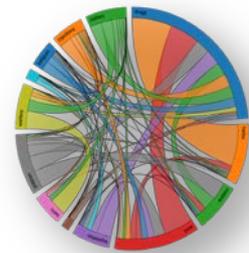


## La recherche publique pourrait rapidement disposer d'un cadre juridique sécuritaire pour porter ses développements TDM ?

---



### En bref :

Le secrétariat d'Etat à l'Enseignement supérieur et à la Recherche et le secrétariat d'Etat à l'Economie numérique organisaient le 6 avril 2016 une conférence d'une demi-journée sous l'intitulé «*Ouvrir la diffusion des savoirs par le numérique : quelles opportunités pour la recherche ?*». Ce débat intervenait à un moment particulier : le gouvernement, au travers d'un amendement déposé au Sénat dans le cadre de la discussion du projet de Loi pour une République numérique, avait rendu officiel le 5 avril son souhait de voir supprimer l'article 18bis voulu par l'Assemblée nationale. Cet article instituait une exception au droit d'auteur visant à sécuriser juridiquement les usages de la fouille de textes et de données (TDM : Text & Data Mining). Dans un mouvement tout à fait inattendu la Commission des lois du Sénat a rejeté le 6 avril l'amendement gouvernemental et a adopté un amendement de Mme Colette Mélot, sénatrice « Les Républicains » de Seine-et-Marne qui réécrit l'article 18bis en évitant toute référence au code de la propriété intellectuelle, mais qui en substance vise les mêmes objectifs. Le gouvernement désavoué sur cette question tant par les députés que par les sénateurs est dans une position inconfortable, que ses membres appellent à dépasser. Un cadre juridique sécuritaire pour le TDM au sein de la recherche publique a donc toutes les chances de voir le jour dans le cadre de la version définitive de la loi pour une République numérique.

### L'analyse d'Intelligence IST :

La conférence du 6 avril visait essentiellement à établir un lien direct entre les politiques (au plus haut niveau, celui des membres du gouvernement) en charge du dossier et la communauté de la recherche afin d'apaiser cette dernière face aux derniers développements - négatifs - résultant de l'arbitrage des services du premier ministre supprimant l'article 18bis de la loi pour une République numérique. Cet arbitrage allait totalement à rebours des arguments très clairement exprimés par la communauté de la recherche et différait *de facto* de plusieurs années la construction d'un cadre juridique sûr pour le développement des usages du TDM au sein de la recherche publique.

Cette matinée très riche a permis de mieux comprendre le très vif débat intra gouvernemental sur l'article 18bis (exception TDM au droit d'auteur) voté à une très large majorité transpartisane à l'Assemblée Nationale en janvier dernier. En raison de l'actualité (la Commission des lois du Sénat se réunissait ce même 6 avril, avec le résultat que l'on a dit) si le TDM n'a pas été le sujet unique de cette matinée (qui portait plus largement sur la « Science ouverte »), il a cristallisé l'attention en raison de la séquence politique serrée qui se joue actuellement sur ce thème.

Au fil des interventions ont été synthétisées les multiples raisons qui justifient l'instauration urgente d'un cadre juridique sécurisant pour les usages du TDM au sein de la recherche publique. La première de ces raisons – que l'arbitrage de Matignon ignore – est que le TDM 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 : à la recherche initiale débouchant sur des articles scientifiques (2 500 000 sont publiés chaque année de par le monde) va se superposer à l'avenir une recherche dérivée qui grâce aux technologies du TDM permettra d'extraire d'immenses corpus de textes et de données, permettant de produire des connaissances scientifiques nouvelles, notamment en ouvrant des approches transdisciplinaires sans précédent. Tout particulièrement celles appréhendant les phénomènes complexes impliquant de multiples interactions que seule une approche analytique des connaissances déjà codifiées permettra de mettre en évidence. Le TDM est d'abord - fait trop peu souligné - la clé d'une meilleure productivité de la recherche publique (dont les enjeux se mesurent donc en milliard d'euros) puisqu'il permettra d'exploiter autrement, dans une perspective holistique, des connaissances déjà produites (et déjà financées).

La matinée, également plus largement consacrée au Libre accès a également mis en évidence cette avancée majeure

que plus personne ne conteste (adoption de l'article 17) : la science ouverte. Elle a notamment permis de rappeler les défis sociétaux majeurs qu'une science ouverte permettra de relever, en associant la société civile à la collecte d'informations utiles à la recherche, au financement, voire à la co-construction d'objets de recherche (science participative). Mais que serait une science ouverte sans un usage libre du TDM ? Le concept même est un non-sens.

## La recherche française handicapée par l'absence d'un cadre juridique sûr pour le TDM

Mais au-delà de cette problématique du « réemploi analytique », le TDM – même si la perspective est une perspective à moyen et long termes – est une (des) technologie(s) qui va permettre (et c'est une réelle inflexion épistémologique) de faire de la science autrement (ce qui ne signifie nullement que la recherche telle qu'elle se pratique aujourd'hui soit menacée) et même de dégager de nouveaux champs de recherche. Or il faut être très net : sans cadre juridique sécurisé, il n'y aura pas de développement des usages du TDM en France. On l'a compris : les enjeux du TDM sont énormes, d'un point de vue économique, scientifique et stratégique. Se greffe donc sur cette question celle de l'excellence et de la compétitivité de la recherche française, puisque dans des cadres juridiques variés, les Etats-Unis, le Royaume-Uni et bien sûr demain la Chine (qui ne s'embarrassera pas d'arguties juridiques) et probablement l'Allemagne offrent ou offriront à leurs chercheurs un cadre juridique adapté ; il nous faut donc jouer à armes égales avec d'autres acteurs économiques majeurs.

Il y a, en effet, un point trop peu évoqué qui est pourtant essentiel dans le cadre d'une vision prospective : si les technologies du TDM sont globalement matures, leur maîtrise (en particulier dans le champ complexe de la science) se fera en suivant une « courbe d'apprentissage » assez longue. Les communautés de recherche vont beaucoup tâtonner pour mettre au point (et la problématique sera différente pour chaque discipline scientifique) le réglage du « grand télescope »<sup>1</sup> que sont les technologies TDM. En d'autres termes si un cadre juridique sécurisé n'est disponible que dans 5 ans (la voie choisie par le gouvernement étant d'attendre la révision annoncée de la directive « droit d'auteur » de 2001- cf. ci-dessous – entraînera inévitablement un tel délai), les 5 années ultérieures (a minima) nécessaires pour apprivoiser ces technologies font que la recherche française accusera un retard irrattrapable de 10 ans dans l'appropriation de ces technologies. La recherche française devra donc affronter une main liée dans le dos la compétition internationale. Est-ce bien raisonnable, pour le gouvernement, de prendre ce risque majeur pour apaiser la « sensibilité » des « industries culturelles » ? Avec tout le respect que l'on doit à ces grandes maisons, on ne voit pas en quoi Gallimard, Hachette, Universal Musique seraient légitimes pour – avec l'appui du gouvernement – stopper, au nom d'un conservatisme d'un autre âge sur les questions du droit d'auteur, la marche en avant de la recherche publique française et de ses retombées économiques ? Le paradoxe est que les quelques éditeurs scientifiques français (EDP Science par exemple) ne sont nullement contre l'article 18bis. Sur cette question, le SNE s'aligne purement et simplement sur les positions de l'association internationale STM (qui regroupe au plan mondial les éditeurs scientifiques et est dominée par les plus grands d'entre eux) : en l'occurrence - on l'a déjà dit ailleurs – le SNE n'est que le faux nez du quarteron de grands éditeurs scientifiques internationaux qui s'approprient l'essentiel des revenus et surtout des profits de l'édition de recherche.

## La voie contractuelle est impraticable et/ou renforcerait les positions des très grands éditeurs.

Les éditeurs privilégient la « voie contractuelle » pour autoriser au cas par cas les usages du TDM. En clair, les organismes de recherche, qui ont déjà signé avec les éditeurs des contrats pour la fourniture de ressources documentaires et l'accès à leurs plates-formes numériques, devraient signer des avenants encadrant les pratiques de Text & Data Mining sur les fonds documentaires acquis. Ces avenants contractuels « types » permettraient unilatéralement aux éditeurs de définir les contraintes imposées aux usagers : par exemple, sur le nombre maximal d'articles téléchargeables, sur l'emploi d'API proposées par les éditeurs comme technologie par défaut, sur le statut juridique des produits des traitements TDM (l'éditeur pourrait se revendiquer co-titulaire des droits de propriété intellectuelle sur ces résultats), etc. Outre qu'elle donnerait aux éditeurs un droit de regard exorbitant sur les orientations de ce front de recherche nouveau que va ouvrir le TDM, la voie contractuelle est tout simplement impraticable. Il y a de par le monde plus de 10 000 éditeurs de scientifiques actifs. Certes, les 5 plus grands éditeurs de recherche (Elsevier, Springer Nature, Wiley, Taylor & Francis, Wolters Kluwer) publient 24,8% des titres de revues

<sup>1</sup>L'image est de Renaud Fabre, directeur de la DIST/CNRS

publiées mondialement (mais 50,1% des revues à facteur d'impact), mais la myriade de petits éditeurs scientifiques qui constituent la « longue traîne » qui caractérise la structuration de ce secteur n'en rendent pas moins compte dans leurs revues d'avancées scientifiques significatives. Il serait tout simplement contraire à la robustesse des résultats scientifiques issus du TDM de les ignorer. Faudra-t-il, pour satisfaire à l'impératif d'exhaustivité et de diversité des corpus, que les organismes de recherche négocient en amont d'une recherche des avenants avec quelques dizaines ou centaines d'éditeurs ? Formuler la question, c'est déjà percevoir l'absurde de la « voie contractuelle ».

Mais avouons-le, cette nécessité de multiplier les contrats n'est qu'une hypothèse d'école. Dans la pratique ce seront les très grands éditeurs (Elsevier dans sa bibliothèque numérique Science Direct donne accès à plus de 14 millions d'articles scientifiques) qui mutualiseront sur leurs plates-formes les contenus d'éditeurs tiers n'ayant pas la taille suffisante pour proposer en direct leurs corpus pour des usages de TDM. Les grands éditeurs deviendront – ils en ont toujours eu l'ambition contrariée – de grands agrégateurs de corpus. La voie contractuelle en pratique risque donc de favoriser les quatre ou cinq très grands éditeurs qui seront à même de construire ces vastes corpus multi-éditeurs, renforçant, sur une dimension essentielle à la science de demain leur position oligopolistique déjà problématique<sup>2</sup>. Mais même si cette hypothèse de mutualisation des corpus disjoints par les grands éditeurs ne se concrétisait pas, les organismes de recherche, ne pouvant négocier des dizaines de contrat d'usage TDM, iraient au plus simple : signer (si la voie contractuelle était la seule praticable) avec les quatre ou cinq éditeurs qui leur fourniraient 80% des références pertinentes. On voit le risque majeur qui en découle : toute une partie du socle de connaissances – celles consignées dans les articles publiés par les plus « petits » éditeurs – risquerait de disparaître de l'écran du grand télescope TDM.

## La position des éditeurs reflète leurs propres ambitions dans le développement de contenus analytiques

On pourrait multiplier les arguments qui s'opposent à la « voie contractuelle » promue par les éditeurs : elle obligerait à travailler sur des corpus disjoints et non interopérables, fragilisant l'approche holistique des technologies TDM ; elle obligerait les scientifiques à passer par les API proposées par les éditeurs pour effectuer les traitements analytiques, privant les chercheurs de la maîtrise de leurs algorithmes ; elle donnerait - ce qui serait hautement problématique - aux éditeurs une visibilité totale sur les questionnements scientifiques des chercheurs mobilisant les techniques du TDM. En privatisant ainsi « la trace » des usages TDM (comme ils le font déjà pour les usages documentaires), les grands éditeurs, en agrégeant ces données d'usage, seraient dans une position compétitive sans équivalent pour développer eux-mêmes et vendre fort cher des contenus « dérivés » de type analytique (par exemple des notes sur les « tendances » des différents champs de recherche). Enfin, bien sûr, les éditeurs autoriseraient par avenant ces usages TDM contre redevances, alors que les sommes dédiées par les organismes de recherche à l'accès aux plates-formes numériques des grands éditeurs sont déjà très élevées en raison des marges hors norme de ces éditeurs. Le tout, il faut le rappeler sans cesse, en valorisant une matière première – la publication scientifique de recherche – qui a été entièrement produite sur budgets publics.

Tous ces arguments sont connus. Ce qui surprend c'est que le gouvernement, tirant un trait sur l'intérêt bien compris de la recherche publique et sur l'optimisation de la dépense publique scientifique, se soit cru tenu de faire droit au lobbying des éditeurs. Car Emeric Brehier, rapporteur du projet de loi à l'Assemblée et qui sur ce dossier est la courroie de transmission du gouvernement, l'a clairement indiqué dans son intervention à la conférence du 6 avril : l'arbitrage de Matignon tient à l'impérieuse nécessité de ne pas se mettre à dos les « industries culturelles » pour qui toute modification du cadre juridique du droit d'auteur, en France très protecteur, est taboue.

Cédant à l'injonction des « industries culturelles » le gouvernement fait de la non-modification du code de la propriété intellectuelle une ligne rouge à ne pas franchir. Et botte en touche en renvoyant une éventuelle exception nouvelle au droit d'auteur en faveur du TDM au sein de la recherche publique au processus annoncé de révision de la Directive européenne de 2001 relative au droit d'auteur.

<sup>2</sup>cf. note Intelligence IST du 30 mars 2016 sur le site [www.eprist.fr](http://www.eprist.fr)

## L'argumentation du gouvernement, visant à raisonner au niveau plus large de la révision de la directive de 2001 sur le droit d'auteur, apparaît fragile

Le gouvernement met en avant le fait que dans la perspective de la révision de la directive de 2001 sur le droit d'auteur, la France devra jouer une partie serrée. Tenante historique d'un droit d'auteur très protecteur elle devra contrer lors des futures négociations européennes l'approche « libérale » des pays comme le Royaume-Uni, les Pays-Bas, et les pays d'Europe du Nord qui voudraient (en résumé) introduire dans le droit européen la souplesse du « Fair Use » anglo-saxon. A quoi s'ajoute la sensibilité exacerbée des industries de contenus françaises (dont les éditeurs), déstabilisées par la révolution numérique, qui redoutent que ces négociations européennes ne mettent fin à l'exception française. Dans ce contexte de négociation prochaine dans le cadre européen, le ministère de la Culture et de la Communication a fait valoir auprès des services du Premier ministre qu'introduire une nouvelle exception au droit d'auteur aurait été un mauvais signal (de plus grande permissivité) envoyé aux autres pays européens mais aussi à des éditeurs français aux nerfs à vif sur cette question. En acceptant demain dans la négociation européenne une nouvelle exception au droit d'auteur en faveur du TDM au sein de la recherche publique, dont l'utilité s'imposera d'elle-même, la France pourra se montrer plus ferme dans son refus d'autres dispositions proposées par les « libéraux ». On remarquera que cette argumentation est une pure construction intellectuelle visant à rallier le monde de la recherche à cette « real politik » qui est le lot de tout gouvernement, mais qui sera demain totalement inopérante. Différer la création d'un cadre juridique sécurisé pour les usages du TDM au sein de la recherche publique ne facilitera en rien les négociations à venir au niveau européen. Sur cette question le gouvernement a accepté de se laisser ligoter par les éditeurs (qui ne seront pas demain plus favorable à une nouvelle exception au droit d'auteur) au risque d'handicaper significativement la recherche publique française.

D'une certaine façon c'est du Sénat qu'est venue ce même 6 avril une raison de penser que l'arbitrage de Matignon sera contourné. Adopté au Sénat, l'amendement Mélot propose une rédaction qui diffère de l'article 18bis introduit à l'Assemblée Nationale. Mais ses finalités sont les mêmes : **« Cet amendement propose d'interdire, dans les contrats conclus entre éditeurs et organismes de recherche ou bibliothèques, toute clause limitant l'accès aux publications scientifiques appartenant à l'éditeur, à des fins de fouille électronique exclusivement pour la recherche publique et à l'exclusion de tout usage commercial. Cette technique ne pourra donner lieu à rémunération ni à limitation du nombre de requêtes autorisées. Enfin, la conservation et la communication des copies techniques issues de ces traitements, aux termes des activités de recherche pour lesquelles elles ont été produites, seraient assurées par des organismes désignés par décret »**. Issu de la majorité sénatoriale, cet amendement sera à coup sûr adopté lors de la discussion plénière qui commencera le 26 avril au Sénat. La commission mixte paritaire qui sera ensuite réunie après l'examen au Sénat en raison du choix par le gouvernement de la procédure accélérée, pourrait sans doute assez facilement se mettre d'accord sur la rédaction retenue au Sénat. Si celle-ci a l'habileté de ne faire aucune référence au Code de la propriété intellectuelle, elle n'en propose pas moins objectivement, sur le point précis du TDM, une modification du droit d'auteur (en l'occurrence, elle restreint le droit des éditeurs).

## L'Assemblée nationale pourrait infliger un désaveu au gouvernement

Si la commission mixte paritaire échoue dans ses travaux (les sénateurs ont introduit, en particulier sur le thème de l'Open Data et de l'accès aux données publiques, des modifications substantielles par rapport au texte voté à l'Assemblée), c'est l'Assemblée Nationale qui aura le dernier mot. Présent au panel de la conférence du 6 avril, Jean-Yves Le Déault, député socialiste de Meurthe-et-Moselle et président de l'Office parlementaire d'évaluation des choix scientifiques et technologiques, a clairement laissé entendre que, comme en première lecture, le parlement fera jouer son droit d'amendement.

La probabilité que le gouvernement soit mis en échec sur l'article 18bis n'est donc plus invraisemblable. Grâce à la « sagesse » des parlementaires qu'il convient de saluer, la recherche publique pourrait rapidement disposer d'un cadre juridique sécurisé pour porter ses développements TDM. Ajoutons à cela que les articles 17 et 17bis qui dans les faits offrent un cadre légal au développement du libre accès ne sont plus remis en cause par personne. La communauté française de la recherche pourrait donc enregistrer dans la formulation définitive de la loi pour une République numérique deux avancées majeures.

Qui, il y a seulement deux mois, se serait risqué à pronostiquer la concrétisation de telles avancées ?