



UNIVERSITÉ DE NANTES

UNIVERSITE DE NANTES  
Faculté de Droit et des Sciences Politiques de Nantes

***Le data mining à des fins de recherche,  
vers une nouvelle exception au droit d'auteur***

Mémoire réalisé sous la direction de Madame le Professeur Carine Bernault  
présenté et soutenu par Anne-Claire Morel  
En vue de l'obtention du diplôme de Master 2 Droit de la propriété intellectuelle  
Année 2014-2015



UNIVERSITÉ DE NANTES

UNIVERSITE DE NANTES  
Faculté de Droit et des Sciences Politiques de Nantes

***Le data mining à des fins de recherche,  
vers une nouvelle exception au droit d'auteur***

Mémoire réalisé sous la direction de Madame le Professeur Carine Bernault  
présenté et soutenu par Anne-Claire Morel  
En vue de l'obtention du diplôme de Master 2 Droit de la propriété intellectuelle  
Année 2014-2015

## Sommaire

Introduction

### **Partie I. Les obstacles au *data mining* à des fins de recherche**

Chapitre I. Une infinité d'autorisations à solliciter

Chapitre II. Des solutions contractuelles insuffisantes

### **Partie II. Vers une nouvelle exception au droit d'auteur**

Chapitre I. Un contexte propice à la création d'une exception

Chapitre II. Le cadre d'une nouvelle exception en faveur du *data mining* à des fins de recherche

## Introduction

**1. Définition.** « Le droit de la propriété intellectuelle n'est pas assez accueillant aux nouvelles valeurs économiques qui émergent dans le secteur de la communication »<sup>1</sup>. Voilà le constat que Monsieur Lucas dressait il y a près de trente ans. C'est le constat que l'on serait enclin à appliquer, à première vue, au *data mining*. Le *data mining* est une technique de recherche algorithmique consistant à explorer et à analyser de manière automatique une masse de données pour en extraire l'information pertinente, mettre en lumière des informations cachées et y découvrir des corrélations significatives. Le *data mining* est un outil qui permet de produire de la valeur par l'extraction de nouvelles connaissances nées du traitement de la masse de publications, données et informations produites quotidiennement dans l'univers numérique. En 2001, le *Massachusetts Institute of Technology*, l'un des leaders mondiaux de l'enseignement et de la recherche scientifique et technologique, présentait le *data mining* comme l'une des dix technologies émergentes qui allaient changer le monde au XXI<sup>ème</sup> siècle<sup>2</sup>. Certains n'hésitent pas à parler de révolution industrielle<sup>3</sup>.

**2. Les raisons de ce succès.** Le *data mining* est effectivement aujourd'hui une activité en plein essor pour plusieurs raisons. L'une d'entre elles réside dans le volume croissant des données disponibles et utilisables. Cette inflation des données est un phénomène relativement nouveau<sup>4</sup>. Les Anglo-Saxons ont inventé un terme pour qualifier cet immense gisement de données numériques : les *big data*, ou données de masse. Ce mouvement est renforcé par l'ouverture progressive des données publiques des institutions, collectivités et entreprises dans le cadre de la politique d'*open data*<sup>5</sup>. On assiste ainsi à « un triple mouvement d'accroissement, de dématérialisation et de diversification des sources d'information documentaire »<sup>6</sup> qui font du *data mining* une technologie si prometteuse.

L'explosion du volume des données et les facilités d'accès à ces innombrables contenus sont des facteurs essentiels mais la dimension nouvelle est avant tout née de l'apparition d'outils informatiques d'analyse de plus en plus performants. C'est l'utilisation d'algorithmes qui permet ici d'exploiter les données et d'en extraire de la connaissance utile, opérations qui n'auraient jamais pu être effectuées de façon manuelle. Or, ces abondantes ressources ont une valeur considérable, ainsi que l'atteste le vocabulaire employé : « Les données sont le carburant de l'économie numérique, le pétrole de l'internet, l'or noir des big data »<sup>7</sup>. Valeur que le *data mining*, mode entièrement nouveau d'exploitation des œuvres, a su mettre en lumière de façon inédite.

**3. Les applications multiples du data-mining.** Le *data mining* trouve des applications dans de nombreux domaines. Le *datajournalism*, ou journalisme de données, initié par les médias anglo-saxons<sup>8</sup>, est une pratique professionnelle reconnue et considérée comme la plus importante de ces dernières années. Le *data mining* offre également des perspectives ambitieuses dans le champ des

---

1 A. Lucas, *Le droit de l'informatique*, Presses Universitaires de France, coll. Themis, 1987, page 360, n°308

2 « *Emerging Technologies that will change the world* » *MIT Technology Review*, janvier 2001  
<http://www2.technologyreview.com/featured-story/400868/emerging-technologies-that-will-change-the-world/>

3 J. Henno, « L'usine 4.0, nouvelle révolution industrielle », *Les Echos*, 8 octobre 2013.

4 K. Cukier et V. Mayer-Schönberger « Mise en données du monde, le déluge numérique », *Le Monde diplomatique* juillet 2013, pages 1, 20 et 21.

5 Exemple du plan d'actions pour l'ouverture des données publiques du Département de Loire-Atlantique  
<http://data.loire-atlantique.fr/accueil/>

6 G. Illien, « Enjeux professionnels. Décrire les objets du savoir, les nouveaux paradigmes du savoir », *Documentaliste-Sciences de l'Information*, 2013/3, (Vol. 50), p. 26-41.

7 L. Marino, « le big data bouscule le droit », *Revue Lamy droit de l'Immatériel*, 2013, page 99.

8 Voir le *datablog* du site *The Guardian* (consulté le 15 avril 2015).

sciences humaines et sociales, en termes d'analyse comportementale et de prise de décision dans des domaines comme le marketing ou la stratégie commerciale et, formidable outil de prévision, il intéresse des secteurs aussi variés que l'industrie pharmaceutique, la banque, la détection de la fraude, l'environnement, les télécommunications, la biologie. A titre d'illustration, citons le suivi des épidémies en fonction des requêtes émises sur internet liées à des maladies ou le suivi des flux journaliers de personnes dérivant des taux d'utilisation des réseaux de téléphonie mobile.

**4. Le *data mining*, un enjeu majeur de la recherche.** Le *data mining* est donc un outil qui permet de passer des données à la connaissance. Il facilite la compréhension des phénomènes et mène au Savoir. Il intéresse donc directement le monde de la recherche. Mais il marque une incontestable évolution des conditions de travail du chercheur et ouvre des possibilités inédites au travail scientifique<sup>9</sup>. Le *data mining* ouvre de nouveaux espaces de recherche. Dans le monde de la recherche, on ne lit plus forcément les documents, on a aussi besoin de traiter et de fouiller des corpus de textes, d'images, de chiffres pour analyser des séries, comparer des résultats issus de sources différentes, interroger des informations sans aucun lien apparent, et ce afin de vérifier des hypothèses, des phénomènes ou des comportements. La science n'est plus seulement le fruit d'observations expérimentales, elle naît aussi de la manipulation à grande échelle des données et publications scientifiques. Il s'agit ici d'un nouvel usage des ressources et d'un profond changement dans les méthodologies de recherche<sup>10</sup>, porteur d'opportunités nouvelles et ambitieuses.

**5. La propriété intellectuelle comme un obstacle au *data mining*.** Le *data mining* nécessite l'extraction automatisée de données de la recherche, de corpus de textes ou d'images, hébergés dans les grandes bases de données des éditeurs scientifiques, dans des archives ouvertes ou même parfois sur le Web. Certains de ces contenus sont protégés par le droit d'auteur, d'autres par le droit *sui generis* des producteurs de bases de données, d'autres encore peuvent être fouillés librement. Mais dans l'immense majorité des cas, l'appropriation de ces contenus par des opérations de *data mining* ne sera licite que si le titulaire des droits y a consenti. Le droit apparaît alors comme un frein au développement d'une technique essentielle au progrès scientifique. Le discours de Madame Neelie Kroes, Vice-Présidente de la Commission européenne dans un sommet sur la propriété intellectuelle et l'innovation le 10 septembre 2012 illustre ce constat :

*« Huge changes have taken place in the research area. Today, new scientific discoveries don't just come from new experiments, new drugs, new clinical trials: in fact, now, we can get new results by manipulating existing data. Data and text-mining techniques now lie behind a huge field of research, like human genome projects, potentially life-saving. They could hold the key to the next medical breakthrough, if only we freed them from their current legal tangle. Research activities are not clearly exempted from the copyright rules and there are many different rules in the 27 member states »*<sup>11</sup>.

9 « Par comparaison aux travaux antérieurs, la première nouveauté de la démarche développée ici est d'avoir cherché à rassembler des sources historiques aussi complètes et systématiques que possible afin d'étudier la dynamique de la répartition des richesses.(...) Nous avons pu, grâce aux possibilités nouvelles offertes par l'outil informatique, rassembler sans peine excessive des données historiques à une échelle beaucoup plus vaste que nos prédécesseurs. Sans chercher à faire jouer un rôle exagéré à la technologie dans l'histoire des idées, il me semble que ces questions purement techniques ne doivent pas être totalement négligées. Il était objectivement beaucoup plus difficile de traiter des volumes importants de données historiques à l'époque de Kuznets, et dans une large mesure jusqu'aux années 1980-1990, qu'il ne l'est aujourd'hui(...). Le présent livre reflète en grande partie cette évolution des conditions de travail du chercheur. » Thomas Piketty, *Le capital au XXIe Siècle*, Paris, Seuil, 2013. Exemple tiré d'un article de G. Colcanap et C. Perales, « CSPLA - Mission relative au *data mining* : l'analyse de Couperin et de l'ADBU » <http://www.abes.fr/>

10 Sur ce sujet, S. Pouyllau, « Enjeux professionnels. Web de données, Big Data, Open Data, quels rôles pour les documentalistes ? », *Documentaliste-Sciences de l'Information*, 3/2013, Vol.50, p. 26-41.

11 [http://europa.eu/rapid/press-release\\_SPEECH-12-592\\_en.htm?locale=en](http://europa.eu/rapid/press-release_SPEECH-12-592_en.htm?locale=en) (consulté le 2 avril 2015)

Les droits de propriété intellectuelle semblent gêner le *data mining* et ne permettent pas en l'état actuel aux chercheurs européens et surtout français d'utiliser cette technique d'exploration des données dans un cadre juridique simple et fiable. A ces difficultés s'ajoutent celles liées à la multiplicité des approches au sein de l'Union Européenne, source de déséquilibres et d'inquiétudes, ainsi qu'une certaine défiance de la communauté des chercheurs.

**6. Un constat partagé par la communauté scientifique.** Dans un rapport publié en 2012 sur la gestion et le partage des données de la recherche, l'Institut National de la Recherche Agronomique (INRA) pose un diagnostic sévère et soulève deux points essentiels :

- « L'indifférence, la méconnaissance, voire la défiance de la plupart des chercheurs relativement aux problèmes juridiques.
- Le développement de la recherche au sein de consortiums internationaux regroupant des organismes n'ayant pas les mêmes environnements juridiques ni les mêmes positions stratégiques. Le problème qui se pose est alors de savoir comment respecter les mesures légales et les politiques des différents organismes ou pays »<sup>12</sup>.

**7. Des questionnements juridiques multiples.** Le *data mining*, s'il interroge le métier de chercheur, questionne aussi incontestablement certaines notions fondamentales du droit d'auteur. La première interrogation réside dans la mise en œuvre ou non du droit d'auteur. Cette question est l'objet de vifs débats entre les différentes communautés confrontées au *data mining*. Comme pour d'autres usages nouveaux, certains réclament ici des espaces de liberté. Comme l'explique Maître Vercken, « Ici, c'est l'objet même de la propriété dont l'existence est déniée (...), il ne s'agit plus, au sein d'un espace de propriété, de reconnaître des zones de liberté, mais de créer des espaces de liberté autonomes ne pouvant être assujettis aux espaces de propriété »<sup>13</sup>.

Ces revendications nouvelles sont apparues en raison des particularités du *data mining*, qui exploite les oeuvres dans leur fonction informationnelle. L'une des questions posées est de savoir si l'on doit prendre en compte cette fonction informationnelle des contenus pour en déduire un changement de leur nature juridique. « Faut-il distinguer l'usage de ces contenus selon la finalité poursuivie – usage de la création ou de la donnée, usage de l'information sur la création ou sur la donnée, usage des métadonnées résultant du traitement massif des créations ou des données à caractère personnel ? »<sup>14</sup>. Dans l'environnement numérique, l'information a acquis une place et une valeur considérables, qui amènent certains auteurs à réclamer la création « d'un droit spécifique qui protège ceux qui manipulent l'information et la mettent à la disposition d'un public dans un contexte différent du contexte initial »<sup>15</sup>. Lorsque l'information est utilisée pour sa valeur documentaire et non dans un objectif premier, ce qui est le cas du *data mining*, elle devrait selon eux obéir à des règles nouvelles permettant aux professionnels de pouvoir s'abstraire des « multiples couches juridiques » qui saisissent l'information et en complexifient l'utilisation.

A ces questions sur la mise en œuvre du droit d'auteur et la remise en cause désormais classique des droits de propriété intellectuelle, s'ajoutent celles posées par la création d'une nouvelle exception en faveur du *data mining*, dans un souci légitime d'incitation à la création et de compétitivité, idées auxquelles se réfère volontiers la Commission européenne.

---

12 Institut National de la Recherche Agronomique (INRA) *Rapport du groupe de travail sur la gestion et le partage de données*, Paris, juin 2012, 62 p. (consulté le 28 avril 2015)

[http://www.pfl-cepia.inra.fr/uploads/gdp\\_docs/Rapport-GestionDonnees-web.pdf](http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf)

13 G. Vercken, « Liberté et propriété : soumission ou autonomie ? », *Documentaliste-Sciences de l'Information*, 2014/4, Vol.51, p.24-26.

14 C. Zolynski, « Réforme du droit d'information : quels défis relever ? », *Documentaliste-Sciences de l'Information*, 2014/4, Vol. 51, p.48-50.

15 C. Manara, « Le droit de l'information : paradoxes et ouvertures », *Documentaliste-Sciences de l'Information*, 2012, Vol. 49, n°3, p.18-19.

**8. La recherche, un monde à part.** Toutes ces questions trouvent une résonance particulière lorsque l'on s'interroge dans le cadre spécifique de la recherche scientifique. Il existe en effet un paradoxe qui anime aujourd'hui la commercialisation de l'information scientifique : bien que la communauté scientifique valorise le partage désintéressé et l'échange libre des connaissances et considère la littérature scientifique comme un bien public, les travaux et données scientifiques issus de la recherche ont progressivement acquis le statut de marchandises. Ce qui fait du *data mining* dans ce domaine spécifique un enjeu stratégique pour les éditeurs commerciaux et les chercheurs. Ce paradoxe donne naissance à une réflexion foisonnante, notamment de la part de la communauté des chercheurs et se révèle pour le juriste un formidable terrain d'étude.

De plus, étudier le *data mining* dans le contexte de la recherche scientifique revient à regarder à la loupe les problématiques posées par le droit d'auteur aujourd'hui. En effet, le mouvement de contestation du droit d'auteur est particulièrement fort au sein de la communauté des chercheurs, et si le mouvement du libre accès a largement investi ce secteur, bouleversant l'environnement qui entoure la création et la publication scientifique, c'est parce que c'est l'un des domaines où le droit d'auteur semble s'être éloigné de sa fonction. Rappelons que le droit d'auteur trouve son origine, son sens dans l'acte de création d'un auteur. Comme le rappelle Monsieur Gautier, les auteurs méritent deux récompenses :

« D'une part, rester les maîtres de leur création, fruit de la conjonction de leur intelligence et de leur sensibilité ; ils doivent donc conserver autorité et puissance sur l'oeuvre (...) ; d'autre part, ils méritent, comme les autres catégories de travailleurs, d'assurer leur subsistance, en conférant la jouissance de leur œuvre au public, moyennant rémunération »<sup>16</sup>.

Or, dans le domaine de la recherche scientifique, il faut avouer que le droit d'auteur, loin d'avoir su protéger les auteurs, apparaît avant tout comme le moyen de rémunérer un investissement, celui des grands éditeurs commerciaux.

**9. Des enjeux essentiels.** Il s'agit donc de s'interroger sur la pertinence du cadre juridique entourant aujourd'hui l'accès et la valorisation des contenus numériques. Il paraît essentiel de se demander de quelle façon le droit peut accompagner le développement du *data mining*, qui fait à l'évidence partie de ces traitements innovants de l'information, porteur d'avenir dans le domaine de la recherche et du progrès scientifique. Il nous semble important de trouver le moyen d'autoriser la réexploitation des contenus protégés tout en préservant une protection efficace et légitime des droits sur ces multiples contenus. Il s'agit en réalité de déterminer les conditions du développement du *data mining* mené à des fins de recherche dans le respect du droit de la propriété littéraire et artistique.

Les questions posées par le *data mining* s'inscrivent dans une réflexion active menée au niveau européen en vue du développement du marché unique du numérique, qui apparaît aujourd'hui comme l'une des grandes priorités de la Commission européenne. Comme le résume Madame Zolynski, « Il s'agit de penser un cadre idoine garantissant la promotion des nouveaux usages de biens culturels dans un environnement à la fois ouvert à la diffusion des connaissances, protecteur des droits sur les contenus et conforme aux besoins du marché »<sup>17</sup>.

---

16 P.-Y. Gautier, *Propriété littéraire et artistique*, PUF, coll. Droit fondamental, 9e éd., 2015.

17 C. Zolynski, « Réforme du droit d'information : quels défis relever ? », *Documentaliste-Sciences de l'Information* 2014/4, Vol.51, p.48-50.

**10. La position de la France.** Si le sujet intéresse l'Union européenne, il est aussi l'objet de discussions en France. Le Conseil Supérieur de la Propriété Littéraire et Artistique (CSPLA) s'est vu confier une mission de réflexion sur l'exploration de données, présidée par Maître Jean Martin et, à la suite de multiples et diverses auditions, a rendu son rapport<sup>18</sup> en juillet 2014. Le rapport recommande le « développement de solutions contractuelles » et privilégie « l'auto-régulation à une intervention législative, afin de favoriser une meilleure prise en compte des enjeux et des besoins d'un phénomène encore émergent, en forte et constante évolution ». Après avoir souligné que les réflexions sur l'adoption d'une nouvelle exception pour le *data mining* à des fins de recherche sont actuellement engagées au niveau européen, le CSPLA clôt le rapport par la phrase suivante :

« La rédaction des contours d'une exception est extrêmement périlleuse, c'est pourquoi il semble nécessaire de fixer un délai de deux années au terme duquel un bilan sectoriel sera dressé et l'éventuelle nécessité d'une intervention législative évaluée. »

La France semble donc adopter une position plus conservatrice que d'autres pays européens, qui réfléchissent à la mise en place d'une exception au droit d'auteur en faveur du *data mining* à des fins de recherche. Cette « frilosité » a été confirmée récemment par le rapport de Monsieur Sirinelli, dans le cadre 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<sup>19</sup>, rendu en octobre dernier, qui à son tour recommande d'éviter la création d'une nouvelle exception. A l'appui de cette préconisation, l'idée qu'une telle exception risquerait de devenir rapidement « obsolète » compte tenu du fait que l'acte d'exploitation pourrait disparaître dans un avenir proche avec les progrès de la technique.

**11. Délimitation du sujet et objectif.** Le *data mining* appliqué à la recherche scientifique est au croisement de plusieurs sujets, droit d'auteur, droit des bases de données, droit de l'information, philosophie du libre accès et il touche au sujet sensible de l'accès aux publications et données. L'objectif de ce mémoire est d'analyser, à travers les différentes questions juridiques soulevées par le sujet, le contexte légal du *data mining* à des fins de recherche en France, afin de pouvoir envisager l'opportunité d'une nouvelle exception au droit d'auteur.

**12. Plan.** Nous établirons que les obstacles juridiques au *data mining* sont nombreux et que le droit positif ne permet pas aujourd'hui de fournir aux chercheurs la sécurité juridique nécessaire au développement des pratiques de *data mining* (Partie I). Puis, nous démontrerons que des aménagements sont possibles et que la création d'une nouvelle exception au droit d'auteur pour le *data mining* à des fins de recherche non commerciale est la voie qu'il convient d'envisager (Partie II).

---

18 Conseil supérieur de la propriété littéraire et artistique (CSPLA), *Rapport de la mission sur l'exploration de données Text and Data mining*, rapport établi par M. Jean MARTIN, président de la mission et Mme Liliane de CARVALHO, rapporteur de la mission, juillet 2014

<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Propriete-litteraire-et-artistique/Conseil-supérieur-de-la-proprieté-litteraire-et-artistique/Travaux-du-CSPLA/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

19 CSPLA, Rapport de 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, P. Sirinelli, octobre 2014.

## Partie I. Les obstacles au *data mining* à des fins de recherche

**13. Objectif.** La question à laquelle il convient de répondre est de savoir si le droit d'auteur est, comme l'affirment certains, un réel obstacle au *data mining* ou s'il permet au contraire, sous certaines conditions, l'exploitation en toute sécurité des travaux, contenus et données de la recherche. Pour cela, il nous faudra mettre en lumière les spécificités de l'édition scientifique, univers complexe, protéiforme et en pleine mutation, qui font du *data mining* un enjeu particulièrement fort mais également une source d'inquiétudes.

Nous analyserons donc la façon dont le droit d'auteur mais également le droit des producteurs de bases de données s'adaptent à cet usage nouveau des œuvres qu'est le *data mining*. Puis nous nous poserons la question aujourd'hui centrale du rôle des contrats dans l'encadrement des pratiques de *data mining*. Ceux-ci permettent-ils réellement la fouille de données et si non, peuvent-ils être adaptés, comme le propose le Conseil Supérieur de la Propriété Littéraire et Artistique dans son récent rapport, afin d'assouplir efficacement le droit exclusif des titulaires de droits ?

**14. Plan.** Ces questions essentielles seront notre fil conducteur pour démontrer les problématiques posées par la variété des ressources explorées (Chapitre 1). Puis nous analyserons les limites des différentes solutions contractuelles actuellement utilisées (Chapitre 2), qu'elles soient proposées par les grands éditeurs scientifiques ou qu'il s'agisse des licences libres.

## Chapitre 1

### Une infinité d'autorisations à solliciter

**15. Position du problème.** L'utilisation d'oeuvres protégées, dans le cadre d'un projet de recherche de *data mining*, suppose pour le chercheur l'obtention des droits d'exploitation de chaque oeuvre, sous peine de se rendre coupable d'actes de contrefaçon. Compte tenu de la variété des sources et de l'immense volume de données à explorer, l'identification des droits à solliciter auprès des auteurs ou titulaires des droits et l'obtention de ces multiples autorisations sont des composantes essentielles et chronophages de tout projet de *data mining*. Or le temps et l'argent dépensés pour solliciter et obtenir les autorisations nécessaires sont à l'évidence incompatibles avec une activité de recherche soumise à des impératifs budgétaires exigeants. L'exemple souvent cité à l'appui de cette idée est celui du projet Text2Genome<sup>20</sup>, qui a nécessité la fouille de trois millions de publications scientifiques provenant de multiples bases éditoriales afin de dresser un état des lieux de l'ensemble des travaux réalisés sur le génome humain. Ce projet a pu être mené en négociant avec chaque éditeur l'accès à leurs données. Ces négociations ont duré trois ans. Monsieur Casey Bergman, chercheur en génomique à l'université de Manchester et l'un des deux initiateurs du projet, résume ainsi l'expérience :

« *We've learned it's a long, hard road with every journal* »<sup>21</sup>.

**16. Plan.** La négociation efficace des contrats suppose la maîtrise des autorisations à solliciter. Le chercheur doit donc être en mesure de distinguer, parmi les différentes ressources qu'il souhaite explorer, celles qui sont protégées au titre de la propriété littéraire et artistique de celles qui ne le sont pas (Section 1). Il doit ensuite identifier les droits d'exploitation à solliciter (Section 2).

#### Section 1. Le problème de l'identification et du statut des ressources explorées

**17. La diversité des contenus.** Les documents et données utilisés par les scientifiques pour faire de la recherche sont aujourd'hui numériques et l'échange, la diffusion et le partage de ces ressources sont au cœur des pratiques des communautés de chercheurs. Le chercheur est désormais confronté à l'exploitation de fonds informationnels très divers, dont il lui faut connaître, s'il souhaite les utiliser, les différents statuts. Il peut s'agir de corpus écrits, de données brutes, d'articles, de chiffres, d'images, d'archives de l'internet, d'informations issues du web ou des réseaux sociaux, de tableaux graphiques ou de cartes. Certaines données peuvent également être des données personnelles, particulièrement utilisées dans le domaine des sciences humaines et sociales, et posent des problèmes spécifiques que nous ne traiterons pas dans ce mémoire.

**18. Plan.** Les conditions d'accès à ces documents sont donc très variables. Le *data mining* se nourrit d'informations susceptibles d'obéir à des régimes distincts : certains contenus seront protégés par le droit d'auteur (§1) ou le droit des bases de données (§2) alors que d'autres seront des œuvres tombées dans le domaine public ou mises librement à la disposition du public (§3), grâce à des dispositifs aujourd'hui bien établis comme l'*Open access* ou les *Creative commons*.

---

20 Pour cet exemple et d'autres exemples d'application du *text mining* à la recherche en sciences exactes, v. R. Van Noorden, « Text mining spa heats up », *Nature*, 20 mars 2013

<http://www.nature.com/news/text-mining-spat-heats-up-1.12636>

21 R. Van Noorden, « Trouble at the text mine », *Nature*, vol. 483, 8 mars 2012

<http://www.nature.com/news/trouble-at-the-text-mine-1.10184>

## §1. Des œuvres couvertes par le droit d'auteur

**19. La protection des œuvres scientifiques sous condition d'originalité.** Les œuvres scientifiques sont des œuvres de l'esprit. Sous réserve d'originalité, elles sont traditionnellement incluses dans le champ de la protection, comme l'attestent l'article 2 de la Convention de Berne qui vise « toutes les productions du domaine littéraire, scientifique et artistique » ainsi que l'article L. 112-2-1° du Code de la propriété intellectuelle<sup>22</sup>. Mais l'exigence d'originalité peut dans certains cas prêter à discussion et, comme le soulignait Desbois en 1966, « les tribunaux ne peuvent pas passer au même crible les œuvres d'imagination et les ouvrages scientifiques : les mailles du filet doivent être plus larges pour celles-ci que pour celles-là »<sup>23</sup>.

**20. La protection des bases de données par le droit d'auteur.** Les ressources fréquemment utilisées par les chercheurs et particulièrement propices à l'exploration de masse sont les compilations d'informations ou de données or, la question de leur protection par le droit d'auteur est « très embarrassante »<sup>24</sup>. Le droit français définit la base de données protégée comme « un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodiques et individuellement accessibles par des moyens électroniques ou par tout autre moyen »<sup>25</sup>. Cette définition d'application large est très proche de celle de la directive du 11 mars 1996.

Si le travail de compilation d'informations n'est pas protégé en soi, comme le rappelle le célèbre arrêt Coprosa de 1989<sup>26</sup>, les créations de compilation peuvent dans certains cas être protégées, les tribunaux appréciant la sélection et le choix des matières, leur disposition, leur regroupement ou encore la réunion, le classement et l'agencement des données. Un droit d'auteur est accordé à l'auteur au titre de la structure de l'information mais le critère de l'originalité dans la forme est ici atténué. Ces bases de données visées par le Code de la propriété intellectuelle concernent aussi bien les bases qui rassemblent des œuvres de l'esprit que des bases recueillant de simples données, non protégées par le droit d'auteur. La protection par le droit d'auteur est donc indépendante du caractère protégeable ou non du contenu des données par le droit d'auteur. Pour admettre la protection, le juge doit caractériser les choix ou la présentation de la base et dire en quoi ces choix reflètent la personnalité de l'auteur. La Cour de Justice de l'Union européenne admet l'originalité de la base si à travers le choix ou la disposition, l'auteur « exprime sa capacité créative de manière originale en effectuant des choix libres et créatifs et imprime sa touche personnelle »<sup>27</sup>. Mais en pratique, il est souvent difficile de faire la distinction entre les données brutes qui ne donnent pas prises au droit d'auteur et les données brutes retravaillées et sur lesquelles la personnalité de l'auteur a pu s'exprimer donnant alors naissance au droit d'auteur. Enfin, notons que la protection par le droit d'auteur est distincte de celle que peut offrir le droit sui generis du producteur. (§2)

**21. Le statut juridique des œuvres couvertes par le droit d'auteur.** L'exploitation par un tiers de ces œuvres et travaux scientifiques protégées par le droit d'auteur, leur introduction dans un corpus de données ou leur fouille via des algorithmes nécessitent l'autorisation du titulaire des droits.

---

22 « Sont considérés notamment comme œuvres de l'esprit au sens du présent code 1° Les livres, brochures et autres écrits littéraires, artistiques et scientifiques ».

23 Henri Desbois, *Le droit d'auteur en France*, Dalloz, 2e éd., 1966, n°33.

24 N. Mallet-Poujol, « Marché de l'information : le droit injustement tourmenté », *RIDA*, 4/1996, n°168, p.93.

25 Article L. 112-3 alinéa 2 du Code de la propriété intellectuelle.

26 Cass. Civ. I, 2 mai 1989, *RIDA* janvier 1990, p. 309.

27 CJUE, 1er mars 2012, Football Dataco c. Yahoo, aff. C.604/10, point 38, *Propriétés Intellectuelles*, Octobre 2012, p. 421, obs. V.-L. Benabou.

## §2. Des œuvres couvertes par le droit *sui generis* du producteur de bases de données.

**22. Principe.** La Loi du 1er juillet 1998 transposant la Directive du 11 mars 1996 (article L. 112-3 du Code de la propriété intellectuelle) complète la protection par le droit d'auteur en instituant une protection spécifique au profit du fabricant de la base de données. Ce droit *sui generis* n'est pas un droit d'auteur car il ne porte pas sur une création mais il protège l'investissement que représente la compilation d'un ensemble informationnel, lorsque la présentation du contenu « atteste d'un investissement financier, matériel ou humain substantiel ». Le droit porte sur le contenu de la base, c'est-à-dire sur des données. La protection est alors accordée au producteur qui doit autoriser toute extraction substantielle ou répétée conformément aux articles L. 342-1 et 342-2 du Code de la propriété intellectuelle. La durée du droit est limitée à quinze ans mais les mises à jour significatives font partir un nouveau délai de protection. Mais cette protection spécifique suscite les réticences d'une partie de la doctrine : « Ce sont toutes sortes d'informations, fussent-elles dénuées d'originalité, appartiennent-elles au fonds commun/domaine public, relèvent-elles des idées et autres méthodes, bref, tout ce qui est traditionnellement repoussé par le droit d'auteur, qui sont ainsi susceptibles d'être protégées par la loi »<sup>28</sup>.

**23. L'objet du droit exclusif.** Le producteur de la base de données est en droit d'interdire ou d'autoriser les actes d'extraction et la réutilisation de la totalité ou d'une partie substantielle de la base de données. L'extraction est définie comme le « transfert permanent ou temporaire »<sup>29</sup>, sur un autre support et s'apparente au droit de reproduction. Et, selon une jurisprudence constante, la finalité et les modalités de l'extraction sont indifférentes. Si le producteur peut invoquer son droit *sui generis* pour interdire l'extraction déloyale et la réutilisation de tout ou partie du contenu de la base, les règles diffèrent selon que les actes litigieux portent sur la totalité ou sur des parties substantielles ou selon qu'ils portent sur une partie non substantielle de la base de données. Et l'appréciation du caractère substantiel de l'extraction s'est révélée particulièrement difficile pour les juges du fond<sup>30</sup>. Les chercheurs, lorsqu'ils ne fouilleront pas la totalité de la base seront très certainement incapables de déterminer si le *data mining* porte sur « une partie qualitativement ou quantitativement substantielle du contenu d'une base de données ».

**24. Articulation avec le droit d'auteur.** Les droits du producteur de la base de données s'exercent, selon les termes de l'article L. 341-1 du Code de la propriété intellectuelle, « sans préjudice » du droit d'auteur auquel la base de données peut donner prise. En effet, le producteur n'est pas toujours titulaire du droit d'auteur sur la base et dans ce cas, il faudra alors appliquer les deux régimes de façon distributive, ce qui compliquera encore la tâche des chercheurs désireux d'obtenir les autorisations nécessaires.

**25. Conclusion sur les bases de données.** Une base de données est souvent un « complexe de plusieurs droits de propriété intellectuelle »<sup>31</sup>. Elle peut engendrer un droit *sui generis* au profit du producteur si les investissements réalisés sont substantiels. Mais nous l'avons vu précédemment, elle peut également être, sous condition d'originalité, une œuvre donnant prise au droit d'auteur. Elle peut encore contenir des informations non protégées comme des œuvres protégées en elles-mêmes par un droit d'auteur. Les bases de données sont particulièrement sollicitées dans le cadre de projets scientifiques de *data mining* et à l'évidence, cette complexité est un obstacle pour les chercheurs.

---

28 P.-Y. Gautier *Propriété littéraire et artistique*, PUF, coll. Droit fondamental, Paris, 2012, n°181 p.181.

29 Article L. 342-1 du Code de la propriété intellectuelle.

30 A. Lucas, H.-J. Lucas, A. Lucas-Schloetter, *Traité de la propriété littéraire et artistique*, LexisNexis, coll. Traités, 4e éd., 2012, n°1195 page 981.

31 F. Pollaud-Dulian, *Le droit d'auteur*, Economica, coll. Corpus, 2e éd., 2014, n°2539, p. 1651.

### §3. Des œuvres du domaine public.

**26. Définition.** Le chercheur peut également explorer des œuvres qui échappent au droit d'auteur. Ce sont les œuvres du domaine public. Nous les classerons en trois catégories :

- Les créations qui n'ont pu accéder à la protection car elles ne remplissent pas les conditions requises ;
- Les œuvres qui ne sont plus protégées par le droit d'auteur en raison de leur ancienneté, leur durée de protection ayant expiré ;
- Les œuvres qui ont été volontairement placées dans le domaine public par leurs auteurs.

**27. Le statut juridique des œuvres et données du domaine public.** Ces données sont de « libre parcours » et alimentent le fonds commun dans lequel les chercheurs peuvent librement puiser. L'œuvre du domaine public devient d'exploitation libre et gratuite et le chercheur peut librement les mettre en mémoire et les fouiller. Sous réserve bien entendu de l'existence du droit moral perpétuel. Hormis cette réserve, ces données semblent donc être aisément exploitables.

Mais il faut cependant souligner les cas où une protection de ces données pourtant « libres » réapparaît. Comme le souligne la doctrine, « y compris à l'égard de matériaux non protégés par le droit d'auteur, il est extrêmement important de s'interroger, chaque fois, sur l'opportunité d'obtenir une autorisation de reproduction, pour ne pas se voir reprocher un acte d'extraction déloyale ou de concurrence déloyale »<sup>32</sup>. En effet, nombreuses sont les décisions ayant jugé qu'est constitutif d'une faute le fait de tirer parti des créations intellectuelles d'autrui, même si ces créations ne font l'objet d'aucun droit exclusif. Et ce courant jurisprudentiel s'est trouvé renforcé par la théorie des agissements parasitaires qui juge fautive l'utilisation d'une création non protégée, lorsqu'elle permet de tirer profit indument des investissements réalisés par un tiers<sup>33</sup>.

**28. La question de la titularité des droits.** De plus, la seule mention que l'œuvre diffusée soit libre de droits ne permet pas de savoir si celui qui la publie a qualité pour le faire, s'il est le titulaire originaire des droits ou s'il les a acquis du véritable titulaire. À ces questions d'identification des sources et de leur accès s'ajoutent donc des questions de titularité car il faut également déterminer qui sont les titulaires des droits d'exploitation. Nous avons choisi de ne pas traiter ce volet mais relevons malgré tout que la détermination du titulaire des droits peut dans certains cas être complexe<sup>34</sup> : elle peut varier selon la loi applicable, selon la qualité de l'auteur<sup>35</sup> (salarié, agent public), selon que l'œuvre est plurale, tous ces critères se combinant. De plus, si aujourd'hui les travaux scientifiques impliquent fréquemment la mise en place de consortiums nationaux regroupant des acteurs publics et privés, de nombreux projets sont également internationaux et naissent d'une collaboration de chercheurs, représentant plusieurs organismes répartis dans le monde. La question de la titularité des droits n'en est que plus complexe.

**29. Bilan.** Cette première étape, qui consiste à identifier l'étendue des autorisations à solliciter avant d'entreprendre un projet de *data mining* se révèle donc particulièrement compliquée d'un point de vue pratique. Et fait l'objet de nombreux articles de la part de la communauté des chercheurs appelant à la simplification du processus. La seconde étape, qui consiste à identifier les droits d'exploitation mis en œuvre par les opérations de fouille est quant à elle l'objet de vifs débats.

32 N. Mallet-Poujol, *La création multimédia et le droit*, Litec, coll. [Droit@Litec](#), 2e éd., 2003 page 20 n°53.

33 A. Lucas, H.-J. Lucas, A. Lucas-Schloetter, *Traité de la propriété littéraire et artistique*, LexisNexis, 4e éd., 2012, page 22 n°21.

34 M. Cornu, « Création scientifique et statut d'auteur », *Hermès, la Revue*, 2010/2, n°57, p. 85-93.

35 Sur le sujet, voir A. Lebois, « Service public et créations intellectuelles », *Revue Lamy Droit de l'Immatériel*, 2008, n°36.

## Section 2. La question de l'identification des droits d'exploitation : *data mining* et droit de reproduction

**30. Exclusion du droit de représentation.** Le droit d'auteur comprend, outre le droit moral, des attributs d'ordre patrimonial. Ces prérogatives d'ordre économique permettent à l'auteur ou à ses ayants droit de décider des conditions d'exploitation de son œuvre et de percevoir une rémunération en contrepartie de l'utilisation de son travail. Il est admis que le *data mining* ne met pas en œuvre le droit de représentation qui « consiste dans la communication de l'œuvre au public par un procédé quelconque »<sup>36</sup>. Les œuvres fouillées ne sont utilisées que pour en extraire de la connaissance et ne font pas l'objet d'une communication au public, ce point n'est pas discuté. Seul le droit de reproduction est donc à envisager.

**31. Les termes du débat.** Selon les éditeurs, le *data mining* met en œuvre le droit de reproduction car il s'agit d'une nouvelle exploitation des contenus protégés. Mais un mouvement porté initialement par les représentants des bibliothèques universitaires, appuyé par d'actifs défenseurs des libertés dans l'espace numérique, met en avant les spécificités du *data mining* qui empêcheraient la mise en œuvre du droit de reproduction.

**32. Deux arguments contre la mise en œuvre du droit de reproduction.** Les arguments avancés avec force par les défenseurs du *data mining* à l'encontre de l'application du droit de reproduction tiennent à deux circonstances particulières :

- il s'agirait d'une simple différence d'échelle :

« Il s'agit avant tout d'une technologie intellectuelle (...) : un outil permettant de faciliter ou d'automatiser certaines opérations mentales. Au lieu de récupérer et de recouper manuellement des milliers, voire des millions de données, il est possible de confier ces tâches peu gratifiantes à des algorithmes et à des systèmes de gestion de bases de données. Si le *content mining* marque un changement d'échelle, il ne fonde pas une activité nouvelle. Extraire et synthétiser des informations préexistantes constituent le labeur quotidien du chercheur depuis que la recherche scientifique existe »<sup>37</sup>.

Certains utilisateurs mettent en avant le droit de lire comme un droit d'extraire. D'autres estiment que le *data mining* n'est que « le prolongement automatisé de pratiques anciennes de traitement de l'information opérées par les chercheurs » et qu'« il n'y a pas de raison que ce qui était libre dans l'environnement analogique devienne soumis au droit d'auteur lors du passage au numérique, sauf à brider drastiquement l'innovation et le développement de la recherche en Europe »<sup>38</sup>.

La lecture, l'analyse et l'exploitation intellectuelle d'un texte n'ont jamais mis en jeu le droit de reproduction. Or le *data mining* ne serait qu'une forme de lecture et l'utilisation de l'outil informatique, s'il provoque bien un changement d'échelle, ne saurait justifier un changement de statut légal. Le droit de lire devrait alors entraîner tout naturellement le droit d'extraire.

36 Article L. 122-2 du Code de la propriété intellectuelle.

37 P.-C.Langlais et L. Maurel, « Quel statut légal pour le data mining ? », Synthèse de Savoirscom1 consécutive à l'audition du 15 janvier 2013 par le CSPLA.

38 Réponse de la quadrature du Net à la consultation publique de la Commission européenne sur la réforme du droit d'auteur en Europe

[https://www.laquadrature.net/files/Réponse\\_de\\_La\\_Quadrature\\_du\\_Net\\_à\\_la\\_consultation\\_publicue\\_de\\_la\\_Commission\\_européenne\\_sur\\_la\\_réforme\\_du\\_droit\\_d\\_auteur\\_en\\_Europe.pdf](https://www.laquadrature.net/files/Réponse_de_La_Quadrature_du_Net_à_la_consultation_publicue_de_la_Commission_européenne_sur_la_réforme_du_droit_d_auteur_en_Europe.pdf)

- les œuvres sont explorées pour leur seule dimension informationnelle :

L'exploration de données utilise, exploite l'œuvre non pour la reproduire ou la représenter au sens du droit d'auteur mais afin de produire de la connaissance et un savoir inédit. Le *data mining* ne restitue pas la forme d'expression originale des œuvres explorées, ne fait pas d'emprunts à la forme originale des œuvres mais se contente de les analyser grâce à des algorithmes imaginés en fonction d'une finalité prédéterminée<sup>39</sup>. Il utilise l'œuvre de façon accessoire, secondaire et non pour elle-même. Or seule la forme d'expression originale est protégée par le droit d'auteur. Cette différence d'objet entre le droit d'auteur et le *data mining* justifierait selon certains que le droit d'auteur ne puisse pas s'appliquer aux activités d'exploration de données, puisque celles-ci ne reprennent pas la forme d'expression des œuvres exploitées.

A l'appui de cette thèse, l'approche fonctionnelle du droit d'auteur parfois retenue par la jurisprudence. Dans l'affaire de la *Place des Terreaux*, la Cour de Cassation a jugé dans un arrêt de la première chambre civile du 15 mars 2005 que le droit de représentation n'était pas en cause au motif que l'œuvre était présentée de façon accessoire « de sorte qu'elle ne réalisait pas la communication de cette œuvre au public »<sup>40</sup>. Or, il faut que la communication de l'œuvre au public soit suffisante pour qu'il puisse entrer en relation avec elle. Ainsi, en appliquant cette jurisprudence au *data mining*, la reproduction jugée accessoire de l'œuvre ne violerait pas le monopole de l'auteur parce qu'il n'y a pas de communication suffisante au public. Ces arguments sont-ils juridiquement recevables ?

## §1. Discussion sur les conditions de la mise en jeu du droit de reproduction

**33. Définition.** Le droit de reproduction est défini à l'article L. 122-3 du Code de la propriété intellectuelle de la façon suivante :

« La reproduction consiste dans la fixation matérielle de l'œuvre par tous procédés qui permettent de la communiquer au public d'une manière indirecte.  
Elle peut s'effectuer notamment par imprimerie, dessin, gravure, photographie, moulage et tout procédé des arts graphiques et plastiques, enregistrement mécanique, cinématographique ou magnétique».

La reproduction nécessite donc la réunion de deux éléments : il faut une fixation de l'œuvre et il faut également que cette fixation soit apte à communiquer l'œuvre au public. Attachons nous à vérifier que le *data mining* remplit bien ces deux conditions.

**34. Un premier élément, une fixation matérielle par tous procédés.** Il faut donc pour qu'il y ait reproduction qu'une « fixation matérielle » de l'œuvre soit réalisée. La fixation matérielle peut être réalisée au moyen de toutes sortes de procédés, la liste de l'article L. 122-3 alinéa 2 n'étant qu'une liste d'exemples. L'article 2 de la Directive 2001/29/CE précise que la reproduction peut être

---

39 « Le monde du Big Data implique en effet pour les acteurs une véritable révolution culturelle. Il leur faut en quelque sorte accepter de lâcher prise, en acceptant de travailler avec des données parcellaires, disponibles, rapidement accessibles : cela implique en retour de ne pas chercher à les conserver, de les considérer comme des produits consommables et même périssables. Cela amène enfin à les transformer pour un nouvel usage. » J.-P. Malle « La triple rupture des Big Data », *Paris Tech review*, 15 mars 2013.

40 C. Caron, « De la représentation accessoire d'une œuvre de l'esprit », *Communication Commerce électronique* n°5, mai 2005, comm. 78.

provisoire ou permanente et peut concerner tout ou partie de l'oeuvre. Une des étapes du *data mining* est la collecte des contenus numériques que le chercheur souhaite exploiter. Cette collecte implique la réalisation d'une copie au moins temporaire afin de permettre l'extraction de données pertinentes à l'aide d'un logiciel d'analyse. Cette copie, aussi fugitive soit-elle, est bien une fixation matérielle de l'oeuvre.

**35. Perspectives.** Cependant, certains auteurs s'interrogent sur l'avenir de ces fixations, condition essentielle à la qualification de reproduction. Car si aujourd'hui le *data mining* requiert bien une reproduction au moment de la collecte des données, cet acte d'exploitation ne sera peut être plus nécessaire dans quelques années<sup>41</sup>, compte tenu des évolutions rapides de la technique. Monsieur Lucas se pose la question de ces copies provisoires et cette interrogation légitime peut s'appliquer aux copies réalisées lors d'une fouille de données : « Qui peut dire si dans quelques années, compte tenu de l'augmentation de la capacité de débit des réseaux, la circulation des œuvres sur les réseaux passera encore par ces fixations provisoires qui mobilisent l'attention aujourd'hui ? »<sup>42</sup>.

**36. Le second élément, la communication de l'oeuvre au public.** A la première condition de la fixation matérielle, il faut ajouter un second élément : la reproduction entre dans le champ du droit d'auteur dès qu'elle est susceptible de permettre de communiquer l'oeuvre au public de manière indirecte. Et cette seconde condition soulève quelques interrogations.

Revenons sur les fondements de cette seconde condition à travers l'interprétation qu'en fait la doctrine. Ainsi selon Monsieur Vivant, « La vérité est qu'il n'y a reproduction qu'aux fins de communication (...) De manière plus fondamentale, cette limite rejoint la distinction faite dans l'ensemble de la propriété intellectuelle entre l'usage intellectuel et l'usage économique de la création, seul le second relevant de l'exclusivité du créateur. Pour toutes ces raisons nous pensons que le droit d'auteur postule bien un usage public de l'oeuvre ou si l'on préfère qu'à défaut, le droit d'auteur n'a pas de raison d'être »<sup>43</sup>.

Monsieur Caron lui aussi appuie cette idée : « En effet, en plus de cet élément matériel, il est indispensable d'être en présence d'un élément moral. Il faut que la reproduction ait pour finalité de communiquer l'oeuvre au public d'une manière indirecte. La référence à la communication indirecte s'explique aisément car le public entre en relation avec l'oeuvre grâce au support sur lequel elle a été fixée matériellement »<sup>44</sup>.

Cet argument est également soulevé par Monsieur Gaudrat au sujet de la copie privée. « Le critère de mise en oeuvre du droit d'exploitation est dans la loi de 1957, la communication de l'oeuvre à un public d'exploitation. A contrario, les actes qui n'impliquent pas la communication au public de l'oeuvre n'affectent pas le droit d'exploitation (...). D'où un certain nombre de prétendues exceptions au droit d'exploitation qui ne font que prendre en compte des hypothèses dans lesquelles le droit n'est pas mis en jeu. C'est ainsi, parce que la copie privée à usage non collectif ne réalise pas une communication au public qu'elle est licite sans cession »<sup>45</sup>.

---

41 CSPLA, Rapport de 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, P. Sirinelli, octobre 2014.

42 A. Lucas, H.-J. Lucas, A. Lucas-Schloetter, *Traité de la propriété littéraire et artistique*, LexisNexis, 4e éd., 2012, n°263.

43 M. Vivant, J.-M. Bruguière, *Droit d'auteur*, Dalloz, coll. Précis, 1ère éd., 2009, n°560.

44 C. Caron, *Droit d'auteur et droits voisins*, Manuel LexisNexis, coll. Les Manuels, 3e éd., 2013, p.281.

45 Ph. Gaudrat « Droit des auteurs (...) Droit de représentation » *J.-Cl. PLA*, Fasc. 1242, 2003.

**37. Des interrogations légitimes.** A la lecture de ces précisions apportées par la doctrine, il est légitime de s'interroger sur ce critère de la communication au public dans le cas précis du *data mining*. Peut-on réellement avancer que cette fixation permette la communication au public ? Le *data mining*, nous l'avons dit, est une exploitation tout à fait particulière des œuvres car la fixation matérielle n'est pas réalisée dans le but de communiquer l'œuvre. De plus, dans la majorité des cas, cette fixation sera dédiée à l'usage privé du chercheur. Nous pourrions aller plus loin et dire que cette fixation n'est dédiée qu'à l'usage de l'algorithme, le chercheur ne lisant probablement même pas le contenu des sources explorées. Et le plus souvent une seule copie de l'œuvre sera effectuée, ce qui est un argument appuyant notre point de vue. En effet, « la réalisation d'une seule fixation matérielle, sur un support précis, prouve que la reproduction ne permet pas, ou difficilement une communication au public »<sup>46</sup>.

## §2. La solution imposée par la conception large du droit de reproduction

**38. Portée du droit de reproduction.** Dans la conception française, le droit de reproduction bénéficie d'une interprétation tellement large qu'il appréhende tous les modes de communication, d'exploitation et d'utilisation de l'œuvre, qui sont dans la continuité de la première fixation de l'œuvre. L'article 2 de la directive 2001/29/CE du 22 mai 2001 va lui aussi dans le sens d'une conception large du droit de reproduction.

Ainsi, on peut concevoir des droits patrimoniaux même sans publicité de l'œuvre, opinion défendue avec force par Monsieur Pollaud-Dulian<sup>47</sup>. C'est ce que dit l'article L. 111-1 du Code de la propriété intellectuelle : « l'auteur d'une œuvre de l'esprit jouit sur cette œuvre, du seul fait de sa création, d'un droit de propriété incorporelle exclusif et opposable à tous ». Les droits moraux et patrimoniaux naissent du seul fait de la création, sans qu'il soit fait référence à une quelconque publicité de l'œuvre. Comme la directive 2001/29/CE du 22 mai 2001 l'y incite, la Cour de justice de l'Union européenne prône une interprétation large des prérogatives de l'auteur. Ainsi, tout acte d'utilisation d'une œuvre devra être autorisé par son auteur.

**39. Position jurisprudentielle.** La jurisprudence a jugé que le droit de reproduction est mis en jeu lorsque l'œuvre est fixée matériellement sans pour autant qu'il y ait eu de communication au public<sup>48</sup>. Il est incontestable que la numérisation des œuvres est un acte de reproduction, quelles qu'en soient les finalités ultérieures. La solution est constante depuis plusieurs années<sup>49</sup>. Le Tribunal de Grande Instance de Paris l'a rappelé dans une décision du 18 décembre 2009 à propos de la numérisation des livres par Google sans autorisation. Google avait avancé l'argument selon lequel la numérisation supposait un acte de manifestation de la volonté de son auteur de la communiquer au public, ce qui n'était pas le cas en l'espèce puisque le site contesté ne permettait pas d'afficher l'intégralité des ouvrages en cause mais seulement de courts extraits couverts par l'exception de courte citation. Cette décision est arrivée « comme un cadeau de Noël pour les auteurs et les éditeurs »<sup>50</sup> :

« Attendu cependant que la numérisation d'une œuvre, technique consistant en l'espèce à scanner l'intégralité des ouvrages dans un format informatique donné, constitue une

46 Caron C., *Droit d'auteur et droits voisins*, LexisNexis, coll. Les Manuels, 3e éd., 2013, p.281.

47 F. Pollaud-Dulian, *Le droit d'auteur*, Economica, coll. Corpus, 2e éd., 2014, p. 712, n°961.

48 Civ. 1re, 24 janvier 1994, *RIDA* 1994/4, p. 433.

49 CJUE 16 juillet 2009 *Communication Commerce électronique*, 2009, comm. 97, note C. Caron

50 F.-M. Piriou « La numérisation des livres sans autorisation constitue un délit de contrefaçon », *Communication Commerce électronique*, n°5, mai 2010, étude 11.

reproduction de l'oeuvre qui requiert en tant que telle, lorsque celle-ci est protégée, l'autorisation préalable de l'auteur ou de ses ayants droit ;  
Que les sociétés Google ne peuvent sérieusement soutenir, sauf à remettre en cause la fonctionnalité même du système Google recherche de Livres, que la constitution d'un fichier numérique ne serait pas un acte de reproduction pour ne pas reproduire en lui-même la forme intelligible de l'oeuvre dès lors que la fixation résultant de la numérisation des ouvrages et leur stockage dans une base de données numériques est toujours apte à communiquer l'oeuvre au public d'une manière indirecte ».<sup>51</sup>

**40. Le *data mining* met en jeu le droit de reproduction.** L'exploration de données impose de prendre possession de mines de contenus et pour cela il est nécessaire de les stocker et donc de les reproduire au sens du droit d'auteur, ou d'en opérer une extraction substantielle au sens du droit des producteurs de bases de données. Compte tenu de la conception extrêmement large du droit de reproduction, la qualification juridique du *data mining* est chose établie : il s'agit bien d'un acte de reproduction et l'autorisation des titulaires de droits est requise.

**41. Appréciation critique.** Mais il n'est pas inutile de s'interroger sur la pertinence de cette qualification. Car, dans le cas particulier du *data mining*, le résultat de la fouille de données ne contiendra pas les données extraites et aucun des éléments caractéristiques qui font l'originalité de l'oeuvre ne sera reconnaissable puisque le but même du *data mining* est de faire surgir un nouveau savoir, des informations inédites et jusqu'alors indécélables. Il est de l'essence même du *data mining* de mettre en lumière des corrélations qui n'apparaissent qu'à travers la confrontation d'un grand nombre de données. La question se pose de savoir si l'on doit considérer qu'un simple procédé technique permettant ensuite l'exploitation du contenu d'une oeuvre justifie que l'on fasse une application quasi-automatique du droit de reproduction. L'interprétation large du droit de reproduction permet de le mettre en oeuvre alors même qu'aucune caractéristique originale de l'oeuvre ne pourra ensuite être identifiée et donc communiquée. Est-ce légitime et n'est-ce pas aller à l'encontre du fondement du droit d'auteur, qui selon nous n'a de sens qu'avec une certaine publicité ? L'acte de reproduction mettant en jeu le droit de reproduction est bien la copie initiale de l'oeuvre, la fixation nécessaire à la fouille réalisée par un algorithme mais peut-on réellement occulter le fait que les oeuvres « disparaissent » ?

Cette interprétation extrêmement large du droit de reproduction correspond à une volonté grandissante et contestable d'appropriation de l'information. Or, comme le souligne Monsieur Lucas<sup>52</sup>, la notion de bien informationnel est source de confusion et le fait que les créations immatérielles soient protégées par le droit d'auteur n'implique pas que l'Information, au sens large, puisse en tant que telle être l'objet d'un monopole. La valeur économique que recèlent les informations ou les données de la recherche, valeur que le *data mining* a formidablement mise en lumière, ne peut suffire à en faire des biens au sens du droit.

**42. Perspectives.** Cette tendance à l'extension des droits patrimoniaux force à s'interroger sur le bien-fondé de ce mouvement, tout particulièrement dans le monde de la recherche fondé sur l'idée de partage du Savoir et des connaissances scientifiques. L'idée parfois évoquée selon laquelle la notion de reproduction devrait être redéfinie dans l'environnement numérique pour que toute copie ne donne pas lieu à l'application du droit d'auteur nous semble une piste intéressante à explorer<sup>53</sup>.

51 TGI Paris, 3e ch., 18 octobre 2009, Sté Editions du Seuil c/ Sté Google Inc. Jurisdate n° 2009-016553.

52 A. Lucas, *Le droit de l'informatique*, PUF, coll. Themis, 1987 pages 347 et s.

53 « Il y a un piège : dès que l'on aborde des questions qui sont marquées par la technicité, on veut voir un décryptage technique. Le droit est un instrument de régulation sociale. Qu'il y ait quelque part une copie, qui signifie reproduction, c'est une chose. Mais est-ce cela que nous devons appréhender en terme de régulation sociale ?  
Audition de Monsieur Michel Vivant devant la mission Lescure du 21 décembre 2012.

Comme le soutient Monsieur Vivant, « Si le droit est un outil de régulation sociale, il est légitime que tout jeu purement mécanique de la règle soit récusé »<sup>54</sup>.

Cette idée a été récemment proposée dans le cadre de discussions menées au niveau européen sur le *data mining* :

« A more promising route could involve reconsideration of the right of reproduction in copyright law, along with the right of extraction in the database regime. These have traditionally been subject to increasingly broad interpretation, but the need to boost TDM in Europe provides impetus to consider a change of emphasis. This would involve the legislator adopting a ‘normative’ approach, designed to ensure that protection is supported by the courts only for acts of reproduction or extraction that entail ‘expressive’ exploitation of the rights-protected material »<sup>55</sup>.

C'est une idée fréquemment avancée mais dont la mise en application semble délicate.

**43. Conclusion.** Utiliser le *data mining* dans un projet de recherche est une entreprise compliquée qui oblige le chercheur, avant même la phase contractuelle, à maîtriser un foisonnement de textes parfois compliqués afin de sécuriser son projet. Il paraît donc essentiel de trouver un moyen simple qui permette aux auteurs et titulaires de droits d'exprimer les conditions de mises à disposition de leurs œuvres et aux chercheurs utilisateurs d'accéder aux œuvres sans devoir négocier au cas par cas une infinité d'autorisations. Le contrat est-il en mesure de répondre à ces attentes ?

---

54 M. Vivant, « Droit d'auteur et théorie de l'accessoire : et si l'accessoire révélait l'essentiel ? », *JCP E* 2011, 1560.

55 Commission européenne, « *Standardization in the area of innovation and technological development, notably in the field of Text and Data Mining* », *Report from the Expert Group*, 2014, page 6.

## Chapitre 2

### Les solutions contractuelles insuffisantes

Avant d'analyser les clauses de *data mining* dans un certain nombre de contrats, il convient de décrire brièvement le modèle économique de l'édition scientifique, qui a connu ces dernières années des changements significatifs et qui comporte certaines spécificités qui ont toute leur importance dans ce chapitre.

**44. La communauté scientifique et le libre accès.** Le mouvement du libre accès est jalonné depuis quinze ans de déclarations face aux difficultés rencontrées pour accéder aux résultats scientifiques. En 2001, la Lettre ouverte de Public Library of Science (PloS)<sup>56</sup> a marqué la première prise de position des chercheurs en faveur du libre accès des résultats publiés de la recherche. Elle a été suivie par ce que l'on nomme les « 3B », Budapest, Béthesda et Berlin. L'initiative de Budapest pour l'Accès Ouvert du 14 février 2002<sup>57</sup> a été une avancée décisive en faveur de la mise à disposition en ligne gratuit et sans restriction de la littérature scientifique à comités de lecture. Cette déclaration rappelait cette tradition ancienne qu'est la volonté des scientifiques et universitaires de « publier sans rétribution les fruits de leur recherche dans des revues savantes, pour l'amour de la recherche et de la connaissance ». La Déclaration de Béthesda du 11 avril 2003 proposait une définition de la publication en libre accès tandis que la Déclaration de Berlin sur le Libre Accès à la Connaissance en sciences exactes, sciences de la vie, sciences humaines et sociales du 22 octobre 2003, signée par un certain nombre d'institutions, élargissait la notion du libre accès à toutes les données issues d'un travail de recherche.

**45. La promotion de nouveaux modes de diffusion des travaux scientifiques.** Ce mouvement a été renforcé par les coûts prohibitifs<sup>58</sup> d'accès aux publications, au point de lancer un débat sur le juste prix de la prestation éditoriale et de promouvoir l'idée auprès des chercheurs de la nécessité d'un accès libre aux résultats de la recherche. Un mouvement de désabonnements spectaculaires aux revues scientifiques (la bibliothèque Interuniversitaire de Santé rattachée à l'Université Paris-Descartes,<sup>59</sup> l'Université Pierre et Marie Curie, Harvard<sup>60</sup>) a contribué à l'émergence de modèles éditoriaux inédits. En réaction aux monopoles des grands éditeurs scientifiques et à la cherté des abonnements, l'accès libre s'est alors développé de manière continue auprès de la communauté scientifique, soit à travers la publication dans des revues en libre accès, soit à travers les archives ouvertes<sup>61</sup>.

- **Le libre accès par les revues.** Dans ces revues scientifiques, les articles sont accessibles sans restriction d'accès et d'utilisation, dès leur publication. A ce jour, le *Directory of Open Access Journals* recense 10 535 revues en libre accès<sup>62</sup>, elles étaient 3 000 en 2007. Il peut s'agir de nouvelles revues qui ont fait le choix d'être uniquement électroniques ou d'anciennes revues qui

---

56 <http://openaccess.inist.fr/?La-lettre-ouverte-de-Public>

57 <http://www.budapestopenaccessinitiative.org/>

58 Le prix des revues a connu une inflation de 273% entre 1986 et 2004. Enquête citée sur le site de site de SPARC, *the Scholarly Publishing and Academic Resources Coalition*.

59 <http://www2.biusante.parisdescartes.fr/wordpress/index.php/contraintes-budgetaires-desabonnements-2014/>

60 [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)

61 C. Bernault, « Archives ouvertes et droit d'auteur : un nouveau mode de diffusion des connaissances scientifiques », *Propriétés Intellectuelles*, octobre 2011, n°41, p. 374-383.

62 <https://doaj.org> Consulté le 18 mai 2015.

sont passées au libre accès. Aujourd'hui, la création de nouvelles revues en *Open Access* représente près de 80% des nouveaux titres lancés sur le marché. Les éditeurs ont largement investi le champ du *Gold Open Access*, démontrant ainsi leur volonté d'être des sociétés de service et de monnayer les nouveaux usages de leurs plates-formes en ligne.

- **Le libre accès par les archives ouvertes.** Le développement des revues numériques a eu pour effet d'entraîner le développement des archives ouvertes, entrepôts de données issues de la recherche scientifique et dont l'accès se veut ouvert. Le chercheur qui publie ses travaux dans une revue scientifique aura également la possibilité de les déposer dans une archive ouverte afin de partager plus largement les résultats de son travail et ainsi favoriser la circulation du savoir. Cette solution permet également aux chercheurs d'organiser eux mêmes la diffusion de leurs travaux, sans passer par les éditeurs, bouleversant ainsi les usages en matière d'édition scientifique. On parle de « Voie Verte » (*Green road*). Créé en 2001 par le CNRS, Hyper Article en ligne (HAL) est une archive ouverte pluridisciplinaire qui compte à ce jour 172 924 documents en intégral.

Si le choix du dépôt dans une archive ouverte appartient à l'auteur, il lui faudra toutefois obtenir l'accord de l'éditeur car, comme mentionné précédemment, celui-ci se fait en général céder à titre exclusif le droit de reproduction de l'oeuvre sur support papier et en ligne. Mais aujourd'hui, de nombreux éditeurs admettent la coexistence de leurs revues avec les archives ouvertes via la technique de l'auteur-payeur même s'ils exigent parfois une période d'embargo, période durant laquelle l'article reste en accès réservé, uniquement accessible dans la revue payante avant de pouvoir être diffusé en archives ouvertes.

- **Le développement de nouveaux modèles de diffusion.** Soulignons que les revues en libre accès et les archives ouvertes ne sont pas les seules modèles de diffusion des publications scientifiques et le modèle du libre accès comporte de multiples formes. Citons des modèles hybrides (certains articles sont en accès libre et d'autres non), le dispositif platinum (le texte est en libre accès mais des services annexes sont commercialisés), ou encore les épi-revues qui représentent également une réponse innovante aux enjeux de la libération des données de la recherche. Il s'agit de revues électroniques en libre accès, alimentées par des articles issus des archives ouvertes en y apportant une valeur ajoutée puisqu'un comité éditorial appose une caution scientifique à chaque article retenu. De nouveaux modèles éditoriaux continuent ainsi à faire leur apparition, preuve que le domaine est en perpétuel mouvement.

**46. Des politiques d'incitation.** Ces nouveaux modes de diffusion de la recherche scientifique ont été développés grâce à des politiques d'incitation en France et au niveau de l'Union européenne. En effet, de plus en plus d'institutions soutiennent l'accès libre et sans restriction aux résultats de la recherche. Parmi les premières à prendre position, citons la fondation *Wellcome Trust (WT)*, une organisation caritative indépendante, qui finance la recherche dans le domaine biomédical. Dès le début de l'année 2006, les chercheurs dont le travail est financé par *WT* doivent déposer le plus rapidement possible une copie de tout article, publié dans une revue à comité de lecture, dans *PubMed Central*, une archive maintenue par la *National Library of Medicine*, le *Wellcome Trust* ayant passé des accords avec l'éditeur *Elsevier*. Cette initiative a été suivie par de nombreuses autres organismes. Même si la France accuse un certain retard par rapport aux politiques ambitieuses d'ouverture des publications scientifiques menées en Angleterre, en Allemagne ou encore aux Pays Bas<sup>63</sup>, l'Agence nationale de la recherche (ANR) et le CNRS ont pris position en faveur du dépôt des articles dans des archives ouvertes.

---

63 A ce sujet, lire le communiqué de presse de l'association des universités néerlandaises.  
<http://vsnu.nl/files/documenten/Domeinen/Onderzoek/Open%20access/Explanation%20press%20release%20for%20scientists%20%282014-11-04%29.pdf>

**47. Le *data mining* dans les contrats, des objectifs différents pour les éditeurs et les chercheurs.** Les modes de diffusion des publications scientifiques et données de la recherche se sont ainsi diversifiés, offrant des opportunités nouvelles à la communauté des chercheurs, en terme de publication mais également d'exploitation. Dans le cas du *data mining*, les chercheurs ont besoin de fouiller des corpus de documents et de données, provenant de différentes bases de données, de différents éditeurs mais également du Web. Comme le soulignent le consortium Couperin et l'Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation (ADBU), « l'innovation se construit le plus souvent aux marges et croisements des disciplines, dans des approches de plus en plus interdisciplinaires : le besoin de croiser des bases de données hétérogènes, issues d'éditeurs différents, est fondamental, et par ailleurs il sera certainement nécessaire d'associer les contenus issus des publications avec des données primaires issues de l'activité scientifique »<sup>64</sup>. De plus, cette exploration doit pouvoir se faire de façon simultanée.

L'éditeur détient les droits d'accès aux publications scientifiques puisque dans la grande majorité des cas et conformément aux usages, l'auteur lui aura cédé ses droits lors de la signature du contrat de publication. Selon les éditeurs, le *data mining*, nouvelle exploitation des contenus protégés, nécessite la mise en oeuvre de licences spécifiques. Ils ont ainsi la volonté de développer des outils contractuels favorables au développement du *data mining*. La solution contractuelle a selon eux l'avantage de permettre la mise en place de conditions différentes selon que le *data mining* est effectué à des fins de recherche ou à des fins commerciales. Le contrat leur apparaît comme un cadre juridique souple parfaitement adapté à une pratique dont les contours sont encore en évolution.

**48. Un équilibre à définir.** L'outil contractuel met-il en place une solution équilibrée ? Le contrat est-il l'outil permettant de garantir aux éditeurs, mais plus largement aux titulaires de droits, la légitime protection de leurs contenus sans freiner les activités de recherche ? Permet-il de réaliser un équilibre entre les objectifs de la recherche et les modèles économiques des éditeurs ? Selon le rapport du Conseil supérieur de la propriété littéraire et artistique rendu en 2014, « l'exploration de données est une activité émergente pour laquelle la solution contractuelle n'a pas encore épuisé l'ensemble de ses capacités ». Nous nous attacherons à vérifier si cette affirmation correspond à la réalité.

**49. Plan.** Pour cela, nous analyserons les solutions contractuelles proposées par quelques grands éditeurs scientifiques, qui se révèlent en réalité inadaptées aux pratiques de *data-mining* (Section 1) avant de nous pencher sur le rôle des licences libres, nouveaux outils juridiques innovants et largement utilisés par la communauté des chercheurs (Section 2).

## **Section 1. Les clauses de *data mining* dans les contrats des grands éditeurs.**

**50. Contexte.** Les dernières années ont été marquées de grands mouvements de concentration dans l'édition scientifique. Alors que le livre tient encore une place importante dans la communication scientifique dans le domaine des Sciences humaines et sociales (SHS), ce sont presque exclusivement les revues qui jouent ce rôle dans le secteur sciences, technique et médecine (STM). Ce marché est aujourd'hui dominé par quatre grands groupes, *Elsevier-Sciences* (2494 titres), *Springer-Kluwer* (environ 2000 titres), *Wiley-Blackwell* (1492 titres) et *Taylor & Francis* (1300 titres), ces quatre grands groupes cherchant en permanence à renforcer leur place dans un

---

<sup>64</sup> CSPLA Mission relative au *data mining* : l'analyse de Couperin et de l'ADBU.

secteur particulièrement rentable<sup>65</sup>. Les logiques marchandes animent fortement l'édition des revues scientifiques et le *data mining*, en raison des enjeux économiques qu'il porte, est rapidement apparu comme un sujet de tensions entre l'édition commerciale et la communauté scientifique. Les grands éditeurs commerciaux y ont vu un moyen de développer leurs activités et de monnayer de nouveaux services. Les pratiques de *data mining* n'étant pas prévues au départ dans les licences négociées avec les institutions et bibliothèques, certains grands éditeurs scientifiques, parmi lesquels *Elsevier*, y ont intégré de nouvelles clauses « afin de répondre aux besoins des chercheurs »<sup>66</sup>. D'autres éditeurs ont fait le choix d'interdire le *data mining*.

## §1. Exemples de clauses autorisant le *data mining*

**51. Exemples.** Nous avons choisi à titre d'exemple d'étudier les conditions fixées par deux majors de l'édition scientifique, *Elsevier*<sup>67</sup> et *Springer*<sup>68</sup>, clauses que nous reproduisons ci dessous :

- « Terms and conditions of text and data mining Elsevier*
- *Please note that text mining access is provided to subscribers for non-commercial research purposes only.*
  - *All access to content for text mining purposes must take place through our APIs and remind you that in order to maintain performance and availability for all users, the terms and conditions of access to ScienceDirect continue to prohibit the use of robots, spiders, crawlers or other automated programs, or algorithms to download content from the website itself. Researchers are able to use their own tools to text mine content accessed via our API.*
  - *Text mining outputs must adhere to the following conditions :*
    - *outputs can contain « snippets » of up to 200 characters, excluding text entity matching or bibliographic metadata*
    - *Where Snippets and/or bibliographic metadata are distributed, they should be accompanied by a DOI link that links back to original content.*
    - *Where images are used you should clear the rights for reuse with the relevant copyright owner and/or rights holder.*
    - *Any TDM Output which is redisplayed should include a proprietary notice in the following form: "Some rights reserved. This work permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited." Researcher should recognize the fact that although the researcher will 'own' his or her research output, the snippets displayed in the output may contain rights of the publisher, the underlying author or third parties ».*

S'ensuit une longue liste d'interdictions dans une rubrique intitulée *Elsevier Text and Data Mining (TDM) License* :

- « The TDM click through agreement does not allow you to:*
- *Use snippets of text from individual full text articles or book chapters of more than 200 characters*

---

65 N. Pignard-Cheynel, « L'édition de revues scientifiques : une forme de marchandisation de la diffusion des connaissances », *Sciences de la société*, n° 66, octobre 2005.

66 « *As a publisher we believe it is our job to help meet the need of researchers and we are committed to reducing the barriers to mining content. To achieve this, we have adopted a license-based approach which both formalizes the right to mine into our academic agreements and delivers a flexible way in which researchers can gain access to our API through a self-service portal.* » Elsevier's policy

67 <http://www.elsevier.com/about/policies/content-mining-policies>

68 <https://www.springer.com/gp/rights-permissions/springer-s-text-and-data-mining-policy/29056>

- *Abridge, modify, translate or create any derivative work based on the Dataset;*
- *Remove, obscure or modify in any way any copyright notices, other notices or disclaimers as they appear in the Dataset;*
- *Substantially or systematically reproduce, retain or redistribute the Dataset;*
- *Extract, develop or use the Dataset in any direct or indirect commercial activity;*
- *Use any robots, spiders or other automated downloading programs, algorithms or devices to search, screen-scrape, extract, or index any Elsevier web site or web application;*
- *Utilize the TDM Output to enhance institutional or subject repositories in a way that would compete with the value of the final peer review journal article, or have the potential to substitute and/or replicate any other existing Elsevier products, services and/or solutions ».*

Si les clauses autorisant le *data mining* sont assez détaillées sur le site de l'éditeur *Elsevier*, elles sont beaucoup plus succinctes sur le site de *Springer*.

« *Springer's text- and data mining policy*

*Springer grants text- and data-mining rights to subscribed content to researchers via their institutions, provided the purpose is non-commercial research. The selection and refinement of desired articles can be conducted by using existing search methods and tools, such as PubMed, Web of Science and Springer's Metadata API, among others.*

*Use of text and data mining results and research output*

*Publications or analyses resulting from TDM of subscribed content may include quotations from the original text of up to 200 characters, or 20 words, or 1 complete sentence. They should cite the original Springer content in the form of a DOI link. Permission to reproduce images may be granted on a case-by-case basis ».*

## §2. Analyse critique

**52. Des conditions restrictives.** Si les contrats d'abonnements autorisent parfois les activités de *data mining* pour la recherche scientifique, parfois même sans coûts supplémentaires, les conditions imposées par les éditeurs sont très restrictives.

- Elles sont en premier lieu limitées aux activités de recherche non commerciales, ce qui a l'avantage de permettre aux éditeurs de mettre en place des conditions différentes pour le *data mining* effectué à des fins commerciales. Cela rassure l'éditeur puisque les travaux publiés ne pourront être exploités par un éventuel concurrent. Cela ne signifie pas pour autant qu'une diffusion dans le cadre d'une exploitation non commerciale ne sera pas susceptible de venir concurrencer celle assurée par l'éditeur.
- Elles interdisent l'utilisation de méthodes automatiques d'aspiration, d'extraction et de recherche sur le site de l'éditeur.
- Sur le plan technique, elles imposent parfois de passer par une interface de programmation ou *API*<sup>69</sup>. C'est le cas du contrat d'Elsevier. De plus, les requêtes via l'*API* sont limitées à 10 000 articles par semaine.
- Ces licences exigent que le chercheur enregistre et détaille son projet de recherche ainsi que sa méthodologie. *Springer* impose par exemple que chaque projet de *data mining* fasse l'objet d'un enregistrement via un formulaire et l'éventail des informations à fournir est large : les chercheurs doivent décrire le projet, préciser sa durée, délimiter le contenu à extraire ...

---

<sup>69</sup> *API* (Application Programming Interface) est une interface de programmation qui permet faire communiquer une application avec ce qui peut être en interaction avec elle.

Ce qui permet à *Springer* d'avoir des informations précieuses sur les projets de recherche en cours et indirectement, d'avoir un pouvoir décisionnel dans l'orientation des politiques de recherche, puisqu'il se réserve le droit d'autoriser ou non la fouille.

- Enfin, les éditeurs décident, en fonction du projet qui leur est présenté, s'ils accordent ou non la fouille à titre gratuit. Souvent, lorsque les recherches n'ont pas de but commercial, le *data mining* est autorisé sans contrepartie financière mais cela n'est pas une règle. Ainsi, dans le cadre du projet Text2genome précédemment mentionné, *Elsevier* a autorisé la fouille gratuite de ses articles alors que pour un autre projet de recherche, BioNØT<sup>70</sup> de l'Université de *Wisconsin-Milwaukee*, les chercheurs ont dû verser une contrepartie financière.

En France le renouvellement des contrats entre le consortium Couperin (représentants des établissements d'enseignement supérieur et de recherche) et *Elsevier* pour la période de 2014 à 2018 a fait grand bruit. Ces négociations difficiles ont été l'occasion de mettre en lumière les conditions extrêmement drastiques mises en place par l'éditeur, et les menaces qu'une telle licence représentent pour les projets de *data mining*. Les termes de cette licence ont finalement été renégociés<sup>71</sup> et selon l'article 6.6 du contrat intitulé *Data et text mining*<sup>72</sup>, « Tous les contenus accessibles et souscrits sur la plateforme du Titulaire dans le cadre de cet accord seront utilisables à des fins de *data et text mining* notamment via une interrogation des données par une AI connectée à la plateforme ScienceDirect®, conformément aux stipulations du Contrat de Licence. ». Les conditions du *data mining* ont ainsi été assouplies : la limite de 200 signes a été retirée et les chercheurs semblent avoir la possibilité d'utiliser une autre API que celle d'*Elsevier*.

**53. Les contrats et l'indépendance de la recherche.** Les solutions contractuelles proposées par les éditeurs restent insuffisantes. Elles ont l'inconvénient de leur laisser une mainmise jugée importante sur les pratiques de la recherche. Compte tenu du modèle économique que représente l'édition scientifique, quelques grands éditeurs privés seront ainsi en possession d'informations précieuses sur les projets en cours, sur ceux en préparation et auront ainsi une vision d'ensemble des projets de recherche actuellement menés, ce qui n'est pas sans susciter des inquiétudes. De plus, ils auront un réel pouvoir de décision, ce qui pour le Consortium Unifié des Etablissements Universitaires et de Recherche pour l'Accès aux Publications Numériques (Couperin) et l'Association des Directeurs et Personnels de direction des Bibliothèques Universitaires et de la Documentation (ADBU) revient « à donner aux éditeurs de contenus, qui ne sont pour rien dans le financement des projets de recherche, le droit de décider quelle recherche pourra ou non voir le jour »<sup>73</sup>. Sans aller jusqu'à parler de science « autorisée »<sup>74</sup>, cela remet très certainement en cause le principe essentiel d'indépendance du chercheur.

**54. Une diversité source de complications.** De plus, ces clauses ne sont en rien une source de simplification pour les chercheurs, qui devront maîtriser les conditions complexes et parfois différentes auxquelles chaque éditeur choisira de soumettre les opérations d'exploration. Le modèle de la licence est donc inadapté, notamment lorsqu'un projet de *data mining* nécessite le croisement de plusieurs bases de données, issues d'éditeurs différents. Les solutions contractuelles imposées par les éditeurs scientifiques sont insuffisantes et ne permettent pas d'appréhender la diversité des situations rencontrées. Comme le souligne le rapport du CSPLA, les négociations contractuelles ralentissent voire découragent les initiatives d'exploration de données.

70 <http://www.biomedcentral.com/1471-2105/12/420>

71 <http://scoms.hypotheses.org/276>

72 [http://rue89.nouvelobs.com/sites/news/files/assets/document/2014/11/marche\\_elsevier.pdf](http://rue89.nouvelobs.com/sites/news/files/assets/document/2014/11/marche_elsevier.pdf)

73 CSPLA Mission relative au data mining : l'analyse de Couperin et l'ADBU

74 P.-C. Langlais et L. Maurel *Quel statut légal pour le data mining ?*, Synthèse de Savoirscom1 consécutive à l'audition du 15 janvier 2013 par le CSPLA.

## Section 2. Licences libres et *data mining*

**55. Définition.** « Les licences libres sont les licences par lesquelles l'auteur autorise la copie, la modification et la diffusion de l'oeuvre modifiée ou non, de façon concurrente, sans transférer les droits d'auteur qui y sont attachés et sans que l'utilisateur ne puisse réduire ces libertés tant à l'égard de l'oeuvre originelle que de ses dérivés. Les œuvres ainsi diffusées sont qualifiées de libres »<sup>75</sup>. Apparues à l'origine pour les logiciels, ces licences se sont étendues à tous types de création.

### § 1. Exemples de licences adaptées au *data mining*

**56. Les licences *Creative commons*.** Les licences *Creative commons* sont des contrats par lesquels les titulaires de droits d'auteur mettent leurs œuvres à la disposition du public à des conditions prédéfinies. L'auteur peut choisir d'autoriser ou non les exploitations commerciales, la modification de l'oeuvre et la diffusion de l'oeuvre sous la même licence, sous la seule réserve du droit à la paternité. Ces licences, dont il existe plusieurs types, se construisent à partir d'une combinaison de pictogrammes et permettent à l'auteur de choisir les actes d'exploitation qu'il souhaite autoriser.

Parmi les différentes licences *Creative commons*, la licence *CC-BY* est souvent recommandée pour la diffusion et l'utilisation maximale des travaux scientifiques car elle permet au titulaire des droits d'autoriser toute exploitation de l'oeuvre, y compris à des fins commerciales, ainsi que la création d'oeuvres dérivées, dont la distribution est permise sans restriction, à la seule condition de respecter le droit à la paternité. *Science-Europe*, association regroupant 52 agences de financement et organismes de recherche européens a ainsi publié en avril dernier une déclaration intitulée *New Science Europe Principles on Open Access Publisher Services*<sup>76</sup> dans laquelle elle incite les éditeurs à diffuser les publications selon les termes de la licence *CC-By* 4.0. Ce type de licences correspond à la volonté de la communauté des chercheurs de favoriser l'existence d'un fonds commun d'oeuvres que chacun peut utiliser librement. La licence *Creative commons BY-SA* qui oblige à citer la source et à partager les données obtenues sous une licence identique à celle qui régit l'oeuvre originale est un autre moyen de s'en assurer.

**57. Les licences libres de type domaine public, licences *CC0* et *PPDL*.** Les licences libres de type domaine public permettent aux auteurs de mettre à disposition leurs œuvres selon les mêmes conditions que les œuvres tombées dans le domaine public. L'auteur peut ainsi renoncer à ses droits et faire rentrer son oeuvre dans le « domaine public volontaire » avant le terme prévu par la loi. Ces licences autorisent donc un accès et une jouissance libres des œuvres. Certains outils ont été créés pour permettre de formaliser cette renonciation et sont aujourd'hui largement utilisés par la communauté scientifique. Ainsi, les chercheurs recommandent fortement que les données de la recherche soient placées dans le domaine public via deux types de licence, l'*Open Data Commons Public Domain Dedication and License*<sup>77</sup> (*PDDL*) de l'*Open Knowledge Foundation* ou la licence *Creative commons zero*<sup>78</sup> (*CC0*).

---

75 M.Clément-Fontaine, *L'oeuvre libre*, Collection Création Information Communication, Larcier, 1ère édition, 2014, p. 89 n°136.

76 [http://www.scienceurope.org/uploads/PressReleases/270415\\_Open\\_Access\\_New\\_Principles.pdf](http://www.scienceurope.org/uploads/PressReleases/270415_Open_Access_New_Principles.pdf)

77 <http://opendatacommons.org/licenses/pddl/1.0/>

78 Cf Les Principes de Pantou publiés le 3 mars 2010. <http://openaccess.inist.fr/?Les-Principes-de-Panton>

Un certain nombre d'institutions culturelles dans le monde ont choisi de placer leurs collections sous licence *CC0*<sup>79</sup>. C'est le cas de bibliothèques nationales comme la *Deutsche National Bibliothek*, *The British Library*, *Europeana*, la Bibliothèque numérique européenne, ou encore, initiative particulièrement remarquable, le *Rijksmuseum* d'Amsterdam<sup>80</sup>. La licence *CC0* est également fréquemment appliquée aux données de la recherche. Ainsi, les licences de BioMed Central<sup>81</sup>, (*Springer*) stipulent que par défaut, les données publiées dans les revues éditées seront diffusées sous licence *CC0* tandis que les articles seront eux, soumis à une licence *Creative Commons 4.0* standard.

**58. Autres licences libres.** D'autres initiatives apparaissent, la dernière en date, le *Public Domain Mark* est un instrument de signalisation créé par *Creative commons* pour permettre aux institutions culturelles d'indiquer à leurs usagers qu'une œuvre appartient au domaine public. Cette marque commence à être utilisée en France, la plateforme MediHal du CNRS la propose comme une option aux chercheurs. Notons que cette licence n'est applicable qu'aux œuvres et non aux données associées. En France, la licence *Open Database License (ODbL)* développée initialement par l'*Open Knowledge Foundation* a été adaptée à un usage national et est utilisée par de nombreuses collectivités locales, les villes de Paris, Toulouse ont ainsi libéré un certain nombre de données sous cette licence. Les données cartographiques collectées dans le cadre du projet international *Open street Map*, qui a pour but de créer une carte libre du monde, sont accessibles à tous sous licence *ODbL*. Il existe aussi des licences nationales, comme la Licence Ouverte (LO) de la mission Etalab<sup>82</sup> (mission gouvernementale en charge de l'ouverture des données en France) qui promeut la réutilisation la plus large des données publiques ou encore l'*Open Government Licence*<sup>83</sup> dont une nouvelle version a été lancée en octobre dernier par le gouvernement britannique.

## §2. Appréciation critique : avantages et limites

**59. L'auteur au centre.** Ces licences s'inscrivent dans la logique du choix de l'auteur, ce dont il faut se féliciter. L'auteur, au lieu de soumettre chaque acte ne relevant pas des exceptions légales à son autorisation préalable, peut ainsi décider en toute liberté d'autoriser à l'avance le public à fouiller ses travaux ou les données qu'il aura collectées dans le cadre de ses recherches. Le choix d'une licence applicable aux publications scientifiques, ou adaptée aux données de la recherche, lorsqu'elles sont publiées et souvent destinées à être librement réutilisées, est devenu une question essentielle pour le chercheur, qui s'interroge sur les actes d'exploitation qu'il souhaite autoriser ou non et l'utilisation plus ou moins large qu'il souhaite permettre de ses travaux.

79 « La personne qui a associé une œuvre à cet acte a dédié l'oeuvre au domaine public en renonçant dans le monde entier à ses droits sur l'oeuvre selon les lois sur le droit d'auteur, droits voisins et connexes, dans la mesure permise par la loi. Vous pouvez copier, modifier, distribuer et représenter l'oeuvre, même à des fins commerciales, sans avoir besoin de demander l'autorisation » Site *Creative Commons Corporation* consulté le 15 mai 2015  
<https://creativecommons.org/publicdomain/zero/1.0/deed.fr>

80 Voir le site *Rijksstudio*, sur lequel le musée propose plus de 125 000 œuvres numérisées en incitant les visiteurs à les réutiliser librement. <https://www.rijksmuseum.nl/en/rijksstudio?ii=0&p=0>

81 « The default use of the *Creative Commons 1.0 Public Domain Dedication waiver (CC0 or CC zero)* for data published within articles follows the same logic: facilitating maximum benefit and the widest possible re-use of knowledge. It is also the case that in some jurisdictions copyright does not apply to data. CC0 waives all potential copyrights, to the extent legally possible, as well as the attribution requirement. The waiver applies to data, not to the presentation of data. If, for instance, a table or figure displaying research data is reproduced, CC BY and the requirement to attribute applies. Increasingly, however, new insights are possible through the use of big data techniques, such as data mining, that harness the entire corpus of digital data. In such cases attribution is often technically infeasible due to the sheer mass of the data mined, making CC0 the most suitable licensing tool for research outputs generated from such innovative techniques. » <http://www.biomedcentral.com/about/license>

82 <https://www.etalab.gouv.fr/licence-ouverte-open-licence>

83 <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

**60. Un rééquilibrage des intérêts.** Ces licences ont le mérite de la simplicité, de la clarté et de la lisibilité et semblent rétablir un certain équilibre entre les intérêts des titulaires de droits et ceux des utilisateurs. Le choix d'une licence précisant les conditions de réutilisation de travaux de recherche et sachant répondre aux souhaits de l'auteur et aux interrogations de ceux qui souhaitent fouiller ces ressources en toute sécurité permettra une réutilisation incontestablement bénéfique à la confrontation, à la critique, à la recherche et au progrès scientifique. Enfin, leur destination internationale est un atout important. Les licences libres ne se limitent pas aux frontières d'un Etat et sont proposées à tout chercheur quelle que soit sa localisation géographique, ce qui, compte tenu du contexte international de la recherche est un atout précieux.

**61. Les licences libres et le respect du droit moral en France.** En France, le droit moral persiste perpétuellement et les œuvres concédées sous licence *PPDL*, *CCO* ou *CC-By* n'en restent pas moins régies par le Code de la propriété intellectuelle, qui prévoit que le droit moral est inaliénable<sup>84</sup>. Une œuvre du domaine public peut certes être reproduite et diffusée librement, y compris à des fins commerciales, mais à la condition de respecter le droit moral dans toutes ses composantes, droit à la paternité, droit au respect de l'œuvre, droit de divulgation et de repentir ou de retrait.

**62. Le droit au respect de l'intégrité des œuvres fouillées.** L'auteur des œuvres fouillées pourra toujours agir en justice en cas d'atteinte à l'intégrité de son œuvre. En effet, le *data mining* en permettant la manipulation des œuvres, présente des risques d'atteinte au respect de l'œuvre, l'auteur pouvant être contrarié par le contexte dans lequel elle est utilisée. Comme vu précédemment, les licences *Creative commons* standard sont fréquemment utilisées pour les articles publiés ou déposés dans des archives ouvertes mais leur utilisation fait parfois débat au sein de certaines revues. A titre d'exemple, citons le cas de la revue *Astronomy & Astrophysics* dont les articles sont tous publiés sous licence *CC-By*. Le fait que l'on puisse autoriser sans contrôle une réutilisation de ces travaux a suscité quelques craintes : certains auteurs redoutent qu'une partie d'un article de cosmologie soit par exemple sortie de son contexte et réutilisée par un auteur de textes quelque peu fantaisistes, ce qui est parfois le cas lorsqu'on parle de l'histoire et de l'évolution de l'univers. Les licences tournées vers une diffusion maximale du Savoir ne permettent pas toujours à l'auteur de garder le contrôle sur ses travaux.

Il faut cependant souligner que ces inquiétudes n'ont pas lieu d'être dans le cas du *data mining* : la question du droit au respect de l'œuvre se pose de façon moins évidente en raison de la « disparition » des œuvres fouillées. Puisqu'elles ne sont pas utilisées en elles même mais pour les informations qu'elles cachent, l'atteinte au respect de l'œuvre sera très certainement plus difficile à caractériser. Dans l'immense majorité des cas, les travaux, données ou informations fouillées seront indétectables dans le résultat du projet de *data mining*.

**63. Le respect du droit à la paternité.** L'obligation de respecter le droit à la paternité des œuvres utilisées semble à première vue, dans le cas du *data mining*, un obstacle difficile à contourner. En effet, dans le cadre d'une fouille de données réalisée à partir de centaines, voire de milliers de ressources, il sera difficile de citer chaque producteur de base de données ou chaque auteur. Cependant, en pratique, là encore, les spécificités du *data mining* font quasiment disparaître les risques d'atteinte au droit à la paternité des œuvres fouillées : rappelons que l'auteur ne peut revendiquer de paternité que sur la partie de l'œuvre qu'il a lui même créée, le droit à la paternité

---

<sup>84</sup> « L'auteur jouit du droit au respect de son nom, de sa qualité et de son œuvre. Ce droit est attaché à sa personne. Il est perpétuel, inaliénable et imprescriptible. Il est transmissible à cause de mort aux héritiers de l'auteur. L'exercice peut être conféré à un tiers en vertu de dispositions testamentaires » Article L. 121-1 du Code de la propriété intellectuelle.

n'étant en cause « qu'à travers une œuvre de l'esprit »<sup>85</sup>. Or, nous l'avons vu, les œuvres ne sont pas reproduites en tant qu'œuvres et elles ne sont utilisées que pour révéler de nouvelles connaissances. On ne retrouvera quasiment jamais les œuvres préexistantes, l'objet du *data mining* étant justement de révéler des informations jusque là cachées et inaccessibles.

On peut alors se demander s'il est toujours justifié d'imposer la mention des auteurs des œuvres préexistantes. Cela correspond toutefois à un usage, qu'il convient de respecter. La communauté scientifique est particulièrement attachée au respect de l'identification de la source des œuvres et la traçabilité des données utilisées et conformément à la Charte nationale de déontologie des métiers de la recherche, l'utilisation des sources, études ou résultats antérieurs, doit apparaître « par un référencement explicite lors de toute production, publication et communication scientifiques »<sup>86</sup>.

**64. Droit de repentir et de retrait/ Droit de divulgation.** L'auteur qui décide de divulguer son œuvre sous licence libre conserve son droit de repentir et de retrait<sup>87</sup>. Mais en raison de la diffusion très large de son œuvre, il sera très certainement dans l'incapacité de modifier la version première de ses travaux ou de mettre fin à l'exploitation de l'œuvre. La mise en œuvre du droit, si elle est possible juridiquement, sera donc délicate voire impossible mais ces questions sont posées par l'environnement numérique en général et ne sont pas spécifiques au *data mining*. Quant au droit de divulgation, il ne pose pas de difficultés particulières.

**65. Une diversité parfois source de complexité et de blocages.** Le système des licences libres a incontestablement « le mérite de stimuler la créativité »<sup>88</sup>, il suffit pour s'en convaincre d'en voir le nombre, qui n'a cessé d'augmenter depuis leur apparition il y a une vingtaine d'années. Toutefois, comme le souligne la doctrine, l'une des difficultés du « libre » dans le domaine des publications scientifiques « tient à ce qu'il n'existe pas un mais des libres »<sup>89</sup>, chaque solution proposée obéissant à des règles différentes parfois même divergentes. A chaque usage correspond une nouvelle licence. Si les licences *Creative commons* semblent adaptées aux articles publiés ou déposés dans des archives ouvertes (cas de *Springer* par exemple<sup>90</sup>), elles ne conviennent pas pour les données de la recherche et leur utilisation est dans ce cas précis déconseillée par la communauté des chercheurs<sup>91</sup>. Les données de la recherche librement accessibles doivent le rester et il faut alors recourir à d'autres licences qui vont dans le sens d'une science ouverte. Et lorsque l'on combine différentes licences avec différents niveaux de protection, les choses peuvent devenir compliquées.

**66. La question de la réservation des résultats du *data mining*.** Si les licences libres de type *PPDL* et *CCO* organisent la diffusion libre des œuvres, elles ne règlent pas les questions de la gestion des droits une fois que l'œuvre est réutilisée. Comme le souligne Madame Clément-Fontaine, qui oppose ces licences de type domaine public aux licences libres pérennes :

---

85 A. Lucas, H.-J. Lucas, A. Lucas-Schloetter, *Traité de la propriété littéraire et artistique*, LexisNexis, coll. Traités, 4e éd., 2012, p. 460 n°535.

86 Charte nationale de déontologie des métiers de la recherche signée le 26 janvier 2015  
[http://www.cnrs.fr/comets/IMG/pdf/charte\\_nationale\\_deontologie\\_signe\\_e\\_janvier2015.pdf](http://www.cnrs.fr/comets/IMG/pdf/charte_nationale_deontologie_signe_e_janvier2015.pdf)

87 Article L. 121-4 du Code de la propriété intellectuelle.

88 M. Clément-Fontaine, « L'entreprise et l'Open source : stratégie de la valorisation », *Revue Lamy Droit de l'Immatériel*, 2014, p. 102.

89 V.-L. Benabou, « Les publications scientifiques : faut-il choisir entre libre accès et libre recherche ? », *Hermès la Revue*, 2010/2, n°57, p. 95-106.

90 *For Open Access Publications from Springer, BioMed Central and SpringerOpen, TDM is usually allowed without restrictions since the majority of our OA content is licensed under CC-BY.* <https://www.springer.com/gp/rights-permissions/springer-s-text-and-data-mining-policy/29056>

91 Les principes de Panton publiés en février 2010 par un groupe de chercheurs <http://openaccess.inist.fr/?Les-Principes-de-Panton>

- « La liberté d'accès et de jouissance de l'oeuvre n'est donc pas la même pour les licences qui tendent à la pérenniser et pour les licences de type domaine public
- Selon les premières, chaque codification ou développement d'une œuvre contribue automatiquement à augmenter le fonds des œuvres libres
  - Selon les secondes, les modifications et les développements ne sont versés au fonds commun que s'ils ne sont pas constitutifs d'une œuvre originale. »<sup>92</sup>

Les licences de type *CC0* ou *PPDL* n'obligent pas à partager les résultats obtenus à partir des données fouillées et le chercheur pourra d'une certaine façon se rapprocher ces nouveaux travaux. Si ces licences de type domaine public semblent à première vue adaptées aux données de la recherche, elles soulèvent cependant des interrogations.

**67. Les spécificités de chaque droit national.** De plus, ces licences ont en principe une vocation internationale. Elles ne sont donc pas toujours très détaillées et se contentent dans certains cas de renvoyer au droit d'auteur ou copyright de chaque pays, par exemple pour ce qui concerne la titularité des droits ou le droit moral. L'entrée dans le domaine public d'une oeuvre a une signification différente selon la portée que chaque pays donne au droit moral de l'auteur. Aux Etats-Unis par exemple, l'oeuvre devient réellement « libre de droits » une fois qu'elle entre dans le domaine public. Il en est de même dans les pays où les durées du droit moral et des droits patrimoniaux sont liées, comme c'est le cas de l'Allemagne. Cette souplesse contraint les chercheurs à se renseigner sur les différentes législations.

**68. Bilan.** Les licences libres, bien qu'imparfaites, correspondent bien à l'esprit de partage des connaissances qui anime le monde de la recherche tout en préservant, compte tenu du caractère très spécifique du *data mining*, les prérogatives essentielles de chaque auteur. Pour toutes ces raisons, la communauté scientifique, soucieuse d'autoriser de nouveaux usages de ses travaux, s'est emparée de ces solutions contractuelles innovantes et tente de promouvoir le plus largement possible leur utilisation. Mais les éditeurs rechignent encore à les utiliser, soucieux de préserver leur monopole. Or, ils sont les principaux titulaires de droit puisqu'ils imposent aux auteurs la cession de leurs droits pour publier leurs travaux.

**69. Perspectives.** Il apparaît aujourd'hui primordial que les chercheurs, au moment de publier les résultats de leurs travaux, restent titulaires de leurs droits d'auteur. Les nouveaux modèles économiques de publication et de diffusion de la recherche et le développement du libre accès sont incontestablement l'occasion de rééquilibrer le rapport de force qui existe entre les éditeurs privés et les acteurs de la recherche publique et de l'enseignement supérieur. Des initiatives vont dans ce sens, incitant les chercheurs à ajouter un avenant au contrat des éditeurs afin de conserver leurs droits lors de la soumission d'articles. On parle d'« *Author Addenda* » et ils sont proposés par différents organismes comme *SPARC*<sup>93</sup> (*Scholarly Publishing and Academic Resources Coalition*), *Sciences Commons* et le *MIT*<sup>94</sup>. Le programme *SCOAP3*<sup>95</sup> (*Sponsoring Consortium for Open access Publishing in Particle Physics*) porté par le CERN et lancé le 1er janvier 2014 est un autre exemple de ce changement en cours : ce projet, fruit d'une collaboration internationale regroupant plus d'un millions de partenaires, permet un accès sans restriction aux articles scientifiques dans le domaine de la physique des hautes énergies, tout en permettant aux auteurs de conserver leurs droits d'auteur, les éditeurs se rétribuant sur la capacité à organiser le système d'évaluation par les pairs.

---

92 M. Clément-Fontaine, *L'oeuvre libre*, Larcier, Coll. Création Information Communication, 2014, p 85 n° 127.

93 <http://www.sparc.arl.org/resources/authors/addendum-2007>

94 <http://libraries.mit.edu/scholarly/files/2010/10/mitamendment2008rev1.pdf>

95 <http://press.web.cern.ch/fr/press-releases/2013/12/lancement-au-1er-janvier-2014-de-scoap3-linitiative-sur-la-publication-en>

Nous pouvons nous féliciter qu'un mouvement de réappropriation par les chercheurs de leurs droits d'auteur soit en cours. Cette tendance aura à l'évidence un impact sur les pratiques de *data mining* car la question de la titularité des droits de propriétés intellectuelles sur les publications scientifiques influe sur le partage de l'information au sein des communautés de recherche et sur l'exploitation des différents corpus scientifiques. Replacer les auteurs au centre ira certainement dans le sens du développement du *data mining*<sup>96</sup>, car les chercheurs sont beaucoup plus ouverts aux nouveaux modes d'exploitation de leurs travaux que ne le sont les éditeurs. En effet, les chercheurs qui publient aujourd'hui sont ceux qui fouilleront les travaux de leurs confrères demain, ce qui incite naturellement à trouver une solution équilibrée.

**70. L'importance d'une solution rapide.** Mais ce rééquilibrage prendra du temps. Or, en matière de *data mining*, les chercheurs eux mêmes appellent à la création d'outils ouverts et à la mise en place de « projets expérimentaux » et « d'échanges de bonnes pratiques au sein de la recherche publique » afin que ces technologies puissent rapidement et réellement être utilisées par l'ensemble de la communauté des chercheurs. La communauté scientifique s'inquiète aujourd'hui de ne pas être en mesure d'utiliser ces technologies, qui semblent pour l'instant réservées au monde du renseignement ou appliquées uniquement à des projets internes de grandes entreprises, au rang desquelles les grands éditeurs<sup>97</sup>.

Si les initiatives de la communauté scientifique nous paraissent une avancée pertinente, correspondant à une réelle nécessité<sup>98</sup>, nous sommes forcés de constater qu'elles risquent d'être limitées, compte tenu des réticences mais également de la puissance économique et de la qualité du lobbying des grands éditeurs. Il faut selon nous aller plus loin et, si l'on veut favoriser le *data mining* dans le cadre de la recherche scientifique, réfléchir rapidement à la création d'une nouvelle exception. C'est l'objet de notre seconde partie.

---

96 Développement encore minime, à titre d'exemple, sur les deux millions et demi d'articles de PubMedCentral, base de données à accès libre d'ouvrages scientifiques en génie biomédical et dans les sciences de la vie, seulement 17% sont placés sous des licences autorisant le *text mining*.

97 *L'édition de sciences à l'heure numérique : dynamiques en cours* CNRS Direction de l'Information Scientifique et Technique (2015) \_ <http://www.cnrs.fr/dist/z-outils/documents/Distinfo2/Distetude2.pdf>

98 V. Marx, « *My data are your data* », *Nature biotechnology*, vol.30, n°6, juin 2012, p.509-511. <http://www.nature.com/nbt/journal/v30/n6/pdf/nbt.2243.pdf>

## **Partie II.**

### **Vers une nouvelle exception au droit d'auteur**

**71. Introduction.** La nécessité de permettre le *data mining* dans certaines circonstances bien établies impose de réfléchir à la création d'une nouvelle exception afin d'admettre que ces opérations puissent, lorsqu'elles sont menées dans le cadre de la recherche, se dispenser du consentement des titulaires de droits.

**72. L'extension des exceptions, un sujet controversé.** Dans le système français, le principe est l'exclusivité du droit et les exceptions au droit d'auteur doivent trouver leur origine dans la loi. Or, le caractère fermé de la liste d'exceptions prévue par la directive 2001/29/CE interdit en principe aux Etats membres de créer une exception qui n'y figure pas. Mais le débat reste ouvert et la question de l'extension des exceptions est un sujet vivement débattu et controversé. Tandis que certains s'inquiètent d'une multiplication des exceptions dans l'environnement numérique, d'autres militent en faveur de plus de souplesse et d'un accès facilité aux œuvres. La doctrine elle-même est divisée<sup>99</sup>.

**73. Une nouvelle exception en faveur du *data mining* à des fins de recherche, un sujet controversé.** S'agissant du *data mining*, les oppositions sont fortes. Il ressort des auditions menées par le CSPLA un certain nombre d'informations : les éditeurs sont opposés à la mise en place d'une nouvelle exception qui risquerait, selon eux, de leur faire perdre tout contrôle sur les contenus téléchargés et qui pose la question de l'usage réel des données, des acteurs commerciaux pouvant alors utiliser librement des contenus leur appartenant. Ils estiment que dispenser les chercheurs de solliciter les autorisations porterait également préjudice aux créateurs. Et rappellent que rendre possible la fouille de textes et de données correspond à un service, qui nécessite de leur part un certain nombre d'investissements. Les utilisateurs, au premier rang desquels les représentants des bibliothèques et organismes de recherche, militent en faveur d'une exception en faveur de la recherche et mettent en avant la nécessité de maintenir la compétitivité de la recherche en France et au sein de l'Union européenne.

**74. Plan.** Nous démontrerons qu'en dépit des oppositions et difficultés, la voie de l'exception est une piste à explorer. Le contexte semble particulièrement propice à l'adoption d'une nouvelle exception au droit d'auteur en faveur du *data mining* (Chapitre 1), contexte législatif en France mais également contexte politique de l'Union européenne. Puis, nous tenterons de dessiner les contours d'une exception légitime (Chapitre 2).

---

<sup>99</sup> « Le fait est qu'à notre sentiment, la liste légale, quoi qu'on en dise, ne clôt pas le nombre des restrictions que peut connaître le droit d'auteur » M. Vivant, J.-M. Bruguière, *Droit d'auteur*, Dalloz, coll. Précis, 1ère éd., p 378 n°567.

## **Chapitre 1.**

### **Un contexte propice à la création d'une nouvelle exception**

**75. Plan.** Avant d'étudier l'opportunité de la création d'une exception, il est indispensable de s'assurer que le cadre légal existant ne contient pas une exception autorisant les pratiques de *data mining* effectuées à des fins de recherche. Nous établirons que les exceptions prévues dans notre droit sont toutes trop restrictives pour couvrir la diversité des pratiques de *data mining* (Section 1) et que le contexte international est dans son ensemble favorable à l'introduction de cette nouvelle exception (Section 2).

#### **Section 1. Les exceptions du droit positif inadaptées au *data mining***

**76. Précision.** Nous cantonnerons notre analyse aux exceptions permettant la reproduction de l'oeuvre tout entière puisque le *data mining* implique la reproduction de l'oeuvre dans sa totalité. En effet, ce sont les algorithmes qui vont traiter les informations et effectuer une sélection parmi les informations explorées mais pour cela, l'oeuvre devra au préalable être reproduite dans sa globalité. Il est important de rappeler que c'est cet acte de reproduction initial qui met en oeuvre le droit d'auteur. Et c'est donc au regard de cet acte de reproduction qu'il convient d'analyser les éventuelles exceptions au droit d'auteur ou au droit du producteur de bases de données. Enfin, rappelons que les exceptions sont d'interprétation stricte et doivent donc s'interpréter en faveur du droit exclusif de l'auteur.

##### **§1. L'exception pédagogique**

**77. Une exception très encadrée.** La directive sur la société de l'information de 2001 a introduit une forme d'exception pédagogique et de recherche à un fin exclusive d'illustration. La transposition de cette disposition, vivement débattue, a finalement eu lieu. La loi du 1er août 2006 est ainsi venue reconnaître une nouvelle exception selon laquelle l'auteur ne peut interdire « la représentation ou la reproduction d'extraits d'oeuvres, sous réserve des oeuvres conçues à des fins pédagogiques, des partitions de musique et des oeuvres réalisées pour une édition numérique de l'écrit, à des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche, à l'exclusion de toute activité ludique ou récréative, dès lors que le public auquel cette représentation ou cette reproduction est destinée est composé majoritairement d'élèves, d'étudiants, d'enseignants ou de chercheurs directement concernés, que l'utilisation de cette représentation ou cette reproduction ne donne lieu à aucune exploitation commerciale et qu'elle est compensée par une rémunération négociée sur une base forfaitaire sans préjudice de la cession du droit de reproduction par reprographie mentionnée à l'article L. 122-10 ».

L'exception pédagogique est soumise à des conditions très strictes. Alors que l'article 5.3 a) de la directive vise l'utilisation d'oeuvres, le Code de la propriété intellectuelle n'autorise que la reproduction ou la représentation d'extraits d'oeuvres. Même s'il est convenu que l'extrait peut être plus conséquent qu'une courte citation, il est donc impossible de reproduire une oeuvre intégralement. Cela revient à priver d'effet l'exception pour de nombreuses oeuvres et à l'empêcher

de jouer s'agissant du *data mining*. De plus, les œuvres réalisées pour une édition numérique de l'écrit ont été explicitement écartées du domaine d'application de l'exception. Même si la formule est assez obscure, elle laisse entendre que les bases de données ainsi que les publications destinées à être exploitées en ligne ne peuvent bénéficier de l'exception et répond ainsi « à la crainte des éditeurs de voir l'exception menacer leurs investissements dans la numérisation des œuvres scientifiques et techniques »<sup>100</sup>. Là encore cela écarte la faculté d'invoquer cette exception pour les pratiques d'exploration qui visent aujourd'hui majoritairement des publications déposées dans des archives ouvertes, des revues numériques en libre accès ou non ou encore des bases de données de la recherche. Les autres conditions posées, celle du public visé, de l'absence d'exploitation commerciale, ou encore celle des fins d'illustration, renforcent encore un peu plus les limites dans lesquelles cette exception est enfermée. Le *data mining* aura bien du mal à s'inscrire dans ce cadre.

**78. Proposition de rattachement.** L'idée selon laquelle le *data mining* pourrait être rattaché à cette exception de recherche est cependant parfois avancée, car présentant l'avantage d'éviter d'ajouter une nouvelle exception à une liste parfois jugée longue. C'est ce que laisse entendre le CSPLA dans son rapport : « il pourrait être utile que la Commission développe des recommandations à l'attention des Etats membres afin de clarifier dans quelle mesure les activités d'exploration de données (...) s'insèrent dans la cadre des exceptions existantes pour les usages à des fins de recherche »<sup>101</sup>. Il semble que cela soit le fondement de la nouvelle exception mise en place au Royaume-Uni.

## §2. L'exception de copie provisoire

**79. Une exception aux interprétations divergentes.** La jurisprudence a toujours considéré que la reproduction provisoire mettait en œuvre le droit de reproduction<sup>102</sup>, même si le Code de la propriété intellectuelle ne le précise pas expressément. L'article 2 de la directive sur la société de l'information du 22 mai 2001 impose aux Etats membres de reconnaître aux auteurs le droit d'interdire la reproduction provisoire ou permanente. Cette solution, favorable aux titulaires de droit, implique que les fixations matérielles fugaces, comme le stockage temporaire de contenus numérisés, mettent en jeu le droit de reproduction.

Le législateur français a transposé l'article 5.1 de la directive sur la société de l'information et a prévu une exception pour « la reproduction provisoire présentant un caractère transitoire ou accessoire, lorsqu'elle est une partie intégrante et essentielle d'un procédé technique et qu'elle a pour unique objet de permettre l'utilisation licite de l'oeuvre ou sa transmission entre tiers par la voie d'un réseau faisant appel à un intermédiaire »<sup>103</sup> et à la condition qu'elle n'ait pas de « valeur économique propre ». Cette exception est obligatoire pour tous les Etats membres. Comme pour l'exception pédagogique, les conditions d'application de cette exception sont particulièrement restrictives.

**80. Une reproduction accessoire.** Tout d'abord, seules les reproductions provisoires transitoires ou accessoires sont visées. Il faudra donc s'assurer que la copie ne sera conservée que le temps de l'exploration et qu'elle a bien un caractère accessoire (le caractère transitoire étant peu probable),

---

100A. Lucas, H.-J. Lucas, A. Lucas-Schloetter, *Traité de la propriété littéraire et artistique*, LexisNexis, 4e éd. , 2012, p. 402 n°440.

101 Conseil supérieur de la propriété littéraire et artistique (CSPLA), *Rapport de la mission sur l'exploration de données Text and Data mining*, rapport établi par M. Jean MARTIN, président de la mission et Mme Liliane de CARVALHO, rapporteur de la mission, juillet 2014, p.44.  
<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux-du-CSPLA/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

102 Cass. 1 re Civ., 15 octobre 1985 : D. 1986, somm. p. 186, obs. C. Colombet.

103 Article L 122-5-6° du Code de la propriété intellectuelle.

cette notion n'ayant pas été définie. Il est légitime de s'interroger sur le caractère accessoire de cette reproduction dans le cas du *data mining* : la reproduction est-elle réellement accessoire alors que l'oeuvre est intégralement reproduite, n'est-ce pas plutôt l'utilisation qui en est faite que l'on pourrait qualifier d'accessoire ? L'utilisation particulière d'une oeuvre peut-elle avoir pour effet de rendre la reproduction accessoire ? Il est permis d'en douter.

**81. Les autres conditions.** En outre la reproduction doit servir un but technique, ce qui signifie qu'elle doit être nécessaire sur un plan technique pour assurer la transmission de l'oeuvre. La reproduction doit également avoir pour finalité une utilisation licite de l'oeuvre, ce qui suppose d'avoir l'autorisation d'accéder à l'oeuvre, par le biais d'un contrat. Ce qui revient à définir la portée d'une exception par référence aux dispositions contractuelles stipulées entre l'utilisateur, le chercheur, et le titulaire des droits, l'éditeur, ce que la doctrine juge critiquable<sup>104</sup>.

Enfin une dernière condition est celle d'une absence de signification économique indépendante de cette reproduction, ce qui n'est pas sans soulever là encore quelques interrogations s'agissant du *data mining*. La Cour de justice de l'Union Européenne a précisé que « lesdits actes ne doivent pas avoir une signification économique indépendante, en ce sens que l'avantage économique tiré de leur mise en oeuvre ne doit être ni distinct ni séparable de l'avantage économique tiré de l'utilisation licite de l'oeuvre concernée et il ne doit pas générer un avantage économique supplémentaire, allant au-delà de celui tiré de ladite utilisation de l'oeuvre protégée »<sup>105</sup>. Dans le cas du *data mining*, le traitement des informations contenues dans cette copie provisoire est doté d'une valeur économique tout à fait indépendante de la valeur des articles, données ou contenus fouillés.

**82. Bilan.** Nous le voyons, la rédaction de cette exception autorise de nombreuses interprétations qui sont une source d'incertitude et d'insécurité et un frein aux activités de *data mining*. De plus, cette exception est encore restreinte par l'exclusion expresse des bases de données de son champ d'application. Cette exclusion, qui ne figurait pas dans la directive, limite strictement les cas où l'exception sera susceptible de jouer. Notons que Monsieur Sirinelli, dans le récent rapport de la mission sur la révision de la directive 2001/29/CE, semble estimer que, dans certaines hypothèses et au vu de la jurisprudence récente de la Cour de justice de l'Union européenne, cette exception pourrait servir de fondement aux opérations de *data mining*<sup>106</sup>. Mais là encore, seuls quelques cas particuliers de *data mining* pourraient bénéficier de l'exploitation.

### §3. L'exception de courte citation

**83. Une exception inadaptée au *data mining*.** Quant à l'exception de courte citation, elle nous paraît être inappropriée au *data mining*. En effet, s'agissant du *data mining*, l'acte de reproduction mettant en jeu le droit d'auteur vise la totalité de l'oeuvre et le fait que le résultat de l'exploration contienne de courts extraits d'une oeuvre protégée ne saurait masquer le fait que c'est l'oeuvre entière qui a initialement été fouillée et donc reproduite. Puisque c'est l'acte de reproduction initiale qui met en jeu le droit de reproduction, il nous semble illogique de vouloir appliquer l'exception de courte citation aux quelques rares extraits ou mots susceptibles de surgir au milieu des résultats de l'exploration de données. L'exception susceptible de jouer doit pouvoir s'appliquer à la reproduction initiale mettant en jeu le droit d'auteur.

---

104A. Lucas, H.-J. Lucas, A. Lucas-Schloetter, *Traité de la propriété littéraire et artistique*, LexisNexis, 4e éd., 2012, page 382 n°412.

105CJUE 17 janvier 2012 Infopaq InternationalA/S contre Danske Dagblades Forening Affaire C-302/10 point 50. <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX:62010CO0302>

106CSPLA, Rapport de 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, P. Sirinelli, octobre 2014, p.46.

#### §4. Une exception prétorienne : la représentation accessoire

**84. Une exception « instrumentalisée ».** Même si le Code de la propriété intellectuelle ne l'évoque pas, la jurisprudence juge de façon constante que l'utilisation accessoire de l'oeuvre est licite. La solution a toujours été admise en matière de droit à l'image<sup>107</sup>, elle l'est également en droit d'auteur et a été consacrée dans la jurisprudence *Etre et avoir*<sup>108</sup>, ou la Cour de Cassation a choisi de faire céder le droit de reproduction devant le caractère accessoire de la reproduction, accordant ainsi à l'utilisateur une véritable exception au droit exclusif. La Cour de Cassation est même allée plus loin puisqu'elle a jugé que la représentation de l'oeuvre était couverte par l'exception d'inclusion fortuite prévue par la directive de 2001. Le droit exclusif est ici mis à l'écart au motif que l'oeuvre est exploitée de façon secondaire et donc n'est pas réellement communiquée au public. Cette exception prétorienne, initialement réservée au cas des oeuvres situées dans des espaces publics, a connu de multiples applications jurisprudentielles, la notion d'accessoire étant interprétée de plus en plus largement. Cette solution va dans le sens selon Monsieur Vivant d' un « affinement de l'analyse des droits de propriété intellectuelle (...) qui consiste à passer d'une vision très mécaniste de ceux ci à une attention toujours plus grande portée à leur fonction ». La théorie de l'accessoire est souvent invoquée par les partisans du *data mining* libre et la proposition est à première vue séduisante.

Le raisonnement qui fonde cette exception semble de prime abord transposable au cas du *data mining*. Il est incontestable que dans le cas du *data mining*, l'utilisation de l'oeuvre dans sa forme protégée est accessoire au résultat de la fouille et ne constitue en aucun cas le sujet principal de la reproduction. Cette reproduction accessoire, bien que mettant en oeuvre le droit de reproduction, devrait alors pouvoir bénéficier de l'exception car elle ne porte pas réellement atteinte au monopole de l'auteur.

La doctrine regrette cependant une « instrumentalisation de la théorie de l'accessoire »<sup>109</sup>, source de confusion et conteste même parfois que l'on puisse admettre l'application d'une exception qui n'a pas été transposée en droit français. Sans aller jusque là, il est légitime de se demander si l'on peut réellement parler d'accessoire lorsque l'on évoque la fouille de données ? L'analyse de la jurisprudence que fait Monsieur Galopin permet de définir de façon plus précise la notion d'accessoire. Il s'agirait des apparitions réellement fortuites, comme l'atteste le nombre de décisions en matière audiovisuelle, mais également des apparitions d'une oeuvre « qui bien que clairement parvenue à la conscience de l'auteur, s'avérait nécessaire afin de saisir un élément plus vaste »<sup>110</sup>. L'exploitation d'une oeuvre via le *data mining* ne semble pas pouvoir bénéficier de cette exception, dont l'esprit ne correspond pas à l'usage qui est fait des oeuvres. Cette exception a selon nous peu de chance de s'appliquer au *data mining* car l'idée d'une incorporation inévitable de l'oeuvre avec son environnement est absente dans le *data mining*. Le rattachement récent de la théorie de l'accessoire à l'exception d'inclusion fortuite renforce cette idée.

#### §5. Les exceptions au monopole du producteur de bases de données

**85.** Comme indiqué précédemment, les bases de données peuvent être protégées par le droit d'auteur et par un droit *sui generis* accordé au producteur à l'initiative de « l'investissement financier matériel ou humain substantiel ». Selon les termes de l'article L. 342-1 du Code de la propriété intellectuelle, le producteur de bases de données a le droit d'interdire :

107CA Paris, 1re chambre, 14 septembre 1999, *JCP* 2000, E, p.1376 obs. C. Magnan.

108Cass. 1re civ., 12 mai 2011, *Bulletin* 2011, I, n°87.

109S. Carre et G. Vercken, « Google et la fortune du droit d'auteur », *Mélanges en l'honneur du Professeur André Lucas*, LexisNexis, 2014, p. 132.

110 B. Galopin, *Les exceptions à usage public en droit d'auteur*, IRPI n°41, LexisNexis, 2012, p. 247, n° 329.

- « 1° L'extraction, par transfert permanent ou temporaire de la totalité ou d'une partie qualitativement ou quantitativement substantielle du contenu d'une base de données sur un autre support, par tout moyen et sous toute forme que ce soit ;  
2° La réutilisation, par la mise à la disposition du public de la totalité ou d'une partie qualitativement ou quantitativement substantielle du contenu de la base, quelle qu'en soit la forme ».

Ainsi, lorsque l'un projet de recherche utilisant le *data mining* implique des extractions substantielles, celles-ci sont soumises à l'autorisation du producteur des bases de données, sauf à bénéficier d'une des exceptions prévues à l'article L. 342-3 du Code de la propriété intellectuelle.

« Lorsqu'une base de données est mise à la disposition du public par le titulaire des droits, celui-ci ne peut interdire :

- 1° L'extraction ou la réutilisation d'une partie non substantielle, appréciée de façon qualitative ou quantitative, du contenu de la base, par la personne qui y a licitement accès ;  
2° L'extraction à des fins privées d'une partie qualitativement ou quantitativement substantielle du contenu d'une base de données non électronique sous réserve du respect des droits d'auteur ou des droits voisins sur les oeuvres ou éléments incorporés dans la base ; (...)  
4° L'extraction et la réutilisation d'une partie substantielle, appréciée de façon qualitative ou quantitative, du contenu de la base, sous réserve des bases de données conçues à des fins pédagogiques et des bases de données réalisées pour une édition numérique de l'écrit, à des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche, à l'exclusion de toute activité ludique ou récréative, dès lors que le public auquel cette extraction et cette réutilisation sont destinées est composé majoritairement d'élèves, d'étudiants, d'enseignants ou de chercheurs directement concernés, que la source est indiquée, que l'utilisation de cette extraction et cette réutilisation ne donne lieu à aucune exploitation commerciale et qu'elle est compensée par une rémunération négociée sur une base forfaitaire. »

Aucune de ces exceptions ne semble pouvoir s'appliquer à un tel projet de recherche. La première vise les extractions non substantielles or comme nous l'avons déjà soutenu, le *data mining* implique par essence une extraction substantielle. Quant à la seconde exception, qui vise l'extraction à un usage privé, elle ne semble pas pouvoir jouer dans le cadre de la recherche, le projet de *data mining* étant le plus souvent initié par une communauté de chercheurs, un laboratoire de recherche. De plus, les bases de données électroniques sont écartées de l'exception. Quant à l'exception à des fins d'enseignement et de recherche, sa mise en jeu se heurte aux mêmes obstacles que ceux que nous avons précédemment mis en lumière.

**86. Conclusion.** Qu'il s'agisse du droit d'auteur ou du droit *sui generis* des producteurs de bases de données, les actes de reproduction ou d'extraction nécessaires au *data mining* ne seront couverts par une exception que dans de rares cas, quasiment anecdotiques. Le contexte législatif français des exceptions est donc peu favorable à la réalisation de ces opérations et confirme donc l'opportunité de réfléchir à une nouvelle exception, position que semblent aujourd'hui partager les instances européennes.

## Section 2. Un environnement international favorable à une nouvelle exception

**87. Les publications scientifiques, un marché mondial.** Les pratiques scientifiques se caractérisent de plus en plus par leur dimension internationale. La majorité des publications scientifiques sont internationales et pour cette raison publiées en anglais, devenu depuis la seconde guerre mondiale la langue d'échange dans ce domaine. Plus encore que dans d'autres secteurs, il faut raisonner à l'échelle d'un marché mondial. Certains pays ont pris l'initiative de légiférer et se sont dotés d'une exception afin de sécuriser ce champ prometteur de nouvelles pratiques scientifiques. Cela a très certainement contribué à mettre en lumière la nécessaire harmonisation au sein de l'Union européenne du cadre juridique entourant le *data mining*.

### §1. Les exceptions au droit d'auteur/copyright en faveur du *data mining* dans le monde

**88. La loi japonaise.** Depuis 2009, le *Japan Copyright Act* comporte un article 47-7 qui a créé une exception dans les termes suivants :

*« 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 such information, and to make a comparison, a classification or other statistical analysis of such information), (...) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation, of a work, to the extent deemed necessary (...) »*

La loi japonaise a adopté une définition extensive des notions d'analyse et d'information, donnant ainsi une portée large à l'exception. Notons que la Corée du sud et Taiwan bénéficient également d'une telle exception.

**89. Le *fair use* américain.** La législation américaine prévoit une exception, le *fair use*, qui permet de s'adapter à de nouveaux modes d'exploitation des œuvres comme le *data mining*. Une décision opposant *Google* au Syndicat des auteurs américains rendue le 14 novembre 2013 par la Cour fédérale de district pour le district sud de l'Etat de New York atteste que le *data mining* à des fins de recherche entre aisément dans le cadre du *fair use*. Ici l'usage a été jugé transformatif car *Google Books* a permis de transformer le texte des livres en données à des fins de recherche<sup>111</sup>. Le juge a estimé que l'autorisation préalable des éditeurs n'était pas nécessaire au motif que *Google Books* n'utilisait pas les œuvres dans un but de lecture mais améliorait l'accès du public au contenu informationnel des livres, lui permettant une utilisation inédite. Appliquée à la recherche scientifique, cette décision semble sécuriser les pratiques de *data mining*, marquant un avantage jugé décisif sur les pays européens.

**90. La réforme irlandaise en cours.** En Irlande, une réforme du *Irish Copyright and Related Rights Act (CRRRA)* est en cours et une nouvelle section 106 F intitulée *Digital analysis and research* prévoit une exception pour le « *text mining, data mining and similar analysis or research* ».

---

<sup>111</sup> Thomas Beaugrand, « Nouvelle victoire de Google face aux auteurs américains dans l'affaire Google Books », *Revue Lamy Droit de l'Immatériel*, 2014, p. 100.

**91. La récente exception du Royaume-Uni.** Mais c'est sans conteste l'initiative du Royaume-Uni qui a été la plus commentée. En effet, alors même que les négociations entre les grands éditeurs scientifiques et les divers représentants institutionnels en France et en Europe s'avéraient difficiles, une réforme du droit d'auteur au Royaume-Uni survint en avril 2014. A la suite du rapport Hargreaves<sup>112</sup> préconisant dès 2011 une réforme du droit de la propriété intellectuelle et après de nombreux et vifs débats entre les éditeurs britanniques et la communauté des chercheurs<sup>113</sup>, une nouvelle exception au droit d'auteur a été introduite au Royaume-Uni et figure à l'article 29 A du *Copyright Act*<sup>114</sup>.

Selon la loi anglaise, les copies techniques nécessaires aux opérations de *text* et *data mining* n'entraînent pas l'application du droit d'auteur. Les chercheurs peuvent explorer librement les données auxquelles l'institution pour laquelle ils travaillent ont un accès licite. Cependant, la copie réalisée doit être entièrement dédiée aux opérations de recherche effectuées dans un cadre non commercial. Il est important de noter que cette exception est gratuite et n'entraîne aucune compensation au bénéfice des titulaires de droits. Les études d'impact réalisées par le gouvernement anglais ont en effet estimé que les bénéfices pour la société étaient supérieurs aux préjudices subis par les titulaires de droits. Cette exception est entrée en vigueur le 1er juin 2014.

Le Royaume-Uni a ainsi décidé de se doter d'une exception législative, que certains ont vue comme une extension de l'exception en faveur de la recherche et de l'enseignement. Il semble en tout cas avoir estimé que le triple test ne s'opposait pas à la mise en place d'une exception gratuite en faveur du *data mining*. Si cette exception est un avantage, elle ne règle pas toutes les questions : puisque les autres pays européens n'ont pas d'exception pour le *data mining* dans leur législation nationale, il reste extrêmement difficile pour les chercheurs britanniques de mener ces projets de recherche au niveau européen. De plus, l'adoption de cette exception a fait apparaître des distorsions de concurrence et de compétitivité entre le Royaume-Uni et les autres états membres. Ainsi l'Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation (ADBU) cite le nom de laboratoires français ayant décidé d'externaliser leurs opérations de *text* et *data mining* au Royaume-Uni ou Etats Unis<sup>115</sup>. Cette initiative du Royaume-Uni a incontestablement relancé le débat sur la nécessaire harmonisation des exceptions au niveau européen.

---

112I. Hargreaves, « *Digital Opportunity, a Review of Intellectual Property and Growth* », mai 2011, p.48.

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/32563/ipreview-finalreport.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf)

113A. Jha, « *Text mining : what do publishers have against this hi-tech research tool ?* », *The Guardian*, 23 mai 2012

<http://www.theguardian.com/science/2012/may/23/text-mining-research-tool-forbidden>

114 **29A Copies for text and data analysis for non-commercial research**

(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 authorised 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 authorised by the copyright owner.*

<http://www.legislation.gov.uk/ukdsi/2014/978011112755/regulation/3>

115 Commentaires des organismes professionnels membres du CSPLA sur le 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, Décembre 2014, page 6.

## § 2. Un sujet au cœur des discussions au niveau européen

**92. En 2013.** Le 4 février 2013, la Commission Européenne lançait « *Licences for Europe* », souhaitant qu'un groupe de travail dresse, au niveau de l'Union européenne, un état des lieux des pratiques et des besoins en matière d'exploration des travaux et données scientifiques à des fins de recherche. L'objectif étant de développer ces activités. Le processus a en vain tenté de trouver un compromis entre les éditeurs scientifiques et les institutions de l'enseignement supérieur et de la recherche. Assez rapidement, les organisations représentant les intérêts des chercheurs ont en effet quitté le groupe de travail, estimant que placer les licences au centre des discussions ne permettait pas un débat équilibré. Les termes du communiqué adressé par la ligue des bibliothèques européennes de recherche à la Commission sont parfaitement clairs :

*« We write to express our serious and deep-felt concerns in regards to Working Group 4 on text and data mining (TDM). Despite the title, it appears the research and technology communities have been presented not with a stakeholder dialogue, but a process with an already predetermined outcome –namely that additional licensing is the only solution to the problems being faced by those wishing to undertake TDM of content to which they already have lawful access. Such an outcome places European researchers and technology companies at a serious disadvantage compared to those located in the United States and Asia<sup>116</sup> ».*

**93. En 2014.** Depuis, la Commission européenne a abordé à nouveau le sujet dans la consultation publique sur la réforme du droit d'auteur<sup>117</sup>. Cette consultation, présidée par Monsieur Hargreaves recommande le changement de loi sur le droit d'auteur afin de permettre aux chercheurs d'utiliser le *text* et *data mining* dans le cadre de leurs recherches scientifiques. Le groupe de travail, composé d'experts universitaires européens est arrivé à la conclusion que, malgré les opportunités économiques, les pratiques de *data mining* en Europe sont en retard par rapport à d'autres pays d'Amérique et d'Asie. Ils proposent en conséquence de réformer le cadre juridique de l'Union européenne<sup>118</sup> en ce qui concerne le droit d'auteur et les bases de données dans le but de soutenir la compétitivité internationale de la recherche européenne.

La Commission européenne a publié en mars 2014 un second rapport sur le sujet, le rapport Triaille<sup>119</sup> sur les conflits juridiques entre droit d'auteur et *data mining*, qui estime à son tour que le droit doit évoluer vers plus de flexibilité. Il recommande que le législateur européen ne se contente pas d'encourager les éditeurs à rédiger des contrats de licence plus permissifs, solution jugée insuffisante « pour résoudre l'insécurité juridique et supprimer des obstacles injustifiés à l'analyse de données ». Au contraire, ce rapport préconise l'adoption d'une nouvelle exception, inspirée de l'exception pour la recherche. Cette exception serait cantonnée au *data mining* pour la seule recherche scientifique, à des fins non commerciales et la mention du nom de l'auteur et de la source serait laissée à la discrétion du chercheur.

<sup>116</sup>Voir la lettre de LIBER, Ligue des Bibliothèques Européennes de Recherches

<http://libereurope.eu/blog/2013/02/26/licences-for-europe-a-stakeholder-dialogue-text-and-data-mining-for-scientific-research-purposes-working-group/>

<sup>117</sup>*Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining* », Report from the Expert Group site <europa.eu>, 8 avr. 2014

<sup>118</sup> « *The best approach to reform is to establish a durable distinction in European law between copyright's longstanding and legitimate role in protecting the rights of authors of 'expressive' works and copyright's questionable role in the digital age of presenting a barrier to modern research techniques and so to the pursuit of knowledge* ».

<sup>119</sup>*Study on the legal framework of text and datamining (TDM)*, Jean-Paul Triaille, Jérôme de Meeüs d'Argenteuil et Amélie de Francquen Commission européenne, mars 2014.

[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)

**94. En 2015.** Dans le cadre d'une table-ronde organisée le 19 février 2015 sous l'égide de Monsieur Oettinger, commissaire européen à l'économie et à la société numérique, LIBER, principale association de bibliothèques européennes, a, au nom de la compétitivité européenne, de nouveau plaidé pour une exception au droit d'auteur en faveur du *text et data mining*, dans la mesure où l'accès à ces documents a déjà été payé via des contrats rémunérant les auteurs.

Madame Reda, députée européenne allemande, a récemment présenté son projet de rapport très controversé devant servir de base à l'élaboration d'une nouvelle directive sur le droit d'auteur et parmi les propositions formulées, deux intéressent notre sujet : la première vise à écarter les œuvres produites par le secteur public de toute protection au titre des droits d'auteur et la seconde vise à autoriser le *data mining et le text mining* pour collecter automatiquement des données des lors que l'utilisateur a la permission de lire l'œuvre exploitée<sup>120</sup>.

Peut-être sous l'influence de l'exemple du Royaume-Uni, le contexte des débats semble ces derniers mois avoir évolué et vise désormais des objectifs plus ambitieux qu'une négociation du périmètre des licences imposées par les grands éditeurs. Ces prises de position récentes des instances européennes vont de pair avec l'objectif lancé en 2001 de bâtir un Espace Européen de la Recherche et plus récemment un marché unique numérique. La Commission Européenne a défini il y a quelques semaines 16 initiatives<sup>121</sup> pour mettre en place un marché unique numérique, parmi lesquelles :

- « donner un caractère moderne et plus européen à la législation sur le droit d'auteur » avant la fin de l'année 2015 afin de réduire les disparités entre les différents régimes de droits d'auteur (point 6) ;
- « proposer une initiative européenne en faveur de la libre circulation des données dans l'Union européenne. Il arrive parfois que de nouveaux services soient entravés par des restrictions liées à l'endroit où sont situées les données ou liées à l'accès aux données, restrictions qui sont souvent sans rapport avec la protection des données à caractère personnel. Cette nouvelle initiative abordera le problème de ces restrictions et encouragera ainsi l'innovation. La Commission lancera également une initiative européenne en faveur de l'informatique en nuage portant sur la certification des services en nuage, sur le changement de fournisseur de services d'informatique en nuage et sur un «nuage pour la recherche (point 14) ».

**95. Perspectives.** Ces recommandations, preuve d'une dynamique européenne en faveur de l'accessibilité et de la diffusion des publications et des données de la recherche, sont peut-être le signe d'une ouverture au *data mining* dans le cadre de la recherche scientifique dans un avenir assez proche. A l'heure où le droit d'auteur dans l'économie numérique est au centre des discussions au niveau européen et face à une évolution très probable de la directive de 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information, on peut regretter la position « pour le moins prudente »<sup>122</sup> du Conseil supérieur de la propriété littéraire et artistique, qui, l'année dernière, s'est clairement prononcé contre l'introduction en France d'une nouvelle exception au droit d'auteur en faveur du *data mining*.

---

<sup>120</sup><https://juliareda.eu/le-rapport-reda-explique/#tdm>

<sup>121</sup>[http://europa.eu/rapid/press-release\\_IP-15-4919\\_fr.htm](http://europa.eu/rapid/press-release_IP-15-4919_fr.htm)

<sup>122</sup>C. Bernault, J.-P. Clavier, *Dictionnaire de droit de la propriété intellectuelle*, Ellipses, coll. Dictionnaires de droit, 2e éd., 2015, p.128.

**96. Proposition.** Après avoir étudié les spécificités de monde de la recherche et de la diffusion des travaux qui en découlent et compte tenu du caractère très particulier de l'exploitation des œuvres réalisées par le *data mining*, nous proposons d'emprunter une autre voie que celle choisie par le CSPLA<sup>123</sup>. Créer une exception au droit d'auteur en faveur du *data mining* à des fins de recherche non commerciale est la solution que nous défendons<sup>124</sup>. La logique du droit européen, mûtiné de la conception personnaliste défendue par la France, peut certainement donner naissance à une nouvelle exception indispensable à la communauté des chercheurs. Qui peut se révéler une formidable opportunité pour le droit d'auteur de montrer ses facultés d'adaptation aux nouvelles pratiques de la communauté scientifique tout en faisant valoir sa légitimité.

---

123 Conseil supérieur de la propriété littéraire et artistique (CSPLA), *Rapport de la mission sur l'exploration de données Text and Data mining*, rapport établi par M. J. MARTIN, juillet 2014.

<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux-du-CSPLA/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

124 Voir, pour un avis contraire <http://www.wu.ac.at/wutv/institutes/iplaw/sciplaw/kamocki>

## **Chapitre II. Le cadre d'une nouvelle exception en faveur du *data mining* à des fins de recherche**

**97. Une exception prévue par la loi.** Même si la doctrine minoritaire réclame une certaine souplesse dans le maniement des exceptions, le principe est qu'une exception au droit d'auteur doit être prévue par la loi. Mais en matière d'exceptions au droit d'auteur, le législateur français voit son pouvoir entravé par un certain nombre de bornes.

**98. La liberté du législateur encadrée par une liste fermée d'exceptions.** La directive du 22 mai 2001 dresse dans son article 5 une liste limitative d'exceptions, afin de garantir un seuil minimum d'harmonisation. Les Etats peuvent puiser dans cette liste fermée. Mais la reconnaissance d'une nouvelle exception au droit d'auteur ne leur appartient pas et ne pourra s'imposer qu'au niveau européen.

**99. La liberté du législateur surveillée par la conformité de l'exception au triple test.** Le test en trois étapes, ou triple test, introduit par l'article 9.2 de la Convention de Berne, constitue une limite à l'adoption d'une exception. Il limite la possibilité de permettre la reproduction des œuvres à « certains cas spéciaux, pourvu qu'une telle reproduction ne porte pas atteinte à l'exploitation normale de l'œuvre ni ne cause un préjudice injustifié aux intérêts légitimes de l'auteur ». Depuis la loi du 1er août 2006 ce texte figure à l'article L. 122-5 du Code de la propriété intellectuelle et les exceptions retenues par le législateur français devront remplir les conditions de ce triple test. Ce test s'adresse également au juge chargé de l'application de la loi à des cas d'espèces concrets.

**100. Le pouvoir des institutions européennes.** Certains plaident pour l'abandon de cette liste d'exceptions strictement interprétées et pour que le triple test soit le seul critère permettant de dire si un acte met en œuvre ou non le droit exclusif. Nous n'en sommes pas là. La politique législative appartient à la Commission européenne, qui devra donc préciser les contours d'une nouvelle exception si elle entend favoriser le développement du *data mining* à des fins de recherche, conformément à ses objectifs d'harmonisation et d'assouplissement de la protection du droit d'auteur.

**101. Plan.** Dans ce chapitre, nous réfléchissons aux circonstances propres au *data mining* à des fins de recherche justifiant que l'on efface l'exclusivité du droit qu'a l'auteur sur son œuvre. Nous analyserons les fondements envisageables de cette nouvelle exception au droit d'auteur (Section 1), avant d'en définir les contours (Section 2).

### **Section 1. Les fondements de cette nouvelle exception**

**102. Plan.** Les exceptions traduisent la volonté du législateur de parvenir à un juste équilibre entre différents intérêts en présence (§1). Mais les limites au monopole de l'auteur doivent toujours être justifiées (§2).

**§1. La nécessité de parvenir à un équilibre entre des intérêts opposés.**

**103. Principe.** Une exception au droit d'auteur naît de la prise en compte d'intérêts autres que ceux des auteurs. Même si le point d'équilibre est plutôt favorable à l'auteur en droit français, le

législateur cherche à établir un équilibre entre des intérêts divergents comme l'indique le considérant 31 de la directive 2001/29/CE qui impose « de maintenir un juste équilibre en matière de droits et d'intérêts entre les différentes catégories de titulaires de droits ainsi qu'entre celles-ci et les utilisateurs d'objets protégés ». Il faut donc se poser la question d'une balance des intérêts, élément indispensable à la création d'une nouvelle exception. Une exception en faveur du *data mining* à des fins de recherche doit répondre à cette préoccupation de juste équilibre entre les bénéficiaires de l'exception, les titulaires de droits et la société en général.

**104. Les intérêts des auteurs ou titulaires de droits.** Les auteurs, mais plus largement les titulaires de droit, sont la première donnée de l'équation : l'exception doit être strictement encadrée afin de ne pas porter une atteinte disproportionnée à leurs intérêts. Comme nous l'avons vu dans la première partie, l'édition scientifique est un modèle économique assez particulier et se poser la question de l'intérêt des titulaires de droits consiste aujourd'hui à prendre en compte l'intérêt des investisseurs que sont les éditeurs. Il s'agit de leur fournir les moyens d'exploiter les œuvres publiées dans de bonnes conditions, en les récompensant de leurs investissements. Rappelons brièvement les spécificités du marché de l'édition scientifique :

- les auteurs des articles publiés ne sont quasiment jamais rémunérés et cèdent en général leurs droits d'auteur sur leurs publications à l'éditeur ;
- la concurrence est très faible dans ce secteur, on dit qu'une revue n'est pas « substituable » ;
- la valeur ajoutée provient des chercheurs car la sélection des articles, leur vérification scientifique et leur relecture sont assurés par des pairs, d'autres scientifiques experts non rémunérés, et non par l'éditeur ;
- les chercheurs publient en anglais et les frais de traduction sont rares pour l'éditeur ;
- ce secteur est très peu dépendant des fluctuations du marché de la publicité ;
- les coûts de publication d'une revue numérique sont sensiblement inférieurs à ceux d'une revue papier, même si ce sujet est l'objet de débats ;
- le chercheur réalise une partie du travail d'édition, en participant à la réalisation du document numérique qui constitue l'article.

Dans ce domaine de l'édition scientifique, le droit d'auteur apparaît comme l'instrument d'un avantage économique injustifié au profit des éditeurs et le contrôle qu'ils entendent exercer sur les activités de *data mining* menées par les chercheurs a encore accru le déséquilibre des rapports de force. Un rééquilibrage s'impose.

**105. Les intérêts des utilisateurs.** Les intérêts des bénéficiaires de l'exception, les chercheurs dans le cas de cette nouvelle exception, doivent également être pris en compte. L'exception a pour objet de profiter à un autre auteur, qui à son tour enrichira la communauté scientifique des résultats de ses travaux. Cela correspond à l'esprit de partage qui anime la communauté des chercheurs et repose sur la notion de contrepartie. Les auteurs doivent accepter de voir leurs œuvres ainsi exploitées car ils auront à leur tour l'opportunité d'utiliser les travaux de leurs collègues, de les améliorer en s'en inspirant, en les utilisant d'une façon originale. L'intérêt de la recherche est étroitement lié à celui de la création.

**106. L'intérêt public.** Enfin, l'intérêt public est l'intérêt fédérateur, supérieur qui légitime le sacrifice de certains intérêts particuliers. Il s'agit ici de s'attacher au bénéfice social, à l'intérêt général que le *data mining* peut représenter pour la collectivité. Deux arguments peuvent ici être évoqués :

- Le mode de financement de la recherche est un élément essentiel qu'il convient de ne pas négliger : le savoir scientifique naît en grande partie du financement public, les chercheurs réalisent un travail de recherche puis publient leurs travaux avec l'aide des éditeurs pour transmettre ce qu'ils ont appris à d'autres chercheurs. Puis ce sont ces mêmes laboratoires de recherche qui doivent ensuite racheter à des coûts parfois prohibitifs les productions de leurs chercheurs. Une nouvelle exception mettrait fin à ce paradoxe, au bénéfice du progrès scientifique, de l'avancement des connaissances et de la science ;
- L'utilisation des méthodes de *data mining* en épidémiologie et santé publique est en forte croissance<sup>125</sup> et démontre les avantages considérables de cette technologie pour la société. Le *data mining* est un outil approprié à l'analyse de très grandes bases de données et l'accessibilité accrue des bases de données médicales s'est révélée un formidable terrain d'investigation pour les chercheurs. Les exemples d'application du *data mining* en matière médicale sont multiples. Pour ne citer que quelques exemples parmi des centaines d'autres, le *data mining* a ainsi été utilisé afin de rechercher les facteurs de risque pour le diabète<sup>126</sup>, les infections nosocomiales<sup>127</sup>, les effets indésirables d'un médicament<sup>128</sup>. Cela démontre, sans qu'il soit besoin d'insister, l'intérêt général que représente le *data mining* pour la collectivité.

## §2. Une diversité de fondements

**107. Introduction.** Certaines exceptions au droit d'auteur ou au droit du producteur de bases de données sont justifiées par la nécessité de préserver les droits et libertés fondamentaux, d'autres sont fondées sur l'intérêt général. D'autres limites sont une façon de répondre à l'impossibilité pratique d'appliquer le monopole du titulaire de droits. Dans le cas du *data mining* à des fins de recherche, il est important d'analyser ce qui permettrait de faire céder le principe d'exclusivité. Et il faut bien reconnaître que la finalité exacte de cette exception n'est pas facile à déterminer.

**108. La liberté de la recherche scientifique.** L'exception de *data mining* à des fins de recherche trouve une justification dans le souci de protéger la liberté de la recherche. Si la liberté de la recherche n'est pas directement protégée par la Constitution française, comme c'est le cas en Allemagne ou en Italie, elle peut être rattachée à la liberté d'expression et de communication des informations<sup>129</sup> telle que l'entend l'article 10 de la Convention européenne des droits de l'homme.

« 1. Toute personne a droit à la liberté d'expression. Ce droit comprend la liberté d'opinion et la liberté de recevoir ou de communiquer des informations ou des idées sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontière. (...) »

2. L'exercice de ces libertés comportant des devoirs et des responsabilités peut être soumis à certaines formalités, conditions, restrictions ou sanctions prévues par la loi, qui constituent des mesures nécessaires, dans une société démocratique (...) à la protection de la réputation ou des droits d'autrui. »

125 Sur des exemples voir la revue *Artificial Intelligence in Medicine*.

126 A. Aljumah et al., « Application of data mining : diabetes health care in young and old patients. », *Journal of King Saul University – Computer and Information Science*, vol. 25, issue 2, July 2013, p. 117-136.  
<http://www.sciencedirect.com/science/article/pii/S1319157812000390>

127 M.A. Maalej et al. « Méthode de data mining pour l'étude des maladies nosocomiales », *Journal de mycologie médicale*, Volume 23 n°1, page 78.

128 E. Chazard et al. « Détection et prévention des effets indésirables liés aux médicaments par data mining », *IRBM*, 2009.

129 Voir aussi sur le rattachement de la liberté de la recherche à la liberté d'entreprendre voir M. Vivant et J.-M. Bruguière *Chronique Autre regard ... Propriétés Intellectuelles*, 2005 n°15 p. 225.

L'article 10 de la Convention protège la liberté d'expression en général et l'on peut considérer que par extension, le droit des scientifiques de chercher est inclus dans cet article. La liberté de la recherche recouvre deux composantes : la liberté de la recherche proprement dite, c'est-à-dire la liberté d'expérimenter, de mener une recherche en exprimant théories, démonstrations et savoir et la liberté de publication au sens large, qui consiste à diffuser les résultats obtenus<sup>130</sup>. Le *data mining* suppose d'avoir accès aux travaux et données de la recherche et débouche sur une variété de communications, sur la production d'informations scientifiques. Il est donc légitime de vouloir le rattacher à l'article 10 de la Convention européenne des droits de l'homme. Ainsi, en application du principe de proportionnalité, les restrictions que les droits d'auteur apportent à la liberté de la recherche ne doivent pas être disproportionnées par rapport au but légitime poursuivi.

**109. Une finalité sociale.** Cette première finalité n'est pas la seule à pouvoir justifier une telle exception, qui répond aussi incontestablement à des intérêts publics. Une limitation au droit d'auteur trouve son fondement dans les besoins de la société d'avoir un accès libre et gratuit à la connaissance. Une exception en faveur du *data mining* permet cet accès, entendu de façon élargie puisqu'il va au delà de la faculté de lire, dans un but d'amélioration des connaissances et de production du savoir au bénéfice de la collectivité toute entière.

**110. Une finalité pratique.** Enfin, les considérations pratiques ne sont pas non plus complètement absentes de la discussion : en effet, est-il raisonnable de vouloir subordonner au consentement des titulaires de droits les milliers de fixations qu'implique la fouille des œuvres, alors que dans l'immense majorité des cas, ils seront dans l'incapacité de reconnaître leurs oeuvres explorées ? Les trois types de fondements sont à l'évidence mélangés et renforcent l'idée que nous défendons de l'opportunité d'une nouvelle exception.

## Section 2 les contours de cette nouvelle exception

**111. Plan.** La grande difficulté consiste désormais à déterminer quelle étendue conférer à cette exception (§1) tout en gardant à l'esprit que toute exception portant une atteinte trop importante au droit des titulaires de droits est susceptible d'être rejetée en vertu du triple test (§2). Enfin nous nous attacherons à prévoir les moyens de garantir l'effectivité d'une telle exception (§3). Nous défendons l'idée d'une exception limitée au *data mining* à des fins exclusives de recherche non commerciale.

### §1. Le caractère restrictif de cette nouvelle exception

**112. Une exception en faveur du *data mining*.** Lorsque que l'on examine la loi japonaise, la loi anglaise et le projet de loi irlandaise, il est intéressant de noter qu'aucun de ces textes n'utilise le terme communément utilisé de *data mining*. Le *Japan Copyright Act* parle d' « *information analysis by using a computer* », la loi anglaise de « *electronic analysis* » et la loi irlandaise, si elle évoque bien les termes *text* et *data mining*, y ajoute la précision suivante : « *and similar analysis and research* ». Il paraît donc essentiel, à défaut de pouvoir utiliser un terme unique, de proposer une définition qui puisse prendre sens dans chaque législation nationale mais aussi au niveau européen et international et qui ne puisse donner lieu à des interprétations trop divergentes.

---

130A. Ovcearenco, « La liberté de recherche des scientifiques au regard de l'article 10 de la Convention Européenne des Droits de l'Homme », *Journal International de Bioéthique*, 2004/1 (Vol. 15), p. 65-83.

**113. Une exception en faveur du *data mining* à des fins de recherche.** Là encore il est essentiel de tenter de définir les termes et la définition de la recherche soulève des interrogations. Qu'entend-on par la recherche ? L'exception s'applique-t-elle à la recherche fondamentale uniquement ou également à la recherche appliquée ? Est-il pertinent de distinguer ces deux aspects de la recherche ?

« La recherche fondamentale a pour principal objectif la compréhension des phénomènes naturels, la mise en place de théories ou de modèles explicatifs. Elle s'intéresse, par exemple, à la façon dont les atomes s'organisent pour former des molécules ou dont les virus trouvent la « clé » des cellules pour les envahir. De son côté, la recherche appliquée se concentre sur la mise au point de nouveaux objets (logiciels, vaccins, médicaments...) ou sur l'amélioration de techniques existantes, comme la téléphonie mobile. Si une telle activité aboutit souvent à des progrès significatifs, c'est pratiquement toujours la recherche fondamentale qui est à l'origine des découvertes réellement innovantes ou des sauts qualitatifs dans les performances techniques »<sup>131</sup>.

Recherche fondamentale et recherche appliquée sont donc étroitement liées et le dénominateur commun semble être le progrès qu'elles sont toutes deux en mesure d'apporter à l'état de la science. Cependant, si la recherche fondamentale se développe principalement dans les universités, grandes écoles scientifiques et grands organismes nationaux de recherche, la recherche appliquée se déroule dans le cadre de grands instituts spécialisés mais surtout dans des laboratoires industriels dépendants de grands groupes industriels. Si à première vue il semble pertinent d'accorder le bénéfice de l'exception à la recherche fondamentale, faut-il élargir l'exception à l'ensemble de la recherche, dans toutes ses applications industrielles ?

**114. Une exception en faveur du *data mining* à des fins « principales » de recherche.** Une seconde question surgit quant à la portée de l'exception. Faut-il élargir l'exception à tous les projets de recherche, même lorsque la recherche n'est qu'accessoire ? La recherche scientifique doit-elle être le seul et unique but du projet de *data mining* ou l'exception peut-elle jouer si, en plus d'une finalité scientifique, la recherche a également un but d'analyse des comportements ou un objectif statistique ? La recherche scientifique intervient dans les domaines scientifiques mais aussi culturels, industriels, économiques et touche à des domaines aussi divers que les mathématiques, la physique, la chimie, les sciences de l'Univers, les sciences du vivant, mais également les sciences et technologies de l'information et de la communication, les sciences humaines et sociales, les sciences de l'environnement ou de l'écologie. La finalité scientifique ne sera parfois qu'accessoire à la finalité sociologique ou économique. Doit-on accorder à ces projets le bénéfice de l'exception ? Ces questions restent ouvertes mais nous pensons que la recherche doit être la finalité principale des opérations de *data mining*. L'ajout d'un critère supplémentaire semble indispensable.

**115. Une exception en faveur du *data mining* à des fins de recherche non commerciale.** Notons qu'en général, les exceptions fondées sur des droits et libertés fondamentales ne se préoccupent pas de la finalité mercantile qui anime les bénéficiaires de l'exception. Ainsi les exceptions en faveur de la presse, les exceptions de parodie ou de citation et d'analyse n'excluent pas une finalité commerciale. Alors qu'au contraire, les exceptions fondées sur des intérêts sociaux ou culturels sont toutes conditionnées en droit français à l'absence d'exploitation commerciale. C'est le cas de l'exception en faveur des personnes handicapées, de l'exception en faveur des bibliothèques et de l'exception de pédagogie et de recherche.

---

131 R. Bimbot et I. Martelly « La recherche fondamentale, source de tout progrès », *La revue pour l'histoire du CNRS* (en ligne) 24/2009 mis en ligne le 5 octobre 2011, consulté le 29 mai 2015.  
<http://histoire-cnrs.revues.org/9141#tocto1n6>

**116. Tentatives de définition de la recherche non commerciale.** Dans le cas du *data mining*, le recours à cette exigence supplémentaire d'une recherche non commerciale permet de restreindre le champ d'application de l'exception et de préserver les intérêts légitimes des auteurs et éditeurs. Mais il est extrêmement difficile aujourd'hui, compte tenu des modes de financement de la recherche publique, de déterminer ce qui est commercial ou non. Deux questions se posent, auxquelles il est délicat de répondre :

- Par rapport à qui faut-il se placer pour apprécier le caractère commercial ou non de la recherche effectuée ? Le chercheur pourra-t-il s'abstenir de solliciter les autorisations s'il ne compte pas tirer profit de l'exploitation de ses travaux de recherche. Mais alors quid si, à la suite de ce projet de *data mining*, la diffusion de ses travaux de recherche constitue une source de revenus pour l'organisme de recherche ou le laboratoire au sein duquel il travaille ? Est-on réellement dans le cadre de l'exception ?

- Que faut-il entendre par la recherche non commerciale ? Il est communément admis qu'un usage commercial inclut un avantage commercial et économique direct ou indirect. Ce critère est difficile à appliquer et plein d'incertitudes. Est-ce que cela signifie une recherche à des fins exclusivement philanthropiques ? Le chercheur peut ne pas toucher de revenus directs mais son travail de recherche et les publications qui en découlent peuvent lui procurer des avantages potentiels, comme se faire connaître auprès de la communauté scientifique, favoriser le rayonnement du laboratoire au sein duquel il travaille, ce qui lui permettra très probablement d'obtenir par la suite des financements ... indirectement cette notoriété peut donc lui profiter. Une conception extensive ne paraît pas devoir être retenue mais ces questions sont la preuve du flou susceptible de surgir d'un tel critère. De plus, dans d'innombrables cas<sup>132</sup>, le développement d'un programme de recherche fondamentale entraînera la conception de nouveaux outils qui trouveront ensuite des applications industrielles inattendues parfois lucratives. Il est tout à fait légitime, compte tenu des enjeux du *data mining*, d'imaginer que certains secteurs chercheront à s'emparer des résultats de ces projets pour les utiliser à des fins commerciales. Enfin, la science peut dans certains cas être utilisée comme un moyen détourné de faire de la publicité et les connaissances scientifiques sont parfois un outil efficace de promotion d'un produit. Cela a-t-il pour effet de rendre la recherche commerciale ?

**117. Le choix d'un critère déterminant permettant de définir la recherche non commerciale.** Certains auteurs proposent de dépasser cette opposition entre usage commercial et non commercial et d'apprécier la proportionnalité des atteintes aux droits des titulaires « à l'aune du test de substitution du marché »<sup>133</sup>, le critère à prendre en compte étant alors l'atteinte portée au marché de l'oeuvre première. Si l'idée est séduisante, nous préférons tenter de délimiter les frontières de l'usage admissible au titre de l'exception à travers la définition d'un critère déterminant.

Le considérant 42 de la directive 2001/29/CE nous donne, à propos de l'exception de recherche, quelques précisions utiles : « Lors de l'application de l'exception ou de la limitation prévue pour les utilisations à des fins éducatives et de recherche non commerciales (...), la nature non commerciale de l'activité en question doit être déterminée par cette activité en tant que telle. La structure organisationnelle et les moyens de financement de l'établissement concerné ne sont pas des éléments déterminants à cet égard. »

---

<sup>132</sup> Pour un grand nombre d'exemples, voir R. Bimbot et I. Martelly « La recherche fondamentale, source de tout progrès », *La revue pour l'histoire du CNRS* (en ligne), 24/2009, consulté le 29 mai 2015.  
<http://histoire-cnrs.revues.org/9141#tocto1n6>

<sup>133</sup> Benabou Valérie-Laure et Zolynski Célia, « Quelle réforme du droit d'auteur pour l'Union européenne ? » *Recueil Dalloz* 2014, p.731.

Le critère déterminant se trouve donc dans la nature non commerciale de la recherche scientifique, indépendamment de l'environnement et des moyens de financement de l'établissement où ces activités ont lieu. Les conditions de diffusion des résultats de la recherche apparaissent alors comme un critère essentiel. Ainsi, si les résultats de la recherche sont destinés à être rendus publics ou publiés en libre accès ou si l'auteur est autorisé par l'organisme de financement à rester titulaire de ses droits d'auteur, cela signifie que l'établissement ayant financé le projet de recherche ne pourra pas exploiter commercialement les résultats et cela atteste du caractère non commercial de la recherche. Et inversement, si les résultats d'une recherche académique sont destinés à faire l'objet d'une publication commerciale, source d'un avantage économique, alors il s'agit d'une recherche commerciale.

Cette distinction correspond à l'esprit de partage du savoir qui anime la communauté des chercheurs car cela va dans le sens d'une large diffusion de la recherche. Si le projet de *data mining* a pour finalité le partage alors il est légitime qu'il puisse bénéficier de l'exception. Si au contraire il cache une volonté de réappropriation des travaux antérieurs dans un but économique, il en est écarté. Les finalités culturelles et sociales attachées à cette exception imposent qu'elle ne puisse pas favoriser l'enrichissement d'un exploitant au détriment d'un auteur.

**118. Une frontière délicate à tracer.** Comme le prouvent toutes ces questions, la frontière entre une recherche de nature commerciale et une recherche de nature entièrement désintéressée est bien difficile à tracer. Cela ne doit pas pour autant faire disparaître ce critère qui, bien qu'imparfait, nous semble, combiné aux autres, une limite utile.

## §2. Une exception conforme au triple test.

**119. Définition.** Il convient maintenant de s'interroger sur la conformité d'une telle exception au triple test. Le test en trois étapes, ou triple test, introduit par l'article 9.2 de la Convention de Berne, constitue une limite à l'adoption d'une exception. Il réserve la possibilité de permettre la reproduction des œuvres « dans certains cas spéciaux, pourvu qu'une telle reproduction ne porte pas atteinte à l'exploitation normale de l'oeuvre ni ne cause un préjudice injustifié aux intérêts légitimes de l'auteur ». Ce test est un mécanisme d'encadrement des exceptions. Depuis la loi du 1er août 2006, il figure à l'article L. 122-5 du Code de la propriété intellectuelle, tiré de l'article 5.5 de la directive du 22 mai 2001.

Le triple test n'est pas un dispositif permettant aux Etats membres de retenir de nouvelles exceptions mais un moyen de préciser la portée des exceptions prévues. La Cour de Justice de l'Union européenne l'a rappelé dans plusieurs décisions<sup>134</sup> et impose donc une interprétation unique du droit européen. Le triple test est un outil au service du législateur qui doit prendre en compte trois critères au moment de la transposition d'une exception. Mais c'est également un outil au service du juge et dans ce cas, il peut devenir « un formidable moyen de contrôle de proportionnalité, très probablement »<sup>135</sup>.

La règle de l'article 9.2 repose sur trois conditions cumulatives. L'exception doit en premier lieu poursuivre un but spécial, une intention particulière. La seconde condition, tenant à l'absence d'atteinte à l'exploitation normale de l'oeuvre a suscité de vives controverses. Les juges du fond devront tenir compte des effets actuels mais aussi potentiels de l'exception, afin d'apprécier le risque

---

134CJUE, 27 février 2014, OSA, affaire C-351-12

<http://curia.europa.eu/juris/document/document.jsf?docid=148388&doclang=FR>

135P.-Y. Gautier, *Propriété littéraire et artistique*, PUF, coll. Droit fondamental, 9e éd., 2015, p.389 n°373.

d'une exploitation concurrente. Enfin, l'exception ne doit pas pas causer « un préjudice injustifié aux intérêts légitimes de l'auteur », c'est ici le principe de proportionnalité qui est mis en place mais souvent difficile à apprécier.

**120. La première condition.** Examinons la conformité au triple test d'une exception en faveur du *data mining* à des fins de recherche non commerciale. S'agissant de la première condition peu contraignante, elle ne soulève pas de questions particulières, une exception restreinte à des activités de recherche répondant à une finalité bien précise.

**121. La seconde condition.** Sur l'absence d'atteinte à l'exploitation normale de l'oeuvre, l'appréciation est plus délicate et se fera au cas par cas. Nous savons que l'exploitation normale de l'oeuvre recouvre non seulement l'exploitation en cours mais aussi celle que le titulaire des droits est susceptible de développer dans l'avenir. Cela vise donc toutes les formes d'exploitation possibles dans le futur. Pour apprécier si le *data mining* entre en conflit avec l'exploitation normale de l'oeuvre, il convient de vérifier si cette nouvelle forme d'exploitation est susceptible de priver les éditeurs d'un gain économique. Une exception permettant à tous de reproduire des oeuvres ou des bases de données en vue d'un projet de *data mining* à des fins de recherche est à l'évidence susceptible de porter atteinte à l'exploitation des oeuvres et donc d'affecter le marché de l'édition scientifique. Compte tenu des perspectives et enjeux prometteurs du *data mining*, la recherche commerciale si elle est couverte par l'exception ne manquera pas de s'y engouffrer, empêchant alors le titulaire des droits de bénéficier à son profit de ce nouveau mode d'exploitation. Restreindre les exceptions aux usages non commerciaux permet de préserver les intérêts des titulaires de droits et donc de respecter les exigences du triple test.

**122. La troisième condition.** Quant aux préjudices injustifiés aux intérêts légitimes de l'auteur, ils pourraient être évités par le paiement d'une rémunération, à titre de compensation pour les revenus que le titulaire de droits ne pourrait percevoir. Cette solution ne nous convainc pas et nous serions plutôt favorables à une exception sans compensation, solution retenue par le Royaume-Uni qui a choisi de ne pas compenser l'exception du fait du caractère minime du préjudice subi. Si l'on parvient à délimiter avec précision les contours de l'exception, notamment sur le caractère non commercial de la recherche, les préjudices portés aux titulaires de droits paraissent justifiés. Même si l'appréciation de ce dernier critère est loin d'être aisée, le caractère minime voire imperceptible des atteintes au droit moral de l'auteur<sup>136</sup>, dans le cas du *data mining*, est un argument qui appuie cette idée.

### §3. Les moyens de garantir l'effectivité de l'exception.

**123. Pour une exception obligatoire pour tous les Etats membres.** Les exceptions de la directive 2001/29/CE sont, hormis celle relative aux reproductions provisoires, facultatives pour les Etats membres, afin de prendre en compte les traditions nationales de chaque Etat membre. Cette souplesse a entraîné des situations divergentes et s'est révélée un obstacle au processus d'harmonisation au sein de l'Union européenne. Dans le récent rapport de la mission sur la révision de cette directive, Monsieur Sirinelli propose dans ses recommandations d'appliquer trois critères cumulatifs justifiant le recours à une exception obligatoire<sup>137</sup>. Ces critères sont les suivants :

---

136 Cf paragraphes 61 et s.

137 CSPLA, Rapport de 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, P. Sirinelli, octobre 2014, p.78.

- l'exception correspond à un besoin socio économique réel ;
- l'usage en cause a un effet transfrontalier ;
- les droits fondamentaux ou un fort enjeu sociologique légitiment la démarche.

Ces trois critères trouvent à s'appliquer s'agissant d'une exception en faveur du *data mining* à des fins de recherche. Nous avons vu que cette exception était au cœur des discussions menées par la Commission européenne en raison des multiples enjeux qu'elle porte au sein de l'Union Européenne mais également au delà des frontières de l'Europe. Compte tenu de l'intérêt social qu'elle présente et de son lien avec la liberté d'expression, il nous semble important de rendre cette exception obligatoire dans son principe, dans ses conditions et dans son régime. Une harmonisation paraît nécessaire et en pratique réalisable, compte tenu du caractère international de la recherche.

**124. Contre la contractualisation de l'exception.** La contractualisation des exceptions est un sujet largement débattu en doctrine. Dans le cas d'une exception de *data mining* à des fins de recherche non commerciale, nous défendons l'idée selon laquelle le contrat ne doit pas pouvoir priver les chercheurs du bénéfice de l'exception. Accepter d'éventuels aménagements contractuels irait à l'encontre de la volonté du législateur de faire prévaloir un intérêt qu'il aura jugé supérieur. Il faudrait alors préciser le caractère impératif de l'exception afin d'empêcher la volonté des parties de contourner les limites d'une liberté fixée par la loi. Dans la mesure où les clauses d'un contrat d'un éditeur voudraient empêcher ou restreindre les possibilités de faire du *data mining* à des fins de recherche non commerciale, ces clauses seraient réputées sans effet. C'est la solution retenue par la loi anglaise<sup>138</sup>.

**125. Une exception dont l'exercice ne peut être gêné par des mesures de protection technique.** Enfin, aux obstacles juridiques au *data mining* s'ajoutent les obstacles techniques et l'un des freins au développement du *data mining* réside dans l'absence fréquente d'interopérabilité des systèmes de gestion des données. Pour permettre le *data mining*, il est essentiel que les données et publications diverses soient diffusées par l'ensemble de la communauté des chercheurs mais surtout par les éditeurs dans des formats ouverts et reconnus. Ainsi, les différents corpus scientifiques pourront être manipulés numériquement aussi facilement et librement que possible, ce que ne permettent pas toujours en l'état actuel des choses les formats propriétaires. Compte tenu de la multiplication et de la diversité des données de la recherche, l'interopérabilité représente un enjeu majeur de l'exploitation des données. Trouver une solution à cette question délicate fait partie des obstacles qu'il faudra parvenir à franchir, préalable indispensable à l'admission d'une exception effective et efficace en faveur du *data mining* à des fins de recherche.

---

138 « *Publishers and content providers will be able to apply reasonable measures to maintain their network security or stability but these measures should not prevent or unreasonably restrict researcher's ability to text and data mine. Contract terms that stop researchers making copies to carry out text and data mining will be unenforceable.* » <https://www.gov.uk/exceptions-to-copyright#text-and-data-mining-for-non-commercial-research>

## Bibliographie

### Ouvrages généraux

Carine BERNAULT et Jean-Pierre CLAVIER, *Dictionnaire de droit de la propriété intellectuelle*, Ellipses, Coll. Dictionnaires de droit, 2e éd., 2015.

André BERTRAND, *Le droit d'auteur et les droits voisins*, Dalloz, 2e éd., 1999.

Christophe CARON, *Droit d'auteur et droits voisins*, LexisNexis, coll. Les Manuels, 3e éd., 2013.

Henri DESBOIS, *Le droit d'auteur en France*, Dalloz, 2e éd., 1966.

Pierre-Yves GAUTIER, *Propriété littéraire et artistique*, PUF, coll. Droit fondamental, 8e éd., 2012.

Pierre-Yves GAUTIER, *Propriété littéraire et artistique*, PUF, coll. Droit fondamental, 9e éd., 2015.

André LUCAS, *Le droit de l'informatique*, PUF, coll. Themis, 1987.

André LUCAS, Henri-Jacques LUCAS et Agnès LUCAS-SCHLOETTER, *Traité de la propriété littéraire et artistique*, LexisNexis, coll. Traités, 4e éd., 2012.

André LUCAS, Pierre SIRINELLI et Alexandra BENSAMOUN, *Les exceptions au droit d'auteur. Etat des lieux et perspectives dans l'Union européenne*, Dalloz, coll. Thèmes et Commentaires, 2012.

Nathalie MALLET-POUJOL, *La création multimédia et le droit*, Litec, Coll. Droit@Litec, 2e éd., 2003.

Frédéric POLLAUD-DULIAN, *Le droit d'auteur*, Economica, coll. Corpus, 2e éd., 2014.

Michel VIVANT et Jean-Marie BRUGUIERE, *Droit d'auteur*, Dalloz, coll. Précis Dalloz, 1ère éd., 2009.

Michel VIVANT et Jean-Marie BRUGUIERE, *Droit d'auteur et droits voisins*, Dalloz, coll. Précis Dalloz, 2e éd., 2013.

*Vers une rénovation de la propriété intellectuelle ?* LexisNexis, coll. IRPI n°43, 2014.

*Mélanges en l'honneur du Professeur André LUCAS*, LexisNexis, coll. Mélanges, 2014.

### Thèses

Mélanie CLEMENT-FONTAINE, *L'oeuvre libre*, Thèse de doctorat Université Montpellier I, Larcier, Coll. Création Information Communication, Bruxelles, 2014.

Benoît GALOPIN, *Les exceptions à usage public en droit d'auteur*, LexisNexis, coll. IRPI n°41, 2012.

## Articles

Thomas BEAUGRAND, « Nouvelle victoire de Google face aux auteurs américains dans l'affaire Google Books », *Revue Lamy Droit de l'Immatériel*, 2014, p. 100.

Valérie-Laure BENABOU, Célia ZOLYNSKI, « Quelle réforme du droit d'auteur pour l'Union européenne ? », *Recueil Dalloz*, 2014, p. 731.

Valérie-Laure BENABOU, « Les publications scientifiques : faut-il choisir entre libre accès et libre recherche ? » *Hermès la Revue*, 2010/2, n°57, p. 95-106.

Carine BERNAULT « Archives ouvertes et droit d'auteur : un nouveau mode de diffusion des travaux scientifiques » *Propriétés intellectuelles*, octobre 2011, n° 41.

Carine BERNAULT, « Accès à la connaissance et droit d'auteur », *Mélanges en l'honneur du Professeur André Lucas*, LexisNexis, 2014.

Murielle CAHEN, « Les aspects juridiques des bases de données : le Web de demain ? », *Documentaliste-Sciences de l'Information*, 2014/4, Vol. 51, p. 52-54.

Christophe CARON, « De la représentation accessoire d'une œuvre de l'esprit », *Communication Commerce électronique* n°5, mai 2005, comm. 78.

Christophe CARON, « Festival communautaire autour du droit de reproduction », *Communication Commerce électronique*, novembre 2009, comm. 97.

Stéphanie CARRE et Gilles VERCKEN, « Google et la fortune du droit d'auteur », *Mélanges en l'honneur du Professeur André Lucas*, LexisNexis, 2014 p.132.

Philippe CHANTEPIE, Alain BENSOUSSAN, « Le droit d'auteur à l'épreuve des nouvelles technologies », *Revue Lamy de la concurrence*, 2007, 11.

Mélanie CLEMENT-FONTAINE, « L'entreprise et l'Open source : stratégie de la valorisation », *Revue Lamy Droit de l'Immatériel*, 2014, p.102.

Mélanie CLEMENT-FONTAINE, « Les licences *Creative Commons* chez les Gaulois », *Revue Lamy Droit de l'Immatériel*, janvier 2005, n°1.

Marie CORNU, « A propos des productions intellectuelles de la recherche, entre logique privative et nécessités publiques », *Propriétés Intellectuelles*, juillet 2006, n°20, p.270-279.

Marie CORNU, « Création scientifique et statut d'auteur », *Hermès la Revue*, 2010/2 n°57 p. 85-93.

François CORONE, « Droit d'auteur et droit à l'image des biens », *Legicom*, 2/2001, n°25, p.25-31.

Thierry DESURMONT, « Qualification juridique de la transmission numérique », *RIDA*, 1996, n°4, p. 55-89.

Emmanuel DREYER, « Les chaises, l'héritier et le publicitaire : petite fable moderne », *Recueil Dalloz* 2001, p.1530.

Philippe GAUDRAT, « Droit des auteurs.- Droits patrimoniaux.- Droit de reproduction. » *J.-Cl. PLA* Fascicule 1246, spéc. N° 43.

Christophe GEIGER, « L'avenir des exceptions au droit d'auteur. Observations en vue d'une nécessaire adaptation et harmonisation du système », *La semaine juridique Edition Générale* n°47, 23 novembre 2005, I 186.

Benjamin JEAN, « Propriété intellectuelle et *Open Innovation* : les frères ennemis ? », *Revue Lamy Droit de l'Immatériel*, 2011, 77 Supplément.

Audrey LEBOIS, « Service public et créations intellectuelles », *Revue Lamy Droit de l'Immatériel*, 2008, n°36.

Agnès MAFFRE-BAUGE, « L'état du droit des bases de données », HAL, 2003.  
<https://halshs.archives-ouvertes.fr/edutice-00000029/document>

Nathalie MALLET-POUJOL, « Appropriation de l'information : l'éternelle chimère », *Recueil Dalloz* 1997, p. 330.

Nathalie MALLET-POUJOL, « Marché de l'information : le droit injustement tourmenté », *RIDA*, 4/1996, n°168, p.93.

Laure MARINO, « le big data bouscule le droit », *Revue Lamy Droit de l'Immatériel*, 2013, p. 99.

Cédric MANARA, « Le droit de l'information : paradoxes et ouvertures », *Documentaliste-Sciences de l'Information*, 2012, Vol. 49, n°3, p.18-19.

Albina OVCEARENCO, « Chapitre 4. La liberté de recherche des scientifiques au regard de l'article 10 de la Convention Européenne des Droits de l'Homme », *Journal International de Bioéthique*, 1/2004 (Vol. 15), p. 65-83.

Florence-Marie PIRIOU, « La numérisation des livres sans autorisation constitue un délit de contrefaçon », *Communication Commerce électronique*, n°5, mai 2010, étude 11.

Gilles VERCKEN, « Liberté et propriété : soumission ou autonomie ? », *Documentaliste-Sciences de l'Information*, 2014/4, Vol. 51, p. 24-26.

Célia ZOLYNSKI, « Réforme du droit d'information : quels défis relever ? », *Documentaliste-Sciences de l'Information*, 4/ 2014, Vol. 51, p. 48-50.

Michel VIVANT, « Droit d'auteur et théorie de l'accessoire : et si l'accessoire révélait l'essentiel ? », *JCP E* 2011, 1560.

Michel VIVANT et Jean-Michel BRUGUIERE, *Chronique Autre regard ... Propriétés Intellectuelles*, 2005 n°15 p. 225.

« L'effectivité des exceptions au droit d'auteur et aux droits voisins : les usages, la loi, la régulation », *Revue Lamy Droit de l'Immatériel*, n°94, juin 2013, Supplément p.1-62.

## Rapports

Commission européenne, 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*, 17 juillet 2012.

[http://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/era-communication-towards-better-access-to-scientific-information\\_fr.pdf](http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_fr.pdf)

Conseil supérieur de la propriété littéraire et artistique (CSPLA), *Rapport de la mission sur l'exploration de données Text and Data mining*, rapport établi par M. Jean MARTIN, président de la mission et Mme Liliane de CARVALHO, rapporteur de la mission, juillet 2014

<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux-du-CSPLA/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

Conseil supérieur de la propriété littéraire et artistique (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*, rapport établi par M. Pierre SIRINELLI, président de la mission et Mme Alexandra BENSAMOUN et M. Christophe POURREAU, rapporteurs de la mission, 18 novembre 2014

<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux-du-CSPLA/Missions/Mission-du-CSPLA-relative-a-l-avenir-de-la-directive-2001-29-Societe-de-l-information>

Ian Hargreaves, « *Digital Opportunity, a Review of Intellectual Property and Growth* », Mai 2011  
<[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/32563/ipreview-finalreport.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf)>

Commission européenne, *Study on the legal framework of text and datamining (TDM)*, rapport établi par M. Jean-Paul Triaille, M. Jérôme de Meeüs d'Argenteuil et Mme Amélie de Francquen, mars 2014 [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)

*Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining* », Report from the Expert Group site <europa.eu>, 8 avr. 2014

Institut National de la Recherche Agronomique (INRA) *Rapport du groupe de travail sur la gestion et le partage de données*, Paris, juin 2012, 62 p. (consulté le 28 avril 2015)  
[http://www.pfl-cepia.inra.fr/uploads/gdp\\_docs/Rapport-GestionDonnees-web.pdf](http://www.pfl-cepia.inra.fr/uploads/gdp_docs/Rapport-GestionDonnees-web.pdf)

*L'Édition de sciences à l'heure numérique : dynamiques en cours* (2015), Rapport du CNRS Direction de l'Information Scientifique et Technique (2015)  
<http://www.cnrs.fr/dist/z-outils/documents/Distinfo2/Distetude2.pdf>

Charte nationale de déontologie des métiers de la recherche signée le 26 janvier 2015  
[http://www.cnrs.fr/comets/IMG/pdf/charte\\_nationale\\_deontologie\\_signee\\_e\\_janvier2015.pdf](http://www.cnrs.fr/comets/IMG/pdf/charte_nationale_deontologie_signee_e_janvier2015.pdf)

Réponse de la quadrature du Net à la consultation publique de la Commission européenne sur la réforme du droit d'auteur en Europe  
[https://www.laquadrature.net/files/Réponse\\_de\\_La\\_Quadrature\\_du\\_Net\\_à\\_la\\_consultation\\_publicque\\_de\\_la\\_Commission\\_européenne\\_sur\\_la\\_réforme\\_du\\_droit\\_d\\_auteur\\_en\\_Europe.pdf](https://www.laquadrature.net/files/Réponse_de_La_Quadrature_du_Net_à_la_consultation_publicque_de_la_Commission_européenne_sur_la_réforme_du_droit_d_auteur_en_Europe.pdf)

### Articles non juridiques sur la recherche scientifique et le *data mining*

Abdullah A. ALJUMAH et al., « *Application of data mining : diabetes health care in young and old patients* », *Journal of King Saud University – Computer and Information Sciences*, vol. 25, Issue 2, p. 117-136, July 2013  
<http://www.sciencedirect.com/science/article/pii/S1319157812000390>

René BIMBOT et Isabelle MARTELLY « La recherche fondamentale, source de tout progrès », *La revue pour l'histoire du CNRS* (en ligne) 24/2009 mis en ligne le 5 octobre 2011, consulté le 29 mai 2015 <http://histoire-cnrs.revues.org/9141#tocto1n6>

Emmanuel CHAZARD et al. « Détection et prévention des effets indésirables liés aux médicaments par *data mining* », *IRBM*, vol. 30, n°4, p.192-196, septembre 2009 (consultation du résumé de l'article uniquement).  
<http://www.em-consulte.com/article/226571/article/detection-et-prevention-des-effets-indesirables-li>

Gregory COLCANAP, Christophe PERALES, « CSPLA - Mission relative au *data mining* : l'analyse de Couperin et de l'ADBU ». <http://www.abes.fr/>

Kenneth CUKIER et Viktor MAYER-SCHONBERGER, « Mise en données du monde, le déluge numérique », *Le Monde diplomatique* juillet 2013, pages 1, 20 et 21.

Jacques HENNO, « L'usine 4.0, nouvelle révolution industrielle », *Les Echos*, 8 octobre 2013.

Agnès HENRI, Bart van TIGGELEN, Marie-Anne LERICHE, « Réflexions sur la voie dorée des publications scientifiques », *Reflets de la Physique* n°30, p.28-31.  
[http://lpm2c.grenoble.cnrs.fr/UserFiles/File/Reflets%2030\\_publications%20scientifiques.pdf](http://lpm2c.grenoble.cnrs.fr/UserFiles/File/Reflets%2030_publications%20scientifiques.pdf)

Gildas ILLIEN, « Enjeux professionnels. Décrire les objets du savoir, les nouveaux paradigmes du savoir », *Documentaliste-Sciences de l'Information*, 2013/3, Vol. 50, p. 26-41.

Alok JHA, « *Text mining : what do publishers have against this hi-tech research tool ?* », *The Guardian*, 23 mai 2012.  
<http://www.theguardian.com/science/2012/may/23/text-mining-research-tool-forbidden>

Pierre-Carl LANGLAIS et Lionel MAUREL, « Quel statut légal pour le *data mining* ? », Synthèse de Savoirscom1 consécutive à l'audition du 15 janvier 2013 par le CSPLA.  
<http://www.savoirscom1.info/wp-content/uploads/2014/01/Synthèse-sur-le-statut-légal-du-content-mining.pdf>

Jean-Pierre MALLE, « La triple rupture des Big Data », *ParisTech review*, 15 mars 2013.  
<http://www.paristechreview.com/2013/03/15/big-data-revolution-culturelle/>

M.A. MAALEJ et al. « Méthode de *data mining* pour l'étude des maladies nosocomiales », *Journal de mycologie médicale*, Volume 23 n°1, page 78.

Vivien MARX, « *My data are your data* », *Nature biotechnology*, vol.30, n°6, juin 2012, p.509-511. <http://www.nature.com/nbt/journal/v30/n6/pdf/nbt.2243.pdf>

Nathalie PIGNARD-CHEYNEL, « L'édition de revues scientifiques : une forme de marchandisation de la diffusion des connaissances », *Sciences de la société* n° 66, octobre 2005.

Stéphane POUYLLAU, « Enjeux professionnels. Web de données, Big Data, Open Data, quels rôles pour les documentalistes ? », *Documentaliste-Sciences de l'Information*, 3//2013, Vol.50, p. 26-41.

Mounce ROSS, « *The right to read is the right to mine : text and data mining copyright exceptions introduced in the UK* », *The London School of Economics and political science*.  
<http://blogs.lse.ac.uk/impactofsocialsciences/2014/06/04/the-right-to-read-is-the-right-to-mine-tdm/>

Richard VAN NOORDEN « Text mining spa heats up », *Nature*, 21 mars 2013, vol. 495, p. 295.  
[http://www.nature.com/polopoly\\_fs/1.12636!/menu/main/topColumns/topLeftColumn/pdf/495295a.pdf](http://www.nature.com/polopoly_fs/1.12636!/menu/main/topColumns/topLeftColumn/pdf/495295a.pdf)

Richard VAN NOORDEN, « *Trouble at the text mine* », *Nature*, 8 mars 2012, vol. 483, p. 134-135.  
[http://www.nature.com/polopoly\\_fs/1.10184!/menu/main/topColumns/topLeftColumn/pdf/483134a.pdf](http://www.nature.com/polopoly_fs/1.10184!/menu/main/topColumns/topLeftColumn/pdf/483134a.pdf)

## Index alphabétique

(Les chiffres renvoient aux numéros des paragraphes)

### A

Accessoire (théorie) 32, 84  
Action en concurrence déloyale 27  
Agissements parasitaires 27

### B

Balance des intérêts 103 et s.  
Bases de données

- contrefaçon appréciation 23
- définition 22
- droit d'auteur 20
- droit *sui generis* 22
- étendue de la protection 25
- exceptions 85, 86
- originalité 20, 22

### C

Cession de droits 69  
Concurrence déloyale 27  
Contrats 47, 51  
Convention de Berne 19, 99  
Convention EDH 108  
CSPLA 10, 95

### D

Domaine public 26, 56, 57  
Droit à la paternité 63  
Droit au respect 62  
Droit d'auteur

- légitimité 5, 6, 8, 96
- fondements 8

Droit de divulgation 64  
Droit de repentir ou de retrait 64  
Droit de représentation 30  
Droit de reproduction 32 et s.

- fixation matérielle 34
- communication de l'oeuvre 36

Droit moral 61  
Droit *sui generis*, bases de données

- coexistence avec le droit d'auteur 20, 24
- droit d'extraction 23

### E

Exceptions

- copie provisoire 79 et s.
- courte citation 83

- fondements 107 et s.
- garantie d'effectivité 123, 124, 125
- inclusion fortuite 84
- liberté d'expression 108
- liste limitative 98
- pédagogique 77
- usage transformatif 89

### H

Harmonisation européenne 92 et s., 100, 123

### I

Interopérabilité 125  
Information 7, 32  
Intérêts (balance) 57, 88 et s.

### L

Législations étrangères 88, 89, 90, 91, 112  
Liberté de création 105  
Liberté d'expression 108  
Libre accès 8, 44, 45  
Licences libres

- *Creative commons* 56
- domaine public 27, 57, 58
- limites 59 et s.
- perspectives 58, 69

### O

Originalité 19, 20, 22

### P

Préjudice 122  
Publications scientifiques 44 et s.

### R

Recherche

- Charte de déontologie 63
- indépendance 53
- liberté 108

### T

Titularité des droits 28  
Triple test

- conditions 119 et s.
- définition 99

## Table des matières

Sommaire	
Introduction	1
<b>PARTIE I. Les obstacles au <i>data mining</i> à des fins de recherche</b>	<b>6</b>
<b>Chapitre 1. Une infinité d'autorisations à solliciter</b>	<b>7</b>
Section 1. Le problème de l'identification et du statut des ressources explorées	7
§1. Des œuvres couvertes par le droit d'auteur	
§2. Des œuvres couvertes par le droit sui generis du producteur de bases de données	
§3. Des œuvres du domaine public	
Section 2. La question de l'identification des droits d'exploitation, <i>data mining</i> et droit de reproduction	11
§1. Discussion sur les conditions de la mise en jeu du droit de reproduction	
§2. La solution imposée par la conception française du droit de reproduction	
<b>Chapitre 2. Les solutions contractuelles insuffisantes</b>	<b>17</b>
Section 1. Les clauses de <i>data mining</i> dans les contrats des grands éditeurs	19
§1. Exemples de clauses autorisant le data mining	
§2. Analyse critique	
Section 2. Licences libres et <i>data mining</i>	23
§1. Exemples de licences adaptées au data mining	
§2. Appréciation critique	
<b>PARTIE II. Vers une nouvelle exception au droit d'auteur</b>	<b>29</b>
<b>Chapitre 1. Un contexte propice à la création d'une exception</b>	<b>30</b>
Section 1. Des exceptions inadaptées	30
§1. L'exception pédagogique	
§2. L'exception de copie provisoire	
§3. L'exception de courte citation	
§4. Une exception prétorienne : la théorie de l'accessoire	
§5. Les exceptions au monopole du producteur de bases de données	
Section 2. Un environnement international favorable à une nouvelle exception	35
§1. Les exceptions en faveur du <i>data mining</i> dans le monde	
§2. Un sujet au cœur des discussions européennes	
<b>Chapitre 2. Le cadre d'une nouvelle exception</b>	<b>40</b>
Section 1. Les fondements	40
§1. La recherche d'un équilibre entre des intérêts opposés	
§2. une diversité de fondements	
Section 2. Les contours de cette nouvelle exception	43
§1. Le caractère restrictif de l'exception	
§2. Une exception conforme au triple test	
§3. Les garanties de l'effectivité de l'exception	
Bibliographie	51
Index alphabétique	49
Table des matières	56