

## Projet de loi « Pour une république numérique »

### Libérer du potentiel économique grâce au TDM

(text and data mining ou fouille de texte)

L'exploitation de l'extraordinaire masse de données aujourd'hui produites, qu'elles soient nativement numériques ou obtenues par numérisation, constitue actuellement pour la recherche probablement la plus prometteuse des perspectives inaugurées par les révolutions digitale et des réseaux, au point que l'on parle même de plus en plus de «data driven innovation» ou de «data driven science». Cette exploitation se fait par le TDM.

Or, cette pratique du TDM ne consiste 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.**

La voie contractuelle n'est de plus pas adaptée aux besoins et aux pratiques de la recherche : les chercheurs ont besoin de croiser dans leur fouille des sources diverses, aussi bien publiques que privées. Il est donc totalement exclu qu'un opérateur ou groupement d'opérateurs privés puisse jamais mettre à disposition, au prix de dépenses d'investissement soutenables, la plateforme technique nécessaire à la fois aux besoins très divers et en pleine expansion de la recherche, comme au contrôle par leur soin exclusif des données présentes dans les produits des ayants droit.

**C'est pourquoi, il est nécessaire d'adopter une exception au droit d'auteur pour le TDM.**

L'adoption rapide d'une telle exception en France est cruciale pour la compétitivité de notre recherche : la pratique du TDM est déjà admise aux USA (jurisprudence HathiTrust), gravée dans la loi en Irlande et en Grande-Bretagne. Comme le montrent différentes études, les bénéfices pour l'ensemble de la société, qu'il s'agisse du secteur public ou commercial, sont très nettement supérieurs au peu probable préjudice que pourraient encourir les titulaires de droits du fait des usages attachés au TDM.

L'absence d'intervention du législateur laissera le champ libre à de gros acteurs du monde de l'édition pour imposer leurs propres termes au monde de la recherche par le biais de licences, renforçant encore un peu plus leur situation dominante et la privatisation des résultats de la recherche.

## Mesure proposée :

- **Créer une exception pour le TDM, à des fins non-commerciales et dans le cadre académique, au titre de l'enseignement et de la recherche, dans le code de la propriété intellectuelle**

La Grande-Bretagne et l'Irlande ont déjà adopté dans leur législation une exception légalisant sans compensation le text and data mining.

Les conclusions de l'*Étude sur le cadre juridique de la fouille de textes et données* de mars 2014, commandée par la Commission européenne sont les suivantes :

*« On the basis of our analysis, our suggestion is to have a new specific data analysis exception which would be inspired from, and contain partly the same conditions than, the scientific research exceptions, but which would have its own peculiarities. The suggestions are made only for TDM, not for scientific research in general.»*

La liste actuelle des exceptions au droit d'auteur autorisées par l'Union européenne ne prévoit pas actuellement le TDM mais la révision de cette directive est annoncée. L'orientation souhaitée va explicitement dans le sens d'une nouvelle exception : *« La Commission analyse actuellement les options envisageables et mènera une réflexion d'ici au printemps 2016 sur l'élaboration de propositions législatives de l'Union relatives à d'autres exceptions, afin de permettre aux organismes de recherche d'intérêt public d'appliquer les techniques de TDM aux contenus auxquels ils ont légalement accès, avec une sécurité juridique totale, à des fins de recherche scientifique »*<sup>1</sup>

Cette position permet à l'instar des britanniques de prendre dès à présent des mesures législatives nationales autorisant le TDM en France sans craindre la moindre opposition de la Commission européenne. On notera qu'aucune action n'a été engagée auprès de la Cour de justice de l'Union européenne depuis la promulgation de l'exception britannique.

---

1 COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ DES RÉGIONS : Vers un cadre moderne et plus européen pour le droit d'auteur. 9 décembre 2015.

## Références

- L'analyse de Couperin et de l'ADBU<sup>2</sup>

[http://adbu.fr/wp-content/uploads/2014/04/Audition\\_CSPLA\\_TDM\\_2014\\_04\\_04\\_final.pdf](http://adbu.fr/wp-content/uploads/2014/04/Audition_CSPLA_TDM_2014_04_04_final.pdf)

- Study on the legal framework of text and data mining (TDM)

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

- Loi anglaise

<http://www.legislation.gov.uk/ukdsi/2014/9780111112755>

- Value and benefits of text mining

<http://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining>

- COMMUNICATION DE LA COMMISSION AU PARLEMENT EUROPÉEN, AU CONSEIL, AU COMITÉ ÉCONOMIQUE ET SOCIAL EUROPÉEN ET AU COMITÉ DES RÉGIONS : Vers un cadre moderne et plus européen pour le droit d'auteur.

[https://www.google.fr/url?](https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjSmdPTiMrLAhWLOxoKHck6DX0QFggiMAE&url=https%3A%2F%2Fec.europa.eu%2Ftransparency%2Fregdoc%2Frep%2F1%2F2015%2FFR%2F1-2015-626-FR-F1-1.PDF&usg=AFQjCNGkyWI4HrX-TKlokfyt4I9-ZbasDQ&sig2=IJGIQS37y_VkOUpbZOzig&bvm=bv.117218890,d.d24&cad=rja)

[sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjSmdPTiMrLAhWLOxoKHck6DX0QFggiMAE&url=https%3A%2F%2Fec.europa.eu%2Ftransparency%2Fregdoc%2Frep%2F1%2F2015%2FFR%2F1-2015-626-FR-F1-1.PDF&usg=AFQjCNGkyWI4HrX-TKlokfyt4I9-ZbasDQ&sig2=IJGIQS37y\\_VkOUpbZOzig&bvm=bv.117218890,d.d24&cad=rja](https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwjSmdPTiMrLAhWLOxoKHck6DX0QFggiMAE&url=https%3A%2F%2Fec.europa.eu%2Ftransparency%2Fregdoc%2Frep%2F1%2F2015%2FFR%2F1-2015-626-FR-F1-1.PDF&usg=AFQjCNGkyWI4HrX-TKlokfyt4I9-ZbasDQ&sig2=IJGIQS37y_VkOUpbZOzig&bvm=bv.117218890,d.d24&cad=rja)

---

2 Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation

## Proposition de formulation pour un article de loi spécifique au TDM

Le texte<sup>3</sup> proposé ci-dessous a été déposé par Couperin dans le cadre de la consultation organisée par le Secrétariat au numérique relative au « Projet de loi pour une République numérique » et a été repris par le Parlement et ajouté au projet de loi proposé par le Gouvernement. Il s'agit du nouvel article 18 bis. Il est essentiel pour la recherche française que cette exception soit conservée dans le projet de loi.

*Le code de la propriété intellectuelle est ainsi modifié :*

*1° Après le second alinéa du 9° de l'article L. 122-5, il est inséré un 10° ainsi rédigé :*

*« 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données pour les besoins de la recherche publique, à l'exclusion de toute finalité commerciale. Un décret fixe les conditions dans lesquelles l'exploration des textes et des données est mise en œuvre, ainsi que les modalités de conservation et de communication des fichiers produits au terme des activités de recherche pour lesquelles elles ont été produites ; ces fichiers constituent des données de la recherche ; »*

*2° Après le 4° de l'article L. 342-3, il est inséré un 5° ainsi rédigé :*

*« 5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données dans un cadre de recherche, à l'exclusion de toute finalité commerciale. La conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche pour lesquelles elles ont été produites, sont assurées par des organismes désignés par décret. Les autres copies ou reproductions sont détruites. »*

### **Argumentaire :**

Outre le droit de fouiller les textes et les données des bases de contenus pour lesquelles un droit d'accès est légalement acquis (licence contractuelle commerciale, dispositions 131-1 à 133-1 du Code du patrimoine relatives notamment au Dépôt légal de l'Internet), dont la nécessité a suffisamment été exposée et qu'il convient d'inscrire dans la loi (dans le cas de la fouille des archives de l'Internet, ni la voie contractuelle ni le recours à un dispositif de gestion collective n'apparaissent praticables : l'exception est la seule voie possible), il est essentiel de conserver les fichiers issus des traitements qui constituent alors des données de la recherche. La capitalisation du travail de curation et de formatage des données, ce que l'on appelle le pré-traitement ou la normalisation est fondamental. Une fois récupérés pour être fouillés, les documents ne sont pas immédiatement exploitables, il faut opérer un certain nombre de traitements qui peuvent être assez lourds. Il est important de pouvoir mutualiser ce travail pour fournir un matériau directement exploitable par d'autres chercheurs, et assurer la reproductibilité des résultats, élément essentiel de la preuve scientifique. Des tiers de confiance permettront de capitaliser et de généraliser le travail de normalisation pour que d'autres chercheurs puissent venir à nouveau exploiter le matériau et rejouer la fouille. Le projet de texte de loi doit prévoir la conservation et la diffusion des jeux de données au terme de la recherche.

<sup>3</sup> <https://www.republique-numerique.fr/projects/projet-de-loi-numerique/consultation/consultation/opinions/section-2-travaux-de-recherche-et-de-statistique/exception-de-fouille-de-texte-et-de-donnees>