

Un cadre juridique clair pour développer les usages de la fouille de texte et de données (FTD) : un préalable indispensable à la compétitivité de la recherche publique française

Synthèse n°1 – Mai 2016

→ Rappel du processus législatif relatif en cours

Dans le cadre de la discussion du projet de loi pour une République numérique discutée puis votée successivement à l'Assemblée nationale et au Sénat, un point essentiel au devenir de la recherche publique française a été abordé via l'introduction d'un article 18bis autorisant la fouille de textes et de données (FTD, parfois aussi désignée de l'acronyme anglais TDM pour « Text & Data mining »).

L'adoption d'un cadre juridique sûr encadrant le développement de la fouille de textes et de données est un préalable indispensable à la compétitivité de la recherche publique française. Dans des formulations différentes, mais adoptées à de larges majorités transpartisanes, l'Assemblée nationale (en adoptant un amendement de Mme Nathalie Kosciusko-Morizet, députée *Les Républicains* de l'Essonne) et le Sénat (en retenant un amendement de Mme Colette Mélot, sénatrice *Les Républicains* de Seine-et-Marne) proposent des dispositions innovantes mais peu compatibles entre-elles en vue de favoriser le développement de la FTD au sein de la recherche publique.

Dans la perspective de la discussion en Commission Mixte Paritaire, et sans méconnaître la volonté novatrice du Sénat sur ces dispositions, ***il apparaît que seule la rédaction votée par l'Assemblée nationale met en place un cadre juridique sans faille, car introduisant une exception au droit d'auteur, pour le développement du FDT au sein de la recherche publique.***

L'objectif du présent argumentaire est de nourrir la réflexion de la Commission Mixte Paritaire sur ce point.

Qu'est-ce que la FTD (ou TDM) ?

La fouille de textes et de données (FTD) ou « Text Data Mining » (TDM) est un ensemble de pratiques appliquant à de très vastes corpus numérisés de textes et de données (polymorphes et pluridisciplinaires) des techniques analytiques puissantes permettant de dégager de nouvelles connaissances (mise en évidence de corrélations, de tendances émergentes...). L'ampleur des gisements de contenus numériques disponibles et la puissance sans précédent des traitements algorithmiques font des technologies de FTD une avancée majeure de l'économie numérique qui ouvre un champ des possibles sans précédent pour la recherche et l'innovation.

→ Pourquoi les technologies de FTD sont-elles si importantes pour la recherche publique ?

Pour la recherche publique, la FTD n'est pas une technologie numérique parmi d'autres : elle est une technologie « habilitante » qui est la clé du développement d'un nouveau paradigme de la recherche.

A la recherche initiale débouchant sur des articles scientifiques (2 600 000 sont publiés chaque année dans le monde) va se superposer à l'avenir une recherche dérivée qui, grâce aux technologies analytiques, permettra d'extraire d'immenses corpus de textes et de données des connaissances scientifiques nouvelles, notamment en ouvrant des approches transdisciplinaires sans précédent ou en permettant d'appréhender les phénomènes complexes impliquant de multiples interactions .

Fait trop peu souligné, la FTD est aussi la clé d'une meilleure productivité de la recherche publique (dont les enjeux se mesurent donc en milliards d'euros) puisqu'elle permettra d'exploiter autrement, dans une perspective holistique, des connaissances déjà produites (et déjà financées).

→ Pourquoi la compétitivité de la recherche publique française est-elle menacée faute d'un cadre juridique adéquat pour la FTD ?

Faute d'une législation adéquate les chercheurs français sont, dès aujourd'hui, dans une position de faiblesse vis-à-vis de leurs collègues étrangers exerçant aux États-Unis, au Canada, au Japon ou en Grande-Bretagne... qui disposent dans leur pays de législations permissives et plus ouvertes. Or, on sait que les travaux s'appuyant sur les techniques de fouille de textes et de données ont un potentiel très important et ont toutes chances d'être un facteur de différenciation qualitative des recherches publiques nationales.

Subordonner à la révision de la directive européenne de 2001 sur le droit d'auteur, comme le fait le gouvernement, l'adoption d'un cadre juridique sûr s'appliquant à la FTD, ferait prendre à la recherche publique française un retard fatal de plusieurs années, lié aux délais incompressibles du processus législatif européen.

→ Pourquoi la modification du Code de la propriété intellectuelle est-elle indispensable à l'établissement d'un cadre juridique sûr pour la FTD ?

L'exploration de données numériques par des outils informatiques sur des 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.

Ce n'est donc que par des dispositions abordant frontalement ces freins induits par le droit d'auteur que peut être établi un cadre juridique sûr favorisant le développement de la FTD. Comme l'ont reconnu les législateurs britanniques dès 2012 (sans que cette option soit contestée par la Commission européenne ou par les éditeurs...), l'exception au droit d'auteur est ici la seule voie praticable.

La Commission européenne a d'ailleurs annoncé qu'elle allait défendre cette mesure dans le cadre du marché unique numérique. L'Allemagne s'apprête à réviser sa loi fédérale sur ces thèmes, sans attendre la révision préalable de la Directive européenne sur le droit d'auteur.

Pour le Gouvernement et les parlementaires, faire le choix de la FTD ne doit pas être vécu comme une entrave au droit d'auteur, mais bien comme un acte de soutien à la recherche et à la souveraineté scientifique de la France.

→ Une exception au droit d'auteur très encadrée

Face à la désinformation menée sur le thème de la FTD (ou TDM) par les grands éditeurs scientifiques internationaux (lire ci-dessous) il apparaît important de rappeler quelques évidences :

- La fouille de textes et de données mentionnée à l'article 18bis du projet de loi pour une République numérique ne concerne que des contenus scientifiques (articles de recherche ou collections de données) qui ont été dans leur immense majorité produits par la recherche publique et financée par des budgets publics. **En aucun cas la FTD prévue à l'article 18bis ne s'appliquerait à des contenus originaux produits par les éditeurs,**
- L'exception prévue dans la version de l'article 18bis approuvée par l'Assemblée nationale (mais aussi dans celle voulue par le Sénat) ne s'applique que :
 - aux usages au sein des organismes de recherche publics,
 - à des contenus que ces organismes ont déjà achetés légalement via leurs abonnements aux éditeurs les ayant publié en revues,
 - à des usages non commerciaux.

Au vu de ces restrictions convergentes il ne peut être de bonne foi soutenu que l'adoption d'une telle exception au droit d'auteur porterait un préjudice économique aux éditeurs scientifiques ou affaiblirait leur maîtrise de leurs contenus.

→ La dimension internationale de l'édition scientifique est la seule pertinente s'agissant de légiférer sur la FTD

La publication scientifique de recherche est d'ores et déjà fortement internationalisée : en 2014 les chercheurs français ont publié 104 739 articles soit 4% du nombre total d'articles publiés mondialement (2 600 000). Moins de 6% de ces articles sont publiés en français. Les grands éditeurs internationaux ont, dans leurs revues, absorbé 89,1% de ces publications. **La question de l'encadrement juridique de la FTD, y compris dans le seul contexte français, ne peut être posée qu'en relation à ce cadre international de l'édition scientifique.**

Or celui-ci est très **concentré** et se caractérise par **des bénéfiques hors-normes alimentés par les budgets publics**. Les six premiers éditeurs (Elsevier, Springer Nature, Wiley) publient 54% des 2 600 000 articles qui paraissent annuellement au plan mondial ; ils contrôlent 50% des revues les plus prestigieuses ; ils détiennent 38% du marché mondial de l'édition scientifique (d'environ 13,8 Md.€), mais captent 65% de ses profits. Elsevier, N° 1 mondial du secteur, à lui seul capte 26% des profits de l'édition de recherche, soit 1 Md. €.

Ces grands éditeurs internationaux « for profit » réalisent des marges bénéficiaires hors normes, supérieures ou égales à 40% de leurs chiffres d'affaires. **Tout cela en publiant des contenus qui sont intégralement issus de la recherche publique.**

Face à cet oligopole, la recherche publique (qui en France consacre plus de 100 M € à des achats de revues scientifiques) est en position de faiblesse, puisqu'elle ne peut se passer des publications scientifiques diffusées par ces grands éditeurs sur leurs vastes plates-formes numériques.

Si dans les faits, faute d'une exception ad hoc au droit d'auteur, les grands éditeurs « for profit » disposaient en plus du droit d'autoriser ou non la fouille de textes et de données, la dépendance de la recherche publique à l'égard de ce cartel d'éditeurs s'en trouverait accrue, tant sur un plan financier que sur un plan stratégique.

→ Pourquoi les éditeurs sont-ils vent-debout contre une exception au droit d'auteur pour l'usage de la TDM par la recherche publique ?

Cette économie très lucrative de la publication scientifique (issue à 95% de la recherche publique) repose sur la captation par les éditeurs des droits d'auteur des chercheurs au moment de la publication en revues des articles retraçant leurs travaux.

Les chercheurs du secteur public, soumis à l'impérative obligation de publier leurs résultats, cèdent **gratuitement** aux éditeurs l'intégralité de leurs droits d'auteurs sur les articles qu'ils publient. **C'est sur la base de cette captation de droits que les éditeurs revendiquent, au-delà de la seule édition dans des revues, une exclusivité sur tout traitement des articles scientifiques qu'ils ont publié**, dont les traitements de fouille de textes et de données.

Edicter une exception en faveur de la FTD c'est limiter les revendications des éditeurs en circonscrivant leur appropriation des droits d'auteur des chercheurs à ce qui était à l'origine sa raison d'être : la rémunération des frais engagés par l'éditeur **pour l'édition** de revues.

Les grands éditeurs internationaux, au travers de l'association STM (www.stm-assoc.org) qui les représente, mènent depuis des années un lobbying acharné contre toute contestation de cette captation étendue des droits d'auteurs, au-delà de la publication de revues. Tout changement qui tend à fragiliser le rapport de force très défavorable imposé par ces grands éditeurs aux communautés de recherche, pourtant uniques bailleuses de fonds de tout le système de la publication scientifique, doit à leurs yeux être contré coûte que coûte.

En France (où Elsevier, via sa filiale Elsevier-Masson est le premier éditeur scientifique), le SNE (Syndicat National de l'Edition) a repris purement et simplement la position et les argumentaires de l'association STM.

→ Pourquoi la position du gouvernement, misant sur la révision de la directive européenne de 2001, est-elle intenable ?

Le gouvernement, tout en reconnaissant les enjeux stratégiques de la FTD au sein de la recherche publique, s'est opposé, tant à l'Assemblée Nationale qu'au Sénat, à l'inclusion dans la loi pour une République numérique d'un article 18bis en faveur du développement de la FTD au sein de la recherche publique.

Il indique privilégier la discussion au niveau européen dans le cadre d'une révision de la directive « droit d'auteur » de 2001 annoncée – mais non encore engagée – par la Commission européenne. Au nom de la défense d'un droit d'auteur « à la française » très exigeant, le gouvernement souhaite, dans la négociation future, pouvoir bloquer de nouvelles exceptions au droit d'auteur (en particulier celle de « Fair Use ») proposées par les pays anglo-saxons. Pour le gouvernement, dans la perspective de ce futur débat, introduire aujourd'hui une exception FTD serait un mauvais signal envoyé à nos partenaires européens.

Ce calcul tactique apparaît perdant. Différer la création d'un cadre juridique sûr pour les usages de la FTD au sein de la recherche publique française ne facilitera en rien les négociations à venir au niveau européen, au contraire. **Sur cette question, le gouvernement a accepté de se laisser ligoter par les éditeurs (qui ne seront pas demain plus favorables à une nouvelle exception au droit d'auteur) au risque d'handicaper significativement la recherche publique française.**

De plus, l'introduction d'une exception FTD est tout à fait possible dans le cadre de la directive de 2001, qui prévoit déjà une exception « recherche ». C'est cette voie qu'ont suivie les législateurs britanniques et qu'explorent actuellement les pouvoirs publics allemands.

→ Pourquoi la rédaction de l'article 18bis adoptée par l'Assemblée nationale semble préférable à celle retenue par le Sénat ?

Un constat s'impose : les deux rédactions de l'article 18bis approuvées par l'Assemblée nationale et le Sénat **participent de la même volonté de favoriser le développement de la FTD au sein de la recherche publique** en limitant la possibilité pour les éditeurs de s'opposer à une réutilisation des contenus scientifiques publiés dans leurs revues par des techniques de FTD.

Cependant, la « rédaction Mélot » adoptée au Sénat limite la question de la FTD aux seuls documents textuels (où sont donc passées les données ?) et procède en instituant des limites aux contrats signés entre la recherche publique et les éditeurs, de façon à interdire à ces derniers de s'opposer aux usages de FTD sur des contenus auxquels les organismes de recherche ont légalement accès. Cette rédaction, pour rassurer les éditeurs, propose que les traitements et les résultats de ces traitements se fassent sur les moyens techniques de tiers de confiance.

Cette approche par le contrat vise clairement à ne pas agiter le chiffon rouge qu'est, pour les éditeurs, toute proposition de modification du code de la propriété intellectuelle. Mais cette approche n'est pas satisfaisante : en passant par la voie du droit des contrats la rédaction adoptée au Sénat ne tranche pas sur le fonds les problèmes incontournables de droit de propriété intellectuelle. En particulier en ne stipulant pas que les copies techniques effectuées dans le cadre d'usages de la FTD ne sont pas couvertes par le droit d'auteur elle élude un point essentiel et ouvre la voie à de nombreux contentieux juridiques. **Faute de clarté sur ce point la question de la propriété des droits d'auteurs sur les résultats des recherches ayant recours aux techniques du FTD n'a pas de réponse claire et satisfaisante.**

De plus les restrictions contractuelles imposées aux éditeurs n'empêcheront nullement ces derniers d'avoir recours à des limitations techniques (imposant par exemple que les usages de la FTD reposent sur leurs logiciels ou API).

Enfin, la rédaction Mélot dessaisit les organismes publics de recherche de la maîtrise opérationnelle des traitements de FTD en invoquant, dans une formulation peu claire, le recours aux moyens techniques de « tiers de confiance ». Or **cette solution mobilisant des « tiers de confiance » (aujourd'hui inexistantes) serait lourde et longue à mettre en place, imposant à la recherche publique française un retard préjudiciable dans la maîtrise des technologies de fouille de textes et de données.**

Parce qu'elle tranche, sous forme d'exception au droit d'auteur (au bénéfice des seuls organismes publics de recherche et avec toutes les limites déjà citées) les questions essentielles de droits de propriété intellectuelle attachés aux usages de fouille de textes et aussi de données, la rédaction de l'article 18bis adoptée à l'Assemblée nationale semble de loin préférable.

→ Une unanimité rare des milieux de l'enseignement supérieur, de la recherche et de l'économie numérique.

Les positions exprimées ci-dessus font l'objet d'un consensus rare au sein des milieux de la recherche et au-delà. Tous les responsables d'organismes publics de recherche français, tous les présidents d'Universités, toutes les associations (EPRIST¹, ADBU²) professionnelles représentant les responsables publics des structures dédiées à l'IST (Information scientifique et technique) se reconnaissent dans cette analyse. Le Conseil National du Numérique s'est aussi prononcé en ce sens. La Commission Mixte Paritaire devrait sans doute tenir compte de cette convergence sans précédent.

Une alternative : modification du Code de la recherche

Compte tenu des enjeux fondamentaux de la FTD (ou TDM) pour la recherche, une alternative à la création d'une exception au droit auteur peut être la création d'un "droit positif" par une modification adaptée du Code de la recherche. Une telle modification est compatible avec le cadre juridique en vigueur (dont la loi DADVSI qui exclut la création de nouvelles exceptions au droit d'auteur) ; elle a une portée plus générale que l'exception au droit d'auteur s'agissant des traitements numériques des matériaux de science (articles de recherche, données, images...).

Un tel amendement au Code de la recherche pourrait être ainsi formulé :

Au chapitre III du titre III du Livre V du code de la recherche, il est inséré l'article L.533-5 ainsi rédigé :

«Les données de la recherche et les écrits scientifiques dans les conditions mentionnées à l'article L. 533-4 du Code de la recherche, peuvent librement et gratuitement faire l'objet d'une copie technique à des fins d'observation, de traitement et d'exploration numérique, pour les besoins de la recherche publique et dans le respect du droit moral de l'auteur».

¹ Association des responsables de l'information scientifique et technique des organismes de recherche français

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