

POSITION D'EPRIST sur la mise en œuvre de la LOI "POUR UNE REPUBLIQUE NUMERIQUE"

1. Contexte : une loi qui renouvelle en profondeur les conditions du travail scientifique

La loi pour une République numérique n°2016-1321 (LRN) est une avancée majeure pour la science française, dans la mesure où elle favorise l'accès ouvert aux publications et aux données de la recherche (article 30), ainsi que la fouille de texte et de données (article 38 - FTD ou en anglais Text and Data Mining / TDM).

Mais il est essentiel de noter qu'au-delà de l'article 30 (libre accès aux écrits scientifiques de la recherche publique) et de l'article 38 (TDM), l'impact de la loi dans son ensemble sur la vie des collectifs de recherche publics et de la production scientifique qui en est issue est bien plus large :

- les jeux et bases de données produites par les administrations doivent être mises en ligne (article 6 de la LRN) ;
- les codes sources des logiciels développés par l'administration doivent être mis à la disposition du public (article 2 de la LRN) ;
- Les licences sur les données, les codes ou les bases de données sont définies dans un décret qui doit prendre en compte les pratiques des scientifiques ;
- les données publiques mises à disposition par les administrations dans le cadre de l'ouverture des données publiques (la LRN pose un principe d'open data par défaut) sont également une source de données des scientifiques.

Une analyse approfondie de l'impact de la loi, dans toutes les directions où elle modifie le travail de la science est donc nécessaire, pour être en mesure d'accompagner demain les scientifiques français. Dans ce cadre, une analyse **systemique** de la loi (qui intègre ses interactions avec les législations antérieures adjacentes) est en cours de réalisation par les services juridiques du CNRS, et sera disponible mi-janvier 2017 ; soutenue par EPRIST, cette initiative pourra nourrir un **guide d'application de la loi** auquel EPRIST propose de contribuer largement, au sein de la BSN, en mobilisant les experts des organismes de recherche et en partenariat avec d'autres acteurs de l'ESR.

Dans l'immédiat, il importe déjà de veiller à ce que les décrets d'application de la loi, en cours d'écriture, prennent en compte effectivement et confirment tous les progrès pour la science française que représente cette loi fondatrice.

2. La rédaction en cours des décrets : premières observations d'EPRIST

En effet, deux décrets relatifs au TDM, associés à l'article 38 de la loi, sont prévus sur lesquels EPRIST souhaite exprimer son point de vue.

Article 38, 1° ; Article L. 122-5 du code de la propriété intellectuelle : 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 publique.

Article 38, 2° ; Article L. 342-3 du code de la propriété intellectuelle : désignation des organismes chargés de la conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche scientifiques.

Si l'objectif de ces décrets est de sécuriser juridiquement, techniquement et commercialement l'utilisation du TDM dans le respect des intérêts et des besoins respectifs des chercheurs (larges extractions des ressources) et des éditeurs (contrôle des copies), le danger d'un retour en arrière est réel compte tenu de certaines imprécisions dans le texte actuel de l'article 38 et des actions de lobbying en cours.

Les membres d'EPRIST tiennent donc à souligner l'importance d'utiliser ces décrets pour lever les éventuelles ambiguïtés ou imprécisions du texte actuel de l'article 38, sans pour autant en trahir le sens et les objectifs initiaux. De fait, EPRIST souhaite que ces décrets ne multiplient pas les précautions dommageables au travail de la recherche et ne dressent pas *de facto* d'obstacles à l'usage du TDM pour la science publique et l'innovation françaises.

3. Co-construire l'innovation pour l'information scientifique : vers un changement de paradigme

Il va de soi cependant que nous pouvons comprendre et partager le souci de certains éditeurs français, mal préparés techniquement pour faire face à ces nouveaux usages de la recherche. Il convient donc aujourd'hui de profiter de l'occasion pour partir sur des bases saines et sereines pour le plus grand bénéfice de tous.

EPRIST tient à rappeler que les scientifiques n'ont nullement l'intention d'utiliser le nouveau cadre juridique que représente la loi « Pour une République numérique » pour contrevioler en quoi que ce soit aux droits des éditeurs privés.

Il faut au contraire comprendre que nous, acteurs publics de la recherche et acteurs privés de l'édition (publication et édition de logiciels), avons tous comme objectifs communs de valoriser les résultats de la recherche scientifique et de contribuer au développement économique de notre pays.

Il est temps de changer de paradigme et de posture tout en entrant résolument dans la modernité : les scientifiques et l'édition privée n'ont pas d'intérêts contradictoires, bien au contraire. Historiquement, ils ont été associés et il importe que l'édition française se positionne comme partenaire privilégiée de la science francophone, notamment en soutien aux SHS. Qu'il faille pour cela l'accompagner, nous le comprenons et trouvons légitime le plan de soutien du gouvernement à l'édition française. Mais nous pensons également que l'édition française se doit de repenser son offre de services à la science pour l'adapter aux pratiques actuelles et futures de la recherche, notamment en permettant à la fois une exploitation des Big Data par la science et en offrant aux productions françaises le moyen de se faire une place honorable dans cette marée de données. A cet égard, nous proposons d'analyser le champ des possibles en terme d'innovation (tous les services à la recherche que l'édition pourrait prendre en charge dans un contexte de bouleversement technologique et culturel de ce secteur), avec tous les éditeurs nationaux et internationaux, français et étrangers, et dans toutes les disciplines, car nous pensons que cette loi doit être lue comme une opportunité et non pas comme une menace par l'édition scientifique dans son ensemble.

L'exploitation des Big Data par le TDM, dont la production scientifique ne représente qu'une infime partie, offre un champ fantastique à investir pour la science, mais aussi un enjeu majeur de visibilité et de compétitivité, face au risque de dilution dans la masse accrue par la concurrence internationale. Travaillons donc de concert, scientifiques et éditeurs pour le bien de notre pays. EPRIST fera des propositions en vue d'atteindre cet objectif et examinera positivement les initiatives allant dans ce sens.

4. La rédaction des décrets relatifs à l'article 38 : recommandations d'EPRIST

C'est en appliquant cette vision des enjeux du futur qu'EPRIST propose une ligne de conduite pour la rédaction des décrets, en complément de sa contribution aux travaux de la Bibliothèque Scientifique Numérique qui ont permis de formuler des propositions de rédaction des décrets et que nous soutenons bien sûr entièrement (cf. annexe).

Mais il importe notamment ici de clarifier les processus liés à la pratique du TDM et de donner les assurances requises aux éditeurs sur les différentes étapes de la production de connaissances nouvelles par la fouille de textes et de données.

Comptons sur la directive européenne à venir pour élargir le périmètre des matériaux et types de documents concernés et profitons, dans l'immédiat, du décret pour clarifier les différentes étapes de traitement et les statuts des fichiers issus de chacune de ces étapes : le fichier d'origine (qui correspond aux données mises à disposition par chaque éditeur), les copies techniques et la constitution du corpus (qui correspondent aux traitements de mise en cohérence du corpus, composé de données issues de plusieurs sources distinctes) et enfin les fichiers finaux de résultats (issus du traitement de fouille de texte et de données appliqué au corpus constitué à l'étape 2). Cette distinction des trois étapes est essentielle car le statut et les droits associés à chaque fichier sont différents.

EPRIST estime crucial, au regard de la diversité et de la complexité des usages, de l'hétérogénéité des matériaux mobilisés, que les décrets respectent les pratiques de recherche et surtout ne les entravent pas. Il est en particulier fondamental pour la recherche :

- de disposer d'un accès sans entrave aux textes et données licites : tout obstacle contractuel ou technique que le titulaire des droits de propriété intellectuelle pourrait être tenté d'imposer aux utilisateurs autorisés du TDM est à proscrire ;
- que les modalités techniques d'accès, possiblement différentes d'un fournisseur à l'autre, reposent sur des standards garantissant un niveau minimum d'interopérabilité (usage d'un outil logiciel générique, respect des formats standards d'échange de données) ;
- que l'on prenne en compte les différentes étapes de traitement des scientifiques dans le cadre de travaux reposant sur le TDM : la réalisation du corpus informationnel doit être considérée comme une étape à part entière du processus scientifique. Par nature, le corpus constitué par le chercheur ne comprend pas uniquement des textes et encore moins des textes ou des métadonnées provenant d'un seul et même fournisseur ; en outre il nécessite des premiers traitements par le chercheur pour rendre exploitables (par fouille) les données issues de sources hétérogènes. En aucun cas, le fichier réalisé dans ce cadre ne doit être assimilé à la simple copie technique de documents originaux qui correspond, quant à elle, à une étape antérieure où le chercheur n'a pas encore apporté de valeur ajoutée ou de travail personnel ; de surcroît, nous tenons à préciser que ces corpus polymorphes ainsi constitués pour les seuls besoins de la

recherche scientifique ne permettent pas une exploitation directe par des tiers et n'offrent pas de possibilité de reconstitution et/ou de dissociation des textes (absence de réversibilité) ; la lecture des textes originaux constituant une partie seulement de ces corpus est par conséquent impossible ; des exemples concrets sur la base de cas réels seront transmis aux éditeurs membres du SNE si besoin ;

- que toutes les étapes correspondant au travail scientifique d'exploration par TDM, puissent être conservées, archivées par le scientifique, le collectif de recherche ou l'institution auquel il appartient, afin de garantir un usage scientifique dans le temps par ce scientifique ou son collectif d'appartenance ;
- que les résultats intermédiaires de fouille ainsi obtenus puissent être communiqués à la communauté scientifique (utilisateurs autorisés) ; on ne parle pas ici du résultat final du traitement par TDM, qui, en tant que données de la recherche (cf. article 38/1°), peut d'ores et déjà être archivé et diffusé (auprès de la sphère scientifique et plus largement) ;
- de ne pas associer le concept de TDM ou de fouille de textes et de données à des techniques ou technologies qui peuvent devenir très vite obsolètes dans un domaine en constante évolution ; l'exploration est le fondement même de la recherche et le TDM désigne l'ensemble des modalités de la fouille, actuelles et futures, au sein d'un ensemble donné, constitué par et pour la recherche.

Conclusion

En France comme ailleurs en Europe et dans le monde, les chercheurs et personnels de recherche affirment la nécessité d'unifier les conditions numériques du travail de la science, autour de règles universelles et partagées, définies à partir de principes d'équité, d'efficacité et d'efficience dans le partage des données scientifiques.

Compte tenu de la complexité des contraintes et de l'imbrication des intérêts anciens et nouveaux de toute nature, faire prévaloir des règles universelles et partagées nécessite une **approche systémique du texte législatif**, décrivant l'architecture des interactions numériques et physiques, personnelles et collectives, publiques et privées, qui sont à l'œuvre dans *la production des résultats scientifiques, sur les plateformes où l'information scientifique prend naissance et se partage, par la confrontation des données, des analyses, des choix documentaires.*

Cette approche systémique doit reposer sur une appréciation claire des **droits et obligations nouveaux** dont la loi numérique dote désormais la recherche française, en la faisant accéder au premier cercle international de la compétitivité à l'ère digitale.

Nous pensons donc que c'est avec cette même vision systémique et intégrée que le législateur doit aborder son analyse de l'ensemble des décrets de la loi, et tout particulièrement ceux qui s'appliquent à l'article 38.

ANNEXE

Travaux de la BSN :

PROPOSITIONS DE REDACTION POUR LES DECRETS D'APPLICATION DE L'ARTICLE 38 DE LA LOI

En tant que membres de la BSN (segments 4 sur le libre accès, 7 sur l'édition et 10 sur les données de la recherche), EPRIST a contribué, avec des scientifiques, des professionnels de l'information scientifique et techniques (membres de Couperin.org, de l'ADBU), des éditeurs scientifiques et des représentants de plateformes d'accès à des ressources numériques, à un groupe de travail mis en place par le Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche. EPRIST soutient donc, en tant que contributeur, la position des membres de BSN qui ont établi des propositions pour ce décret, dont nous re prenons ici les points essentiels :

Article 1 du décret – Les modalités de traitement des fichiers

Il nous paraît essentiel d'indiquer dans le décret lui-même (et pas seulement dans les commentaires) les trois étapes du traitement des fichiers (fichiers d'origine, copies techniques, fichiers finaux de résultats) car leur statut et les droits associés sont différents pour chacun d'eux (cf. document propositions). **Il est particulièrement important de ne pas assimiler le statut des fichiers d'origine (dont la copie est à détruire) à celui des copies techniques ayant déjà bénéficié de l'expertise du chercheur et de traitements préalables à la fouille** (que le chercheur peut conserver et communiquer à des utilisateurs autorisés, notamment dans le cadre de collaboration de recherche).

Article 2 du décret - Les conditions d'application et de régulation du TDM

Accord avec la proposition MENESR. Juste un avis concernant les définitions à éventuellement introduire : les réserver aux termes techniques et surtout ne pas redéfinir des concepts déjà bien cadrés sur le plan juridique (*source licite, commercialisation ...*).

Article 3 du décret - La sécurisation technique des plateformes des éditeurs et fournisseurs

Il ne nous paraît pas opportun de rappeler le droit de protection des réseaux et bases de données qui relève de la liberté d'entreprise des éditeurs. **Par contre, il convient de préciser que les mesures prises à cet effet ne doivent pas entraver le libre exercice d'exploration et de fouille par le chercheur et doivent garantir l'interopérabilité des données mises à disposition, sans laquelle la fouille perd tout sens.** Cette notion d'**interopérabilité**¹, que nous ajoutons au texte MENESR, est fondamentale et est devenue un des piliers de la loi actuelle "*Pour une république numérique*".

Dans le cas où un éditeur² se trouverait dans l'incapacité technique de répondre à une demande de téléchargement légitime mais importante, il devra pouvoir bénéficier de l'aide d'un ou plusieurs des organismes désignés à l'article 4 pour l'accompagner dans la mise en place d'un dispositif permettant la

1 Introduite via le concept de standard ouvert par la loi du 21 juin 2004 "*Pour la confiance dans l'économie numérique*" (<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000801164>)

2 Ce ne sera jamais un problème pour les grands éditeurs internationaux, mais cela peut l'être pour un éditeur français dont les capacités de plateforme de diffusion sont moins développées.

pratique du TDM.

Le groupe de travail de la BSN propose d'intégrer dans le décret l'obligation, pour un éditeur qui mettrait en place des contraintes techniques trop fortes incompatibles avec le TDM, de transmettre son corpus à un tiers habilité qui prendrait en charge la distribution de ses données.

En outre, les délais de livraison des données par l'éditeur doivent être compatibles avec la temporalité du projet de recherche concerné.

Enfin, on restera vigilant sur le fait que la loi annule toutes dispositions particulières qui pourraient être contractualisées par le titulaire des droits de propriété intellectuelle en vue d'entraver la pratique du TDM, notamment par une interdiction ou une limitation de l'étendue du corpus à fouiller.

Article 4 du décret - Désignation des organismes régulateurs du TDM

La seule solution qui nous paraisse compatible avec les pratiques de la recherche pour le stockage et la mise à disposition des usagers autorisés des copies techniques (étape 2 du TDM), compte-tenu des disparités volumétriques des données suivant les disciplines, est de confier cette charge aux établissements les plus proches des laboratoires et des chercheurs-producteurs de ces copies techniques, à savoir l'ensemble des établissements de recherche publique et les opérateurs de soutien à la recherche publique (ABES, ...), quel que soit leur statut (EPSCP, EPST, EPIC,...). **Chaque établissement pourra ainsi déléguer cette charge à ses différents laboratoires dans un cadre de garantie d'un usage strictement légal de ces copies et de leur sécurisation informatique ; ce qui est déjà majoritairement la pratique actuelle.**

On ne peut imaginer que le TDM puisse fonctionner à grande échelle dans un délai raisonnable sans cette solution distribuée. La BNF³ pourrait assurer une fonction centrale d'identification des corpus et d'aiguillage vers leur lieu de stockage dans la mesure où ces corpus seraient munis de descripteurs (métadonnées alimentant un répertoire ou annuaire central). Mais cette fonction centrale ne peut pas être un prérequis à l'exercice du TDM.

Il ne nous paraît pas opportun de détailler dans un décret ce mode de fonctionnement, susceptible d'évoluer au fil de la pratique. Il nous paraît beaucoup plus réaliste de proposer l'élaboration d'une **charte nationale**, signée par ces établissements, précisant leur rôle d'organisme régulateur (le terme *tiers de confiance* n'étant en fait jamais mentionné), en tant que garant de la destruction des copies initiales des originaux (étape 1 du TDM) et chargé de la conservation et de la mise à disposition des copies techniques (étape 2) aux usagers autorisés. Les fichiers résultant de la fin des travaux (étape 3), dans la mesure où ils constituent, aux termes de la loi, des données de la recherche, ne sont pas concernés par ce décret.

Ces différents points seraient à préciser lors de la rédaction de la charte nationale.

³ Ou tout autre opérateur portant des missions à caractère national.