture des données *raisonnablement*.

D'une façon générale, le chercheur public doit poursuivre un idéal de partage et d'échange entre pairs et participer à la diffusion des données obtenues sur fonds public tout en respectant les exceptions de nature contractuelle dans lesquelles il serait engagé. Inversement, les modèles d'accords de consortium impliquant des partenaires public et privés (notamment dans les Pôles de compétitivité) sont souvent trop restrictifs en matière d'ouverture de données: ils doivent désormais être négociés en amont par les chercheurs publics de façon à ne pas conduire à une confiscation des données non exploitées par les partenaires privés en fin de projet.

Bien que soumises aux principes de partage et d'ouverture, les données ne sont pas *libres*: structurées ou non en banque de données, elles représentent une valeur marchande (économique) ou non marchande (éthique). Comme les publications, elles relèvent de plus en plus du droit d'auteur. Il est donc nécessaire que le producteur définisse les restrictions ou exceptions de façon explicite aux chercheurs ré-utilisateurs. De plus, quand elles sont sensibles, les données doivent suivre strictement les politiques de protection des données personnelles tout au long de leur traitement pour ne pas mettre en difficulté les usages ultérieurs.

---

<sup>133</sup> "Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences", *Committee on Responsibilities of Authorship in the Biological Sciences, National Research Council*. National Academy of Sciences.

<sup>134</sup> Cette directive PSI est transposée par une ordonnance complétée d'un Décret le 30/12/2005 pris en l'application de la loi CADA (*Commission d'Accès aux Données Administratives*) de 1978.

Les chercheurs ont découvert que l'ouverture des données - mais aussi des logiciels, ontologies et métadonnées qui en permettent l'exploitation-, impliquait une nouvelle responsabilité : celle d'être particulièrement soucieux de la qualité des informations et des données *offertes* ainsi que de la clarté de la documentation qui les accompagne. Pour permettre à d'autres de répliquer ou de réutiliser des données, il est nécessaire de vérifier le caractère intègre et interopérable des données, l'identification de leurs sources, leurs dates de recueil ou de traitement, ainsi que l'examen détaillé des différentes étapes de la constitution de dépôts de données : collecte, classification, standardisation, mise à disposition, réutilisation, conservation, destruction. Les questions relatives au droit à l'image, à la confidentialité et à la sécurité soulèvent ainsi des questions juridiques et éthiques qui, si elles existaient avant le *data sharing*, deviennent plus délicates à interpréter à l'heure d'un partage généralisé et international des résultats de la recherche.

Ainsi l'organisation, la maintenance et l'accessibilité de données interopérables et de qualité deviennent des moments fondamentaux pour garantir l'intégrité des données scientifiques à l'heure numérique et créent de nouvelles responsabilités juridiques et éthiques *entre* les chercheurs<sup>135</sup>. A qui appartiennent les données ? Au laboratoire ? Aux chercheurs ? Aux agences ? Dispose-t-on de tous les moyens (logiciels, algorithmiques) pour les utiliser et les reproduire ? Le changement d'échelle impose également de véritables infrastructures internationales de données, ce qui complique encore leur gouvernance. Ainsi le droit des données et des banques de données n'est pas homogène, ni même harmonisé au niveau européen. Il y a donc un déplacement des centres de gravité de l'activité scientifique qui appelle à poursuivre une réflexion non seulement au plan stratégique mais aussi au plan éthique. Comme il a été justement remarqué dans un article sur ce sujet en 2014, « la tendance actuelle vers la valorisation marchande des résultats scientifiques, avec l'accent mis sur la propriété intellectuelle, va dans le sens opposé à celui du partage des données »<sup>136</sup>. C'est sur ces contradictions pratiques auxquelles est confronté le monde de la recherche publique que cet avis entend créer une alerte.

### III - Recommandations

1. Le COMETS rappelle que le CNRS est signataire de la Déclaration de Berlin (2003), comme la plupart des grands organismes de recherche internationaux. Celle-ci engage les chercheurs dans le mouvement mondial de *partage de données ouvertes* (*data sharing*). Le COMETS invite tous les chercheurs du CNRS à s'associer à ce mouvement dans le respect des pratiques propres à chaque discipline

---

<sup>135</sup> « Ensuring the integrity, accessibility and stewardship of research data in the digital age », *Report of Committee of Science, Engineering and public policy*, Washington, The National Academies Press, 2009.

<sup>136</sup> M. Vito, « Partageons nos données », *Le Monde*, 28 mai 2014.

2. La contribution aux travaux de partage des données doit être reconnue dans l'évaluation et les décisions de promotion des chercheurs. Pour faciliter cette reconnaissance, le COMETS préconise que des indicateurs pertinents soient créés et qu'une rubrique sur ces activités soit ajoutée dans le rapport d'activité et la fiche annuelle d'activité des chercheurs

3. Les chercheurs et les personnels du monde de la recherche doivent être formés à la gestion éthique de données (ce qu'on appelle « *privacy, accuracy, property, accessibility* ») et informés sur les *règles de bonne pratique*, ainsi que sur les règles juridiques concernant le partage responsable de données incluant le recueil Les chercheurs et les personnels du monde et proportionné des données personnelles ou susceptibles de ré-identifier des personnes.

4. Les pratiques de partage de données doivent être encouragées dans les politiques de publication auprès des revues scientifiques et dans l'organisation des colloques, tant vis-à-vis des auteurs que des évaluateurs. Le COMETS recommande aux auteurs de refuser que leurs données fassent l'objet d'une tarification spéciale de la part des éditeurs scientifiques et/ou d'une exploitation ultérieure séparée par ceux-ci (revente ou *paywall*)

5. Le COMETS préconise que les archives ouvertes HAL soient privilégiées pour le dépôt des données sur lesquelles s'appuient les publications des résultats de la recherche et que le chercheur puisse choisir, par des licences ouvertes telles Creative Commons, les conditions de leur réutilisation

6. Il recommande au CNRS de veiller à l'existence d'infrastructures pérennes permettant de gérer les plateformes de données dans la durée au niveau d'une équipe, d'un laboratoire ou d'un réseau. Il suggère que le CNRS incite ses chercheurs à participer à la mise en place et à l'activité d'instances internationales pour traiter des métadonnées utilisant des identifiants uniques et persistants pour ces données

7. Les coûts du partage des données, l'aide à la création et à la maintenance d'entrepôts de données ou de bases de données, la construction et la maintenance de plateformes multi-usages ou d'archives ouvertes doivent être pris en compte dans l'attribution par l'organisme de moyens appropriés (subventions, bourses...) aux équipes, cela sans préjudice d'une tarification d'un traitement de données à la demande et personnalisé.

8. Le COMETS préconise qu'une réflexion soit menée avec la Commission Informatique et Libertés et le correspondant Informatique et libertés du CNRS, ainsi qu'avec ETALAB, afin de tenir compte de la spécificité des données et de leur traitement dans le monde de la recherche. Il suggère la création d'un Comité consultatif

d'administration des données scientifiques, impliquant diverses disciplines dans cette réflexion

9 Il souligne enfin l'importance pour les communautés scientifiques de recenser de façon ouverte et collaborative les obstacles juridiques au partage éthique des données (propriété intellectuelle des données et statut *sui generis* des banques de données), afin de promouvoir de véritables communs scientifiques, d'intégrer le concept de *fair use research* et de participer à l'adaptation du droit des données aux *intérêts légitimes* de la recherche.

## 5. Guide d'entretien des auditions - contribution de la Recherche aux thèmes du projet de loi numérique

### I- Contexte

Le numérique transforme en profondeur les modes de production et de diffusion des résultats scientifiques : données, publications, analyses sont accessibles désormais sur des plateformes. Cette disponibilité du matériau scientifique contient un potentiel d'exploitation et de partage des connaissances dont la loi doit pouvoir définir les conditions, les modalités, les limites.

### II- Les propositions du CNRS : conditions, modalités, et limites du partage de l'Information Scientifique

#### A. Conditions

Le libre accès aux résultats de la science a pour conditions principales l'abolition des limites que peut poser le droit éditorial (droit de la publication, droit des auteurs), au regard de l'exploration des corpus numériques de publications ou de données.

**Question 1 : Quelle est votre opinion sur les adaptations nécessaires du droit de la publication (éditeurs et/ou auteurs) et de l'exploration des corpus (techniques de Text & Data Mining, API, ...) ?**

#### B. Modalités

Les modalités de partage des résultats de la science doivent, à l'heure numérique, assimiler de nouvelles contraintes : le partage des résultats entre acteurs de la recherche publique d'une part et d'autre part entre usagers et bénéficiaires de la science publique.

**Question 2 : Comment tracer la frontière entre la captation légitime et la captation abusive des résultats disponibles sur une plateforme de science publique, comment protéger ces résultats ?**

#### C. Limites

Les plateformes de science contiennent aujourd'hui une IST de forme, de contenu et de statut juridique très hétérogène. Cette hétérogénéité nuit à la visibilité des plateformes de science. De la même façon qu'il existe aujourd'hui une notion de données d'intérêt générales (choix récent de la Ministre de l'économie numérique), il y a lieu de réfléchir à

la dénomination des plateformes de science publique et, éventuellement, à un régime juridique spécifique.

**Question 3 : Doit-on aller vers une dénomination d' « infrastructure essentielle » pour les grandes plateformes d'amont de la recherche, dans les cas où ces plateformes occupent une fonction inédite et irremplaçable ?**

## 6. Compte-rendu d'auditions

### a. UPMC, Jean CHAMBAZ et Paul INDELICATO, 9 juin 2015

#### Participants

Pour l'UPMC :

- Monsieur Jean CHAMBAZ, Président
- Monsieur Paul INDELICATO, Vice-président Recherche et innovation

Pour le Cabinet BENSOUSSAN :

- Maître Alain BENSOUSSAN, Avocat à la Cour spécialisé en droit des technologies avancées
- Maître Laurence TELLIER- LONIEWSKI, Avocate, directrice du pôle Propriété intellectuelle

Pour la DIST CNRS :

- Renaud FABRE, Directeur
- Charlotte AUTARD, Chargée du projet IA ISTEEX

#### Démarche

La démarche de rédaction d'un Livre blanc « Science Numérique et Droit », portée par le CNRS en collaboration avec le Cabinet Alain BENSOUSSAN, s'inspire des problématiques soulevées par la sécurisation juridique du Projet d'Investissement d'Avenir ISTEEX ANR-10-IDEX-0004-02 ([www.istex.fr](http://www.istex.fr)) qui soulève de nombreuses questions juridiques par ses possibilités de TDM, d'interdisciplinarité, d'agrégation de contenus et son objectif de rendre explorable et traversable ses bases de données.

Ce Livre Blanc a pour objectif de **proposer un cadre juridique pour les données scientifiques afin de répondre aux préoccupations des communautés scientifiques** (publications et usages des données, droit des plateformes de Science, fouille de texte, ...) en étant force de proposition dans la **participation au projet de loi numérique**.

Dans un premier temps, le cabinet BENSOUSSAN et la DIST réuniront des groupes de travail et mèneront des auditions pour **recueillir les usages de la communauté et l'état de l'art des pratiques de valorisation**.

Le recueil des usages et de l'état de l'art permettront au CNRS et au Cabinet BENSOUSSAN d'établir une **grille de pertinence des analyses et des pratiques qui sera comparée au cadre normatif actuel** afin d'évaluer les écarts et **élaborer des propositions de lois**.

### Droit d'auteur et Droit de la Science

A ce jour, aucun texte, en dehors du droit d'auteur, n'a préempté la Science. La Science a été emprisonnée par le droit d'auteur dans son droit, dans son évaluation, dans sa diffusion. Le droit d'auteur n'intègre pas seulement la science passée (article) mais aussi la science future (analyse des résultats).

Le droit d'auteur a été conçu pour privatiser la forme de la création et permettre aux éditeurs de trouver denier et de participer à la culture par sa diffusion. Les choses se sont plus tard inversées et l'auteur est aujourd'hui soumis à cet intermédiaire qui récolte 90% de la valeur de sa création.

Le droit d'auteur apparaît aujourd'hui inadapté à la démarche scientifique numérique, l'objectif du Livre blanc Science Numérique et Droit est de proposer des solutions pour la création d'un nouveau droit de la Science.

Il est très difficile d'isoler la réflexion sur le droit d'auteur pour les données scientifiques de la problématique générale du droit d'auteur qui fait l'objet de longue date de débats vigoureux. Il pourrait être plus productif de se concentrer sur la notion de donnée et du droit de la donnée, problématique distincte du droit d'auteur et qu'on pourrait faire ainsi émerger sans malentendu.

La donnée scientifique évolue, et bien que la science des données ne se substituera pas à la Science en tant que démarche, la donnée et plus encore sa réutilisation se place au cœur de cette nouvelle démarche.

Certains domaines de recherche s'appuient sur la collecte d'une masse importante de données sur un sujet défini ou dans le cadre d'activités humaines (données publiques, gouvernementales, de santé, ...). Ces données ont une valeur dans l'activité de recherche en cours mais peuvent aussi être réutilisées pour d'autres projets de recherche. Il faut déterminer à qui appartiennent ces données collectées : à la personne qui les recueille et les étiquette, à l'organisme, le laboratoire qui les stocke ?

La donnée doit être considérée comme un bien public libre de réutilisation. Si nous parvenons à définir un statut particulier de la donnée scientifique (ou donnée à matière scientifique) le droit d'auteur serait alors caduc.

### **Statut particulier de la donnée**

Jean CHAMBAZ et Paul INDELICATO différencient deux types de données :

- Les données observables
- Les données construites : recueil de plusieurs données sur lesquelles sont appliqués des méthodes ou des algorithmes particuliers qui permettent de retirer des observations à partir d'un regroupement de données particulières.

La démarche intellectuelle qui donne du sens aux données collectées reste la propriété intellectuelle du chercheur et de l'organisme financeur mais la donnée produite au cours de cette recherche n'a aucune raison pour ne pas devenir un bien public avec un statut particulier permettant sa réutilisation. Un statut particulier des données pourrait stimuler l'activité de recherche au service de la société dans son ensemble.

Les comportements autour de la collecte des données sont dépendants des disciplines. Pour certaines la collecte et le partage de données en masse sont essentiels pour conforter leur discipline mais pour d'autres, comme en science expérimentale, la réutilisation des données collectées n'est pas toujours pertinente.

Paul INDELICATO précise que, selon les disciplines, les données scientifiques, pour être pertinentes, doivent être accompagnées de la méthode, de la procédure qui les a générées.

### **Bibliothèque Nationale (ou Européenne) des Données**

Aujourd'hui les données collectées lors de projets de recherche sont pour la plus part en déshérences ou déposées sur des bases de données ou dans des annuaires dispersés. Si nous envisageons un dépôt des données dans une Bibliothèque Nationale (ou Européenne) définie il faudra aussi y déposer les publications associées.

Une Bibliothèque Européenne aurait une valeur supérieure pour les disciplines, comme la recherche médicale, pour lesquelles un partage des données dans un périmètre national n'est pas pertinent.

Le délai pour la mise en place d'une Bibliothèque Européenne des données pourrait s'avérer très long, Jean CHAMBAZ propose d'insérer cette possibilité dans la réflexion en cours autour du plan JUNKER et s'associe à la réflexion du CNRS dans ce sens.

### **Éléments nouveaux de droit pour la démarche scientifique**

Aujourd'hui la morale, la religion, l'éthique, n'influent qu'inégalement sur les nouveaux droits dans le domaine scientifique.

Si nous défendons la création d'un entrepôt global de données libres, il faut que celui-ci soit justifié comme bien commun universel. La possibilité de faire du TDM sur les données scientifiques est essentielle. Si cette possibilité dépend des éditeurs, les chercheurs seront soumis à un contrôle important qui ne pourra pas être maîtrisé.

La solution aux dangers que pourrait représenter l'utilisation abusive du TDM sur des larges bases de corpus (aspiration, piratage des bases de données, ...), crainte actuelle des éditeurs, ne se trouve pas dans le contrôle des accès mais dans la formation des utilisateurs. Les établissements, les organismes, les laboratoires, doivent sensibiliser et former leurs personnels, chercheurs, étudiants à la pratique du TDM comme enjeu scientifique, économique et culturel. Pour permettre le TDM les données doivent être étiquetées intelligemment pour être réutilisables et largement partagées.

Paul INDELICATO souligne qu'il faut attacher au dépôt obligatoire des données, les conditions de réutilisation de celles-ci.

### **Valorisation économique des données**

Une donnée est brute : le recueil de la donnée n'est pas neutre et se fait dans un cadre précis mais la donnée ne contient pas en elle-même toutes les idées que l'on pourra avoir pour la réexploiter. On pourrait considérer que la personne qui a recueilli la donnée doit bénéficier de toute l'exploitation possible de celle-ci. Ce constat faciliterait le dépôt des chercheurs dans une Bibliothèque des données et apporterait une vaste plus-value aux établissements/organismes qui financent la recherche. Cela permettrait d'apporter de la valeur au recueil, au traitement, à l'étiquetage des données dans le but de les rendre exploitables.

Dans le cas où le chercheur souhaiterait une période d'embargo avant le dépôt des données collectées (données construites) une notion de réservation scientifique et/ou économique des données pourrait être envisagée.

Jean CHAMBAZ indique que la réutilisation libre de droit des données par la communauté scientifique pourrait faire consensus mais il faut prévoir un droit qui encadrerait la réutilisation des données par les entreprises et les autres bénéficiaires d'usages de la recherche publique. Il faut trouver le juste équilibre entre la réutilisation économique et scientifique des données.

Exemple : La récolte de données liées aux comportements des voyageurs SNCF pouvant servir aussi bien à une équipe scientifique pour une étude sociologique qu'à la SNCF elle-même pour optimiser de nouveaux services.

Les entreprises pourraient avoir accès aux données pour le développement de leurs activités sous condition d'un retour organisé vers la société dans son ensemble. Une clause d'accès aux données par les entreprises privées pourrait prévoir un reversement sur la valorisation de la donnée afin que la Bibliothèque Nationale (ou Européenne) des données puisse s'autofinancer par le biais par exemple d'une taxation tirée de l'utilisation des données.

### **Instrument juridique des données**

En termes d'acceptabilité sociale, les données pourraient être libres de réutilisation lorsqu'elles sont réutilisées par la communauté scientifique. Pour une réutilisation par les entreprises, une différenciation devra être faite entre données observables (libres de réutilisation) et données construites. Pour les données construites l'équipe de recherche pourrait utiliser un instrument juridique (brevet, licence d'utilisation, ...) pour lui permettre de maximiser le ROI (retour sur investissement) sur les données qu'elle aura récolté et étiqueté.

Cet instrument juridique pourrait justifier, dans un nombre de cas limitativement énumérés, une période d'embargo de 5 ans (notion d'amortissement). Cette durée doit être testée pour que le système soit assez incitatif pour motiver les chercheurs dans cette démarche de dépôt de leurs données.

### **Données éthiques**

Le traitement et l'étiquetage des données créent une nouvelle mission du chercheur public. Celui-ci doit non seulement faire de la science mais aussi déposer des données traitées ce qui représente un coût supplémentaire en temps et en moyens humains.

Le chercheur doit donc non seulement faire de la recherche mais aussi assurer une fonction de reversement numérique vers la communauté publique.

Alain BENSOUSSAN indique qu'un nouveau concept est actuellement en train de naître : celui de la donnée « éthique ». Une donnée loyale, recueillie loyalement, une donnée autorisant un développement durable.

Une donnée pour le développement durable ne doit pas être détruite mais au contraire suivie et déposée.

Les communautés scientifiques doivent être sensibilisées et formées à un comportement de développement durable à l'égard des données qu'elles récoltent pour leurs projets de recherche.

Il faut que les chercheurs acceptent que le traitement et le dépôt de leurs données ne représentent pas un coût supplémentaire mais une valeur ajoutée nouvelle. Il faut valoriser le protocole de traitement et de dépôt des données dans le travail des chercheurs. Il faut que celui-ci devienne une part intégrante de leur métier tout comme le sont l'évaluation et la publication de leurs recherches.

## **b. Conseil Scientifique du CNRS, Bruno CHAUDRET, Claire LEMERCIER, François BONNAREL, François TRONCHE, 30 juin 2015**

### **Participants :**

Pour le groupe de travail du Conseil Scientifique :

- François BONNAREL, Ingénieur CNRS, Centre de données astronomiques de Strasbourg
- Bruno CHAUDRET, Président du Conseil Scientifique du CNRS
- Claire LEMERCIER, Directrice de recherche au CNRS en histoire, Centre de sociologie des organisations, Paris
- François TRONCHE, Directeur de recherche au CNRS, Institut de Biologie Paris-Seine, CNRS, INSERM, UPMC

Pour le Cabinet BENSOUSSAN :

- Maître Alain BENSOUSSAN, Avocat à la Cour spécialisé en droit des technologies avancées
- Maître Laurence TELLIER- LONIEWSKI, Avocate, Directrice du pôle Propriété intellectuelle

Pour la DIST :

- Renaud FABRE, Directeur
- Charlotte AUTARD, Chargée du projet IA ISTEEX

### **Introduction de la démarche :**

Dans le cadre du Projet IA ISTEEX, à la demande de la DIST, la DAJ du CNRS a missionné un Cabinet juridique pour accompagner le projet. Le Cabinet BENSOUSSAN a été sélectionné pour accompagner le Comité Exécutif dans la sécurisation juridique du projet. Cette collaboration a d'ores et déjà permis d'associer les réflexions portées par ISTEEX pour le projet de loi numérique.

Cette collaboration a amené le Cabinet BENSOUSSAN à accompagner la DIST du CNRS pour la rédaction d'un Livre Blanc avec pour objectif construction d'un ensemble de références, largement repris par le projet de loi numérique autour de l'Open Access et du principe de fouille de texte (TDM pour *Text and data mining*).

La stratégie numérique du gouvernement reprend les conclusions du CNRS à l'attention de l'ESR en matière numérique pour la science et réaffirme l'Open Access et le TDM comme éléments clés de cette démarche.

Le projet de loi numérique sera présenté avant la fin du mois de juillet. Le Livre Blanc a donc toute son importance. Les conclusions tirées du recueil des pratiques des usages auprès du CS et des autres partenaires auditionnés (CNN, UPMC, COEPIA, Université de Strasbourg, Comité Exécutif ISTEEX) viendront s'inscrire dans le projet de loi au moment de la discussion parlementaire. La DAJ et la DIST, accompagnées du Cabinet BENSOUSSAN, s'associeront pour présenter les positions de la communauté de recherche appuyées par le CS du CNRS.

Les premiers retours du Conseil Scientifique devront être transmis à la DIST (pour transmission au cabinet BENSOUSSAN) avant la fin du mois d'août 2015 pour un délai suffisant de construction du Livre blanc avant le dépôt du projet de loi à la session parlementaire début octobre.

La contribution du CS permettra de centrer le débat autour du Livre Blanc lors des prochaines auditions. Des modifications mineures pourraient être apportées au texte proposé par le CS suite à sa prochaine réunion plénière le 24 ou 25 septembre.

### **Présentation des objectifs**

La collaboration du Comité Exécutif ISTEEX, piloté par la DIST, de la DAJ du CNRS et du Cabinet BENSOUSSAN a pour objectif de trouver un cadre juridique au projet ISTEEX. Une réflexion préalable a été menée lors de deux réunions du Comité d'éthique du CNRS (COMETS).

La consultation autour du projet ISTEEX a permis l'émergence de deux projets :

- Un projet « restreint » autour du projet ISTEEX et du droit des plateformes et du TDM (utilisation plus ou moins libre des données scientifiques).
- Un projet « large » pour un cadre juridique plus général de l'IST : un Droit de la Science. Toujours au travers d'une démarche scientifique d'utilisation des données.

Le groupe de travail du CS est invité aujourd'hui à donner son avis sur ces deux projets « restreint » et « large ».

Sur le projet « large », la DIST et le Cabinet BENSOUSSAN ont d'ores et déjà recueillis les avis de deux experts de premier rang : Laurent CYTERMANN, Maître des requêtes au Conseil d'Etat et Alain ABECASSIS, chef du service de la coordination stratégique et des territoires.

La démarche du Livre Blanc proposée ce jour s'intéresse au recueil des avis des membres du CS en tant que scientifiques pour comprendre les besoins juridiques liés à leurs domaines de recherche.

## Publications et données

### Définition des données

Les données peuvent être décomposées en 5 niveaux :

- **Brutes** : avant tout travaux humains et en l'état disponibles.
- **Instrumentalisées** : obtenues à travers un instrument donné (exemple : un télescope en astronomie). Celui qui contrôle l'instrument contrôle la donnée.
- **Interprétées analytiquement** : résultent de calculs, de traitements. La propriété devant être définie entre celui qui fournit la données et celui qui fournit l'algorithme.
- **Données scientifiques** : interprétées à partir d'un cerveau humain ( $E=MC^2$ ). Mise en relation de données. Ce sont souvent ces données qui font l'objet d'une interprétation dans le cadre d'une théorie.
- **Données sur les données ou métadonnées** : ensemble des éléments, des relations entre les données qui permettent parfois d'interpréter la données elle-même.

au cœur de plusieurs utilisations :

- **Open Data** (gouvernance ouverte des données)
- **Open Access** (possibilité d'accès librement ou non à la données, avec pour certains domaines de recherche une période d'embargo)
- **Open Process**
- **Open Format**
- **Open Use**

## - **Open Business**

Trouver le bon niveau de définition de la donnée peut parfois s'avérer compliqué comme par exemple pour la photographie d'une œuvre d'art.

Au premier abord cette photographie est une donnée instrumentalisée mais l'objet pris en photo existe en dehors de la photographie, dispose de ses propres droits, est une donnée brute. Des conflits juridiques existent sur la réutilisation de données brutes représentées sous forme d'image.

De nombreux domaines scientifiques travaillent sur des données qui existent sans être issues des sciences, des données brutes, dont il faut aussi pouvoir déterminer les usages.

### **Différentiation données et articles**

Un article, rédigé dans le cadre de travaux de recherche subventionnés par l'Etat, est financé par des fonds publics. A partir des travaux menés les chercheurs communiquent selon un standard international : l'article scientifique. Le chercheur devient alors propriétaire de cet article en tant qu'auteur protégé par le droit de propriété artistique (propriété morale et patrimoniale).

Le chercheur paie pour publier et paie de nouveau pour accéder aux revues dans lesquelles ses articles et ceux de ses collègues sont publiés. Et cela le plus souvent sans possibilité de TDM.

Suite à ce constat la question suivante est posée : le système actuel est-il encore de nature à satisfaire l'ouverture / le partage de la science ? La privatisation de l'article est-elle encore le moyen de diffusion adapté ? Le droit d'auteur est-il encore adapté à la science ?

### **Recueil des usages**

**Pratiques et partage : principe de Fair Use (TDM)**

**Pratiques et partage : principe de Fair Use (TDM)**

**En biologie**, l'édition numérique est généralisée et la recherche d'articles se fait via des plateformes développées par des institutions académiques, comme le National Institute of Health (NIH, USA) permettant d'accéder librement aux résumés d'articles. L'accès à la totalité des articles est généralement payant, le transfert du copyright à l'éditeur étant

une pratique encore courante. Il est à noter que le NIH, s'y est opposé et propose en accès libre, via PubMed, une version non formatée de tout article publié par un éditeur décrivant des travaux financés par le NIH. L'Open Access se développe rapidement.

La production d'articles de recherche diffusés sous forme numérique, s'accompagne non seulement de la production de données brutes numériques mais également de matériel (collections d'ADN, anticorps, cellules, appareillages, logiciels...) et d'organismes vivants (microorganismes, lignées de souris... sélectionnées ou génétiquement modifiées) utilisés dans la démarche scientifique présentée dans l'article de recherche. Selon la nature de ces productions, leur propriété relève du droit d'auteur ou des pratiques de l'organisme de tutelle. Si on prend en exemple, l'établissement de souris génétiquement modifiées, la propriété relève en pratique de la tutelle et d'éventuels revenus sont partagés entre inventeurs, tutelle et laboratoire. Lorsqu'un article publié décrit une recherche générant des données brutes numériques ou du matériel biologique, l'éditeur demande généralement un engagement de l'auteur pour la donation de ces données, ou de ce matériel, à tout chercheur d'organisme public en faisant la demande. Cette pratique essentielle pour la validation des résultats par reproduction est facilitée par l'existence de plateformes internationales pour le libre accès aux données et la distribution du matériel biologique.

Le TDM sur les textes n'est pas prioritaire pour la plupart des domaines de la biologie en termes d'outil de découverte, même s'il l'est en termes de collecte documentaire. Le TDM sur les données prend une place de plus en plus importante. Le libre accès à ces données est répandu, à l'instar de ce qui s'est passé pour le génome humain.

Au-delà de la question des données en masse, plusieurs éditeurs dont *Nature* songent à mettre en place un système permettant d'accéder, via leur site, aux données brutes ayant conduit à l'élaboration des figures d'un article. Si cela permettra au lecteur de s'assurer de la bonne interprétation des résultats, la question se pose de la propriété de ces données et de leur cession éventuelle. Cette extension prévisible des éditeurs vers les données est inquiétante et pourrait représenter une nouvelle barrière pour l'accès à la Science.

Dans un second temps, lorsque l'on compare les usages, l'**astronomie** semble être le domaine de recherche le plus organisé en matière de données et d'interopérabilité de l'accès à ces dernières.

Un consensus a été trouvé pour mettre en place une période d'embargo pour la diffusion et la réutilisation des données instrumentalisées. Cette période d'embargo est réservée

jusqu'à la publication de l'article et après publication selon un délai qui fluctue pour la libre réutilisation.

François BONNAREL indique que l'astronomie est un secteur de recherche où la valeur commerciale des données est le plus souvent nulle ce qui facilite grandement les partages de données et où l'échange de ces dernières est nécessité par la très grande variété des techniques d'observation (liées aux différents domaines de longueur d'onde) potentiellement disponibles pour un même objet.

François TRONCHE précise qu'**en biologie** l'accès aux données est gratuit pour la communauté académique avec parfois une période d'une ou deux années de gel, indépendamment de la publication (cette période d'embargo démarrant alors à la découverte de la donnée).

A ce jour il n'existe pas de document spécifique qui détermine cette période de gel. **En astronomie** la période « propriétaire » est définie autour de l'instrument ou autour du projet complet (plusieurs instruments).

La période d'embargo permet au chercheur un droit d'exclusivité sur ses résultats et ses données à des fins de publication.

Nous sommes donc dans une économie double : une économie du don (données disponibles gratuitement aux chercheurs d'une même communauté) et une économie de marché (accès aux publications).

Claire LEMERCIER précise que les chercheurs doivent travailler parfois avec de grands groupes tels qu'Amazon pour avoir accès à des espaces de stockage importants (exemple donné par Pierre BINETRUY lors d'une précédente réunion) mais toujours en contrepartie de quelque chose (paiement, accès aux données, ...). Ces grands groupes disposent également pour la plupart de laboratoires de recherche.

Si nous envisageons que les données soient ouvertes à la communauté scientifique comment déterminer si elles ne seront pas utilisées à des fins lucratives par ces laboratoires de recherche privés? Dans quelles mesures pouvons-nous déterminer qu'une réutilisation est à but Scientifique, lucratif ou autre (par exemple à des fins pédagogiques ou de formation).

Si un historien décide d'utiliser des œuvres d'art à des fins de recherche, de formation, il faut pouvoir différencier ces usages avec l'utilisation de données pour l'édition de « beaux livres » d'œuvres d'art (à but lucratif).

On pourrait envisager, pour définir les champs de possibilité de réutilisation des données, d'inventer un principe de *Fair Use*. La réutilisation des données est très demandée par la communauté scientifique et dans l'enseignement (droit de citation, de reproduction à des fins de recherche et de formation).

**Le concept d'un droit du TDM pourrait alors s'élargir pour devenir un principe de *Fair Use*.**

### Droit d'auteur

L'échange est un élément majeur de la recherche. Si celui-ci est restreint dès l'accès aux données par des contrôles et identifiants (Application Programming Interface - API) ou des paiements (abonnements, Article Processing Charges,...) alors l'élan scientifique lui-même pourrait être freiné.

Le droit de citation est celui-ci : droit de citer un texte à condition que celui-ci soit un extrait, entre guillemets, et que l'auteur soit indiqué. Mais ce droit de citation n'est pas valable pour les images. Même entre guillemets avec citation de la source et de l'auteur, l'utilisation d'image n'est pas possible. Il en va de même pour la réutilisation d'extraits sonores.

Cette pratique, qui avait du sens pour les technologies analogiques, n'en a plus à l'heure du numérique.

Le droit d'auteur est encore à ce jour intouchable en France. Si l'on souhaite faire avancer les choses ce n'est pas vers une exception au droit d'auteur qu'il faut aller mais vers la donnée. Il faut pour l'instant laisser de côté l'article et se focaliser sur la donnée. La science pourrait alors un jour revenir vers une exception au droit d'auteur par le biais d'un droit de la donnée.

Il devient donc essentiel de trouver un consensus entre les communautés scientifiques pour libérer l'exploitation des données. Si la donnée n'est pas libre, il faut qu'elle s'inscrive dans un principe de *Fair use* qui permette de la ré-exploiter.

Claire LEMERCIER indique que le principe de *Fair use*, appliqué à la citation d'extraits sonores ou d'images, ne pourrait être possible que sur des données produites par des scientifiques. La problématique en SHS est la réutilisation de données qui ne sont pas « scientifiques » par nature (pour ne prendre que quelques exemples, on peut étudier des objets trouvés lors de fouilles, la communication des entreprises sur leurs sites web, ou encore des chansons populaires...). Une autre partie des données utilisées en SHS

ne peuvent pas être mises en libre accès car elles sont soumises à la protection de données personnelles, ou à des clauses de non divulgation. Dans certains cas, comme dans le domaine des statistiques publiques au niveau le plus fin, une solution a été trouvée par le biais notamment de procédures d'anonymisation.

Certains éditeurs demandent aujourd'hui aux publiants en SHS de joindre leurs données aux articles, notamment en économie mais la plupart du temps les données SHS ne sont pas diffusables (données personnelles, données de test en psychologie, photographies de documents d'archives que les chercheurs ne sont autorisés à prendre par les archivistes que s'ils ne les diffusent pas ensuite, etc.).

La communauté des SHS serait favorable à des plateformes de partage mais il faut trouver les moyens juridiques de les mettre en place.

Nous pourrions par exemple envisager une plateforme qui permettrait d'échanger librement des données entre chercheurs avec une anonymisation des données, comme sur les plateformes d'échanges de données statistiques (des projets sont en développement pour les données plus « qualitatives » mais ce n'est pas simple).

Les plateformes (ensemble de données sur un projet donné mise à disposition sur un ordinateur donné) pourraient être reconnues comme Infrastructures Essentielles.

## **Adaptations juridique attendues**

### **Attentes des communautés**

Le Cabinet BENSOUSSAN, missionné par la DIST, souhaite interroger le Conseil Scientifique sur ces attentes pour la rédaction d'un texte pour un droit de la Science et un texte pour un droit des plateformes.

Aujourd'hui le droit numérique de la science est flou. La Science produit des plateformes alors qu'il n'y a pas encore de droit des plateformes. Le droit de la Science est un enjeu majeur pour la reconnaissance des pratiques des communautés scientifiques. La Science numérique a aujourd'hui besoin d'un droit qui lui permette d'évoluer dans ses usages, dans ses pratiques sans être bloquée par le droit d'auteur ou par les captations abusives des éditeurs.

Le débat de l'Universalité de la Science est déjà acquis mais il n'est plus envisageable avec le développement du numérique de rester dans un parallèle entre marché du don et marché de l'économie.

Aujourd'hui, au sein même du CNRS, il existe un paradoxe entre la politique d'Open access et celle du bureau de valorisation.

Claire LEMERCIER souligne qu'aujourd'hui la Science fait face à la problématique de la captation abusive des résultats et des données. Est-ce que cette captation pourrait devenir légitime ? A quel moment juge-t-on que la captation est abusive ? Pas seulement une captation par les éditeurs mais aussi par d'autres extérieurs.

La démarche du Livre Blanc doit s'appuyer sur des règles données par la communauté scientifique.

Claire LEMERCIER indique que si cette démarche est séduisante, notamment pour empêcher les captations abusives et augmenter les possibilités d'accès, deux risques sont visibles :

- L'uniformisation : quel degrés d'uniformisation impliquerait un droit de la Science ? Il faut respecter les différences de pratiques entre les communautés scientifiques, qui ont souvent de bonnes raisons d'être.
- Le Droit de la Science numérique qui s'opposerait à la non-science : comment empêcher les grands groupes tels qu'Amazon, Google, mais aussi des entreprises d'autres domaines, par exemple les assureurs, qui ont eux aussi leurs centres de recherche, de capter gratuitement tous les contenus scientifiques. Il y a des données que l'on rêve de pouvoir mieux partager entre scientifiques mais dont l'exploitation à but lucratif poserait des problèmes éthiques graves. Ce n'est pas la donnée qui est intrinsèquement scientifique ou non, c'est l'usage, mais comment délimiter l'usage scientifique ?

Nous pouvons envisager que ce droit de la Science permettrait :

- un principe de *Fair Use*
- de considérer que les scientifiques peuvent s'échanger sans contrainte tous les éléments d'une démarche scientifique
- de mettre en place une période d'embargo avant diffusion des données. Le chercheur ayant ainsi une priorité d'accès aux données pendant une période définie (selon les rythmes de publication d'articles ou les actions de valorisation).

## Construction du droit numérique de la Science

Le droit numérique de la Science doit se construire, il est donc possible d'imaginer son contenu mais celui-ci doit être appuyé par les communautés pour être viable. Si l'on crée un texte partagé par la communauté sa diffusion n'en sera que plus rapide.

François TRONCHE rappelle que tout ce qui est décrit dans un article mais n'est pas breveté est reproductible par tous. De nombreux chercheurs ne déposent pas de brevet pour permettre une communication large.

Il serait par ailleurs possible d'envisager de ne pas publier, et donc de ne pas privatiser l'article ou la donnée, sans en limiter l'accès.

Les éditeurs commencent à s'intéresser aux métadonnées. Celles-ci doivent également être protégées des captations abusives des éditeurs.

François BONNAREL s'interroge sur le risque d'un manque de compétitivité des chercheurs qui ne pourraient utiliser des outils proposés par des services pour classer, utiliser les métadonnées de façon intelligente si ces services étaient privés et payants, par rapport à ceux qui seraient en capacité de se soumettre à la privatisation de leurs métadonnées.

## Période d'embargo

La période « propriétaire » peut difficilement être déterminée en nombre d'années car la durée des projets varient de quelques mois à des dizaines d'années. On pourrait envisager une période d'embargo jusqu'à la date de publication d'un article. Une période liée à la notion de fin de projet sous réserve d'arriver à déterminer la période du projet (début et fin). Ou bien encore on pourrait envisager une période d'embargo définie par une convention.

## Création d'un modèle de publication alternatif

### Revue financées par les institutions publiques

Si nous ne pouvons pas toucher au droit d'auteur, des modèles alternatifs pourraient être développés pour permettre de publier sans captation abusive.

En France, des revues SHS souvent financées par des institutions publiques et basées sur un modèle OA existent et sont lus comme par exemple celles présentes sur la

plateforme Open Edition (gérée par des Universités et le CNRS). Ces nouveaux modes de publication en OA et sans captation abusive pourraient devenir les modèles de demain.

### Impact factor

Le problème que posent encore en France les revues en OA est celui de l'impact factor. L'évaluation des chercheurs en Europe et en France, dans beaucoup de disciplines, se base essentiellement sur leurs publications dans des revues à fort Impact Factor. Ce système est promu activement pour l'évaluation de la recherche. Les journaux qui attirent les soumissions sont donc ceux qui disposent d'un Impact Factor important. L'Impact Factor dépendant de la sélectivité du journal. Le système a des effets pervers et certains journaux en OA ont pu augmenter nombre d'articles payants publiés l'année où leur impact factor était bon. Ce qui conduit à augmenter leur bénéfice, quitte à faire baisser leur IF les années suivantes.

Ce mode d'évaluation à partir de l'Impact Factor évolue même aux Etats-Unis. La publication des chercheurs français est formatée par celui-ci.

### Partage de l'information

Si dans la plupart des domaines de la biologie, les chercheurs ne voient pas pour l'instant un intérêt aux TDM sur les articles, au-delà de leur consultation, la communauté a par ailleurs besoin d'un accès libre aux articles le plus rapidement possible. En biologie le partage est aussi un partage d'éléments manufacturés. De nombreux éditeurs demandent que l'accès de la communauté à l'élément décrit dans l'article soit garanti afin que n'importe qui puisse reproduire l'expérience. Au-delà, plusieurs éditeurs mettent actuellement en place un système d'accès aux données expérimentales brutes synthétisées dans les figures d'articles.

S'il semble qu'en biologie la communauté soit celle qui partage le plus, le partage n'est pas uniquement basé sur les données mais aussi sur le matériel produit dans le cadre d'un article.

En SHS, la communauté a besoin de se protéger des captations et d'avoir un accès libre aux articles avec ainsi une diminution des coûts d'accès à l'information. Le TDM est également important avec un besoin d'accès à de large corpus de ressources. Le partage des données, n'est pas encore assez avancé, avec un manque de culture et de

moyen humains pour le partage. Si un droit de la Science publique ou des plateformes est créé cela permettrait peut-être aux communautés SHS de mettre en place des plateformes de partage qui n'existent pas encore largement, sur le modèle des plateformes de données statistiques.

En SHS, il serait aussi intéressant de trouver un principe d'Open Data, de partage, mais sans doute pour la plupart des données limité au partage entre scientifiques, avec tous les problèmes de définition que cela pose, qui perdurerait après la fin du projet.

En astronomie la communauté partage beaucoup mais le taux de réemploi reste sans doute encore faible. Les algorithmes doivent être liés aux articles publiés en OA pour permettre de vérifier le résultat et de le rendre réutilisable.

**L'ensemble du groupe de travail conclut que la préoccupation partagée est celle de permettre un échange d'informations sans contrainte entre les communautés scientifiques.** Pour cela il faut trouver un moyen intelligent de partager les données.

### Partage des logiciels

En astronomie, dans le résultat scientifique lorsqu'on parle d'OA on se réfère au partage de la données instrumentalisées (ou brute) sur laquelle est construit un algorithme qui permet de l'utiliser. Cet algorithme et ce logiciel pourraient-ils également être partagés ?

François BONNAREL indique que le partage des logiciels est souhaitable et existe déjà mais il n'existe pas de système général et standardisé d'échanges de logiciels. Il existe néanmoins toute une gamme de logiciels interopérables dans le cadre de l'observatoire virtuel qui ne représente cependant qu'une petite partie du logiciel en astronomie.

François TRONCHE souligne qu'en biologie en plus de la diffusion du matériel on peut y associer le partage du logiciel si celui-ci est décrit dans l'article publié. La plupart du temps les logiciels développés par la communauté académique sont en freeware. Certains centres ou structures académiques importantes développent leurs propres logiciels en libre accès.

## **Formes particulières de partage de pratiques et de connaissances**

De nombreuses plateformes de partage sont créées puis disparaissent en fonction des besoins.

L'EMBO (organisation internationale de recherche et éditeur) a été la première à vouloir mettre en place un service d'accès aux données brutes ou traitées derrière les graphes.

(Thomas LEMBERGER - <http://www.embo.org/> )

François BONNAREL indique qu'un papier récent, référencé sur le site de l'INSU, parle des publications qui ne sont plus basées sur des observations dans le cadre de travaux de recherche mais sur des données d'archives. La réutilisation des données oubliées ou en jachères. Ce courant se développe en astronomie où parfois le chercheur n'observe plus mais publie des articles à partir de données d'archive en libre accès.

François TRONCHE ajoute qu'en biologie une partie importante de l'expérimentation d'un laboratoire est souvent sous-traitée, dans le cadre de collaboration ou de prestations de services. En pratique la personne qui s'occupe de la sous-traitance peut faire celle-ci gratuitement s'il est cité dans l'article en tant que co-auteur ou contre paiement s'il ne l'est pas.

### **c. Université de Strasbourg, Paul-Antoine HERVIEUX et Françoise CURTIT, 10 juillet 2015**

#### **Participants**

Pour l'Université de Strasbourg :

- Paul-Antoine HERVIEUX, Vice-président délégué Partenariats avec les EPST et les collectivités
- Françoise CURTIT, CNRS, Chargée de mission « Open Access » à l'Université de Strasbourg

Pour le Cabinet BENSOUSSAN :

- Maître Alain BENSOUSSAN, Avocat à la Cour spécialisé en droit des technologies avancées
- Maître Laurence TELLIER- LONIEWSKI, Avocate, directrice du pôle Propriété intellectuelle

Pour la DIST :

- Renaud FABRE, Directeur
- Charlotte AUTARD, Chargée du projet IA ISTEEX

#### **Présentation de la démarche**

La démarche de rédaction d'un Livre blanc « Science Numérique et Droit », portée par le CNRS en collaboration avec le Cabinet Alain BENSOUSSAN, s'inspire des problématiques soulevées par la sécurisation juridique du Projet d'Investissement d'Avenir ISTEEX ANR-10-IDEX-0004-02 ([www.istex.fr](http://www.istex.fr)) qui soulève de nombreuses questions juridiques par ses possibilités de TDM, d'interdisciplinarité, d'agrégation de contenus et son objectif de rendre explorable et traversable ses bases de données.

Le droit nouveau de la science à une exception particulière dans le droit d'auteur : un statut juridique des résultats scientifiques et plus particulièrement des données et des métadonnées.

Ce Livre Blanc a pour objectif de **proposer un cadre juridique pour les données scientifiques afin de répondre aux préoccupations des communautés scientifiques**

(publications et usages des données, droit des plateformes de Science, fouille de texte, ...) en étant force de proposition dans la **participation au projet de loi numérique**. La Science porte à l'amont la loi numérique, et elle le fait sous des formes qui doivent s'inscrire concrètement.

Dans un premier temps, le cabinet BENSOUSSAN et la DIST réuniront des groupes de travail et mèneront des auditions pour **recueillir les usages de la communauté et l'état de l'art des pratiques de valorisation**. L'association forte avec les universités et les organismes permettra l'audition de grands témoins et une parole plurielle sur les évolutions qui se préparent dans le cadre du projet de loi numérique.

Le recueil des usages et de l'état de l'art permettront au CNRS et au Cabinet BENSOUSSAN d'établir une **grille de pertinence des analyses et des pratiques qui sera comparée au cadre normatif actuel** afin d'évaluer les écarts et **élaborer des propositions de lois**.

Ce Livre blanc sera anonyme et les personnes auditionnées auront un contrôle complet si elles souhaitent ou non être citées.

La DIST et le Cabinet BENSOUSSAN recontacteront les personnes auditionnées et soumettront un premier texte pour avis aux Présidents d'universités.

Le Cabinet BENSOUSSAN, missionné par la DIST, souhaite recueillir aujourd'hui l'avis de l'Université de Strasbourg sur d'éventuelles propositions d'écriture de textes, de préconisations.

### Droit d'auteur

Paul-Antoine HERVIEUX exprime que dans les communautés universitaires il y a une méconnaissance quasi générale du droit d'auteur et de ses modalités d'application, quel que soit le type de documents ou de données. Les aspects soulevés par la loi numérique représentent un chantier énorme en termes de formation.

La démarche du Livre blanc a pour objectif de recueillir l'avis des communautés scientifiques qui sont témoins des carences. La DIST et le Cabinet BENSOUSSAN souhaitent savoir si un droit des résultats scientifiques et des données serait à même d'être partagé par les communautés.

Selon Paul-Antoine HERVIEUX un droit des données et des résultats devient indispensable avec l'évolution du numérique. Les usages se modifient et les données de la recherche (BSN10) sont en train de changer le paradigme. Les chercheurs, et

enseignants-chercheurs, commencent à se rendre compte de la valeur de leurs données. Dès que nous parlons de valeur, il faut réfléchir au droit. Nous sommes aujourd'hui dans un monde où les entreprises privées sont prédominantes et s'intéressent de plus en plus aux données pour en privatiser les usages.

### **Captations légitimes et abusives**

L'Université de Strasbourg est connue pour ses recherches dans le domaine de la chimie et possède des ressources telles que des données d'expériences ou plus spécifiquement des banques de données dans lesquelles sont répertoriées les propriétés d'un nombre considérable de réactions chimiques. Ces dernières peuvent avoir un grand intérêt pour les entreprises privées : par exemple, l'industrie pharmaceutique s'intéresse de plus en plus aux catalogues de réactions chimiques pour sa propre réutilisation.

Ces données doivent être protégées des captations. Les communautés n'ont pas encore conscience de ce danger, notamment dans les universités.

Renaud FABRE souligne qu'une différenciation doit être faite entre captation légitime et captation abusive.

Un texte particulier, autour de l'Open Science, pourrait définir les règles de la captation abusive.

### **Modèle économique des données**

Paul-Antoine HERVIEUX souligne que l'Université de Strasbourg dispose, dans ses laboratoires, de données précieuses pour le secteur privé. Un système économique pourrait être créé autour des données et permettrait d'accompagner et de financer la recherche fondamentale. Les données devraient être ouvertes, même au privé, mais pour cela il faut définir un modèle économique, éditer des règles précises (un guide de la donnée) ou encore un outil simple d'utilisation pour les chercheurs.

La recherche peut être différenciée par :

- Un financement par des fonds publics
- Un financement par des fonds privés

Ces aspects doivent entrer en jeu pour définir un modèle économique et proposer les données à des organismes privés.

Une des suggestions possibles est, dans les suites du Livre blanc, de proposer un Abécédaire, une foire aux questions à l'attention des chercheurs pour qu'ils puissent s'orienter dans le dédale du droit des données.

### **Période d'embargo**

Les données en astronomie sont issues d'instruments particuliers : données d'instrumentation. La communauté gèle ses données pendant un délai d'un ou deux ans. L'idée d'une période d'embargo pourrait-elle être partagée avec l'Université de Strasbourg ?

Paul-Antoine HERVIEUX répond que la clé de la science est l'instantanéité de la recherche. Une donnée à un cycle de vie plus ou moins long, mais plus elle est fraîche plus elle est intéressante. Pour que la Science avance, l'embargo ne doit pas faire obstacle à la diffusion des résultats. Cependant, en fonction de la communauté scientifique concernée et pour des raisons diverses (e.g. compétition, processus de valorisation en cours...) une période d'embargo pourra être instaurée.

### **Logiciels et algorithmes**

Pour avoir l'effet maximum de réutilisation potentielle des données, il faut avoir accès au matériel qui permet de les traiter. Lorsque celles-ci sont ouvertes et mises à disposition des communautés, les logiciels et les algorithmes doivent-ils également être mis en ligne ?

Paul-Antoine HERVIEUX indique que les chercheurs sont aujourd'hui dans une logique de compétitivité. Chacun doit produire un certain nombre de publications par an pour rester compétitif. A partir de ce constat, partager le fruit de ses travaux, les codes de calcul, qui sont le cœur de sa recherche, n'est pas envisageable. On peut par ailleurs partager des informations sur la structuration du code, mais pas l'algorithme lui-même, ou alors au bout d'un certain temps, sous forme d'un article décrivant la méthodologie employée, ou bien dans une revue éditant des codes de calcul. Mais la plupart des chercheurs seront réticents à la publication d'emblée de leurs codes, de leurs algorithmes.

Il existe déjà aujourd'hui des revues comme par exemple *Computational Physics*, qui permettent aux physiciens de déposer leurs codes de calcul.

Il pourrait également être envisageable de déposer en libre accès l'exécutable avec un copyright sur le code.

Si les communautés souhaitent une période d'embargo pour le dépôt des logiciels et algorithmes, une durée d'un à deux ans pourrait convenir.

### **Publications scientifiques et archive de dépôt**

Paul-Antoine HERVIEUX indique que la relation avec les éditeurs est au centre des préoccupations des chercheurs, notamment au regard du TDM. Le TDM sera l'outil par excellence dans les années à venir.

Les éditeurs cherchent à se protéger en contrôlant la possibilité de TDM via les API (Application Programming Interface), des clauses contractuelles (licences Creative Commons, limitation du nombre de mots « explorables »).

Le principe du libre accès dans le respect du droit d'auteur est une évidence, mais il faut dans un premier temps faire la distinction entre l'exploitation des corpus et la réutilisation de données déjà compilées. Différencier les données brutes des données construites.

Lorsqu'une institution se dote d'une archive ouverte, elle accède à la possibilité de mettre en place sa propre politique. A l'heure de l'autonomie des établissements, à l'heure où ceux-ci doivent asseoir leurs stratégies, les archives ouvertes et le TDM représentent un enjeu majeur.

L'Université de Strasbourg a pour ambition de construire sa propre archive pour référencer sa production scientifique et permettre de la localiser dans un emplacement unique. Cette archive permettra à la communauté universitaire strasbourgeoise d'avoir un droit de regard et de recherche sur sa propre production sans verrous (sauf officiels tels que Sherpa ou droits d'embargo).

Selon Paul-Antoine HERVIEUX, l'idéal serait que le CNRS se positionne sur le principe d'un dépôt de toutes les productions scientifiques (CNRS et partenaires) dans une archive commune (mais pas forcément unique) avec des règles semblables.

En tant qu'interlocuteur des EPST à l'Université de Strasbourg, Paul-Antoine HERVIEUX signale que les partenariats Universités/CNRS fonctionnent très bien malgré des politiques distinctes. Partager les mêmes données avec les mêmes droits apparaît donc comme une évidence.

## **Pistes d'actions**

Un article du texte proposé à la rédaction par le Cabinet BENSOUSSAN pourrait exprimer que : « les données issues de la recherche scientifique et les résultats associés sont libres d'usage et peuvent être mis à disposition librement aux communautés scientifiques des publications. Et ceci sans restrictions sauf pour les logiciels qui pourront être mis à disposition de la communauté après une période d'embargo d'un à deux ans».

Renaud FABRE rappelle que le développement de la plateforme HAL a pour objectif d'inciter les chercheurs du CNRS et des organismes à déposer toutes leurs publications.

Paul-Antoine HERVIEUX précise que le grand projet de l'Université de Strasbourg pour la construction d'une archive ouverte de la connaissance prévoit également un moyen de connecter l'archive de l'université avec la plateforme HAL pour permettre un reversement automatique des publications. Le CNRS appuie ce projet. L'Université de Strasbourg veut mettre l'accent sur la création d'un corpus correspondant à la production de ses chercheurs principalement à des fins de valorisation mais tient à être en accord avec la politique du CNRS et la politique européenne via ce reversement automatique dans HAL.

## **Projet de loi sur le numérique**

Françoise CURTIT s'interroge sur le croisement entre le calendrier de la loi numérique en France et celui des directives européennes en cours de refonte. Le droit européen ne risque-t-il pas de s'imposer à la loi numérique française ?

Alain BENSOUSSAN indique que le Président a annoncé cette réglementation comme création d'une loi française sur le numérique semblable aux droits de l'Homme. Le Premier ministre appuie cette initiative et un texte circule déjà au Cabinet de la Secrétaire d'Etat, Axelle LEMAIRE.

Le code de la recherche scientifique traite de l'organisation de la science mais pas des données de la science. La loi sur le numérique pourrait prévoir un article qui libellerait que le code de la recherche scientifique est modifié et prend en compte les recommandations du Livre blanc.

Renaud FABRE indique que la feuille de route numérique du gouvernement a été rendue publique par le Premier ministre le 18 juin dernier, les dispositions en matière de science sont définies vers le développement de l'Open Access.

Un deuxième point de ces dispositions concerne l'autorisation de l'exploration du texte et des données (TDM) et à cela il est prévu une insertion au code de la propriété intellectuelle.

Dans les deux cas, cette feuille de route inscrit clairement les éléments nécessaires pour un droit de libre parcours.

### **Open Process**

La question de la démarche d'Open Process et la question de libre requête met en avant les aspects de captations légitimes et abusives. C'est autour de ce processus que la DIST et le Cabinet BENSOUSSAN souhaitent travailler collectivement avec les représentants des communautés auditionnés. Les auditions ont d'ores et déjà permis d'enrichir la définition de cette pratique. Le Conseil Scientifique du CNRS se prononcera sur ces sujets au mois de septembre.

Il apparaît également indispensable de recueillir en amont l'avis de grandes Universités comme celles de Strasbourg ou encore de l'UPMC (Université Paris 6) pour avoir une vision d'ensemble de l'écosystème de la recherche française.

### **Régulation des plateformes**

La question posée est la suivante : faut-il réguler les plateformes et les qualifier d'Infrastructures Essentielles de la recherche scientifique ?

Selon Paul-Antoine HERVIEUX il serait indispensable d'associer aux plateformes un guide juridique d'utilisation. Un code de conduite permettant aux utilisateurs des plateformes de connaître les règlements, les usages, et d'adapter leurs comportements. L'Université de Strasbourg réfléchit déjà à la rédaction d'un document sur l'utilisation des plateformes.

Renaud FABRE indique que les Infrastructures Essentielles sont soumises à deux conditions : elles doivent être indispensables et ne peuvent être mises en place par d'autres moyens que celui de la puissance publique (ex : aéroport).

Les plateformes numériques de Science sont indispensables et ne peuvent être que les moyens du partage des connaissances.

La TGIR Huma-Num par exemple, est un objet qui n'a pour l'instant pas de statut mais qui dessert de nombreux besoins vers des bénéficiaires de la science. La labéliser en tant que Infrastructures Essentielles pourrait permettre de l'inscrire comme structure d'amont de la science publique.

Paul-Antoine HERVIEUX rejoint l'avis de Renaud FABRE sur la nécessité de reconnaître les plateformes numériques de Science comme Infrastructures Essentielles.

Dans un premier temps, il faut définir pour les plateformes le droit de conservation, de consultation et de partage des données entre acteurs de la science publique avant diffusion vers tous les usagers et tous les bénéficiaires.

Françoise CURTIT précise que ses travaux menés autour de la Science ouverte lui ont permis de découvrir la variété des usages des communautés, et leurs incidences sur les aspects juridiques. Un modèle large risquerait de ne pas convenir à l'ensemble des disciplines. La variété des usages doit également être prise en compte pour la définition des droits des plateformes.

La DIST et le Cabinet BENSOUSSAN proposent un renvoi à l'Etat de l'art des communautés scientifiques afin de permettre d'institutionnaliser les usages. Pour cela un travail en amont de recensement des usages des communautés est nécessaire.

Paul-Antoine HERVIEUX et Françoise CURTIT indiquent que cette démarche est indispensable. Il faut que le Livre blanc tienne compte des usages pour définir des orientations communes sans diviser les communautés. Le Livre blanc permettra de définir des lignes directrices communes avec intégration des usages.

#### d. Conseil National du Numérique, Benoit Thieulin et Yann Bonnet, 1 septembre 2015

### Rencontre avec la DIST et le Cabinet BENSOUSSAN dans le cadre de la démarche de rédaction d'un « Livre Blanc Science Numérique et Droit »

#### La démarche :

#### Rédaction d'un Livre blanc en vue (notamment) du projet de loi numérique

- Démarche portée par le **CNRS** en collaboration avec le **Cabinet BENSOUSSAN**.
- S'inspire des problématiques juridiques soulevées par le **Projet d'Investissement d'Avenir ISTE**X ([www.istex.fr](http://www.istex.fr)), qui consiste en **2 actions** :
  - **Vaste programme d'acquisition de ressources scientifiques**, sous forme de licences nationales.
  - **La création d'une bibliothèque numérique** accessible à distance par tous les membres des établissements de l'enseignement supérieur et de la recherche.
- Ce projet soulève de nombreuses questions juridiques par ses possibilités de *text and data mining*, d'interdisciplinarité, d'agrégation de contenus et son objectif de rendre explorable et traversable ses bases de données.
- Le Livre Blanc a pour objectif de **proposer un cadre juridique pour les données scientifiques afin de répondre aux préoccupations des communautés scientifiques** (publications et usages des données, droit des plateformes de Science, fouille de texte, ...) en étant force de proposition dans la **participation au projet de loi numérique**.
- Le cabinet BENSOUSSAN et la DIST réunissent des groupes de travail et mènent des auditions pour recueillir les usages de la communauté et l'état de l'art des pratiques de valorisation, dans le but d'établir une grille de pertinence des analyses et des pratiques qui sera comparée au cadre normatif actuel afin d'évaluer les écarts et élaborer des propositions de lois qui seront soumises.

## Réponses au guide d'entretien

### I. Contexte

Le numérique transforme en profondeur les modes de production et de diffusion des résultats scientifiques : données, publications, analyses sont accessibles désormais sur des plateformes. Cette disponibilité du matériau scientifique contient un potentiel d'exploitation et de partage des connaissances dont la loi doit pouvoir définir les conditions, les modalités, les limites.

Un problème aujourd'hui : il n'existe pas de statut juridique spécial pour l'information scientifique.

Les études et travaux issus de la recherche scientifique sont appréhendés comme des œuvres écrites, susceptibles donc d'être protégées par le droit d'auteur.

- Pour des raisons de visibilité et de carrière les chercheurs sont contraints de publier les résultats de leurs recherches dans certaines revues, et pour cela, de céder tous leurs droits à des éditeurs.

L'éditeur se trouve dès lors seul investi du droit d'exploiter, reproduire, diffuser l'article.

→ Dès lors ni l'organisme de recherche, ni la communauté scientifique, ne peuvent disposer de l'étude, sauf à passer un accord avec l'éditeur.

- En plus de céder tous leurs droits, les organismes de recherche sont ainsi soumis à un double paiement :
  - Les frais de publication en amont ;
  - Les frais de consultation des documents en aval : les établissements d'enseignement supérieur et de recherche dépensent annuellement plus de 80 M€ pour avoir accès aux ressources électroniques. Les prix d'accès ont d'ailleurs continuellement augmenté : de 7% par an depuis 10 ans.

→ Un poids pour les finances publiques et la productivité de la recherche public, déjà fortement concurrencée au niveau international.

- Conséquence : un déséquilibre important entre le chercheur et l'éditeur, aggravé par l'émergence, ces dernières années, d'oligopoles de faits dans le secteur de l'édition scientifique (Elsevier, Springer, Wiley...).

**Face à ce constat, il est nécessaire d'affirmer un droit à l'open access, une science ouverte pour répondre aux préoccupations de la communauté scientifique.**

**Définition de l'open access.** Les publications en accès libre ("Open Access" ou OA) désignent les articles accessibles sous forme électronique et lisibles gratuitement via Internet. L'open access autorise les lecteurs, dans le respect du droit d'auteur, à lire, télécharger, copier, distribuer, imprimer, rechercher ou lier le texte intégral des articles, à les indexer, à les récupérer pour les traiter informatiquement et à s'en servir pour tout autre usage légal, sans barrière financière, juridique ou technique.

- **Objectifs**
  - **Desserrer la pression des coûts** des revues numériques sur les budgets des établissements publics ;
  - **Faciliter l'accès des connaissances scientifiques** à la communauté des chercheurs, mais aussi à la société civile ;
  - **Offrir un accès plus large aux résultats de la recherche scientifique aux entreprises**, et notamment aux petites et moyennes entreprises, qui pourront ainsi améliorer leur capacité d'innovation.
- La Commission européenne a d'ailleurs invité les Etats membres à "définir des politiques claires en matière de diffusion des publications scientifiques issues de la recherche financée par des fonds publics et du libre accès à ces dernières " (2012).

### **Propositions du CNNum sur l'Open Access**

- **La conservation et la diffusion des résultats de la recherche sont des missions de service public.** Un mouvement de publication ouverte existe déjà via les entrepôts d'archive ouverte des universités et organismes, ou via la plateforme HAL.
- Aujourd'hui un cadre législatif adapté permettrait d'encourager ce mouvement amorcé par des membres de la communauté scientifique.
  1. Reconnaître un droit à l'exploitation secondaire, à l'image du droit allemand. La version de l'auteur déposée dans une archive institutionnelle reste en accès libre, quelles que soient les suites éditoriales données à ces travaux ;
  2. Rendre accessible librement les publications scientifiques financées sur fonds publics, après un court délai d'embargo, permettant l'activité commerciale de l'éditeur, soit dans des revues ouvertes soit dans un dépôt institutionnel (à l'instar du droit allemand et italien). Cette obligation ne doit pas reposer sur les chercheurs, mais sur les organismes de recherche.

→ Attention à ne pas créer une simple **faculté** de publier ces documents en open access, qui reviendrait à condamner ce droit à l'ineffectivité (comme c'est le cas en Allemagne). Il est nécessaire d'en faire une **obligation**.

→ **L'exemple anglais est éclairant** Le HEFC<sup>137</sup>, chargé de répartir entre les différents établissements d'enseignement supérieur la somme qui leur est allouée globalement par l'Etat, a annoncé qu'à compter de l'année prochaine, toute publication qui ne sera pas disponible en Open access ne sera pas prise en compte pour évaluer l'activité de ces établissements.

3. **Encourager les chercheurs à mettre en accès libre des données brutes et anonymisées de la recherche** à chaque fois que cela ne se heurte pas à des questions déontologiques ou de vie privée.

## II. Les propositions du CNRS : conditions, modalités, et limites du partage de l'Information Scientifique

### A. Conditions

**Position du CNRS** : Le libre accès aux résultats de la science a pour conditions principales l'abolition des limites que peut poser le droit éditorial (droit de la publication, droit des auteurs), au regard de l'exploration des corpus numériques de publications ou de données.

**Question 1** : Quelle est votre opinion sur les adaptations nécessaires du droit de la publication (éditeurs et/ou auteurs) et de l'exploration des corpus (techniques de Text & Data Mining, API, ...) ?

**Définition du *text and data mining***. Le *text and data mining* désigne différents outils d'extraction et d'analyse qui permettent une exploration automatisée de contenus numériques, qui peuvent inclure des textes, des données, des sons, des images ou d'autres éléments, ou une combinaison de ceux-ci, afin de découvrir de nouvelles connaissances ou des idées.

---

<sup>137</sup> Higher Education Funding Council.

## Les techniques de *text and data mining* permettent de propulser la recherche française à l'heure du big data.

- Ces techniques sont qualifiées par le MIT comme l'une des 10 technologies émergentes qui "changeront le monde" au XXI<sup>e</sup> siècle. Exemples d'applications :
  - Le projet Text2Genome a permis de cartographier le génome humain en compilant automatiquement plus de 3 millions de publications.
  - S'agissant de la presse, l'exploration de données est sans doute l'un des modèles économiques de l'information de demain.
  - Le TDM permettrait à Amazon de réaliser 20% de son chiffre d'affaires.

### Le TDM n'est pas en soi une activité nouvelle.

- **Il revient à lire et dégager des informations, du sens, à partir de documents.** Il ne se distingue pas fondamentalement du relevé manuel d'informations, pratiqué par la recherche depuis la nuit des temps.

**Toutefois le TDM nécessite fréquemment la création de copies et le stockage des contenus, donc une reproduction au sens du droit d'auteur.** Opérer des traitements informatisés de gisements de contenus, quelle que soit leur nature (textes, données, images fixes ou animées, musique, sons...) impose, pour les fouiller, d'en prendre possession et de les stocker ou d'en opérer une extraction substantielle, au sens du droit des producteurs de bases de données.

### Le cadre normatif ne permet pas au TDM de livrer tout son potentiel :

- **La gestion individuelle des droits est peu adaptée pour un ensemble aussi massif de données.** Face à la multiplicité des sources et l'importance des volumes traités, il serait vain pour le data miner d'aller chercher les autorisations unes à unes.
- **La solution contractuelle n'est pas satisfaisante.** Les accords proposés par les éditeurs scientifiques encadrent sévèrement les usages autorisés, et posent parfois des contraintes et des interdictions importants :
  - Obligation pour le chercheur de déclarer sa recherche, ce qui va à l'encontre de toute éthique scientifique (secret de la recherche, etc).
  - Les chercheurs dépendants des éditeurs scientifiques, qui proposent, sur les plateformes qu'ils mettent à disposition, des services très limités et décriés par les chercheurs. La contractualisation rend difficile, voire impossible, la fouille entre différents corpus et jeux de données.

## Propositions du CNNum sur le *text and data mining*

Introduire une exception au droit d'auteur, sans compensation, de droit d'usage à des fins de TDM dans un cadre de recherche.

→ Cela ne contrevient pas aux exceptions posées par l'UE au droit d'auteur. Les anglais le font avec leur exception de fair dealing.

### B. Modalités

**Position du CNRS** : Les modalités de partage des résultats de la science doivent, à l'heure numérique, assimiler de nouvelles contraintes : le partage des résultats entre acteurs de la recherche publique d'une part et d'autre part entre usagers et bénéficiaires de la science publique.

**Question 2 : Comment tracer la frontière entre la captation légitime et la captation abusive des résultats disponibles sur une plateforme de science publique, comment protéger ces résultats ?**

- **Posture du CNNum dans le rapport sur la question des ressources ouvertes et de la captation :**

- **Les bénéfices de la diffusion ouverte sont aujourd'hui encore trop peu exploités par la société dans son ensemble.** Bien souvent, ils sont captés pour l'essentiel par les plus grands acteurs, notamment les plateformes web déjà bien établies, qui, en croisant les apports des communs avec leurs propres ressources, font de la prédation un risque bien réel.

- **La réponse à ce risque ne peut être celle d'un retour en arrière, d'une renonciation au mouvement d'ouverture.** Il s'agit bien davantage de viser le développement de la capacité d'un plus grand nombre d'entreprises, associations, acteurs publics, chercheurs, médias, etc. à contribuer et participer aux communs, et surtout à en utiliser les ressources.

→ **Recours possible à des licences de réutilisations non commerciales, ou de partage à l'identique** : il y a tout un éventail de solutions contractuelles pour encadrer la réutilisation des résultats de recherche. Creative commons a su développer des solutions à la fois facilement compréhensibles et lisibles par une machine.

→ **Il faut néanmoins se poser la question de l'objectif de la recherche publique : n'est-elle pas en partie financée publiquement pour alimenter l'économie et la société dans son ensemble ?** La multiplication de licences différentes peut être un frein

à une exploitation des données et compliquer l'utilisation des plateformes. Il s'agit donc finalement d'un choix plus politique que technique sur le modèle de la recherche publique que nous défendons.

### C. Limites

**Position du CNRS** : Les plateformes de science contiennent aujourd'hui une IST de forme, de contenu et de statut juridique très hétérogène. Cette hétérogénéité nuit à la visibilité des plateformes de science. De la même façon qu'il existe aujourd'hui une notion de données d'intérêt générales (choix récent de la Ministre de l'économie numérique), il y a lieu de réfléchir à la dénomination des plateformes de science publique et, éventuellement, à un régime juridique spécifique.

**Question 3 : Doit-on aller vers une dénomination d' « infrastructure essentielle » pour les grandes plateformes d'amont de la recherche, dans les cas où ces plateformes occupent une fonction inédite et irremplaçable ?**

- **Les plateformes deviennent les nouveaux gatekeepers de l' information. Il est nécessaire de repenser les concepts de droit de la concurrence sur les "infrastructures essentielles" pour imposer des obligations d'ouvertures à ces acteurs.**
- il nous paraît évident qu'il y a un enjeu d'indépendance qui se joue. Est-ce qu'on souhaite qu'un opérateur privé régisse les règles d'accès à la connaissance de la recherche publique, qu'il monétise cet accès? Je ne crois pas.

→ Les communs font ici pleinement sens.

- Importance du libre.

*Pour la connaissance et l'accès à la connaissance comme bien commun voir :*

*Charlotte Hess et Elinor Ostrom, (coord.) Understanding knowledge as a commons, MIT Press, dec. 2006.*

*Gaëlle Krikorian et Amy Kapczynski (coord.), Access to Knowledge in the Age of Intellectual Property , Zone Books, 2010.*

## **e. Comité ISTEEX, Laurent Schmitt, Jean-Marie Pierrel et Grégory Colcanap, 24 septembre 2015**

### **Participants :**

#### **Pour le Comité Exécutif ISTEEX :**

- Pour le Consortium COUPERIN : Grégory COLCANAP, coordonnateur, et Monique JOLY, responsable du département services et prospectives
- Pour l'Inist : Laurent SCHMITT, responsable département projets et innovation, représentant de Raymond BERARD au Comex pour la durée de son absence
- Pour l'Université de Lorraine : Jean-Marie PIERREL, Professeur

#### **Pour le MENESR :**

- Marie-Pascale LIZEE, Département de l'Information Scientifique et Technique et Réseau Documentaire (DISTRD)

#### **Pour le Cabinet BENSOUSSAN :**

- Maître Alain BENSOUSSAN, Avocat à la Cour spécialisé en droit des technologies avancées
- Maître Laurence TELLIER- LONIEWSKI, Avocate, directrice du pôle Propriété intellectuelle
- Maître Sarah LENOIR, Avocate, pôle Propriété intellectuelle

#### **Pour la DIST :**

- Renaud FABRE, Directeur
- Laurence EL KHOURI, Directrice adjointe
- Charlotte AUTARD, Chargée du projet IA ISTEEX

### **Présentation de la démarche**

Comme évoqué lors du séminaire juridique ISTEEX de juillet 2014 et en Comité Exécutif, le Cabinet BENSOUSSAN, en collaboration avec la DIST et la DAJ du CNRS, entreprend, suite aux issues soulevées par le projet ISTEEX, la rédaction d'un Livre Blanc « Une science ouverte dans une république numérique ».

Cette démarche s'inspire des problématiques soulevées lors du séminaire de juillet 2014 sur la sécurisation juridique d'ISTEEX sur les thèmes: objectifs, gestion des bases de

données, fouille de texte et de données (*Text and data mining* – TDM), interdisciplinarité, agrégation de contenus.

Thèmes évidemment présents dans le projet de loi numérique porté par la Secrétaire d'Etat chargée du numérique, Axelle LEMAIRE.

Le Livre blanc a pour objectif de contribuer à **évoquer le cadre juridique pour les données scientifiques afin de répondre aux préoccupations des communautés scientifiques** (publications et usages des données, droit des plateformes de Science, fouille de texte, ...) en étant force de proposition dans la **participation au projet de loi numérique**.

La démarche proposée par le Cabinet BENSOUSSAN est dans un premier temps **mener des auditions** afin de **recueillir les usages de la communauté scientifique et l'état de l'art des pratiques de valorisation de l'IST**.

Le recueil des usages et de l'état de l'art permettront d'établir une **grille des analyses et des pratiques qui sera construite au regard du cadre normatif actuel** afin d'évaluer les écarts et **élaborer des propositions**, qui seront soumises à l'ensemble des acteurs participant à l'élaboration de ce Livre blanc.

### **TDM et droit d'observation**

L'expérience du projet ISTEEX a permis de constater que la mise en place des outils et des droits de TDM en France est plus difficile dans la situation actuelle de la loi et dans la relation avec les éditeurs.

Selon Grégory COLCANAP, obtenir une exception du droit d'auteur comme en Angleterre semble aujourd'hui un minimum pour permettre le développement des pratiques de TDM en France et la reconnaissance du droit de lire.

En France l'édition est majoritairement liée à la culture : une exception allant contre l'édition serait vue comme allant à l'encontre de la culture.

Jean-Marie PIERREL rappelle que le TDM est une lecture automatisée sans différence avec la lecture normale à la différence qu'elle s'effectue au travers d'une machine.

Le droit de TDM est aussi le droit de produire des données de la Recherche accessible à l'ensemble de la communauté.

Au-delà de la possibilité de TDM, il faut garantir la conservation des copies intermédiaires, des documents annotés et modifiés. A ce jour certains laboratoires font

déjà ce travail de conservation mais doivent maintenant développer les outils de diffusion et de partage de ces données mais aussi des métadonnées.

Les métadonnées peuvent, selon les disciplines et les usages, être définies en deux catégories :

- Les informations associées au document (métadonnées)
- Les informations associées aux usages du document (données d'utilisation)

Les Contenus Générés par les Utilisateurs (CGU), sous-produit de l'utilisation des bases d'article, utiles et industriellement reconnus, doivent être aujourd'hui intégrés dans cette réflexion afin d'évaluer la valeur ajoutée de ces données d'usage et de définir leur intégration dans la valorisation du travail scientifique.

Renaud FABRE indique que le TDM est une pratique largement connue et reconnue qui fait l'objet de textes et de prises de positions fortes de la part d'organismes tels que l'ADBU, les Bibliothèques universitaires, l'ensemble des organismes de recherche au niveau des directeurs et de leurs associations d'IST.

Dans le cadre de la consultation sur le projet de loi numérique, les dispositifs de fouille (TDM) dans l'immensité des données et des publications, rendus possibles par leurs traitements électroniques, constituent un enjeu crucial pour l'avenir de la recherche. La plupart des grands pays de recherche (Allemagne, Canada, Etats-Unis, Royaume-Uni,...) ont adopté de telles dispositions : la France ne saurait adopter des dispositions qui la mettent à l'écart de la communauté scientifique internationale.

Jean-Marie PIERREL, indique que les chercheurs se dotent d'outils de fouille de textes ou d'un ensemble de données afin d'expliquer finement des phénomènes et de pouvoir utiliser des procédés électroniques pour les rendre évident.

Les besoins des chercheurs à ce jour sont :

- L'acquisition de données primaires au vue de l'état du marché
- L'accès à ces données primaires pour les observer sans limitation.

Ainsi, plus qu'un droit de lecture, le TDM doit être un droit d'observation. Le TDM est un droit d'observation des objets scientifiques, nécessaire à la Science. Si les scientifiques disposent d'un droit universel et fondamental d'observation pour faire avancer la science alors une exception au TDM n'est plus nécessaire car le TDM devient un droit fondamental.

Un droit d'observation des publications scientifiques est nécessaire pour synthétiser une multitude d'observations sur des domaines communs. L'informatique n'est rien de plus qu'un appareillage spécifique pour observer des données.

Le monopole des éditeurs vise à limiter ce droit d'observation des bases scientifiques numériques en imposant leurs propres outils de TDM et en contrôlant les usages (API, limitation de la recherche en nombre de mots, ...).

**A partir du moment où l'on fait l'acquisition d'un contenu pour acquérir un droit de lecture le droit d'observation ne doit pas être limité.**

### **Infrastructures essentielles et tiers de confiance**

Le droit du TDM étant autorisé dans ISTEEX, la plateforme pourrait être reconnue comme un modèle de tiers de confiance pour le dépôt/ l'archivage des copies intermédiaires.

ISTEEX pourrait également être reconnue comme Infrastructure essentielle.

D'autres organismes, comme la BNF, pourraient être envisagé comme tiers de confiance.

### **Période d'embargo et dépôt**

Jean-Marie PIERREL indique qu'idéalement il ne devrait pas exister de période d'embargo sur les publications scientifiques mais sur les données scientifiques.

Cette période d'embargo permet en effet de garantir la pleine utilisation des données scientifiques par les équipes sur la longueur d'un projet de recherche ou d'une thèse.

Par ailleurs, les métadonnées doivent être accessibles dès le départ afin de garantir le partage d'information dans le cadre des projets en cours afin d'éviter de doubler les dépenses dans le cadre par exemple de projets où un même ensemble de corpus est nécessaire.

Marie-Pascale LIZEE, indique que la période d'embargo recommandée par la Commission européenne de 6 mois après la date de 1<sup>ère</sup> publication et 12 mois pour les SHS est soutenue par le MENESR mais une levée de bouclier des éditeurs dans le cadre du projet de loi numérique a amené à un retour en arrière sur le projet de loi avec une proposition d'embargo de 12 mois et 24 mois pour les SHS.

Jean-Marie PIERREL indique que pour certains domaines de recherche qui nécessitent une période d'embargo sur les données de 2 à 4 ans cela doit être possible (tout en étant argumenté).

La période d'embargo de 24 mois sur les SHS exprime clairement la crainte des éditeurs français mais concernent surtout les ouvrages et non les articles. Une période d'embargo sur les articles pour les éditeurs de revues scientifiques en SHS ne les impacterait pas. Les éditeurs français doivent être éduqués à cette réalité pour freiner cette levée de bouclier.

Grégory COLCANAP propose que sur la mise à disposition des données de recherche soit introduit dans le mécanisme de financement une mention indiquant que lorsque les données sont produites dans le cadre d'un projet financé par un organisme public alors celles-ci doivent devenir publiques à la fin du projet (ou d'une période d'embargo).

La période d'embargo doit être supprimée sur les publications scientifiques mais pas sur les ouvrages.

## f. ABES, Jérôme KALFON, 5 octobre 2015

### Participants :

#### Pour l'ABES :

- Jérôme KALFON, Directeur de l'ABES

#### Pour le Cabinet BENSOUSSAN :

- Maître Alain BENSOUSSAN, Avocat à la Cour spécialisé en droit des technologies avancées
- Maître Laurence TELLIER- LONIEWSKI, Avocate, directrice du pôle Propriété intellectuelle
- Maître Sarah LENOIR, Avocate, pôle Propriété intellectuelle

#### Pour la DIST :

- Laurence EL KHOURI, Directrice Adjointe

### Présentation de la démarche

Comme évoqué lors du séminaire juridique ISTEEX de juillet 2014 et en Comité Exécutif, le Cabinet BENSOUSSAN, en collaboration avec la DIST et la DAJ du CNRS, entreprend, suite aux issues soulevées par le projet ISTEEX, la rédaction d'un Livre Blanc « Une science ouverte dans une république numérique ».

Cette démarche s'inspire des problématiques soulevées lors du séminaire de juillet 2014 sur la sécurisation juridique d'ISTEEX sur les thèmes: objectifs, gestion des bases de données, fouille de texte et de données (Text and Data Mining – TDM), interdisciplinarité, agrégation de contenus.

Thèmes évidemment présents dans le projet de loi numérique porté par la Secrétaire d'Etat chargée du numérique, Axelle LEMAIRE.

Le Livre blanc a pour objectif de contribuer à **évoquer le cadre juridique pour les données scientifiques afin de répondre aux préoccupations des communautés scientifiques** (publications et usages des données, droit des plateformes de Science, fouille de texte, ...) en étant force de proposition dans la **participation au projet de loi numérique**.

La démarche proposée par le Cabinet BENSOUSSAN est dans un premier temps **de mener des auditions** afin de **recueillir les usages de la communauté scientifique et l'état de l'art des pratiques de valorisation de l'IST.**

L'objectif est de définir le statut que souhaite voir appliquer les communautés scientifiques à la Science ouverte et plus particulièrement quel statut appliquer aux données, aux métadonnées, aux articles ou encore aux résultats scientifiques.

### **Données et publications : l'idéal**

Pour Jérôme KALFON l'idéal de l'ouverture des données et des publications est le suivant :

Dans tous les cas il est nécessaire d'éviter les captations.

**Pour les Données** : l'ouverture des données est le principe de base, qui doit comprendre des exceptions motivées et la plupart du temps temporaires. Il convient de prendre des précautions dans l'ouverture des données en définissant des exigences éthiques et des règles en termes de confidentialité. Les barrières à la captation doivent être :

- la sécurité éthique ;
- la confidentialité ;
- la définition d'un monopole temporaire et d'une période d'embargo sur certaines données, la période pouvant varier au cas par cas ;

Alors que l'édition scientifique était un secteur très modérément rentable, la stratégie des grands groupes éditoriaux en a fait l'un des secteurs les plus profitables en saisissant deux leviers :

- La publication scientifique est un domaine vital et stratégique de toute structure académique, que ce soit pour le chercheur ou pour la ou les institutions dont il dépend : s'agissant chaque fois d'un produit non substituable et exclusif par construction, l'institution sera contrainte d'accepter une inflation du prix de la publication, même si elle est sans lien avec l'évolution des coûts de production ;
- La bascule technologique bouleverse l'organisation traditionnelle et favorise la concentration des acteurs au point de les rendre incontournables.

**Pour les Publications** : l'objectif même d'une publication est de rendre publics des contenus et par conséquent aucun filtre ou aucune période d'embargo ne devrait empêcher ou ralentir la publication. Il ne faut pas qu'il y ait de captation sur la publication de quelques manières que ce soit. L'essentiel de l'effort économique de la publication (rédaction, relecture, secrétariat de rédaction, ..) a toujours été fourni par la communauté scientifique. L'éditeur y jouant un rôle - marginal si on le compare au travail de l'édition littéraire ou artistique - de composition, impression et diffusion (gestion des abonnements).

Une grande part de la diffusion se faisait par échanges de publications et diffusions de tirés à part. L'éditeur y jouait un rôle de prestataire de services et une part des éventuels gains de la diffusion payante revenait aux sociétés savantes propriétaires du titre. La passation de contrat entre les auteurs et les éditeurs est une chose récente et a pris son essor avec la diffusion numérique.

Jamais, si l'on devait inventer maintenant un système de publication scientifique, on ne choisirait un système aussi coûteux, lent, contre-performant et complexe que celui qui est en vigueur aujourd'hui. Au moment où la rareté (limite physique du nombre d'exemplaires et de leur diffusion) laisse la place à l'abondance naturelle, il est absurde et paradoxal que l'on crée de la rareté artificielle.

Cet anachronisme a perduré trop longtemps. Il repose sur un modèle économique basé sur la cession de droits exclusifs par l'auteur à l'éditeur (phénomène récent qui s'est développé à la fin des années 1990 - peu de contrats écrits entre auteurs et éditeurs auparavant) .

Par conséquent, **les publications dans le monde de la Science doivent être complètement ouvertes, sans période d'embargo et ce sous la seule réserve du droit de paternité.**

Toutefois, si l' « on quitte le principe du désir pour revenir au principe de réalité » il convient de définir les conditions d'une transition vers un modèle conforme aux exigences de la construction des savoirs et connaissances

La définition d'une période d'embargo peut être temporaire dans le cadre d'une période de transition et ne doit être gouvernée que par l'objectif d'une mise à disposition immédiate des publications. Les périodes d'embargo doivent être les plus courtes possibles et peuvent par compromis avec les éditeurs être au maximum de 6 mois et de 12 mois pour les SHS.

## Rémunération de l'éditeur

L'édition scientifique doit trouver un nouvel équilibre économique.

L'éditeur doit être rémunéré sur le travail qu'il effectue : rémunération de la valeur ajoutée, du "travail à façon".

La publication est financée par les bailleurs de fonds de la recherche. La relation contractuelle entre le chercheur (la collectivité) et l'éditeur doit se traduire par la rémunération d'une **prestation de service**. Cette prestation doit être transparente du droit d'auteur.

Il faut en effet dissocier la propriété littéraire et artistique (PLA) du droit de la Science :

- d'un côté, l'auteur académique n'est pas rémunéré par la publication (il est d'ailleurs aujourd'hui rarissime que l'auteur touche des droits d'auteur - au contraire il lui arrive plus souvent, au travers des APC, d'avoir à payer pour être publié). L'auteur est rémunéré pour son activité académique dont la publication est une partie intégrante.
- de l'autre côté, la PLA est un système qui protège l'auteur d'une création littéraire et artistique et lui assure une rémunération. Ce dispositif traditionnel ne doit pas s'appliquer à la Science.

Pour la publication scientifique, il doit y avoir la transparence de l'acte de valeur ajoutée rémunéré de l'éditeur

Un scientifique peut parfaitement publier d'une part des ouvrages qui relèvent de la propriété littéraire et artistique et de l'autre des publications de libre parcours.

Plusieurs droits afférents à la propriété intellectuelle coexistent aujourd'hui et varient en fonction du support, du type de produit (audio, vidéo, musique, brevets, marques, etc.). Ne pas distinguer la publication scientifique de la propriété littéraire et artistique est actuellement gênant pour la Science, mais le modèle de publication scientifique ne peut qu'évoluer dans un sens conforme à ses exigences. A ne pas faire cette distinction, c'est plutôt la PLA qui est menacée.

## **Définition d'un droit spécifique de la Science**

Il est par conséquent nécessaire de **définir un droit spécifique de la Science** (comme il y en a pour le dépôt de brevets, pour les bases de données...)

Pourquoi un droit spécifique ?

- parce qu'il s'agit d'une forme de 'commerce' spécifique ;
- L'activité de recherche dont la publication est partie intégrante bénéficie dans quasiment tous les pays d'un régime fiscal spécifique (déductibilité)
- la Science est un facteur d'enrichissement collectif et de développement : les données de la science, notamment les publications doivent être des données, des contenus ouverts
- l'utilisation de ces contenus ouverts est un facteur de développement économique
- la création de richesse liée à l'ouverture de ces contenus est bien supérieure à celle induite par leur captation
- la Science est un objet particulier qui répond à une démarche spécifique : les chercheurs ont besoin de disposer et de discuter les textes, les résultats d'une recherche, de refaire les expériences (ce qui n'empêche en rien de déposer des brevets)...
- le droit de retrait, qui existe dans le droit d'auteur est incompatible avec l'éthique de la publication et du débat scientifique

## **Distinction Open access / Open process :**

Ce n'est pas parce qu'une publication est en Open access qu'il doit y avoir un transfert de propriété de l'auteur vers l'éditeur.

L'Open access permet d'avoir accès à la publication sur le site de l'éditeur. Il s'agit d'un droit de lecture chez l'éditeur. Dans un article scientifique :

- l'idée doit être de libre parcours intégral ;
- l'article scientifique peut être lu sur le site de l'éditeur (lieu de l'Open access choisi par l'éditeur). L'éditeur limite l'échange des idées à un droit de lecture.

L'Open access sur le site de l'éditeur doit être à distinguer du dépôt en archive ouverte. Ce dépôt est basé sur le concept du dépôt de la version finale remise à l'éditeur (quid par ailleurs du gap avec la version finale publiée).

Pour surmonter les problématiques liées à la propriété intellectuelle, les propositions sont les suivantes :

- définir un principe de dépôt systématique par l'éditeur dans une archive ouverte (avec rémunération du travail utile de l'éditeur), ce dispositif d'archive ouverte est une transition indispensable mais doit rester transitoire ;
- la Science en tant que règles spécifiques du développement des savoirs, devrait être dépourvue du droit patrimonial, seul le droit moral doit être maintenu.
- l'IST doit être défini comme un bien commun en tant que tel.

### **La définition d'un nouvel écosystème**

Le modèle redéfini serait le suivant :

- l'éditeur serait un prestataire de service : ses prestations seraient notamment l'organisation peer review, la labellisation, l'évaluation, la mise en libre accès et en libre traitement ; la prestation de l'éditeur ne serait pas la même selon les communautés,
- l'éditeur serait rémunéré pour cette prestation ;
- le service final proposé serait un accès et une réutilisation sans limite des publications dans tout espace, public / privé, et y compris à des fins commerciales : il s'agit d'un gold access généralisé mais sans réversibilité et sans cession de droits patrimoniaux ;
- le financement de la diffusion de la connaissance sera intégré dès le départ dans les financements de l'activité académique; (en vue macro c'est un jeu à somme nulle : cela correspond au transfert des sommes consacrées à l'achat des publications -notamment des bibliothèques - vers le financement de l'acte de publier intégré au projet académique-recherche)
- la publication scientifique serait libre de droit, sans propriété patrimoniale ; elle serait un objet différent : un commun.

Ce système maintient le métier d'éditeur en tant que tiers prestataire de service au sein de la chaîne de diffusion de l'IST. L'éditeur cesse de détenir des droits patrimoniaux sur les contenus qu'il contribue à diffuser. Le modèle est assaini par la suppression d'une captation au seul profit de l'un des acteurs, du contrôle économique et technique de l'ensemble d'une chaîne de valeur.

Tous les autres systèmes sont des systèmes de compromis. Les systèmes d'embargo, les revues hybrides, recours au TDM au travers de règles contractuelles doivent être des dispositifs transitoires.

## g. CCSD, Claude Kirchner, 15 octobre 2015

### Participants :

#### Pour le comité de pilotage du CCSD :

- Claude KIRCHNER, Président en exercice du comité de pilotage du CCSD, Conseiller du Président Inria et Directeur de recherche

#### Pour le Cabinet BENSOUSSAN :

- Maître Alain BENSOUSSAN, Avocat à la Cour spécialisé en droit des technologies avancées
- Maître Laurence TELLIER- LONIEWSKI, Avocate, directrice du pôle Propriété intellectuelle
- Maître Sarah LENOIR, Avocate, pôle Propriété intellectuelle

#### Pour la DIST :

- Renaud FABRE, Directeur
- Charlotte AUTARD, Chargée du projet IA ISTEEX

### Présentation de la démarche

La démarche de rédaction d'un Livre blanc « Une science ouverte dans une république numérique », portée par le CNRS en collaboration avec le Cabinet Alain BENSOUSSAN, s'inspire des problématiques soulevées par la sécurisation juridique du Projet d'Investissement d'Avenir ISTEEX ANR-10-IDEX-0004-02 ([www.istex.fr](http://www.istex.fr)) qui soulève de nombreuses questions juridiques par ses possibilités de TDM, d'interdisciplinarité, d'agrégation de contenus et son objectif de rendre explorable et traversable ses bases de données.

Le droit nouveau de la science à une exception particulière dans le droit d'auteur : un statut juridique des résultats scientifiques et plus particulièrement des données et des métadonnées.

Ce Livre Blanc a pour objectif de **proposer un cadre juridique pour les données scientifiques afin de répondre aux préoccupations des communautés scientifiques** (publications et usages des données, droit des plateformes de Science, fouille de texte, ...) en étant force de proposition dans la **participation au projet de loi**

**numérique.** La Science porte à l'amont la loi numérique, et elle le fait sous des formes qui doivent s'inscrire concrètement.

Dans un premier temps, le cabinet BENSOUSSAN et la DIST réuniront des groupes de travail et mèneront des auditions pour **recueillir les usages de la communauté et l'état de l'art des pratiques de valorisation.** L'association forte avec les universités et les organismes permettra l'audition de grands témoins et une parole plurielle sur les évolutions qui se préparent dans le cadre du projet de loi numérique.

Le recueil des usages et de l'état de l'art permettront au CNRS et au Cabinet BENSOUSSAN d'établir une **grille de pertinence des analyses et des pratiques qui sera comparée au cadre normatif actuel** afin d'évaluer les écarts et **élaborer des propositions de lois.**

Ce Livre blanc sera anonyme et les personnes auditionnées auront un contrôle complet si elles souhaitent ou non être citées.

La DIST et le Cabinet BENSOUSSAN recontacteront les personnes auditionnées et soumettront un premier texte pour avis aux Présidents d'universités.

### **Consultation publique pour le projet de loi « Pour une république numérique »**

Dans le cadre de la consultation publique sur le projet de loi « Pour une république numérique », Claude KIRCHNER a rencontré la Secrétaire d'Etat au numérique et son directeur de Cabinet et a alerté les correspondants au MENESR en particulier autour des points suivants :

- Les données scientifiques doivent être sous la maîtrise des scientifiques ;

Les données scientifiques comprennent aussi bien les textes, articles, des pages web, des données de calculs, les carnets de laboratoires, les programmes, etc.

Maitriser n'est pas synonyme de posséder. L'objectif n'est pas de posséder les données mais d'en avoir la maîtrise au sens d'une pleine possibilité d'utilisation. Les scientifiques doivent pouvoir accéder au texte intégral, à l'intégralité des données pour pouvoir lire, réutiliser, refaire, reproduire, exécuter des programmes et pouvoir référencer tout ou partie des données appropriées en veillant pour l'ensemble de ces actions à respecter les règles de déontologie des domaines scientifiques concernés.

- Les services autour des données doivent être ouverts à la concurrence : il faut donc en particulier donner un accès complet au TDM.

Le CCSD envisage par exemple de créer des services de fouille et d'analyse des données. Les chercheurs doivent pouvoir accéder aux données et surtout les vérifier, or aujourd'hui le secteur privé retient les informations nécessaires à cette vérification.

L'ouverture de services adaptés permettrait aux communautés d'avoir accès aux meilleurs outils possibles et ainsi de ne pas perdre leur souveraineté scientifique : être capable de développer pour une discipline, un laboratoire, un pays donné, la meilleure science au niveau international.

L'intérêt concurrent des éditeurs dans la gestion des données de science pourrait ne pas les mener à livrer l'ensemble des données nécessaires aux chercheurs.

Or, être chercheur signifie être amené à reproduire une expérience, à vérifier des données et leurs processus d'exploitation.

Aujourd'hui le biais possible des éditeurs sur les requêtes est évident pour les chercheurs et informaticiens expert en algorithmes et conscients que certaines stratégies appliquées à des algorithmes peuvent être prouvées justes alors que d'autres peuvent être biaisées. L'affaire récente de Volkswagen a récemment montré combien l'encadrement précis de l'accès aux données et aux algorithmes est fondamental.

A ce jour il n'y a pas d'évidence que les éditeurs biaiserait sciemment les accès aux données mais la possibilité de vérification lèverait tout doute immédiat ou futur.

L'article 4 du projet de loi numérique sur la «Création d'un service public de la donnée (garantie de qualité des données publiques de référence)» permet ce droit de justification essentiel à la publication scientifique.

Le droit de transparence exprimé dans l'article 4 doit s'étendre sur les requêtes et s'étendre à toutes les traces qui traduisent les interrogations et travaux des chercheurs ;

Les requêtes sont elles même des données et doivent être accessibles lorsqu'elles concernent les objets scientifiques. Toujours dans le respect de la déontologie, les données de discussions sur les réseaux sociaux scientifiques doivent aussi être accessibles et diffusables

Sur l'article 9 sur projet de loi numérique les remarques sont les suivantes, la version de l'article concerné étant la suivante à la date de cet entretien :

« Lorsque un écrit scientifique, issu d'une activité de recherche financée au moins pour moitié par des fonds publics, est publié dans un périodique, un ouvrage paraissant au moins une fois par an, des actes de congrès ou de colloques ou des recueils de mélanges, son auteur, même en cas de cession exclusive à un éditeur, dispose du droit de mettre à disposition gratuitement sous une forme numérique, sous réserve des droits des éventuels coauteurs, la dernière version acceptée de son manuscrit par son éditeur et à l'exclusion du travail de mise en forme qui incombe à ce dernier, au terme d'un délai de douze mois pour les sciences, la technique et la médecine et de vingt-quatre mois pour les sciences humaines et sociales, à compter de date de la première publication. Cette mise à disposition ne peut donner lieu à aucune exploitation commerciale. Le texte du projet de loi numérique – Article 9 »

Cette rédaction appelle les commentaires suivants.

Sur la partie « ... *Même en cas de cession exclusive à un éditeur, ...* » : comme l'indique Roberto di Cosmo dans le cadre de la consultation publique sur le projet de loi numérique sur la proposition d'Olivier Morin « Accès libre intégral et obligatoire » si les chercheurs ne cèdent pas leurs droits à un éditeur, ils ne peuvent alors pas remplir les contrats actuellement rédigés par ceux-ci qui imposent la cession des droits pour la publication des articles dans des journaux prestigieux. Il est donc essentiel que la loi garantisse la libre diffusion en ligne ailleurs et immédiatement sans cession des contrats.

« la dernière version acceptée de son manuscrit par son éditeur et à l'exclusion du travail de mise en forme qui incombe à ce dernier » : Claude KIRCHNER insiste d'abord sur un élément important du processus de publication dans une revue. Un document scientifique soumis à l'évaluation du comité éditorial d'une revue comporte trois versions successives principales qu'il est important de bien identifier. La **version auteur initiale** encore appelée **version auteur soumise** est la version du document lors de son envoi pour évaluation au comité éditorial. La **version auteur acceptée** est celle qui après évaluation par les pairs est déclarée acceptée par le comité éditorial. La **version éditeur** est celle qui, basée sur la version auteur acceptée, a été éditée par la maison d'édition (le *publisher*) pour en améliorer la présentation et la rédaction et la mettre au format de la revue si ce n'était pas déjà fait par les auteurs. C'est cette dernière version qui sera publiée par l'éditeur dans la revue.

Le travail sur la version auteur acceptée ayant été réalisé par les pairs, il ne convient pas d'y appliquer une période d'embargo : la rédaction de l'article ici est donc incorrecte. La version sur laquelle peut s'appliquer une période d'embargo est la version éditeur, c'est-à-dire celle que la maison d'édition a contribué à mettre en forme.

Ces réflexions donnent à Claude Kirchner l'occasion de préciser également que la transformation actuelle de la communication scientifique permet d'identifier trois éléments fondamentaux :

- **La publication** d'un écrit, d'un résultat, d'une étude, de données, consiste par définition à les rendre public. Cette publication peut être précédée ou suivie de qualification et de validation comme explicité ci-dessous, mais il est fondamental de noter qu'aujourd'hui, tout un chacun peut publier. Publier un tweet, un texte sur son blog ou sa page web, un article dans une revue, une communication à une conférence, une opinion ou un texte élaboré dans un réseau social scientifique, etc.
- **La qualification** : Un document, un texte, un ensemble de donnée, qu'il soit publié au sens précédent ou pas, peut être relu pour en évaluer la qualité, l'originalité, l'apport relativement à l'état de l'art, le référencement, la qualité d'écriture, ... Cette qualification peut être faite dans le cadre du comité éditorial d'une revue ou du comité de programme d'une conférence, on parle alors de revue par les pairs (*peer-review*). Mais cette qualification est aussi souvent conduite ailleurs : dans des courriers échangés entre quelques personnes, dans les réseaux sociaux scientifiques, dans les forums de discussion spécifiques à une communauté ou dans des workshops. Elle peut aussi être effectuée lors de l'évaluation de propositions de recherche dans le cadre de l'ANR, de l'ERC ou d'H2020 par exemple. Dans le cadre de science citoyenne, elle peut être le résultat de processus différents, tels que des votes.
- **La validation ou certification** : Suite à un processus de qualification, une communauté de chercheurs comme par exemple un comité éditorial de revue ou un comité de programme de conférence peut décider de valider (on peut aussi dire certifier) le document ou l'ensemble de données. Dans le cas d'un texte scientifique, cette validation par un comité éditorial se traduit par l'acceptation de l'article dans une revue. Par abus de langage, mais aussi parce qu'avant la révolution numérique, publier en dehors d'une revue était difficile voir impossible, on confond encore validation et publication. Il est crucial de noter que ce sont en fait deux éléments fondamentalement différents.

Dans ce cadre, indépendamment de la consultation et suite à un processus de maturation entamé depuis le tout début des années 2000, on peut noter que le Président d'Inria a diffusé très récemment une note qui s'applique à l'ensemble des scientifiques effectuant leur recherche dans une équipe-projet d'Inria. Elle demande à ce que toute publication soit faite en texte complet (*full text*) dans une archives ouverte comme HAL et que pour l'évaluation de l'institut, de ses équipes-projets et de ses chercheurs, seront uniquement prises en compte ces publications. Dans les cas où ces publications sont également faites dans une revue ou une conférence, la note demande à ce que la version déposée dans HAL soit ou la version initiale de l'article ou bien (non exclusivement) la version auteur acceptée. La note indique également que si un éditeur va à l'encontre de ce dépôt sur HAL, en contestant par exemple la mise en ligne libre de la version auteur acceptée, Inria s'engage à prendre en charge cette responsabilité.

La disposition : « Cette mise à disposition ne peut donner lieu à aucune exploitation commerciale » appelle le commentaire suivant. Le contenu d'un écrit scientifique est bien entendu la source potentielle d'innovations dont l'exploitation commerciale peut être considérable. La valorisation des avancées scientifiques, et par conséquent des écrits qui les décrivent, est une des missions fondamentales des scientifiques dans le cadre des organismes de recherche et des universités. Interdire l'exploitation commerciale d'un article scientifique par ses auteurs et leurs employeurs est donc contraire aux missions fondamentales des écoles, organismes de recherche et universités..

« II. – Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite. Elles ne s'appliquent pas aux contrats en cours. » : À ce jour les contrats en cours avec les grands éditeurs sont signés pour une durée courant jusqu'en 2018. L'application de la loi ne permettrait donc pas d'agir avant la fin de ceux-ci.

La loi doit inscrire une application aux contrats en cours ou cette mention doit être retirée pour laisser la porte ouverte.

### **Captation légitime et abusive**

Claude KIRCHNER a indiqué que toutes les données scientifiques, au moins avant leur version éditeur, doivent être mises à disposition de tous : public et privé.

Lors des précédentes auditions un consensus a pu être noté sur le libre parcours des données mais pas sur la période d'embargo. Celle-ci est différente en fonction des

usages de chaque discipline et en fonction de la nécessité au chercheur de pouvoir valoriser leur recherche avant sa diffusion et son partage.

Selon Claude KIRCHNER, aujourd'hui l'élément essentiel lorsqu'on rend public un résultat est de déterminer en amont si celle-ci doit ou non être protégé.

Inria est attentif avant la mise en ligne publique d'un théorème, d'un procédé, d'un article, de données, d'un programme ou d'un algorithme sur la volonté du chercheur et des co-auteurs de publier cet objet.

Dans certains cas :

- L'institut peut émettre un avis défavorable sur la publication mais l'auteur a la décision finale en toute connaissance de cause.
- L'institution peut empêcher la diffusion si le contenu va à l'encontre de l'éthique. En effet, un certain nombre de publications ou de programmes ne peuvent pas être rendus directement publics comme par exemple un programme qui permettrait de pirater des méthodes de chiffrement et permettraient des détournements bancaires

L'organisation des différentes entités qui encadrent l'activité de recherche doit permettre la vérification des publications en OA (respect éthique, possibilité de valorisation) et d'informer les chercheurs avant le dépôt dans une archive ouverte.

### **Services et innovation**

Afin de garantir la maîtrise des données, le CCSD via la plateforme HAL souhaite créer des services de qualité internationale. Si l'Etat accompagne cette loi et la rédaction de cet article sur le libre accès, notamment au travers de la proposition de la DIST sur l'article 9 « Une durée d'embargo plus courte, ne pas entraver le TDM (fouille de texte et de données) et ne pas interdire une exploitation commerciale », il faut également mettre en place des services dans le cadre du public qui soient compétitifs par rapport aux offres du privé.

La mise en place de ces services publics pourrait, par exemple, prendre la forme d'une reconnaissance de certaines plateformes de science comme Infrastructures essentielles.

## 7. Tableau récapitulatif des législations étrangères : Open access et TDM

Pays	Principes	Références
<p><b>Union Européenne (depuis 2012)</b></p>	<p>Réflexions autour de l'introduction du TDM et de l'Open access dans la réglementation européenne.</p>	<p>Recommandation de la Commission européenne « <i>Access to and preservation of scientific information</i> » du 17 juillet 2012, C(2012) 4890.</p> <p>Publication par un groupe d'experts de la Commission européenne d'un rapport intitulé « <i>Standardisation in the area of innovation and technological development, notably in the field of Text and data mining</i> » en avril 2014.</p> <p>Mise en place du projet "OpenAire" et modification des lignes directrices « <i>Open Access Guidelines for research results funded by the ERC</i> » en décembre 2014.</p> <p>Adoption du rapport Reda par le Parlement européen (Rapport sur la mise en œuvre de la directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information (2014/2256(INI)). Prône :</p> <ul style="list-style-type: none"> <li>- L'impératif d'évaluer avec soin la mise à disposition des techniques analytiques automatisées des textes et des données (par exemple la "fouille de textes et de données") à des fins de recherche.</li> </ul>
<p><b>Allemagne (2013)</b></p>	<p>Reconnaissance d'un droit des auteurs de contributions scientifiques de rendre leurs manuscrits disponibles au public à l'expiration d'un délai de 12 mois.</p>	<p>Art. 38 (4) du German copyright Act<sup>138</sup>, 2013.</p>

<sup>138</sup> "The author of a scientific contribution which is the result of a research activity publicly funded by at least fifty percent and which has appeared in a collection which is published periodically at least twice per year has the right, even if he has granted the publisher or editor an exclusive right of use, to make the contribution available to the public in the accepted manuscript version upon expiry of 12 months after first publication, unless this serves a commercial purpose. The source of the first publication shall be indicated. Any deviating agreement to the detriment of the author shall be ineffective."

Pays	Principes	Références
<b>Royaume-Uni (depuis 2013)</b>	1/Introduction d'une exception pour l'exploration de données	1/Copyright, Designs and Patents Act, Article 29A "Copies for text and data analysis for non-commercial research", octobre 2014 <sup>139</sup> .
	2/Mise en place d'une solution duale de « green » et de « gold »	2/Publication d'un rapport sur l'Open access par le Parlement Britannique en septembre 2013.
<b>Espagne (2011)</b>	Dépôt d'une copie de la version finale de l'article des chercheurs dont les recherches sont principalement financées par le gouvernement, le plus rapidement possible, sans dépasser 12 mois après la publication dans une archive institutionnelle ou thématique.	Loi sur la Science, la Technologie et l'Innovation, article 37, 2011.
<b>Italie (2013)</b>	Publication des travaux des chercheurs, dont les recherches sont au moins financées pour moitié par des fonds publics, dans des revues en libre accès ou dépôt du manuscrit final dans une archive institutionnelle ou thématique dans un délai fixé par la loi.	Loi relative à la valorisation de la culture, article 4, 8 août 2013.
<b>Etats-Unis (depuis 2008)</b>	1/ Introduction de dispositions légales relatives à la mise à disposition du public des travaux de recherches financés par le National Institutes of Health (NIH). Cette loi prévoit que tous les articles publiés dans des revues faisant suite à des travaux financés par le NIH, doivent être versés dans l'archive ouverte en ligne propre au NIH, le « National Library of Medicine's	1/Consolidated Appropriations Act, 2008. 2/ Loi sur l'accès équitable à la recherche scientifique et technologique) (FASTR) au Congrès en février 2013. 3/Affaire Authors Guild v. Google (14 novembre 2013) dans le cadre de la mise en œuvre d'un vaste programme de numérisation des livres dans le but de constituer une bibliothèque numérique accessible au plus grand nombre.

<sup>139</sup> "(1) The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that :

(a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non commercial purpose, and

(b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).

(2) Where a copy of a work has been made under this section, copyright in the work is infringed if

(a) the copy is transferred to any other person, except where the transfer is authorized by the copyright owner, or

(b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorized by the copyright owner.

(3) If a copy made under this section is subsequently dealt with

(a) it is to be treated as an infringing copy for the purposes of that dealing, and

(b) if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.

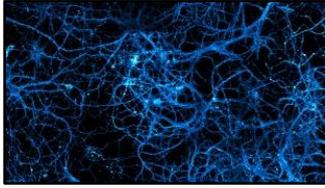
(4) In subsection (3) "dealt with" means sold or let for hire, or offered or exposed for sale or hire.

(5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable."

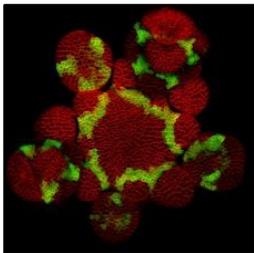
Pays	Principes	Références
	<p>PubMed Central ». Les contrats avec les éditeurs doivent explicitement le permettre.</p> <p>2/ Présentation du “Fair Access to Science and Technology Research Act”.</p> <p>3/Reconnaissance que les opérations de TDM peuvent être couvertes par l’exception de « fair use ».</p>	
<b>Canada</b>	Reconnaissance de principe des opérations de TDM dans le cadre de l’exception de « fair dealing ».	Exception de « fair dealing »
<b>Japon (2009)</b>	Introduction d’une exception pour l’exploration de données.	Article 47 septies of the Japan Copyright Act <sup>140</sup> introduit en 2009.
<b>Israël</b>	Introduction de la notion de « fair use » dans la législation israélienne et réflexion sur l’application du « fair use » aux opérations de TDM.	Amendements de 2007 à l’Israël Copyright act.
<b>Argentine (2013)</b>	Création de réservoirs institutionnels par les institutions de recherche, dans lesquels sont déposés les résultats de recherches financées par des fonds publics.	Loi initiée par le ministère de la Science, Technologie et Innovation portant sur la création d’archives ouvertes numériques et institutionnelles, loi n°26899, 2013.

<sup>140</sup> “For the purpose of information analysis (‘information analysis’ means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other much information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use by a person who makes an information analysis.”

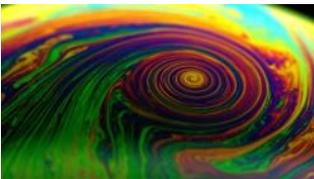
## **Sources images :**



20120001\_1920  
© Sébastien MARAIS/Daniel  
CHOQUET/Bordeaux Imaging  
Center/CNRS Photothèque



20090001\_0466  
© Pradeep DAS/ENS Lyon/CNRS  
Photothèque



20140001\_0904  
© Hamid KELLAY/LOMA/CNRS  
Photothèque