

Mission sur l'exploration de données

(« Text and Data mining »)

Président de la mission :

Jean MARTIN,

Avocat à la cour,

personnalité qualifiée au CSPLA

Rapporteur de la mission :

Liliane de Carvalho,

Avocat à la cour,

Juillet 2014

Préface et recommandations

La tension entre protection et exploitation de la création est un équilibre dynamique, c'est-à-dire instable, donc en définition permanente. L'exploration de données (« *Text and Data mining* ») en donne encore une illustration, tout en renouvelant le genre.

Comme bien souvent avec le numérique, le centre de gravité s'est déplacé. La nouveauté est là. L'exploration de données utilise, exploite, l'œuvre, non pour la reproduire ou la représenter, au sens du droit d'auteur, mais pour produire de la connaissance. De cette connaissance pourront naître de nouvelles connaissances, de nouveaux produits et services à valeur ajoutée.

La revue de presse, manuelle puis automatisée, suggère l'exploration de données mais elle n'en est pas. Certes elle se réalise par une démarche similaire d'identification de multiples sources puis leur fouille afin d'en extraire de l'information. Toutefois cette information contient des « extraits » de l'œuvre, que ceux-ci relèvent ou non des exceptions de citation ou de revue de presse. Il y a emprunt à l'œuvre par l'extraction et la restitution de l'objet même de la protection du droit d'auteur : la forme originale.

Le résultat de l'exploration de données ne fait pas d'emprunt à la forme originale de l'œuvre. Il n'a pas pour objet, sauf exception, à la restituer, pas plus que le contenu des bases de données. De l'une ou de l'autre, il ne dégage que de la « *connaissance* », du « *sens* », par l'analyse de l'objet prédéfini ou aléatoire opérée grâce à des algorithmes élaborés pour combiner et traiter une multitude de sources et de problématiques en fonction de la finalité recherchée.

C'est un nouveau défi pour le droit de la propriété littéraire et artistique : protecteur de la forme de l'œuvre et du contenu des bases de données est-il apte à les protéger d'appropriations de leur sens, apparent ou caché, à supposer que l'analyse ne soit pas libre, fût-elle automatisée ?

Deux constats sont de nature à éclairer le chemin de la réflexion.

Le premier est technique. Il est reconnu que l'exploration de données, qui consiste à opérer des traitements informatisés de gisements de contenus, quelle que soit leur nature (textes, données, images fixes ou animées, musique, sons...) impose, pour les fouiller, d'en prendre possession et de les stocker, 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. Les exceptions existantes n'apportent ni la sécurité ni une réponse appropriée au développement de l'exploration de données. L'autorisation des titulaires de droits est donc requise. Le « droit d'auteur » n'est dès lors pas impuissant et doit s'appliquer.

Le second est économique, si ce n'est moral. On ne peut ignorer que l'activité d'exploration de données se développe sur la création et l'investissement d'autrui, des bases de données et

des œuvres. Sans cet apport initial, l'exploration de données ne pourrait ni exister, ni développer de valeur. En ce sens, dans l'ordre des organismes vivants, l'exploration de données est de nature « parasitaire ». Mais comme certains parasites, l'exploration de données ne se nourrit de l'organisme hôte que pendant une partie de son cycle vital. Dans d'autres phases, il crée, produit et utilise ses propres outils et « substances », aux coûts intellectuels et financiers parfois très significatifs (algorithmes), pour générer ses résultats et de la valeur ajoutée.

La multiplicité des sources et le caractère considérable des volumes traités dans une opération d'exploration de données nécessitent un mode de gestion juridique adapté, minimisant les coûts de transaction et les délais.

La gestion individualisée des droits est peu adaptée, à moins d'un mode opératoire simplifié et quasiment automatisé. Les autres réponses sont connues : la gestion collective, volontaire ou obligatoire, laquelle a parfois été mise en place pour répondre à des besoins similaires ; l'exception est l'ultime recours dans notre tradition juridique, à défaut de toute autre solution.

La mission a conduit ses travaux avec le précieux concours de son comité de pilotage constitué de personnalités expertes et grâce aux auditions multiples et diverses, de façon à couvrir le spectre du domaine et à garantir l'expression de la diversité des opinions et des intérêts.

Le rapport de mission, que madame Liliane de Carvalho a accepté de prendre en charge restituée, avec finesse et maîtrise malgré les difficultés d'une telle tâche, toute l'ampleur du sujet, la variété des secteurs concernés, l'importance des enjeux et fournit un cadre à l'analyse juridique susceptible d'être faite du phénomène à la lumière des derniers travaux européens et mondiaux.

Enfin, les « Recommandations » reposent sur l'idée maîtresse que la diversité des secteurs d'activité concernés et la spécificité des besoins à prendre en considération ainsi que tout l'intérêt de développer le « DTM » pour l'économie du numérique et de la connaissance imposent de rechercher des équilibres intelligents et vertueux qu'une solution radicale ne peut apporter.

LES RECOMMANDATIONS

- 1- Favoriser un développement de l'exploration de données qui ne soit pas parasitaire de la création et des investissements dans les gisements de contenus, dans l'intérêt même de l'exploration de données.**
- 2- L'exploration de données qui se caractérise par un traitement de masse d'une multitude de sources est peu compatible avec un exercice individuel des droits de propriété littéraire et artistique par leurs titulaires.**
- 3- Agir pour favoriser un cadre juridique et économique pertinent.**
- 4- Initier une politique publique pour mettre en œuvre et adapter les modes opératoires de gestion des droits de PLA afin de répondre aux besoins spécifiques du traitement de masse de l'exploration de données.**
- 5- Privilégier 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.**
- 6- Engager un processus de concertation et d'auto-régulation sectorielle pour permettre la recherche et l'élaboration des modalités opératoires et des équilibres les plus adaptés et pertinents aux divers secteurs concernés.**
- 7- Fixer un délai de deux années au terme duquel un bilan sectoriel sera dressé et l'éventuelle nécessité d'une intervention de nature législative évaluée.**
- 8- Inscire ce processus dans l'agenda des instances nationales sectorielles appropriées, avec mission de rendre un rapport spécial annuel.**
- 9- Inscire l'exploration de données dans la politique d'accès aux données publiques.**
- 10- Inscire l'exploration de données dans les plans de soutien à l'économie numérique, dont le « Big Data » et aux investissements d'avenir.**

11- Faire partager cette approche au sein de l'Union Européenne et prévenir toute initiative qui ne s'inscrirait pas dans la démarche recommandée.

12- Faire de même au sein de l'OMPI et dans les instances internationales concernées.

Jean Martin

Introduction

En juillet 2013, le président du Conseil supérieur de la propriété littéraire et artistique a confié à M. Jean Martin une mission sur l'exploration de données.

La numérisation massive des données, le développement du « *Big data* » et la création d'algorithmes et de systèmes de gestion de bases de données ont permis l'émergence de cette activité.

L'exploration de données¹ permet d'analyser des gisements de données, présents dans des œuvres protégées ou non, afin d'en extraire de l'information et de produire des connaissances. Elle a été présentée en 2001 par le Massachusetts Institute of Technology comme l'une des dix technologies émergentes qui « *changeront le monde* » au XXI^{ème} siècle².

Afin de recueillir la diversité des analyses et des expertises dans les différents secteurs des industries de la création et de l'information concernés par les problématiques de l'exploration de données, M. Jean Martin s'est entouré d'un comité de pilotage³ qui a procédé, de septembre 2013 à avril 2014, à de nombreuses auditions⁴.

La réflexion engagée avec l'appui du comité de pilotage s'est inscrite dans le contexte des travaux ouverts sur le sujet par la Commission européenne⁵. Elle s'est également nourrie des réponses apportées ou envisagées à cette problématique dans les pays d'Europe ou sur les autres continents.

La mission s'est fixée pour objectif d'envisager les conditions du développement de l'activité d'exploration de données dans le respect du droit de la propriété littéraire et artistique, alors que s'expriment d'une part, les inquiétudes des ayants droits relatives à la dissémination des œuvres et aux bouleversements du partage de la valeur propres à l'univers numérique et d'autre part, les souhaits des utilisateurs⁶ relatifs à la sécurité juridique et à la faisabilité technique des opérations d'exploration de données.

Après une présentation de l'exploration de données (I), le rapport engage une réflexion sur l'application du droit de la propriété littéraire et artistique aux actes nécessaires à cette activité (II) et sur les outils juridiques et techniques adaptés à son développement (III).

¹ Il est souvent fait référence à la terminologie anglaise « *text and data mining* » (ci-après TDM).

² Data mining et statistique décisionnelle – L'intelligence des données, Stéphane Tufféry, Editions Technip 2012.

³ Annexe 1. Composition du comité de pilotage

⁴ Annexe 1. Liste des personnes auditionnées

⁵ Au travers de l'initiative des « *licences pour l'Europe* » et de la consultation publique lancée dans le cadre des travaux visant à revoir et à moderniser les règles européennes sur le droit d'auteur. http://europa.eu/rapid/press-release_IP-13-1213_fr.htm

⁶ Chercheurs, bibliothèques, entreprises commerciales.

I. Présentation de l'activité d'exploration de données

L'exploration de données est une activité protéiforme (1) dont le développement soulève de nombreuses problématiques notamment en droit de la propriété littéraire et artistique (2).

1. L'exploration de données : une activité protéiforme

Le caractère protéiforme de l'exploration de données conduit à l'utilisation d'une terminologie variable (1.1) mais n'empêche pas de dégager un certain nombre d'éléments permettant de définir cette activité (1.2).

I.1. Une terminologie variable

Selon le Journal officiel du 27 février 2003⁷, l'équivalent anglais du terme « *exploration de données* » est « *data mining* ». Dans la pratique, il est également souvent fait référence aux termes « *text mining* », « *text and data mining* », « *web mining* », « *media mining* », « *video mining* », « *sounds mining* » ou encore « *content mining* ».

Le rapport « *Study on the legal framework of text and data mining (TDM)* »⁸ remis en mars 2014 à la Commission européenne engage une réflexion sur la terminologie utilisée pour l'activité d'exploration de données et considère finalement que le « *text mining* », le « *data mining* » et le « *text and data mining* » sont des sous-ensembles d'un concept plus général qui serait « *l'analyse de données* » (« *data analysis* »).

Pour notre part, nous retiendrons le terme « *exploration de données* ». En effet, il a l'avantage d'être suffisamment large pour être à la fois neutre d'un point de vue technologique et non discriminant par rapport à la source ou à la nature des données concernées.

I.2. Les éléments de définition de l'exploration de données

L'utilisation d'une terminologie variable en la matière a pour effet de conduire à la multiplication des définitions. Il est possible de trouver différentes définitions pour les activités de « *data mining* »⁹, « *text mining* »¹⁰ ou encore de « *text and data mining* »¹¹.

⁷ <http://www.dglf.culture.gouv.fr/cogeter/27-02-03-internet.htm>

⁸ « *Study on the legal framework of text and data mining (TDM)* » mars 2014. Jean-Paul Triaille, De Wolf & Partners, lecteur avec Jérôme de Meeûs d'Argenteuil, De Wolf & Partners et avec la collaboration de Amélie de Francquen, associate, De Wolf & Partners

⁹ L'ouvrage de référence sur le sujet « *Data mining et statistique décisionnelle – L'intelligence des données* » de Stéphane Tufféry (Editions TECHNIP) définit dans son avant-propos le « *data mining* » comme « *l'application des techniques de statistiques, d'analyse des données et d'apprentissage automatique à l'exploration et à l'analyse sans a priori de grandes bases de données informatiques, en vue d'en extraire des informations nouvelles et utiles pour le détenteur de ces données* ».

¹⁰ « *Text mining is the process that turns text into data that can be analysed* ». « *Text mining and Scholarly Publishing* », (PRC) Jonathan Clark

¹¹ « *Text and Data Mining means to perform extensive automated searches of Publisher's Content, the sorting, parsing, addition or removal of linguistic structures, and the selection and inclusion of content into an index or*

Des textes de lois en vigueur ou en cours d'élaboration à l'étranger offrent également quelques éléments de définitions.

Le Japon est le seul pays à prévoir une exception au droit d'auteur pour les activités assimilées à l'exploration de données. En effet, le texte de loi utilise encore une autre terminologie, l'« *analyse d'information* ». Il précise que, pour cette activité, l'utilisation d'un ordinateur permet des actes d'extraction, de comparaison, de classification ou tout autre acte d'analyse statistique sur des informations pouvant provenir du langage, des sons, images ou de tout autre élément constituant de l'information¹².

En Grande-Bretagne, la réforme du *Copyright, Designs and Patents Act* de 1998 actuellement en cours de discussion, prévoit une exception pour la reproduction dans le cadre du « *text and data analysis* »¹³. L'étude d'impact réalisée par le Gouvernement britannique¹⁴ précise la définition de cette activité¹⁵.

En Irlande, un projet de réforme du droit d'auteur est également à l'étude et pourrait prévoir une exception pour l'exploration de données en faveur de la recherche. Les activités de « *text mining* » et « *data mining* » y sont mentionnées mais sans être clairement définies.

Dans le cadre de la réflexion menée par la Commission européenne sur le sujet, le rapport « *Study on the legal framework of text and data mining (TDM)* » remis en mars 2014 propose une définition de « *l'analyse de données* » (« *data analysis* ») qui serait « *le traitement automatisé de contenus numériques, qui peuvent inclure des textes, des données, des sons, des images ou d'autres éléments, ou une combinaison de ceux-ci, afin de découvrir de nouvelles connaissances ou des idées* »¹⁶.

database for purposes of classification or recognition of relations and associations ». http://www.stm-assoc.org/2012_03_15_Sample_Licence_Text_Data_Mining.pdf

¹² Article 47 septies « *For the purpose of information analysis ("information analysis" means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other much information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use by a person who makes an information analysis* ». <http://www.cric.or.jp/english/clj/cl2.html>

¹³ Nous reviendrons sur l'économie de cette exception dans la troisième partie du rapport

¹⁴ <http://www.ipo.gov.uk/consult-ia-bis0312.pdf>

¹⁵ « *Text and data and data analytics methods extract data from existing electronic information, to establish new facts and relationships, building new scientific findings from prior research. These new methods involve copying of prior works as part of the process to extract data.* »

¹⁶ « *The automated processing of digital materials, which may include texts, data, sounds, images or other elements, or a combination of these, in order to uncover new knowledge or insights* ». Study of the legal framework of text and data mining. Page 17.

En France, le Journal officiel du 27 février 2003¹⁷ définit l'exploration de données comme un « *processus de recherche dans un ensemble de données destiné à détecter des corrélations cachées ou des informations nouvelles* ».

Pour notre part, nous retiendrons que l'activité d'exploration de données est un processus de recherche automatisé (1.2.1) qui porte sur un ensemble de données numériques (1.2.2) dans le but de découvrir de nouvelles connaissances ou de nouvelles idées (1.2.3).

I.2.1. Un processus de recherche automatisé

L'utilisation d'outils informatiques et d'algorithmes pour l'exploration de données est l'un des éléments qui permet de distinguer cette activité de la collecte et de l'analyse traditionnelles de données effectuées avec un crayon et un papier. Le volume croissant des données propose une quantité d'informations potentiellement pertinentes mais qui ne sont pas exploitables manuellement.

En effet, les possibilités nouvelles offertes par les outils informatiques permettent de collecter des données à une échelle beaucoup plus vaste et l'automatisation du traitement rend possible l'analyse simultanée de plusieurs jeux de données, de sources et de natures différentes, ce qui rend l'exploration de données très efficace pour la découverte de nouvelles connaissances.

Sur le plan technique, le processus de recherche automatisé implique de nombreux actes tels que l'extraction, la copie, le tri, la comparaison, le classement, la suppression, l'interprétation des données.

Chaque opération d'exploration de données peut impliquer des actes différents (copie ou non, classement ou non des données) et l'accomplissement ou non de ces actes peut dépendre notamment de l'objet de la recherche ou des outils informatiques utilisés.

I.2.2. Sur un ensemble de données numériques

L'exploration de données permet de traiter, d'analyser et de croiser des données de natures différentes, de sources diverses et pouvant relever de statuts juridiques distincts.

Sur le plan technique, l'exploration de données ne peut être réalisée que sur des contenus numériques. Par conséquent, les données traitées doivent être nativement numériques ou être préalablement numérisées.

Tous les contenus numériques peuvent potentiellement faire l'objet d'une exploration de données : l'écrit, la musique, les images, animées ou non, les données, sous forme ou non de bases de données.

¹⁷ Vocabulaire de l'informatique

Dans la pratique, l'accès aux contenus numériques peut être libre (ex : données web) ou limité par un statut juridique particulier (ex : données personnelles) ou par des conditions générales d'utilisation (ex : sites internet) ou par un contrat (ex : abonnement).

Certaines initiatives permettent d'élargir le corpus de données numériques disponibles et d'encourager ainsi le développement de l'activité d'exploration de données. Il s'agit par exemple des données publiques¹⁸ et de la politique d'« *open data* ». À cet égard, il est possible de citer le catalogue Joconde qui regroupe près de 500 000 notices d'objets de toute nature (archéologie, beaux-arts, ethnologie, histoire, sciences et techniques) ou encore l'initiative de la Bibliothèque nationale de France qui depuis le 1^{er} janvier 2014 met à disposition plus de 12 millions de notices bibliographiques et 2,5 millions de notices d'autorité pour tout type de réutilisation sous licence ouverte de l'État. Il faut préciser que l'« *open data* » implique une ouverture de données pour toute réutilisation y compris commerciale mais cela ne signifie pas qu'il faille s'affranchir d'un encadrement contractuel rigoureux. En effet, tout acte de réutilisation de données publiques culturelles doit faire l'objet d'un contrat de licence.

Il faut également noter que si certains contenus numériques susceptibles de faire l'objet d'une opération d'exploration de données, peuvent bénéficier d'une protection par le droit de la propriété littéraire et artistique, un certain nombre d'œuvres sont par leur nature dans le domaine public ou peuvent être mises à la disposition du public grâce à des dispositifs tels que le « *libre accès* »¹⁹ ou les « *Creatives Commons* »²⁰. Par exemple, actuellement plus de 50% des articles scientifiques européens sont publiés sous une forme de libre accès²¹.

I.2.3. Dans le but de découvrir de nouvelles connaissances et de nouvelles idées

Les différentes auditions menées au cours de la mission ont confirmé que tous les secteurs sont concernés par le développement de l'activité d'exploration de données. Cependant, les acteurs concernés ne rencontrent pas les mêmes problèmes ; leurs demandes à l'égard des outils d'exploration de données, comme l'état de leurs réflexions, leurs engagements dans diverses procédures diffèrent largement.

¹⁸ Le régime juridique français du droit à la réutilisation des données publiques est défini par le chapitre II de la Loi du 17 juillet 1978 modifiée. Au niveau européen, la directive 2013/37/UE relative aux informations du secteur public a étendu son champ d'application aux bibliothèques, musées et archives et a rappelé à cette occasion que le principe est l'ouverture et l'exception est la redevance.

¹⁹ Le « *libre accès* » permet la mise à disposition en ligne des articles scientifiques afin de favoriser la circulation du savoir. Dès lors, il est notamment possible de lire, télécharger, copier, transmettre, imprimer les contenus sans barrière financière, légale ou technique.

²⁰ Les licences « *Creatives commons* » sont destinées à faciliter l'utilisation des œuvres mises à la libre disposition du public. Elles impliquent l'autorisation de reproduire, de distribuer et de communiquer la création au public à condition de le faire à titre gratuit. Il existe des options que le titulaire de droit peut choisir (ex : pas d'utilisation commerciale).

²¹ Communiqué de presse de l'Union européenne, « Open access to research publications reaching 'tipping point' », 21 août 2013, http://europa.eu/rapid/press-release_IP-13-786_en.htm

L'étude entreprise pour le Publishing Research Consortium sur la situation et le potentiel concernant l'exploration de textes et de données²² a permis d'identifier quelques acteurs qui ont recours à ces méthodes : les éditeurs qui enrichissent ainsi leurs contenus, les chercheurs qui identifient de nouvelles informations, les entreprises notamment pharmaceutiques qui accélèrent la découverte de nouveaux médicaments et les utilisateurs commerciaux qui développent des produits basés sur cette activité. Pour les bibliothèques, le développement de l'exploration de données constitue un enjeu considérable dans la mesure où elles sont productrices et détentrices de données.

Pour la recherche scientifique, l'exploration de données permet notamment de sélectionner plus rapidement la littérature scientifique utile pour un sujet. Au cours des auditions, le projet Text2genome²³ qui a permis de cartographier le génome humain en compilant automatiquement trois millions de publications a souvent été très souvent cité en exemple.

S'agissant de la presse, l'exploration de données est sans aucun doute l'un des modèles économiques de l'information de demain. L'activité de veille média qui consiste à fournir une information utile, triée, contextualisée pourrait tirer de grands profits du développement de l'activité d'exploration de données.

Les entreprises commerciales considèrent également l'exploration de données comme un levier marketing potentiel à explorer. Par exemple, M6 publicité a présenté aux membres du comité de pilotage un outil de « *data mining* » permettant d'évaluer et d'améliorer les performances publicitaires²⁴. Par ailleurs, selon certaines informations, l'exploration de données permettrait à Amazon de réaliser 20% de son chiffre d'affaires²⁵.

Dernièrement, les services du Ministère des Finances et Comptes publics, après une délibération de la CNIL²⁶, se sont également dotés d'un outil de « *data mining* » pour lutter contre la fraude fiscale.

L'exploration de données est une activité en plein essor dont la nouveauté et le caractère évolutif soulèvent de nombreuses problématiques notamment en droit de la propriété littéraire et artistique.

²² Etude entreprise par Eefke Smit pour le Publishing Research Consortium (le PRC comprend les syndicats britannique et américain d'éditeurs, les presses universitaires américaines et les associations STM et ALPSP/ Association of Learned and Professional Society Publishers) sur la situation et le potentiel concernant l'exploration de textes et de données : http://www.publishingresearch.net/documents/PRCSmitJAMreport20June2011VersionofRecord_000.pdf

²³ <http://text2genome.smith.man.ac.uk>

²⁴ <http://www.journaldunet.com/solutions/saas-logiciel/m6-publicite-et-datamining-avec-quinten.shtml>

²⁵ <http://www.actualitte.com/usages/donnees-personnelles-le-livre-numerique-mouchard-de-luxe-47205.htm>

²⁶ http://www.legifrance.gouv.fr/affichTexte.do;jsessionid=8CA83991992B4D7235CA3B7390FBCE12.tpdjo14v_2?cidTexte=JORFTEXT000028685085&dateTexte=&oldAction=rechJO&categorieLien=id&idJO=JORFCONT00028683573

2. Les problématiques soulevées par le développement de l'exploration de données en droit de la propriété littéraire et artistique

Les problématiques soulevées par le développement de l'exploration de données concernent plusieurs branches du droit²⁷. Le présent rapport concentre son attention sur celles relatives au droit de la propriété littéraire et artistique qui se présentent lorsque des contenus protégés sont soumis à une opération d'exploration de données.

La question de l'application des règles du code de la propriété intellectuelle aux opérations d'exploration de données ou des évolutions nécessaires du droit dans ce domaine suscite des points de vue différents de la part des utilisateurs²⁸ (2.1) et des ayants droit (2.2).

2.1. Le point de vue des utilisateurs

D'une part, les utilisateurs considèrent que le droit d'auteur ne devrait pas empêcher le développement de l'exploration de données (2.1.1) et d'autre part, ils précisent que la solution contractuelle proposée par certains éditeurs est inadaptée à cette activité (2.1.2) qui nécessite la mise en place d'une exception (2.1.3).

2.1.1. Le développement de l'activité d'exploration de données et le droit d'auteur

Les utilisateurs considèrent que le droit d'auteur ne doit pas être frein au développement de l'activité d'exploration de données pour les raisons suivantes.

- L'exploration de données est un traitement dont l'objet est informationnel, or le droit d'auteur protège la forme d'expression et non les idées

Il est admis que l'objet des opérations d'exploration de données est de fouiller un contenu, ici par définition une œuvre, pour en extraire de l'information, de la connaissance (faits, idées, analyses).

Or, il est unanimement reconnu par la doctrine et la jurisprudence que « *le droit d'auteur ne protège pas les idées, mais seulement la forme dans laquelle celles-ci sont exprimées* »²⁹, dans la mesure où elle est originale³⁰.

²⁷ Données personnelles (informatique et libertés), droit de la consommation, droit de la concurrence

²⁸ Chercheurs, bibliothèques, entreprises commerciales

²⁹ Civ. 1^{ère} 29 novembre 2005, n°04-12.721, Marie-Claire c/Hamel, Bull civ 1. n° 458 p.384

³⁰ Les législations de nombreux pays étrangers insistent sur le fait que la protection accordée par le droit d'auteur ne peut s'étendre aux idées, ce qui n'est pas le cas de la législation française. Cependant, l'article 9.2 de l'accord Adpic, qui est d'application directe en France et l'article 2 du Traité de l'Ompi sur le droit d'auteur du 20 décembre 1996 précisent dans les mêmes termes que : « *La protection du droit d'auteur s'étendra aux expressions et non aux idées, procédures, méthode de fonctionnement ou concepts mathématiques en tant que tels* ».

Les membres du collectif SavoirComm1, désireux de faciliter le développement de l'exploration de données ont souligné cette dichotomie d'objet entre le droit d'auteur et l'exploration de données et que « *l'invocation d'un principe fondamental de la législation sur la propriété intellectuelle pourrait régler en partie la question : la distinction entre information et expression. Par définition, le droit d'auteur ne porte pas sur des informations brutes mais sur leur expression originale* ».

Cette analyse conduit à considérer que l'opération d'exploration d'une ou plusieurs œuvres n'ayant ni pour objet ni pour résultat de restituer la forme d'expression originale, seule à protéger, mais ayant pour objet de générer de l'information sur les œuvres explorées n'est pas tributaire du droit d'auteur.

Les linguistes partagent également cette vision et souhaitent explorer les contenus sans acquérir de licence, au motif que ce n'est pas l'œuvre elle-même qui les intéresse. Un représentant de l'Association Européenne des Ressources relatives à la Langue (ELRA) a signalé au comité de pilotage ainsi qu'au Syndicat national de l'édition (SNE) le souhait des linguistes d'avoir accès, sans en faire l'acquisition, à des œuvres fournies avec leur résumé, à plusieurs traductions d'un même livre (ex. : « *Harry Potter* » dans plusieurs langues), à des dictionnaires ou encore à des éditions bilingues.

Certaines activités d'exploration de données comportent la reprise partielle de la forme d'expression des œuvres explorées (ex : extraits). Dans ces cas l'emprunt effectué entre dans la sphère d'application du droit d'auteur, avec ses obligations et éventuellement le bénéfice de ses exceptions (ex : droit de citation).

- L'approche « fonctionnelle » du droit d'auteur veut que l'exploration de données sur une œuvre ne corresponde pas à l'exploitation que le droit d'auteur a pour fonction de protéger

Certains utilisateurs auditionnés ont souhaité évoquer l'approche « *fonctionnelle* » du droit d'auteur déjà retenue dans plusieurs décisions de justice³¹.

À titre d'exemple, dans le cadre de l'affaire concernant le documentaire « *Être et avoir* », les ayants droit souhaitaient obtenir une rémunération au titre de la reproduction et de la représentation non autorisée d'une méthode pédagogique protégée par le droit d'auteur. La Cour de cassation a rejeté cette demande en soulignant que « *les illustrations ne sont à aucun moment présentées dans leur utilisation par le maître et font corps au décor dont elles constituent un élément habituel, apparaissant par brèves séquences mais n'étant jamais représentées pour elles-mêmes* ».

³¹ Cass 1^{ère} civile, 12 mai 2011, pourvoi n° 08-20651 et Cass 1^{ère} civile, 15 mars 2005.

Cette approche fonctionnelle du droit conduit à considérer que l'utilisation des œuvres protégées par le droit d'auteur à d'autres fins que celles pour lesquelles elles ont été créées et protégées est libre.

Dans cette optique, l'exploration de données, en tant que pure fouille pour extraire de l'information sur l'œuvre, est une opération non seulement étrangère à l'objet même de la protection du droit d'auteur, en raison de l'absence d'emprunt formel, mais également à sa finalité puisque l'utilisation de l'œuvre en tant que forme élaborée protégée est pour le moins accessoire (pour reprendre la démarche jurisprudentielle), voire absente du résultat de la fouille.

Tout comme la critique d'une œuvre n'est pas considérée comme l'exploitation de l'œuvre, les opérations d'exploration de données peuvent également être regardées comme une forme d'analyse critique de l'œuvre et donc ne pas relever du droit d'auteur.

L'approche fonctionnelle conduit à considérer que la protection légale ne s'applique pas hors du périmètre de la fonction pour laquelle elle a été instaurée, ce qui est le cas lorsque la représentation de l'œuvre est accessoire ou que l'œuvre n'est pas utilisée pour elle-même en tant qu'objet de droit. La fonction de protection par le droit d'auteur a donc un objet limité qui ne lui permet pas d'appréhender l'exploration de l'œuvre.

Le processus d'exploration de données sert à extraire du « *sens* » et à créer une nouvelle connaissance. Dans ce cadre, il serait possible de considérer que les œuvres sont explorées pour leur dimension « *informationnelle* » qui devient décisive pour celui qui opère le traitement et non plus en tant que création intellectuelle.

➤ Le droit de lire entraîne le droit d'explorer les données

Comme nous l'avons relevé précédemment, l'utilisation d'outils informatiques et d'algorithmes dans l'activité d'exploration de données permet de distinguer cette activité de l'analyse traditionnelle des données.

Certains soulignent que l'exploration de données constitue « *le prolongement automatisé de pratiques anciennes de traitement de l'information, opérées par les chercheurs (indexation, référencement, repérage des occurrences, etc.)* » 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* »³².

En effet, il est possible de s'interroger sur le point de savoir si l'automatisation et le traitement massif et simultané des données nécessaires à l'exploration de données caractérisent une

³² 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

activité nouvelle ou si la technologie n'introduit finalement qu'une différence de degré dans l'activité traditionnelle d'analyse de données.

A cet égard, le collectif SavoirComm1 a souligné que l'exploration de données marque un changement d'échelle, mais ne fonde pas une activité nouvelle. « *Extraire et synthétiser des informations préexistantes constituent le labeur quotidien du chercheur depuis que la recherche scientifique existe* ».

Le communiqué commun du Consortium Unifié des établissements Universitaires et de recherche pour l'accès aux publications numériques (Couperin) et de l'Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation (ADBU)³³ précise que « *certes, la technologie peut introduire des différences de nature : c'est la base même qui a par exemple présidé en France à la Loi informatique et libertés. Le législateur a considéré que la possibilité de croiser aisément, grâce à l'informatique, des fichiers contenant des données personnelles facilitait l'atteinte aux libertés publiques et justifiait des dispositions législatives contraignantes. Mais dans le cas présent, la technologie n'introduit qu'une différence de degré dans l'activité, non une différence de nature : on montrera que le TDM n'est à l'origine d'aucun préjudice à l'égard des ayants droits. La pratique du TDM ne consiste donc pas pour les chercheurs à exercer un nouveau droit, mais à poursuivre par des moyens technologiques modernes une activité très ancienne, intrinsèquement liée à l'activité de lecture savante. Le TDM n'est rien d'autre qu'une manière de lire et d'exploiter l'information, caractéristique des pratiques de lecture intensive propres au monde académique* ».

Suivant ce raisonnement, l'accès légal aux œuvres qui permet de les lire, de les écouter, de les voir devrait également permettre de les explorer. « *L'extraction du texte et des données est une forme de lecture, qu'elle soit ou non réalisée par un humain ou par une machine. Le droit de lire devrait entraîner le droit d'extraire* ».

2.1.2. Une solution contractuelle inadaptée

A l'heure actuelle, les utilisateurs doivent conclure des accords contractuels qu'ils perçoivent comme un frein au développement des activités d'exploration de données.

- Les négociations contractuelles ralentissent voire découragent les initiatives d'exploration de données

L'exploration de données nécessite le traitement simultané de plusieurs jeux de données issues de plusieurs bases de données ou de plusieurs documents.

Un projet comme Text2genome³⁴ qui impliquait l'exploration de millions d'articles a pris trois ans car l'accès aux données pour l'exploration a été négocié avec chaque éditeur au cas par cas.

³³ http://adbu.fr/wpcontent/uploads/2014/04/Audition_CSPLA_TDM_2014_04_04_final.pdf

³⁴ Voir *supra*

Pour les utilisateurs, la nécessité d'obtenir de multiples autorisations pour les activités d'exploration de données entraîne des négociations très chronophages qui ralentissent les processus voire découragent les initiatives. Ce qui sera autorisé par un titulaire de droits pourra ne pas l'être par un autre. En outre, les utilisateurs peuvent également se trouver confrontés à des incertitudes quant à la titularité des droits sur les contenus.

C'est pourquoi la majorité des représentants des bibliothèques et des instituts de recherche ont souligné l'exploration de données sur des contenus pour lesquels un accès légal a été obtenu ne devrait pas nécessiter de nouvelle licence ni de nouvelle rémunération.

➤ Les négociations contractuelles introduisent des dispositifs pouvant menacer l'indépendance de la recherche

Pour les chercheurs, la voie contractuelle soulève des questions quant à la légitimité des procédures entourant la signature des licences encadrant les activités d'exploration de données.

En effet, certains éditeurs comme Springer souhaitent connaître le projet de recherche et la méthodologie utilisée³⁵. Pour de nombreux chercheurs, cette pratique peut poser des difficultés quant au respect de l'indépendance de la science.

2.1.3. La nécessité d'une exception pour le développement de l'exploration de données

Les utilisateurs considèrent que l'exception est la solution la plus souple pour encourager le développement de l'activité d'exploration de données.

Tout d'abord, ils estiment qu'elle permettrait aux chercheurs français de ne pas être pénalisés par rapport aux chercheurs étrangers qui peuvent bénéficier d'un cadre juridique favorable au développement de l'exploration de données³⁶.

S'agissant des contours de l'exception, alors que certains souhaitent qu'elle soit limitée aux activités de recherche, d'autres soulignent qu'elle devrait être étendue aux usages à finalité commerciale.

Sur le plan technique, les utilisateurs sont soucieux de ne pas voir l'exercice de l'exception empêché par des mesures techniques de protection. En outre, certains ont exprimé la nécessité de mettre en place certains standards techniques pour faciliter l'accès aux contenus.

Par ailleurs, ils font valoir qu'elle ne devrait pas faire l'objet d'une compensation financière au profit des titulaires de droit dans la mesure où l'exploration de données ne constitue pas un préjudice dont pourraient se prévaloir les titulaires de droits pour exiger un paiement³⁷.

³⁵ https://ec.europa.eu/licences-for-europe-dialogue/sites/licences-for-europe-dialogue/files/Publishers-Perspective-Initiatives_0.pdf

³⁶ Au Japon, bientôt en Grande-Bretagne, Irlande. Aux États-Unis, le « *fair use* » semble favorable au développement de l'exploration de données. (voir développements III)

³⁷ <http://www.savoirscom1.info/wp-content/uploads/2014/01/Synthèse-sur-le-statut-légal-du-content-mining.pdf>

2.2. Le point de vue des ayants droit

A titre liminaire, il faut préciser que si les ayants droit souhaitent un développement de l'activité d'exploration de données, l'état de leur réflexion et leurs engagements dans diverses procédures diffèrent largement en fonction des secteurs d'activités. Par exemple, les éditeurs ont souligné vouloir mettre en place des outils juridiques et techniques adaptés (2.2.1) et respectueux de leurs contenus (2.2.2).

2.2.1. La mise en place d'outils juridiques et techniques adaptés au développement de l'exploration de données

Les éditeurs souhaitent gérer les demandes d'exploration de données par la voie contractuelle et sont opposés à la mise en place d'une exception pour l'exploration de données.

➤ La volonté de développer la solution contractuelle

Au cours des auditions, les ayants droit ont fait part, à de nombreuses reprises, de leur souhait de voir l'exploration de données se développer et de leur volonté de proposer des solutions lorsque le droit d'auteur est en jeu.

À l'heure actuelle, les groupes d'édition les plus importants disent recevoir moins de 10 demandes par an pour des usages d'exploration de données.

Sur le plan juridique, lorsque des contenus protégés font l'objet d'une opération d'exploration de données, les ayants droit partent du principe que les différents actes nécessaires (ex : téléchargement, reproduction, structuration, analyse des données, utilisation des résultats, parfois à des fins commerciales) relèvent du droit d'auteur et nécessitent leur autorisation.

La faiblesse de la demande et le fait que certaines activités d'exploration de données peuvent passer rapidement de la sphère de la recherche à la sphère commerciale, voire concerner les deux simultanément, incitent à trouver des solutions souples et non figées qui permettent d'apporter des réponses adaptées aux différentes opérations d'exploration de données.

Les ayants droit prônent la gestion des demandes d'autorisation par contrats (avenant aux licences pour les contenus déjà acquis par les établissements) ou par le biais de la gestion collective, uniquement pour les usages non commerciaux, concernant les contenus auxquels sont abonnés les chercheurs et sans permettre la réutilisation de résultats pouvant se substituer aux contenus eux-mêmes.

Les éditeurs soulignent que les sociétés de gestion collective, en particulier aux Etats-Unis et au Royaume-Uni, réfléchissent également à la possibilité de conclure des conventions avec

des utilisateurs ayant un accès légal à ces œuvres et voulant pratiquer l'exploration de textes et de données.

En outre, les éditeurs techniques se sont engagés à faciliter dans un premier temps l'exploration de textes et de données pour les contenus faisant l'objet d'un abonnement à des fins de recherche non commerciale³⁸.

➤ L'opposition à la mise en place d'une exception pour l'exploration de données

Les ayants droits et en particulier les éditeurs sont opposés à la mise en place d'une exception même à but non lucratif. Ils considèrent qu'elle pourrait faire perdre tout contrôle aux éditeurs sur les contenus téléchargés.

Les éditeurs font valoir que sur le plan pratique, l'exploration de données nécessite souvent qu'ils réalisent une copie de la base d'origine et qu'ils mettent en place un environnement technique spécifique, tel qu'une API (Application Programming Interface) permettant de télécharger le contenu avant de l'explorer, ce qui n'est pas sans coûts, en particulier pour les maisons d'édition. Une exception ne serait d'aucune utilité pour gérer ces questions techniques.

2.2.2. Un développement de l'exploration de données respectueux des contenus éditoriaux

Les éditeurs craignent un développement des pratiques allant au-delà des usages d'exploration de données, par exemple par la mise en ligne incontrôlée des contenus de l'éditeur. Il existe une véritable insécurité sur l'usage réel des données après autorisation.

Par exemple, des utilisateurs mal intentionnés pourraient profiter d'un droit d'extraire les données non encadré pour diffuser une reproduction substantielle de publications ou de bases de données scientifiques.

En outre, les ayants droit sont inquiets des bouleversements du partage de la valeur propres à l'univers numérique qui s'opèrent au détriment de ceux qui prennent le risque de la création. Il serait inéquitable que les acteurs commerciaux de ce secteur génèrent leur propre valeur par l'utilisation des contenus de tiers, sans autorisation ni rémunération de ceux-ci. Ils sont soucieux de prévenir des comportements parasites qui donneraient lieu à un transfert de valeur vers un autre opérateur, notamment à travers la mise au point de nouveaux produits.

Les éditeurs souhaitent mettre en place des outils techniques permettant de faciliter l'exploration de données, tout en préservant l'intégrité de leurs bases de données. Ils sont confrontés à la nécessité de devoir effectuer des investissements importants pour permettre un

³⁸

accès opérationnel aux gisements de données.

Ils souhaitent connaître l'identité des utilisateurs qui ont accès au contenu pour l'exploration de données et savoir exactement quel est le contenu qui fait l'objet de traitement.

Le développement de l'activité d'exploration de données engendre des tensions objectives entre les utilisateurs et les ayants droit qui nous conduisent à engager une réflexion sur l'application du droit de la propriété littéraire et artistique aux actes nécessaires à l'exploration de données.

II. Réflexion sur l'application du droit de la propriété littéraire et artistique aux actes nécessaires à l'exploration de données

Une description sommaire du déroulement d'une opération d'exploration de données (1) nous permettra d'identifier, ensuite, les actes susceptibles d'entrer dans le champ d'application du droit d'auteur et des droits voisins (2) et du droit *sui generis* du producteur de bases de données (3).

1. Le déroulement d'une opération d'exploration de données

Chaque opération d'exploration de données peut impliquer des actes de nature différente et l'accomplissement ou non de ces actes peut dépendre notamment de l'objet de la recherche ou des outils techniques utilisés.

Il est cependant possible d'identifier un certain nombre d'étapes nécessaires à la bonne réalisation d'une opération d'exploration de données qui sont les suivantes :

1.1. La définition des objectifs

Cette première étape consiste d'abord à choisir le sujet et elle peut également servir à planifier le projet, à prévoir l'utilisation opérationnelle des informations extraites et des modèles produits et à spécifier les résultats attendus.

Afin d'éclairer notre propos, nous allons prendre l'exemple simple d'une opération d'exploration de données qui aurait pour objectif d'identifier les noms des villes citées dans un ensemble d'articles scientifiques.

1.2. L'inventaire des contenus numériques utiles

La deuxième étape consiste à recenser les sources de données utiles, fiables, légalement accessibles et techniquement exploitables. Les conditions d'accès ainsi que le statut juridique des contenus numériques qui renferment les données utiles auront un impact direct sur la réalisation de la troisième étape du processus.

Dans notre exemple, les articles scientifiques pourront bénéficier ou non d'une protection par le droit d'auteur ou être intégrés dans une base de données protégée³⁹. Ils pourront également être en libre accès ou non.

1.3. La collecte des contenus numériques

Il ressort des différentes auditions que l'étape de collecte des données implique d'effectuer une copie au moins temporaire de tout ou partie des contenus numériques.

En effet, les outils informatiques mis en œuvre dans l'exploration de données doivent accéder une fois au moins aux contenus numériques afin de procéder à des extractions.

Dans notre exemple, si l'objectif de l'opération d'exploration de données consiste à identifier les noms des villes citées dans un ensemble d'articles scientifiques, les outils informatiques devront procéder au moins une fois à une « lecture » de ces articles. Cette opération de « lecture » ne peut s'effectuer que si les articles en questions sont copiés au moins de manière temporaire sur le même système que celui où fonctionnent les outils informatiques.

1.4. La préparation des données

Une fois que l'inventaire et la collecte des contenus numériques sont opérés, il peut être nécessaire de transformer temporairement les données afin qu'elles puissent être exploitables par les outils informatiques.

Dans notre exemple, si les articles scientifiques sont au format PDF, il pourra être nécessaire de les faire passer au format XML.

La préparation des données pourra consister également dans certains cas à créer une base de données qui servira ensuite à l'analyse.

1.5. Le traitement des données

Cette étape consiste à appliquer les outils informatiques et les algorithmes utiles sur les données. Les outils informatiques et algorithmes font appel à trois grandes disciplines à savoir les mathématiques, la linguistique et la visualisation⁴⁰.

³⁹ Protection par le droit d'auteur et/ou par le droit *sui generis* des producteurs de base de données.

Dans notre exemple, il s'agira d'appliquer un logiciel d'analyse de texte afin d'extraire automatiquement l'ensemble des noms des villes citées dans les articles. Ce logiciel qui porte également le nom d'extracteur de connaissance contient au cœur de son programme des lexiques de noms de villes ainsi que des modèles informatiques⁴¹ lui permettant par exemple d'identifier un nom de ville ne faisant pas partie de ses lexiques et de désambigüiser le nom de ville lorsque qu'il est en doublon (Paris-France et Paris-Texas-USA).

1.6. Le résultat de l'exploration des données

Le résultat de l'exploration de données est une nouvelle connaissance. Dans la plupart des cas, et même si elle est le résultat du processus d'exploration des données, cette nouvelle connaissance ne contiendra pas les données extraites.

Dans notre exemple, le résultat pourra être une infographie ou un histogramme indiquant en abscisse le nom des villes et en ordonnée le nombre de fois où elles sont citées dans l'ensemble des articles. Cette nouvelle connaissance a bien été produite à partir de la lecture des articles, néanmoins le contenu de ces derniers n'est pas nécessaire à la compréhension.

Par conséquent, l'objet principal de l'exploration de données n'est pas la restitution des contenus numériques initiaux. Cependant, si les contenus initiaux sont protégés par la propriété littéraire et artistique, certains actes nécessaires au bon déroulement de l'exploration de données sont susceptibles d'entrer dans le champ d'application du droit d'auteur et des droits voisins et du droit *sui generis* du producteur de bases de données.

2. Les actes susceptibles d'entrer dans le champ d'application du droit d'auteur

L'article L. 111-1 du code de la propriété intellectuelle dispose que « *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. Ce droit comporte des attributs d'ordre intellectuel et moral ainsi que des attributs d'ordre patrimonial* ».

L'article L.112-3 du code de la propriété intellectuelle précise que « *les auteurs de traductions, d'adaptations, transformations ou arrangements des œuvres de l'esprit jouissent de la protection instituée par le présent code sans préjudice des droits de l'auteur de l'œuvre*

⁴⁰ Les mathématiques, modélisées sous une forme algorithmique, sont utilisées pour tout ce qui relève des statistiques simples, de l'analyse des données, de la classification automatique des objets, de profilage et enfin d'analyse prédictive. La linguistique, dont une partie de la transcription informatique, connue sous le vocable TAL (Traitement Automatique des Langues) constitue les bataillons lourds des outils informatiques du TDM à travers des expressions diverses telle que les moteurs de recherche, l'analyse sémantique, les ontologies, l'extraction d'information. Enfin la visualisation qui permet de gommer d'une certaine façon les complexités des modèles ainsi que des algorithmes via des représentations de cartographie géographique, d'histogrammes et autres camemberts, des réseaux de connaissance exprimant la proximité entre des représentations conceptuelles de la connaissance comme des graphes.

⁴¹ Règles d'extraction ou pattern

originale. Il en est de même des auteurs d'anthologies ou de recueils d'œuvres ou de données diverses, tels que les bases de données, qui, par le choix ou la disposition des matières, constituent des créations intellectuelles. On entend par base de données un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique, et individuellement accessibles par des moyens électroniques ou par tout autre moyen ».

Par conséquent, lorsque des œuvres protégées par le droit d'auteur sont soumises à une opération d'exploration de données, il faut s'interroger sur la mise en jeu des droits patrimoniaux (2.1) et du droit moral (2.1).

2.1. L'exploration de données et la mise en jeu des droits patrimoniaux

Il est nécessaire d'identifier si les actes nécessaires à l'exploration de données mettent en jeu le droit d'exploitation (2.1) appartenant à l'auteur ou à ses ayants droit et si, le cas échéant, ils pourraient bénéficier d'une exception au droit d'auteur (2.2) dispensant de toute obligation d'autorisation préalable.

2.1.1. La mise en jeu du droit d'exploitation

L'article L.122-1 du code de la propriété intellectuelle dispose que « *le droit d'exploitation appartenant à l'auteur comprend le droit de représentation et le droit de reproduction* ».

Il convient donc d'aborder successivement la question de la mise en jeu, dans le cadre d'une opération d'exploration de données, du droit de représentation (2.1.1.1) et du droit de reproduction (2.1.2.2)

2.1.1.1. La mise en jeu du droit de représentation

L'article L 122-2 du code de la propriété intellectuelle dispose que « *la représentation consiste dans la communication de l'œuvre au public par un procédé quelconque, et notamment : 1° Par récitation publique, exécution lyrique, représentation dramatique, présentation publique, projection publique et transmission dans un lieu public de l'œuvre télédiffusée ; 2° Par télédiffusion. La télédiffusion s'entend de la diffusion par tout procédé de télécommunication de sons, d'images, de documents, de données et de messages de toute nature. Est assimilée à une représentation l'émission d'une œuvre vers un satellite* ».

Comme nous l'avons évoqué précédemment, la finalité de l'exploration de données est d'extraire du sens, de l'information et de nouvelles connaissances – et non de communiquer au public l'œuvre explorée en totalité ou en partie.

L'objet de l'exploration est de créer une nouvelle connaissance. Par ailleurs, dans la plupart des cas⁴², le résultat de l'exploration de données, ne contiendra pas les contenus objets du traitement.

Par conséquent, aucun des actes nécessaires à l'exploration de données ne nécessite la communication de l'œuvre au « *public* ». Il peut en aller différemment du droit de reproduction.

2.1.1.2. La mise en jeu du droit de reproduction

En principe, les actes de lecture d'un texte, ainsi que l'exploitation intellectuelle de son contenu (analyse, réflexion, mise en relation) ne sont pas soumis aux droits exclusifs des titulaires de droit. Sauf que la description d'une opération d'exploration de données montre que l'utilisation d'outils informatiques conduit à effectuer, en général, un acte de copie au moins provisoire de tout ou partie des contenus numériques objets du traitement pour effectuer cette lecture informatique.

La collecte des contenus numériques (étape 3) nécessite des actes de reproduction préalables au traitement et à l'exploration de données (étape 5). Dans la plupart des cas, les impératifs techniques inhérents à l'univers numérique conduiront à la réalisation d'une copie au moins temporaire de toute ou partie des contenus numériques sur le système où fonctionnent les outils informatiques qui permettent l'exploration de données. En outre, la préparation des données (étape 4) peut nécessiter un changement de format qui implique une copie préalable des contenus numériques.

La lecture de la loi en vigueur au Japon⁴³ et des projets législatifs en cours sur le sujet en Grande-Bretagne⁴⁴ et Irlande⁴⁵ qui prévoient une exception au droit d'auteur pour les activités

⁴² Il pourra contenir dans certains cas très limités des extraits de l'œuvre. S'il s'agit de la reprise partielle de la forme d'expression des œuvres explorées (extraits), l'emprunt effectué entre dans la sphère d'application du droit d'auteur, avec ses obligations et éventuellement le bénéfice de ses exceptions (ex : droit de citation).

⁴³ Article 47 septies « *For the purpose of information analysis ("information analysis" means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other much information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use by a person who makes an information analysis* ». <http://www.cric.or.jp/english/clj/cl2.html>

⁴⁴ Projet de l'article 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)* ». <http://www.legislation.gov.uk/ukdsi/2014/978011112755>

⁴⁵ Projet d'article 106F. *Digital analysis and research*. (1) It is not an infringement of the rights conferred by this Act for a person to reproduce a work for a purpose to which this section applies if (...) This section applies to (a)

d'exploration de données confirme que leurs rédacteurs considèrent que l'acte de copie⁴⁶ des contenus numériques est nécessaire à ces activités.

Selon l'article L.122-3 du code de la propriété intellectuelle « *la reproduction consiste dans la fixation matérielle de l'œuvre par tous les 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. Pour les œuvres d'architecture, la reproduction consiste également dans l'exécution répétée d'un plan ou d'un projet type* ». L'article 2 de la directive 2001/29/CE⁴⁷ précise que la reproduction peut être « *provisoire ou permanente* » et peut concerner « *tout ou partie* » de l'œuvre.

La Cour de Cassation a pu rappeler qu'il y a reproduction de l'œuvre dès lors qu'il y a communication au public des traits caractéristiques originaux de la création⁴⁸ mais la définition du droit de reproduction⁴⁹ continue d'être interprétée très largement c'est-à-dire que tout acte de reproduction d'une œuvre ou d'un objet protégé doit faire l'objet d'une autorisation préalable.

Dans l'univers numérique, cette interprétation large du droit de reproduction est souvent remise en cause⁵⁰ mais en application de la directive 2001/29/CE et des règles du code de la propriété intellectuelle, si l'exploration de données porte sur des œuvres protégées par le droit d'auteur et que la technique employée nécessite d'effectuer une copie de toute ou partie du contenu objet du traitement, l'acte de copie sera qualifié d'acte de reproduction.

text-mining, data-mining, and similar analysis or research.
http://www.djei.ie/science/ipr/crc_consultation_paper.pdf

⁴⁶ Utilisation des termes « recording », « copy » et « reproduce »

⁴⁷ « Les États membres prévoient le droit exclusif d'autoriser ou d'interdire la reproduction directe ou indirecte, provisoire ou permanente, par quelque moyen et sous quelque forme que ce soit, en tout ou en partie: a) pour les auteurs, de leurs œuvres; b) pour les artistes interprètes ou exécutants, des fixations de leurs exécutions; c) pour les producteurs de phonogrammes, de leurs phonogrammes; d) pour les producteurs des premières fixations de films, de l'original et de copies de leurs films; e) pour les organismes de radiodiffusion, des fixations de leurs émissions, qu'elles soient diffusées par fil ou sans fil, y compris par câble ou par satellite ».

⁴⁸ Civ.1^{ère}, 1 juill. 1987

⁴⁹ Tant dans code de la propriété intellectuelle français que dans la directive 2001/29/CE.

⁵⁰ Les membres du collectif Savoircomm1 défendent l'idée suivant laquelle dans l'univers numérique il n'y a pas forcément prise au droit d'auteur à chaque fois qu'une copie est réalisée pour des raisons techniques. Ils citent les propos de Michel Vivant qui lors de son audition dans le cadre de la mission Lescure a abordé la question de la définition des droits et en particulier celle de la reproduction, avec le sujet des copies techniques. « Il y a un piège : Dès que l'on aborde des questions qui sont marquées par la technicité, on veut avoir 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 termes de régulation sociale ? ». Dans la même logique, le rapport « Standardisation in the area of innovation and technological development, notably in the field of Text and Data mining », propose notamment de revoir la définition et l'interprétation du droit de reproduction en droit de l'Union européenne. Il s'agirait d'avoir une interprétation plus restrictive du droit de reproduction qui n'aurait vocation à jouer que lorsque l'acte a pour objet restituer la forme d'expression originale de l'œuvre. http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf

L'article L 122-4 du code de la propriété intellectuelle précise que « *toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droits ou ayants cause est illicite. Il en est de même pour la traduction, l'adaptation, ou la transformation, l'arrangement ou la reproduction par un art ou un procédé quelconque* ».

Par conséquent, toute reproduction de tout ou partie d'une œuvre dans le cadre d'une opération d'exploration de données sans autorisation préalable des auteurs constitue une contrefaçon. Dès lors, il faut se poser la question de savoir si les actes nécessaires à l'exploration de données peuvent bénéficier d'une exception au droit d'auteur.

2.1.2. La question du bénéfice d'une exception

La liste des exceptions au monopole de l'auteur est prévue par l'article L.122-5 du code de la propriété intellectuelle. Les exceptions sont d'interprétation stricte ce qui signifie, en droit d'auteur, qu'elles doivent toujours s'interpréter en faveur du principe qui est le monopole.

L'application de ce principe d'interprétation conduit à écarter, dans la plupart des cas et en particulier, l'application de l'exception de copie transitoire (2.1.2.1) de l'exception de citation (2.1.2.2) et pour les activités de recherche, de l'exception prévue à l'article L.122-5 3° e) (2.1.2.3). Enfin, nous évoquerons, pour le cas spécifique des bases de données l'exception prévue à l'article L.122-5 5° du code de la propriété intellectuelle (2.1.2.4).

2.1.2.1. L'application de l'exception de copie provisoire transitoire

La numérisation des œuvres entraîne un certain nombre de reproductions qui sont techniquement nécessaires pour permettre leur diffusion sur les réseaux. Afin de prendre en considération certains impératifs techniques⁵¹, l'article 5 paragraphe 1 de la directive 2001/29/CE⁵² a imposé aux États membres une exception, transposée en droit français à l'article L. 122-5 alinéa 6 du Code de la propriété intellectuelle. Avant d'envisager l'application de l'exception aux actes de reproduction nécessaires à l'exploration de données, il est nécessaire de rappeler ses conditions d'application.

➤ Les conditions d'application de l'exception

L'article L.122-5 alinéa 6 du code de la propriété intellectuelle dispose que, lorsque l'œuvre a été divulguée, l'auteur ne peut interdire « *la reproduction provisoire présentant un caractère*

⁵¹ Cette exception a pour objectif d'assurer le développement et fonctionnement des nouvelles technologies, ainsi que de maintenir un juste équilibre entre les droits et les intérêts de titulaires de droits, d'une part, et d'utilisateurs d'œuvres protégées qui souhaitent bénéficier de ces nouvelles technologies, d'autre part.

⁵² « Les actes de reproduction provisoires visés à l'article 2, qui sont transitoires ou accessoires et constituent une partie intégrante et essentielle d'un procédé technique et dont l'unique finalité est de permettre: a) une transmission dans un réseau entre tiers par un intermédiaire, ou b) une utilisation licite d'une œuvre ou d'un objet protégé, et qui n'ont pas de signification économique indépendante, sont exemptés du droit de reproduction prévu à l'article 2 ».

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'œuvre ou sa transmission entre tiers par la voie d'un réseau faisant appel à un intermédiaire ; toutefois, cette reproduction provisoire qui ne peut porter que sur des œuvres autres que les logiciels et les bases de données ne doit pas avoir de valeur économique propre ».

Une disposition similaire existe pour les droits voisins⁵³. En droit français, cette exception est restreinte par l'exclusion expresse des logiciels et des bases de données de son champ d'application.

Cette exclusion, qui n'est pas prévue par la directive, empêche de faire jouer l'exception dans de nombreux cas. En effet, il est impossible d'appliquer cette exception d'une part, aux œuvres présentes sur les réseaux qui ont une dimension logicielle et d'autre part, aux bases de données qui pourraient faire l'objet d'exploration de données.

Le considérant 33 de la directive 2001/29/CE éclaire l'interprétation de l'article 5 paragraphe 1 de la directive 2001/29/CE en précisant que *« le droit exclusif de reproduction doit faire l'objet d'une exception destinée à autoriser certains actes de reproduction provisoires, qui sont transitoires ou accessoires, qui font partie intégrante et essentielle d'un processus technique et qui sont exécutés dans le seul but de permettre soit une transmission efficace dans un réseau entre tiers par un intermédiaire, soit une utilisation licite d'une œuvre ou d'un autre objet protégé. Les actes de reproduction concernés ne devraient avoir par eux-mêmes aucune valeur économique propre. Pour autant qu'ils remplissent ces conditions, cette exception couvre les actes qui permettent le survol (browsing), ainsi que les actes de prélecture dans un support rapide (caching), y compris ceux qui permettent le fonctionnement efficace des systèmes de transmission, sous réserve que l'intermédiaire ne modifie pas l'information et n'entrave pas l'utilisation licite de la technologie, largement reconnue et utilisée par l'industrie, dans le but d'obtenir des données sur l'utilisation de l'information. Une utilisation est réputée être licite lorsqu'elle est autorisée par le titulaire du droit ou n'est pas limitée par la loi ».*

La Cour de justice a eu l'occasion d'interpréter les dispositions de l'article 5 paragraphe 1 de la directive du 22 mai 2001 dans les arrêts Infopaq I⁵⁴ et Infopaq II⁵⁵. La CJUE a répondu que *« l'article 5, paragraphe 1, de la directive 2001/29/CE, doit être interprété en ce sens que les actes de reproduction provisoires effectués au cours d'un procédé dit «d'acquisition de données», tels que ceux en cause au principal, satisfont à la condition selon laquelle ces actes doivent constituer une partie intégrante et essentielle d'un procédé technique, nonobstant le fait qu'ils introduisent et clôturent ce procédé et impliquent une intervention humaine ; sont conformes à la condition selon laquelle les actes de reproduction doivent poursuivre une finalité unique, à savoir de permettre une utilisation licite d'une œuvre ou d'un objet protégé ; satisfont à la condition selon laquelle ces actes ne doivent pas avoir une signification économique indépendante pour autant, d'une part, que la mise en œuvre de ces*

⁵³ Article L.211-3 alinéa 5 du Code de la propriété intellectuelle

⁵⁴ CJUE, 16 juillet 2009, Infopaq I C-5/08

⁵⁵ CJUE, 17 janvier 2012, Infopaq II C-302/10

actes ne permette pas de réaliser un bénéfice supplémentaire, allant au-delà de celui tiré de l'utilisation licite de l'œuvre protégée et que, d'autre part, les actes de reproduction provisoires n'aboutissent pas à une modification de l'œuvre »⁵⁶.

La Cour de justice a précisé que les conditions posées sont cumulatives et qu'elles sont d'interprétation stricte dans la mesure où elles dérogent à un principe général établi par cette même directive⁵⁷. Elles sont les suivantes :

➤ La reproduction provisoire doit être transitoire ou accessoire

Le caractère transitoire de la copie impose que *«sa durée de vie est limitée à ce qui est nécessaire pour le bon fonctionnement du procédé technique concerné, étant entendu que ce procédé doit être automatisé de sorte qu'il supprime cet acte d'une manière automatique, sans intervention humaine, dès que sa fonction visant à permettre la réalisation d'un tel procédé est achevée. »⁵⁸*

➤ La reproduction ne doit pas avoir une signification économique indépendante

La Cour de Justice précise que les actes ne doivent pas avoir une signification économique indépendante, en ce sens que *« l'avantage économique tiré de leur mise en œuvre ne doit être ni distinct ni séparable de l'avantage économique tiré de l'utilisation licite de l'œuvre concernée et il ne doit pas générer un avantage économique supplémentaire, allant au-delà de celui tiré de ladite utilisation de l'œuvre protégée »⁵⁹.*

« Un avantage tiré d'un acte de reproduction provisoire est distinct et séparable si l'auteur de cet acte est susceptible de réaliser des bénéfices en raison de l'exploitation économique de reproductions provisoires elles-mêmes. Il en va de même si les actes de reproduction provisoires aboutissent à une modification de l'objet reproduit, tel qu'il existe au moment du déclenchement du procédé technique concerné, car lesdits actes visent ainsi à faciliter non plus son utilisation, mais l'utilisation d'un objet différent »⁶⁰.

Les actes de reproduction provisoires effectués au cours d'un procédé d'acquisition de données satisfont à la condition selon laquelle ces actes ne doivent pas avoir une signification économique indépendante pour autant, d'une part, que la mise en œuvre de ces actes ne

⁵⁶ Rapport du CSPLA, « *Référencement des œuvres sur internet* », Valérie-Laure Benabou, Joëlle Farchy, Cécile Méadel.

⁵⁷ CJCE, 4e ch., 16 juill. 2009, aff. C-5/08, Infopaq International A/S c/ Danske Dagblades Forening, préc. n° 13 – points 55 et 56

⁵⁸ CJUE, 16 juillet 2009, Infopaq I C-5/08 – point 64

⁵⁹ CJUE, 17 janvier 2012, Infopaq II C-302/10 – point 50

⁶⁰ CJUE, 17 janvier 2012, Infopaq II C-302/10 – point 52

permette pas de réaliser un bénéfice supplémentaire, allant au-delà de celui tiré de l'utilisation licite de l'œuvre protégée et que, d'autre part, les actes de reproduction provisoires n'aboutissent pas à une modification de cette œuvre.

- La reproduction doit être un élément essentiel d'un procédé technique dans laquelle elle est intégrée

Selon la Cour de justice⁶¹, cette condition requiert que « *les actes de reproduction provisoires soient entièrement effectués dans le cadre de la mise en œuvre d'un procédé technique et qu'ils ne soient donc pas réalisés, totalement ou partiellement, en dehors d'un tel procédé* ». Par ailleurs, « *les actes de reproduction ne sauraient excéder ce qui est nécessaire pour le bon fonctionnement du procédé technique en question* ». En outre, « *rien dans cette disposition n'indique que le procédé technique doive n'impliquer aucune intervention humaine* ».

- La reproduction doit avoir pour finalité exclusive une utilisation licite de l'œuvre

Comme le souligne le considérant 33 de la directive du 22 mai 2001, une utilisation est réputée être licite lorsqu'elle est autorisée par le titulaire du droit ou n'est pas limitée par la loi.

- Application des conditions de l'exception aux actes de reproduction nécessaires à l'exploration de données

La première condition « *une reproduction provisoire transitoire ou accessoire* » semble ne pas pouvoir être toujours remplie dans le cadre d'une opération d'exploration de données.

En effet, dans certains cas, l'étape de collecte des contenus peut nécessiter une opération de transmission de contenus protégés (données, images, sons, vidéo) d'un serveur vers le système où fonctionnent les outils informatiques qui permettent l'exploration de données (téléchargement, récupération du contenu grâce à une API⁶²). Pour éviter cela, il serait possible sur le plan technique d'envisager un accès direct aux données qui ne nécessiterait pas de téléchargement du contenu protégé. Par ailleurs, la préparation des données peut impliquer de transformer les données afin qu'elles puissent être exploitables par les outils informatiques d'exploration de données. Par exemple, le changement de format par exemple implique une copie préalable du contenu protégé. En outre, l'étape de traitement des données qui consiste à appliquer les outils informatiques et les algorithmes utiles sur les données peut nécessiter en fonction des logiciels utilisés des copies temporaires des contenus. Enfin, dans des cas très restreints, le résultat de l'exploration de données pourra contenir des extraits des œuvres protégées objets du traitement.

⁶¹ CJUE, 17 janvier 2012, Infopaq II C-302/10 – points 30, 32.

⁶² En informatique, une interface de programmation (souvent désignée par le terme API pour *Application Programming Interface*) est un ensemble normalisé de classes, de méthodes ou de fonctions qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels.

De nombreuses copies peuvent être nécessaires et si l'une de ces copies possède un caractère permanent il est impossible de faire jouer l'exception prévue à l'article L. 122-5 alinéa 6 du Code de la propriété intellectuelle⁶³.

Le caractère transitoire de la copie peut faire défaut dans de nombreux cas. En effet, la suppression de la reproduction nécessaire par exemple pour la collecte des données, dépendra de la seule volonté de l'utilisateur, dont il n'est nullement certain qu'il veuille s'en défaire, ce qui a pour conséquence que la dite reproduction risque de subsister pendant une période prolongée en fonction des besoins de l'utilisateur.

Dans l'arrêt Infopaq I, cet élément a été pris en considération par la Cour de justice pour constater dans l'affaire en cause que « *le dernier acte du procédé d'acquisition de données, au cours duquel Infopaq imprime les extraits composés de onze mots, ne constitue pas un acte transitoire au sens de l'article 5, paragraphe 1, de la directive 2001/29* ».

La deuxième condition « *une reproduction sans signification économique indépendante* » semble ne pas pouvoir être remplie dans de nombreux cas.

En effet, le résultat issu du processus d'exploration de données crée une valeur spécifique ajoutée à celle des données premières grâce, par exemple, à la détection des corrélations entre ces données.

Par conséquent, l'application de l'exception de copie provisoire transitoire dans le cadre de l'exploration sera très difficile. Elle n'apportera pas la sécurité juridique nécessaire au développement des activités d'exploration de données.

2.1.2.2. L'application de l'exception de citation

Comme nous l'avons évoqué précédemment, dans des cas très restreints, le résultat de l'exploration de données pourra contenir des extraits des œuvres protégées objets du traitement. Dès lors, il faut s'interroger sur le point de savoir si ces extraits relèvent du droit de reproduction ou peuvent bénéficier de l'application d'une exception de citation.

Dans l'arrêt *Infopaq I*, la Cour de justice a rappelé que la reprise d'un extrait d'une œuvre protégée, constitue une reproduction partielle, si un tel extrait contient un élément de l'œuvre qui, en tant que tel, exprime la création intellectuelle propre à l'auteur⁶⁴. Si tel est le cas, la reproduction partielle nécessite l'autorisation préalable des ayants droit sauf à bénéficier d'une exception de citation.

L'article L.122-5 du code de la propriété intellectuelle précise que la citation doit s'inscrire dans une œuvre citante, elle doit être courte, à la fois au regard de l'œuvre citée et de l'œuvre citante, elle doit mentionner la source dont elle émane et le nom de l'auteur et doit se faire après que l'œuvre a été divulguée. Enfin, elle doit s'inscrire dans une finalité particulière, à

⁶³ Par exemple : le téléchargement initial des données dans l'étape de collecte des contenus

⁶⁴ CJUE, 16 juillet 2009, Infopaq I C-5/08 – point 48

savoir être justifiée par le caractère critique, polémique, pédagogique, scientifique ou d'information de l'œuvre à laquelle elle est incorporée. L'article L.122-5 du Code de la propriété intellectuelle évoque l'œuvre à laquelle est incorporée la citation. Il faut donc normalement une œuvre citante qui accueille l'œuvre citée.

La Cour de cassation dans une célèbre affaire *Microfor* en a, par deux fois, décidé autrement. Selon l'arrêt d'Assemblée plénière du 30 octobre 1987, « *dès lors que l'œuvre a été divulguée, l'auteur ne peut interdire, sous réserve que soient indiqués clairement le nom de l'auteur et la source, les courtes citations justifiées par le caractère d'information de l'œuvre à laquelle elles sont incorporées, c'est en violation de l'article L.122-5 du Code de la propriété intellectuelle qu'un arrêt a, au motif qu'ils n'étaient pas incorporés dans une œuvre décidé que les résumés signalétiques insérés dans l'index des articles d'un journal publié par un tiers ne pouvaient tenir lieu de courtes citations permises sans le consentement de l'auteur, alors que les résumés constitués uniquement de courtes citations de l'œuvre ne dispensant pas le lecteur de recourir à celles-ci étaient indissociables de la section analytique de la publication par le jeu des renvois figurant dans cette section et que cet ensemble avait le caractère d'une œuvre d'information* ». Cette jurisprudence a par ailleurs été confirmée par la Cour de justice⁶⁵.

L'utilisation du terme « *œuvre d'information* » permet de s'interroger sur l'application de cette jurisprudence aux activités d'exploration de données. En tout état de cause, il semble que cette décision soit trop isolée pour constituer un fondement juridique offrant une sécurité juridique suffisante pour justifier l'application de l'exception de citation dans le cadre d'une activité d'exploration de données.

2.1.2.3. L'application de l'exception de l'article L.122-5 3° e) du code de la propriété intellectuelle

L'article L.122-5 3° e) dispose que, sous réserve que soient indiqués clairement le nom de l'auteur et la source, l'auteur ne peut interdire « *la représentation ou la reproduction d'extraits d'œuvres, sous réserve des œuvres conçues à des fins pédagogiques, des partitions de musique et des œuvres 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* ».

Il s'agit là de la transposition de l'article 5.3 a) de la directive du 22 mai 2001 mais l'exception ainsi créée est précisément circonscrite. En effet, elle est limitée à l'utilisation

⁶⁵ CJUE, 1^{er} décembre 2001.

d'extraits d'œuvres et ne joue pas pour certaines créations. On ne peut revendiquer son bénéfice qu'à « *des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche* » et non à des fins d'analyse en visant un public « *composé majoritairement d'élèves, d'étudiants, d'enseignants et de chercheurs directement concernés* ». Elle ne peut autoriser une « *exploitation commerciale* » et doit être compensée sous la forme d'une rémunération.

Par conséquent, il semble que cette exception en raison de son caractère très limité ne pourra pas être invoquée par les chercheurs dans le cadre des activités d'exploration de données.

2.1.2.4. L'application de l'exception de l'article L.122-5, 5° du code de la propriété intellectuelle

Les bases de données originales peuvent par le choix des matières et/ou leur présentation originale bénéficier de la protection par le droit d'auteur (article L.112-3 du code de la propriété intellectuelle). Elles peuvent recevoir une protection supplémentaire ou alternative par un droit *sui generis*⁶⁶.

Tous les contrats d'accès à des bases de données contiennent des clauses relatives à la propriété intellectuelle des bases de données, ainsi qu'à l'utilisation qui peut être faite des données ou des résultats obtenus après leur interrogation.

L'article L.122-5 5° du code de la propriété intellectuelle autorise « *les actes nécessaires à l'accès au contenu d'une base de données électronique pour les besoins et dans les limites de l'utilisation prévue par contrat* ».

Cette disposition ne sera utile que si l'exploration de données est expressément prévue au contrat. Dans ce cas, l'ayant droit ne pourra interdire les actes que l'utilisateur légitime aura accomplis en raison de la licence obtenue.

2.2. L'exploration de données et la mise en jeu du droit moral

Même s'il a peu été évoqué lors des auditions, le droit moral⁶⁷ peut être perçu par les utilisateurs comme un frein au développement de l'activité d'exploration de données. En effet, même lorsque l'œuvre est publiée et que les droits patrimoniaux sont cédés à un exploitant, l'auteur conserve des prérogatives qui ont un caractère perpétuel, inaliénable et imprescriptible. Il s'agit du droit de divulgation (2.2.1), du droit de retrait et de repentir (2.2.2), du droit à la paternité (2.2.3) et du droit au respect et à l'intégrité de l'œuvre (2.2.4).

⁶⁶ Article L. 341-1 et suivants du code de la propriété intellectuelle

⁶⁷ Le droit moral réunit l'ensemble des prérogatives extrapatrimoniales qui permettent à l'auteur de défendre sa personnalité telle qu'elle s'exprime dans l'œuvre qu'il a créée, il est souvent perçu comme le symbole du droit d'auteur à la française.

2.2.1. Le droit de divulgation

La divulgation est le don de l'œuvre au public⁶⁸. C'est parce que l'œuvre a été préalablement divulguée qu'elle va pouvoir être exploitée. Une activité d'exploration de données ne saurait avoir lieu sur une œuvre qui n'aurait pas été préalablement divulguée.

2.2.2. Le droit de retrait et de repentir

Sur les réseaux, la mise en œuvre du droit de retrait et de repentir peut s'avérer problématique. Il peut être techniquement difficile de retirer définitivement une œuvre car l'œuvre diffusée a été généralement copiée et elle est généralement toujours accessible grâce aux copies-caches effectuées par les moteurs de recherches. L'activité d'exploration de données n'a pas, en soi, d'impact sur cette problématique.

2.2.3. Le droit à la paternité

La question du droit à la paternité de l'auteur sur son œuvre ne se pose que si l'utilisateur reprend tout ou partie de l'œuvre. Ce ne sera généralement pas le cas.

2.2.4. Le droit au respect et à l'intégrité de l'œuvre

En raison de sa qualité, l'auteur peut s'opposer à toute déformation, démantèlement, mutilation ou autre modification de son œuvre. Certains pourraient considérer que les actes nécessaires à l'exploration de données constituent dans certains cas une dénaturation de l'œuvre.

3. Les actes d'exploitation susceptibles d'entrer dans le champ d'application du droit *sui generis* des producteurs de bases de données

Dans la société dite « *de l'information* », les bases de données ont par essence une très grande importance et une grande valeur⁶⁹. Les bases de données peuvent être protégées par le droit d'auteur mais également par le droit *sui generis* du producteur de bases de données.

⁶⁸ La doctrine s'oppose de longue date sur la question de l'épuisement du droit de divulgation de l'auteur. Le Conseil constitutionnel, dans la décision n° 2013-370 QPC du 28 février 2014, a précisé que le droit de divulgation, selon la jurisprudence de la Cour de cassation, s'épuise par le premier usage qu'il en fait.

⁶⁹

Cette protection peut avoir un impact très important sur le développement des activités d'exploration de données, dans la mesure où le producteur qui peut prouver un investissement, bénéficie d'un monopole sur la base de données qui peut être perpétuel et lui permet de s'opposer à plusieurs extractions et réutilisations.

Il est nécessaire d'identifier si les actes nécessaires à l'exploration de données mettent en jeu le monopole du producteur de bases de données (3.1) et si, le cas échéant, ils pourraient bénéficier d'une exception (3.2).

3.1. La mise en jeu du monopole du producteur de bases de données

L'article L.341-1 du code de la propriété intellectuelle précise que c'est le « *contenu de la base* » qui est protégé, y compris si ce dernier est constitué de données insusceptibles d'être protégées par un droit de propriété intellectuelle. Dans l'hypothèse de données protégées en tant qu'œuvre, elles ne pourront pas être extraites, même une fois tombées dans le domaine public.

Pour être protégée, la base de données⁷⁰ doit être le résultat d'un « *investissement financier, matériel ou humain substantiel* ». Dans ce cas, le monopole accordé permet au producteur de la base de données d'interdire certaines extractions du contenu de sa base.

Selon les dispositions de l'article L.342-1 du code de la propriété intellectuelle, le producteur de bases de données a le droit d'interdire :

- « *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* » ;
- « *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* ».

La jurisprudence a explicité deux critères : « *l'extraction quantitative* » doit être appréciée au regard des données extraites et par rapport au contenu total de la base⁷¹, sans que le volume de la base qui accueille les données extraites soit pris en considération⁷². « *L'extraction qualitative* » doit, quant à elle, être appréciée au regard de l'investissement réalisé pour l'obtention et la vérification des données en cause⁷³.

⁷⁰ La base de données est définie à l'article L. 112-3, alinéa 2 du Code de la propriété intellectuelle comme « *un recueil d'œuvres, de données ou d'autres éléments indépendants, disposés de manière systématique ou méthodique, et individuellement accessibles par des moyens électroniques ou par tout autre moyen* ».

⁷¹ CJUE, 9 novembre 2004, *The British Horseracing Board Ltd e.a.*, affaire C-203/02 : Comm.com.electr.2005, note Ch.Caron ; D.2005, somm. p.1495, obs. Sirinelli.

⁷² CJUE, 5 mars 2009, *Apis-Hristovich EOOD c/ Lakorda AD*, affaire C-545/07 : Comm.com.electr. 2009, comm.44, note Ch Caron.

⁷³ CJUE, 5 mars 2009 : Comm.com.electr. 2009, comm.44, note Ch Caron.

L'article L. 342-2 du Code de la propriété intellectuelle interdit également les extractions répétées qui ne sont ni qualitativement ni quantitativement substantielles mais qui « *excèdent manifestement les conditions d'utilisation normale de la base de données* ». En outre, le producteur de la base de données peut s'opposer à ces extractions même si elles ne sont pas effectuées à partir de la base elle-même mais d'une copie de cette dernière⁷⁴.

Dans le cadre d'une opération d'exploration de données, la collecte des contenus numériques (étape 3) nécessite des actes de reproduction préalables au traitement et à l'exploration de données (étape 5). Si les utilisateurs souhaitent explorer des contenus présents dans des bases de données, il leur faudra procéder à des extractions qualitativement ou quantitativement substantielles. Une même opération d'exploration de données pourra nécessiter l'interrogation de plusieurs bases de données en simultanée.

Par conséquent, il sera possible de soumettre au jeu du droit *sui generis*, les opérations d'exploration de données qui nécessiteront des extractions quantitativement ou qualitativement substantielles d'une base de données protégée.

3.2. Le bénéfice d'une exception

L'article L.342-3 du Code de la propriété intellectuelle prévoit plusieurs exceptions au monopole du producteur de bases de données qui sont, par ailleurs, soumises depuis la loi du 1^{er} août 2006, au test des trois étapes. Il convient de noter que ces exceptions ne sont pas les mêmes qu'en droit d'auteur ou en droits voisins. L'exception de copie provisoire transitoire n'est pas prévue en matière de bases de données, il en va de même pour les exceptions à des fins d'actualité ou les revues de presse.

L'article L.342-3, 1^o du code de la propriété intellectuelle prévoit que lorsqu'une base de données est mise à la disposition du public par le titulaire des droits, celui-ci ne peut interdire, l'extraction ou la réutilisation d'une partie non substantielle, appréciée de façon quantitative ou qualitative, du contenu de la base, par la personne qui y a licitement accès. En matière d'exploration de données, même si l'utilisateur possède un accès licite à la base de données, il semble qu'il n'y a jamais d'extraction qui ne soit pas quantitativement ou qualitativement substantielle. Par conséquent, cette exception ne semble pas pouvoir s'appliquer.

L'article L. 342-3, 2^o du code de propriété intellectuelle précise qu'il est possible d'extraire une partie quantitativement ou qualitativement substantielle de la base de données dès lors qu'elle est réalisée à des fins privées. Si dans le cadre de l'activité d'exploration de données, les données sont mises à disposition d'une communauté de chercheurs ou d'une entreprise, cette exception ne pourra pas jouer.

L'article L. 342-3, 4^o du code de la propriété intellectuelle prévoit une exception pour extraction et réutilisation à des fins d'enseignement et de recherche. L'exception ne s'applique ni aux bases de données conçues à des fins pédagogiques ni aux bases de données réalisées pour une édition numérique de l'écrit. L'extraction et la réutilisation sont destinées à un public

⁷⁴ CJUE, 9 novembre 2004 : Comm.com.electr.2005, note Ch.Caron ; D.2005, somm. p.1495, obs. Sirinelli.

composé majoritairement d'élèves, d'étudiants, d'enseignants ou de chercheurs. La source doit être indiquée et il ne doit pas y avoir d'exploitation commerciale. En outre, elle doit être compensée par une rémunération négociée sur une base forfaitaire. Les conditions de mise en œuvre de cette exception étant très stricte, elle sera très difficile à mettre en œuvre dans le cadre des activités d'exploration de données à des fins de recherche.

Si aucune des exceptions n'est applicable, le producteur de la base de données pourra mettre en jeu le droit *sui generis* pour s'opposer à des activités d'exploration de données. Par conséquent, une demande d'autorisation d'exploration de la base de données sera nécessaire et impliquera ou non le versement d'une rémunération suivant les conditions négociées.

S'agissant du droit d'auteur, les actes de copies effectués sur des contenus protégés dans le cadre d'une opération d'exploration de données peuvent être qualifiés d'actes de reproduction nécessitant l'autorisation des ayants droit sauf à bénéficier d'une exception. L'interprétation restrictive des exceptions pertinentes conduit à écarter, dans la plupart des cas, leur application aux activités d'exploration de données.

S'agissant du droit *sui generis* des producteurs de bases de données, les opérations d'exploration de données peuvent conduire à des extractions quantitativement ou qualitativement substantielles d'une base de données protégée. Celles-ci nécessitent l'autorisation du producteur de la base de données sauf à bénéficier d'une exception. L'interprétation restrictive des exceptions pertinentes conduit à écarter, dans la plupart des cas, leur application aux activités d'exploration de données.

III. Réflexion sur les outils juridiques adaptés au développement de l'exploration de données

Le présent rapport n'entend pas proposer des solutions définitives ni trancher les points de désaccord entre les différentes parties intéressées. Les différents acteurs concernés ne rencontrent pas les mêmes problèmes ; leurs demandes à l'égard des outils d'exploration de données, comme l'état de leurs réflexions, leurs engagements dans diverses procédures diffèrent largement. Par conséquent, ce travail se propose d'offrir une réflexion générale sur les outils juridiques pouvant être adaptés au développement de l'exploration de données.

La diversité des activités d'exploration de données et des acteurs concernés conduit à rechercher des solutions souples et non figées. Le présent rapport encourage donc le développement de solutions contractuelles (1) et engage une réflexion sur l'exception à des fins de recherche (3).

1. La solution contractuelle

L'exploration de données est une activité émergente pour laquelle la solution contractuelle n'a pas encore épuisé l'ensemble de ses capacités. Le programme « *Des licences pour l'Europe* » (1.1) avait privilégié cette voie qui fait l'objet de nombreuses initiatives (1.2) susceptibles d'être améliorées (1.3).

1.1. L'initiative « Des licences pour l'Europe »

Le 4 février 2013, la Commission européenne a lancé l'initiative « *Des licences pour l'Europe* »⁷⁵. Il s'agissait d'établir un dialogue structuré entre toutes les parties prenantes, pour identifier des solutions concrètes, opérationnelles à court terme, à quatre questions des citoyens et des entreprises à l'heure du numérique⁷⁶.

L'exploration de données était l'une d'entre elles⁷⁷. La perspective était claire : favoriser le développement de cette activité, à des fins de recherche scientifique, en facilitant les relations entre les détenteurs de droits sur les contenus et les utilisateurs d'exploration de données.

Partant du constat que l'exploration de données exige actuellement des stipulations contractuelles établissant les modalités d'accès technique aux données entre les utilisateurs⁷⁸ et les titulaires des droits⁷⁹, la Commission européenne avait souhaité qu'un groupe de travail dresse un diagnostic des pratiques et des besoins en matière d'exploration de publications et de données scientifiques à des fins de recherche au niveau de l'Union européenne et propose des moyens juridiques et techniques pour développer cette activité.

L'objectif du dialogue « *Des licences pour l'Europe* » était de proposer fin 2013, des solutions pratiques initiées par l'industrie, sans préjudice d'une éventuelle action publique ultérieure, y compris des réformes législatives. Le groupe de travail relatif à l'exploration de données⁸⁰ était notamment composé des organisations représentatives des titulaires de droits, des utilisateurs, des prestataires techniques.

Le 4 juillet 2013, lors du point d'étape de l'initiative « *Des licences pour l'Europe* », le groupe de travail relatif à l'exploration de données a précisé la demande pour l'exploration de données à des fins de recherche au niveau de l'Union européenne était en train de croître.

⁷⁵ http://europa.eu/rapid/press-release_SPEECH-13-97_fr.htm

⁷⁶ L'initiative était menée conjointement par les commissaires Michel Barnier (Marché intérieur et services), Neelie Kroes (Stratégie numérique) et Androulla Vassiliou (Education, culture, multilinguisme et jeunesse).

⁷⁷ Les trois autres points étaient : Accès transfrontière et portabilité des services ; Contenus générés par les utilisateurs et octroi de licences aux petits utilisateurs de matériel protégé ; Secteur audiovisuel et institutions de gestion et de conservation du patrimoine culturel

⁷⁸ par exemple, les instituts de recherche

⁷⁹ par exemple, des éditeurs de revues scientifiques

⁸⁰ groupe de travail 4

Au cours de l'initiative, dix organisations représentant les intérêts des chercheurs ont quitté le groupe de travail⁸¹. Bien qu'ils se soient félicités de l'initiative de la Commission européenne et de l'engagement subséquent d'adapter le droit d'auteur au numérique, ils considéraient que le fait de placer les licences au centre de la discussion n'était pas la bonne méthode.

La Commission européenne a regretté ses départs car ils ont eu un impact sur l'efficacité du groupe de travail. Le point d'étape a donc été l'occasion pour la Commission de faire part de son inquiétude quant à la poursuite des travaux du groupe 4. Elle a demandé à l'ensemble des parties d'être pragmatique et constructif afin d'aboutir à un bon résultat.

Le 13 novembre 2013, l'initiative «*Des licences pour l'Europe*» s'est achevée à Bruxelles. Le groupe de travail 4 n'a pas réussi à trouver de consensus pour proposer un engagement conjoint de tous les membres. Cependant, l'association internationale des éditeurs scientifiques, techniques et médicaux (STM) s'est engagée à faciliter dans un premier temps l'exploration de textes et de données pour les contenus faisant l'objet d'un abonnement à des fins de recherche à but non commercial⁸² :

- proposition d'un modèle de contrat-type permettant ces usages ;
- mise au point de « licences en un clic » ;
- développement d'un portail permettant l'accès aux sites des éditeurs pour leur exploration, tout en préservant l'intégrité technique des systèmes de ces derniers, grâce à des clés pour des interfaces de programmation (API) fournies par les éditeurs (projet PROSPECT de CrossRef⁸³).

Dans son discours du 25 novembre 2013⁸⁴, le Commissaire Michel Barnier soulignait que, « le «*mining-hub*» proposé par les éditeurs, offre une réponse pragmatique aux chercheurs qui souhaitent pouvoir faire des fouilles automatiques sur des publications scientifiques dont l'accès n'est pas actuellement ouvert. L'enjeu est important : permettre des avancées concrètes dans ces activités – de «*text and data mining*» - à fort potentiel pour la recherche européenne ».

Si l'exploration de données n'a pas pu bénéficier d'un engagement conjoint de tous les membres du groupe de travail, l'initiative «*Des licences pour l'Europe*» a permis d'identifier les attentes de la recherche et la volonté des éditeurs de développer des outils contractuels favorables au développement de l'exploration de données.

⁸¹ <http://www.libereurope.eu/news/licences-for-europe-a-stakeholder-dialogue-text-and-data-mining-for-scientific-research-purpose>

⁸² <http://ec.europa.eu/licences-for-europe-dialogue/sites/licences-for-europe-dialogue/files/10-Text-data-mining.pdf>

⁸³ <https://ec.europa.eu/licences-for-europe-dialogue/sites/licences-for-europe-dialogue/files/CrossRef.pdf>

⁸⁴ «*Propriété intellectuelle et modernisation de l'environnement juridique des entreprises : deux leviers pour la compétitivité de l'Europe*»

1.2. Les initiatives contractuelles en cours

L'association internationale des éditeurs scientifiques, techniques et médicaux STM et l'Association of Learned and Professional Society Publishers et l'industrie pharmaceutique ont mis au point un contrat-type et trois groupes d'édition, Elsevier, Springer et Wiley ont mis en place des procédures sur l'exploration de données.

La clause qui est insérée dans les contrats d'abonnements existants autorise les activités d'exploration de données pour la recherche scientifique non commerciale et sans coûts supplémentaires. Sous réserve de ne pas utiliser de méthodes automatiques d'aspiration, d'extraction, de recherche, d'indexation sur le site de l'éditeur, d'effacer les copies faites après la fin du projet, de ne pas utiliser les contenus ou les résultats de l'exploration de données dans un cadre commercial ou avec des partenaires commerciaux.

Sur le plan technique, il est nécessaire de passer par une API⁸⁵ et d'enregistrer son projet de recherche. Les requêtes via l'API sont limitées à 10 000 articles par semaine. Le résultat issu de l'exploration de données peut comprendre des extraits de 200 caractères au maximum du texte original, il doit être publié sous une licence non commerciale (CC-BY-NC) et doit inclure un lien DOI⁸⁶ vers le contenu original.

En France, Elsevier a récemment passé un accord avec Couperin et l'Agence Bibliographique de l'Enseignement Supérieur (ABES) au sujet de l'abonnement au bouquet de périodiques « *Freedom Collection* » en licence nationale, incluant une clause sur l'exploration de données⁸⁷.

Il est important de noter que ces procédures ne font pas l'unanimité auprès des chercheurs, car les systèmes apparaissent relativement complexes⁸⁸. D'abord, les chercheurs regrettent que les requêtes soient limitées à 10 000 articles par semaine. Ensuite, ils sont opposés à l'idée de remplir un formulaire qui détaille le projet de recherche afin de pouvoir obtenir une clé d'accès. Le passage par une API permet alors aux éditeurs de constituer un vaste corpus de métadonnées de la recherche.

Il est également possible de prendre l'exemple du projet ISTEEX qui est un vaste programme d'acquisition de ressources scientifiques visant à créer une bibliothèque numérique, accessible à distance par tous les membres des établissements de l'enseignement supérieur et de la

⁸⁵ En informatique, une interface de programmation (souvent désignée par le terme API pour *Application Programming Interface*) est un ensemble normalisé de classes, de méthodes ou de fonctions qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels.

⁸⁶ Le DOI, « identifiant d'objet numérique » (« *Digital object identifier* ») est un mécanisme d'identification de ressources numériques, comme un film, un rapport, ou des articles scientifiques. Un DOI permet notamment de pouvoir retrouver l'emplacement d'un document en ligne si son URL a changé.

⁸⁷ <http://www.couperin.org/images/stories/editos/Communiqu%20COUPERIN%20Bilan%20ngociation%20Elsevier%202014%2002%2010.pdf>

⁸⁸ <http://www.libereurope.eu/news/liber-response-to-elsevier's-text-and-data-mining-policy>

recherche. Lancé en novembre 2012, ISTEEX est financé par l'Etat à hauteur de 60 millions d'euros.

La plateforme accueillant les données sera hébergée à l'Inist-CNRS. Initié par le Ministère de l'Enseignement Supérieur et de la Recherche, ce projet novateur est porté par le CNRS, l'ABES, le Consortium Couperin et l'Université de Lorraine agissant pour le compte de la Conférence des Présidents d'Université. Cette plateforme a notamment pour objectifs de proposer, outre l'accès aux données et métadonnées disponibles, des services à valeur ajoutée basés sur le traitement des données en texte intégral (par exemple et de manière non limitative, interrogation en texte intégral sur les objets numériques indexés dans leur totalité, production de synthèse documentaire par analyse de sous-corpus individualisés pour l'occasion et auxquels sont appliquées des méthodes de text-mining, lexicographie...).

Les ressources acquises en licence nationale font l'objet d'une licence type qui prévoit que les droits spécifiques concédés dans le cadre de la mise en place de la plateforme ISTEEX sont notamment de : « *concéder aux bénéficiaires et aux utilisateurs autorisés le droit d'utiliser les données hébergées sur la plateforme de l'Editeur ou sur la Plateforme ISTEEX pour se livrer à des opérations de fouilles de texte (text-mining et data-mining), conformément aux missions des Bénéficiaires* »⁸⁹.

Suivant les prévisions de la licence, les bénéficiaires ou utilisateurs autorisés ont le droit de mettre à disposition des tiers dans un but scientifique et non commercial, dans le cadre d'une licence libre permissive de type *Creative Commons*, des extraits enrichis ou annotés, dans la limite de 5% d'un Titre, et de 15 % de la Base de données. Ils ont également le droit de diffuser dans le cadre d'une licence libre de type *Creative Commons*, les résultats de recherches issus de l'exploitation des données d'ISTEEX (lexiques, ontologies, réseaux de connectivité lexicaux, thématiques, d'auteurs ou d'institutions par exemple), à l'exclusion des textes eux-mêmes. En outre, ils peuvent diffuser dans le cadre d'une licence libre de type *Creative Commons*, les métadonnées des documents support d'une recherche en vue de la présentation ou de la valorisation de recherches exploitant le corpus de données d'ISTEEX.

Pour les activités commerciales, la solution contractuelle permet aux ayants droit de conserver un contrôle sur leurs contenus et d'éviter d'être défavorisés par les bouleversements du partage de la valeur propres au numérique alors qu'ils prennent le risque de la création.

1.3. Les améliorations proposées

L'avantage de la solution contractuelle est de permettre la mise en place de conditions différentes pour les activités d'exploration de données. En effet, la frontière entre les opérations d'exploration de données effectuées à des fins de recherche et à des fins commerciales est difficile à tracer. Il est également possible qu'une opération d'exploration de

⁸⁹ http://www.istex.fr/IMG/pdf/2012_23_10_Licence_type_ISTEX.pdf

données soit effectuée à la fois dans un but de recherche et dans un but commercial.

La licence prévue par Istex semble une clause favorable au développement de l'activité d'exploration de données qui doit être autorisée dans les contrats, sans rémunération supplémentaire. Les éditeurs ont le droit de vouloir préserver l'intégrité technique de leurs bases de données.

En France, le CFC travaille actuellement, en particulier avec les éditeurs de presse en ligne, à l'élaboration de licences à destination des acteurs commerciaux.

En effet, la mise en place d'un système de gestion collective pourrait permettre de contrecarrer un certains nombres de difficultés de la voie contractuelle. Les utilisateurs pourraient s'adresser directement à la société de gestion collective qui pourrait redistribuer les sommes aux ayants droit. Cette solution pourrait correspondre aux demandes d'exploration de données de masse car l'organisation collective des titulaires de droits et des utilisateurs permet de rechercher un optimum, notamment par secteurs d'activités ou par finalités d'exploration de données.

Cependant, pour certains utilisateurs, la mise en place d'un système de gestion collective pour l'exploration de données semble inenvisageable. *« Une SACEM du content-mining aboutirait rapidement à une situation ingérable. Les frais d'administration de cette instance risquent d'être considérables. Le droit d'extraire serait ainsi potentiellement surévalué, au risque d'avoir un effet négatif sur le développement de ce type d'outils en France — et, de fait, d'accroître le retard pris sur les États-Unis en la matière. Par ailleurs, pour les contenus en ligne, il sera très difficile d'identifier les titulaires de droits à qui reverser les sommes collectées. Ce système risque donc de générer des montants importants de sommes irrépartissables, ce qui n'est pas de bonne politique⁹⁰ ».*

Deux rapports⁹¹ remis à la Commission européenne proposent notamment d'encourager les initiatives contractuelles.

Le rapport « Standardisation in the area of innovation and technological development, notably in the field of Text and Data mining », souligne que la voie contractuelle est une première étape avant la mise en œuvre d'une réforme sur le droit d'auteur.

Il faudrait donc encourager le développement de la solution contractuelle dans la mesure où celle-ci offre un cadre juridique souple pour une pratique aux contours évolutifs. Il est possible de recommander d'engager un processus de concertation et d'autorégulation sectorielle pour permettre la recherche et l'élaboration des équilibres les plus pertinents et les plus adaptés aux secteurs concernés.

⁹⁰ <http://www.savoirscom1.info/wp-content/uploads/2014/01/Synthèse-sur-le-statut-légal-du-content-mining.pdf>

⁹¹ http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf et http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf

2. La solution d'une exception à des fins de recherche

Après un bref regard sur les dispositifs législatifs d'autres pays (2.1) nous examinerons la solution d'une exception pour l'exploration de données à des fins de recherche (2.2).

2.1. Les dispositions législatives dans les autres pays

Le Japon (2.1.1) est le seul pays dont la législation prévoit une exception pour les activités assimilées à l'exploration de données. Aux Etats-Unis (2.1.2), la doctrine du *fair use* semble plutôt favorable au développement des activités d'exploration de données à des fins de recherche. Alors que la Commission européenne a introduit la question d'une exception dans le cadre de la consultation sur le droit d'auteur (2.1.3), l'Irlande (2.1.4) et la Grande-Bretagne (2.1.5) ont déjà engagé des initiatives qui pourraient déboucher sur des modifications de leur législation sur le droit d'auteur.

2.1.1. Le Japon

Le Japon est le seul pays où la loi prévoit une exception au droit d'auteur sur le sujet. Le texte précise que, pour cette activité, l'utilisation d'un ordinateur permet des actes d'extraction, de comparaison, de classification ou tout autre acte d'analyse statistique sur des informations pouvant provenir du langage, des sons, images ou de tout autre élément constituant de l'information⁹². L'exception s'applique à toutes les activités d'exploration de données sans distinction.

2.1.2. Les Etats-Unis

À titre liminaire, il est intéressant de noter qu'aux États-Unis la doctrine du « *fair use* » semble plutôt favorable au développement de l'exploration de données.

Contrairement au droit français qui dresse une liste limitative d'exceptions au monopole de l'auteur, la doctrine du « *fair use* » posée à l'article 107 du *Copyright Act* constitue une exception à caractère général, faisant application de quatre critères non exclusifs, appréciés *in concreto* par les juridictions américaines. Les juges confrontés au moyen de défense invoquant le *fair use* doivent prendre en considération le but de l'utilisation de l'œuvre (à des fins commerciales ou non), la nature de l'œuvre protégée, la quantité et le caractère

⁹² Article 47 septies « For the purpose of information analysis ("information analysis" means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other much information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use by a person who makes an information analysis ». <http://www.cric.or.jp/english/clj/cl2.html>

substantiel de la part de l'œuvre utilisée, l'impact de cette utilisation sur le marché potentiel de l'œuvre utilisée.

Dans sa décision rendue le 14 novembre 2013⁹³, le juge Denny Chin a considéré que la numérisation par Google des fonds des bibliothèques universitaires américaines sans autorisation des titulaires des droits relevait du « *fair use* ». Cette notion permet aux moteurs de recherche de numériser les livres, pour en extraire des contenus par mots-clés en raison des nombreux avantages que Google Books accorde à la collectivité par rapport aux effets limités des atteintes portées aux titulaires de copyright.

Par ailleurs, le juge a fait clairement référence à l'exploration de données en considérant que dans le cadre du projet Google Books elle relevait de la doctrine du « *fair use* » : « *Google Books est transformatif dans le sens où il a transformé le texte des livres en données à des fins de recherche, y compris pour de la fouille de données ou de texte (data mining et text mining), ouvrant ainsi de nouveaux champs à la recherche. Les mots dans les livres ont ainsi pu être utilisés d'une manière complètement différente par rapport à ce qui existait avant. Google a créé quelque chose de nouveau dans la manière d'utiliser le texte des livres, la fréquence des mots et les tendances dans leur utilisation fournissant des informations substantielles*⁹⁴ ».

Dans le cadre de l'exploration de données, la copie des données peut relever du « *fair use* » dans la mesure où l'activité ne pourrait exister sans elle. Pour autant, la copie ne saurait être rendue publique.

Si la doctrine du « *fair use* » est synonyme de souplesse pour les activités d'exploration de données, elle n'est pas nécessairement un gage de sécurité juridique dans la mesure où ce qui a été jugé pour un cas précis ne vaut pas nécessairement pour un cas voisin.

2.1.3. La Commission européenne : la consultation sur le droit d'auteur

Afin d'adapter à l'économie numérique la législation et les pratiques de l'Union européenne en matière de droit d'auteur, la Commission européenne a annoncé le 18 décembre 2012, dans sa Communication sur le contenu du marché unique numérique⁹⁵, poursuivre deux lignes d'actions parallèles.

D'un côté, pour un certain nombre de questions identifiées, elle a annoncé le lancement de l'initiative baptisée « *Des licences pour l'Europe* » et d'un autre côté elle a indiqué vouloir conduire à son terme l'effort qu'elle a engagé pour revoir et moderniser le cadre législatif de l'UE sur le droit d'auteur.

⁹³ On peut également citer la décision *HathiTrust* consultable à l'adresse : http://www.tc.umn.edu/~nasims/HathivAG10_10_12.pdf

⁹⁴ Texte de la décision précédente, traduit par Lionel Maurel, « Verdict dans l'affaire Google Books : une grande leçon de démocratie ? », *S. I. Lex*, <http://scinfolex.com/2013/11/15/verdict-dans-laffaire-google-books-une-grande-lecon-de-democratie>

⁹⁵ COM (2012) 789 final, 18.12.2012.

Le 5 décembre 2013, la Commission européenne a lancé une consultation publique dans le cadre de ses travaux visant à revoir et à moderniser les règles européennes sur le droit d'auteur, elle s'est clôturée le 5 mars 2014. Les services de la Commission ont indiqué avoir reçu plus de 10 000 réponses à la consultation.

Une partie de la consultation était consacrée à l'exploration de données (« *text and data mining* ») et posait la question de savoir si l'exploration de données devait être traitée par la voie contractuelle ou si la mise en place d'une exception était nécessaire.

2.1.4. L'Irlande

Un rapport irlandais⁹⁶ suggère la mise en place d'un cadre dit de *fair dealing* pour les activités d'exploration de données. Il ne s'agit pas d'énoncer des principes généraux valables pour tous les usages mais de prévoir un nombre limité d'usages strictement définis par la loi. Le rapport irlandais précise par exemple :

- La nécessité de ne pas attenter aux intérêts des détenteurs des textes ou des données ;
- L'exploration de données doit être effectuée dans un but non commercial ;
- L'association du droit d'extraction au droit de lire : le lecteur ne peut pas extraire un contenu qu'il ne peut pas lire ;
- La présentation de contenu protégé dans la limite du droit de courte citation ;
- L'autorisation de réaliser une copie pour le temps des opérations d'exploration de données.

2.1.5. La Grande-Bretagne

Au Royaume-Uni le *Copyright, Design and Patents Act* prévoit une série d'exceptions au *Copyright* qui doivent être interprétées par le juge à la lumière du *fair dealing*⁹⁷.

Depuis 2010, le gouvernement britannique prépare une réforme du droit d'auteur avec l'objectif de le rendre mieux adapté au numérique et de créer davantage de croissance économique. Les modifications en cours ont été entreprises à la suite d'un rapport⁹⁸ du professeur Ian Hargreaves et d'une vaste consultation des parties intéressées. Après leur soumission au Parlement, elles entreront en vigueur le 1^{er} juin 2014.

⁹⁶ Modernising Copyright, The report of the Copyright Review Committee, Dublin 2013. P85 à P88.

⁹⁷ Le « *fair dealing* » est un critère d'appréciation de la validité des exceptions. L'appréciation se fait *in concreto* en veillant à ce que l'utilisation n'empêche pas l'exploitation normale de l'œuvre ou ne cause une perte financière au titulaire de droit. Le « *fair dealing* » s'approche de l'esprit du *fair use* américain sans être comme lui une limitation à part entière au monopole conféré à l'auteur car son application est liée à des exceptions expressément prévues par la loi.

⁹⁸ <http://www.ipo.gov.uk/ipreview-finalreport.pdf>

Le gouvernement britannique a introduit une exception pour l'exploration de données (« *text and data mining* ») à des fins de recherche non-commerciale qu'il semble fonder sur l'article 5 3. a) de la directive 2001/29⁹⁹.

La réforme ouvre la possibilité, pour les chercheurs en possession légitime d'une œuvre ou y ayant un accès légitime, de réaliser une copie de cette œuvre afin de mener des opérations d'exploration de données à des fins de recherche non-commerciale.

Par ailleurs, une mesure d'ordre public rend nulles les stipulations contractuelles qui empêcheraient la réalisation d'une copie pour l'exploration de données à des fins de recherche non-commerciale.

Enfin, la faculté pour les éditeurs de prendre des mesures pour garantir la stabilité et la sécurité de leurs réseaux de diffusion est limitée afin de permettre la réalisation des opérations d'exploration de données.

La réforme ne modifie pas le droit applicable aux bases de données car selon le gouvernement britannique, l'exception pour la recherche prévue par le *Copyright and Rights in Databases Regulations* de 1997 autorise les extractions nécessaires.

2.2. La solution d'une exception à des fins de recherche « non-commerciale »

Comme nous l'avons vu précédemment, aucune des exceptions prévues par la directive 2001/29/CE ni par l'article L.122-5 du code de la propriété intellectuelle ne semble offrir de sécurité juridique suffisante pour le développement de l'exploration de données. Il en va ainsi de l'exception de copie transitoire, de l'exception de citation et pour les activités de recherche de l'exception prévue à l'article L.122-5, 3° e). Il en va de même pour le droit *sui generis* des bases de données.

En droit de l'Union européenne ou en droit national, l'exception « recherche » ne pourra vraisemblablement pas être invoquée par les chercheurs dans le cadre des activités d'exploration de données en raison de son caractère très limité.

À ce titre, il pourrait être utile que la Commission développe des recommandations à l'attention des États-membres afin de clarifier dans quelle mesure les activités d'exploration de données sont couvertes par le droit de la propriété intellectuelle et dans quelle mesure elles s'insèrent dans le cadre des exceptions existantes pour les usages à des fins de recherche¹⁰⁰.

Au cours des auditions, certains utilisateurs ont appelé de leurs vœux une modification de la

⁹⁹ « Les États membres ont la faculté de prévoir des exceptions ou limitations aux droits prévus aux articles 2 et 3 dans les cas suivants: a) lorsqu'il s'agit d'une utilisation à des fins exclusives d'illustration dans le cadre de l'enseignement ou de la recherche scientifique, sous réserve d'indiquer, à moins que cela ne s'avère impossible, la source, y compris le nom de l'auteur, dans la mesure justifiée par le but non commercial poursuivi »

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2001:167:0010:0019:FR:PDF>

¹⁰⁰ En particulier en regard de la directive sur la société de l'information et de la directive sur le droit des bases de données.

directive 2001/29/CE voire de la directive 96/9/CE relative aux bases de données ne serait-ce que pour les activités de recherche non-commerciales. Il pourrait s'agir d'une exception au droit de reproduction mais également au monopole du producteur de base de données.

À l'heure actuelle, il ne semble pas envisageable de modifier le droit national dans la mesure où la liste des exceptions prévue par les directives 2001/29/CE est limitative. Par conséquent, la prévision d'une exception ne peut être effectuée qu'au niveau européen.

Deux rapports¹⁰¹ remis à la Commission européenne proposent une modification de la législation européenne (directive 2001/29/CE voire de la directive 96/9/CE) visant à introduire une exception pour favoriser le développement de l'exploration de données. Ils considèrent notamment que l'exception devrait porter principalement sur les travaux de recherche scientifique, et encore, à condition qu'ils ne donnent lieu à aucune valorisation marchande.

Le rapport « *Standardisation in the area of innovation and technological development, notably in the field of Text and Data mining* », souligne que si une exception est mise en œuvre, il ne devrait pas être possible d'y déroger par contrat ou de la rendre inapplicable par la mise en place de mesures techniques de protection. Le rapport précise également qu'il est difficile dans la pratique de distinguer entre les activités de recherche commerciales et les activités de recherche non-commerciales et propose de prévoir une exception à des fins de recherche sans distinction.

Le rapport « *Study on the legal framework of text and data mining (TDM)* » propose la mise en œuvre d'une exception pour les activités d'exploration de données à des fins de recherche non-commerciale.

Par conséquent, la réflexion sur la mise en place d'une exception pour les activités d'exploration de données à des fins de recherche est largement engagée.

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.

¹⁰¹ http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf et http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf

Composition du comité de pilotage

BLACHE Catherine	Syndicat national de l’édition (SNE)
BERTHAULT Denis	Groupement français de l'industrie de l'information (GFII) – Lexis Nexis
CAVANAGH Jean Franck	Groupement français de l'industrie de l'information (GFII) – Reed Elsevier
DASCIER Angélique	Chambre syndicale de l'édition musicale (CSDEM)
DELBOUIS Emmanuel	Ministère de la Culture et de la Communication (MCC)
DUPLAA Bruno	Institut national de la propriété industrielle (INPI)
FLÜRY-HERARD Elisabeth	Autorité de la concurrence – Vice-présidente
HUOT Charles	Groupement français de l'industrie de l'information (GFII) – TEMIS
LEDUC Philippine	Chambre Syndicale des Editeurs de Musique de France (CEMF)
MASSERON Philippe	Centre français du droit d’exploitation de copie (CFC)
MIGNAN Denis	Hachette livre

ZOLYNSKI Célia

Université de Versailles Saint Quentin - Professeur agrégé en poste, directrice du pôle Propriété intellectuelle du réseau Trans Europe Experts.

Liste des personnes auditionnées

BERGMAN Anne	Fédération européenne des éditeurs (FEE) – Directrice
CHOUKRI Khalid	European Language Resources Association (ELRA) – Secrétaire général
MANARA Cédric	Google – Copyright Counsel
DEBARNOT Jean-François	Institut national de l’audiovisuel (Ina) – Directeur juridique
MUTZ Barbara	Institut national de l’audiovisuel (Ina) – Juriste
BEAUFORT Arnaud	Bibliothèque nationale de France (BnF) – Directeur général
MAUREL Lionel	Lionel Maurel, conservateur de bibliothèques, juriste, co-fondateur du Collectif SavoirsCom1
LANGLAIS Pierre-Carl	Collectif Savoircom1
DELAHOUSSE Jean	Consultant indépendant - Expert des technologies sémantiques

AUBIN Sophie	Institut national de la recherche agronomique (INRA) – Chargée de ressources
BERARD Raymond	Institut de l’information scientifique et technique (INIST) – Centre national de recherche (CNRS) – Directeur
GILBERT Yan	Le Nouvel Observateur – Chargé de la diffusion numérique Groupement d’intérêt économique Panoramas presse Magazine et Professionnelle (PPMP) – fondateur
BONNET Vincent	European Bureau of Library Information and Documentation Associations (EBLIDA) – Directeur
GUIBAULT Lucie	Institut du droit de l’information d’Amsterdam – Professeur associé
BATTISTI Michèle	European Bureau of Library Information and Documentation Associations (EBLIDA) – Expert
DOMANGE Camille	Ministère de la Culture et de la Communication (MCC) – Chef du département des programmes numériques
DICKES Christophe	Kantar Média – Directeur Europe droit d’auteur Fédération des entreprises de veille média (FEVEM)
DONOT Alexis	Argus de la Presse – Président-directeur général Fédération des entreprises de veille média (FEVEM)
PIERREL Jean-Marie	Centre national des ressources textuelles et lexicales – Université de Lorraine & CNRS – Responsable
POTTECHER Richard	EDD – Président-directeur général

GALOPIN Benoît	Institut de recherche en propriété intellectuelle (IRPI) – Docteur en droit
COLAS Alain	Ministère de l’Enseignement Supérieur et de la Recherche - Chef du Département de l'Information Scientifique et Technique et Réseau Documentaire
COLACANAP Grégory	Consortium unifié des établissements universitaires et de recherche pour l'accès aux publications numériques (COUPERIN) – Coordinateur
PERALES Christophe	Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation (ADBU) – Directeur
BOUDILLET Yvan	Parlophone Music France – Warner - Directeur Marketing digital
de FRESSENEL Ronan	M6 Publicité - Directeur Général adjoint en charge du Marketing et de Etudes
GOLDSMITH Frédéric	Association des producteurs de cinéma – Délégué général
TILLIET Hubert	Société des Auteurs et Compositeurs dramatiques – Directeur Juridique
BOLOGNE Jean Claude	Société des Gens de Lettres – Président

Table des matières

1.L'exploration de données : une activité protéiforme.....	7
<i>1.1. Une terminologie variable.....</i>	7
<i>1.2. Les éléments de définition de l'exploration de données.....</i>	7
<i>1.2.1. Un processus de recherche automatisé.....</i>	9
<i>1.2.2. Sur un ensemble de données numériques.....</i>	9
<i>1.2.3. Dans le but de découvrir de nouvelles connaissances et de nouvelles idées....</i>	10
2. Les problématiques soulevées par le développement de l'exploration de données en droit de la propriété littéraire et artistique.....	12
<i>2.1. Le point de vue des utilisateurs.....</i>	12
<i>2.1.1. Le développement de l'activité d'exploration de données et le droit d'auteur. .</i>	12
<i>2.1.2. Une solution contractuelle inadaptée.....</i>	15
<i>2.1.3. La nécessité d'une exception pour le développement de l'exploration de données.....</i>	16
<i>2.2. Le point de vue des ayants droit.....</i>	17
<i>2.2.1. La mise en place d'outils juridiques et techniques adaptés au développement de l'exploration de données.....</i>	17
<i>2.2.2. Un développement de l'exploration de données respectueux des contenus éditoriaux.....</i>	18
1. Le déroulement d'une opération d'exploration de données.....	19
<i>1.1. La définition des objectifs.....</i>	19
<i>1.2. L'inventaire des contenus numériques utiles.....</i>	20
<i>1.3. La collecte des contenus numériques.....</i>	20
<i>1.4. La préparation des données.....</i>	20
<i>1.5. Le traitement des données.....</i>	20
<i>1.6. Le résultat de l'exploration des données.....</i>	21
2. Les actes susceptibles d'entrer dans le champ d'application du droit d'auteur.....	21
<i>2.1. L'exploration de données et la mise en jeu des droits patrimoniaux.....</i>	22
<i>2.1.1. La mise en jeu du droit d'exploitation.....</i>	22

2.1.1.1.	La mise en jeu du droit de représentation.....	22
2.1.1.2.	La mise en jeu du droit de reproduction.....	23
2.1.2.	<i>La question du bénéfice d'une exception.....</i>	<i>25</i>
2.1.2.1.	L'application de l'exception de copie provisoire transitoire.....	25
2.1.2.2.	L'application de l'exception de citation.....	29
2.1.2.3.	L'application de l'exception de l'article L.122-5 3° e) du code de la propriété intellectuelle.....	30
2.1.2.4.	L'application de l'exception de l'article L.122-5, 5° du code de la propriété intellectuelle.....	31
2.2.	<i>L'exploration de données et la mise en jeu du droit moral.....</i>	<i>31</i>
2.2.1.	<i>Le droit de divulgation.....</i>	<i>32</i>
2.2.2.	<i>Le droit de retrait et de repentir.....</i>	<i>32</i>
2.2.3.	<i>Le droit à la paternité.....</i>	<i>32</i>
2.2.4.	<i>Le droit au respect et à l'intégrité de l'œuvre.....</i>	<i>32</i>
3.	Les actes d'exploitation susceptibles d'entrer dans le champ d'application du droit sui generis des producteurs de bases de données.....	32
3.1.	<i>La mise en jeu du monopole du producteur de bases de données.....</i>	<i>33</i>
3.2.	<i>Le bénéfice d'une exception.....</i>	<i>34</i>
1.	La solution contractuelle.....	36
1.1.	<i>L'initiative « Des licences pour l'Europe ».....</i>	<i>36</i>
1.2.	<i>Les initiatives contractuelles en cours.....</i>	<i>38</i>
1.3.	<i>Les améliorations proposées.....</i>	<i>39</i>
2.	La solution d'une exception à des fins de recherche.....	41
2.1.	<i>Les dispositions législatives dans les autres pays.....</i>	<i>41</i>
2.1.1.	<i>Le Japon.....</i>	<i>41</i>
2.1.2.	<i>Les Etats-Unis.....</i>	<i>41</i>
2.1.3.	<i>La Commission européenne : la consultation sur le droit d'auteur.....</i>	<i>42</i>
2.1.4.	<i>L'Irlande.....</i>	<i>43</i>
2.1.5.	<i>La Grande-Bretagne.....</i>	<i>43</i>
2.2.	<i>La solution d'une exception à des fins de recherche « non-commerciale ».....</i>	<i>44</i>