



CONSEIL SUPERIEUR DE LA PROPRIETE LITTERAIRE ET ARTISTIQUE
- RAPPORT DE MISSION -

ANNEXES

**Transposition des exceptions de fouille de textes et de données :
enjeux et propositions**

Alexandra BENSAMOUN

Professeure des universités
Personnalité qualifiée au CSPLA
Présidente de la mission

et

Yohann BOUQUEREL

Maître des requêtes en service extraordinaire au Conseil d'Etat
Rapporteur de la mission

Ce document annexe regroupe certaines des contributions écrites à la mission sur la transposition en droit français des exceptions relatives à la fouille de textes et données des articles 3 et 4 de la directive 2019/790 du 17 avril 2019.

Ont été reproduites les contributions des organismes suivants :

- la Bibliothèque nationale de France,
- le Conseil national du numérique,
- l'Hadopi,
- l'INA,
- le MESRI.

CONSEIL SUPERIEUR DE LA PROPRIETE LITTERAIRE ET ARTISTIQUE (CSPLA)

Mission sur l'exception de fouille de textes et de données en vue de la transposition en droit interne des articles 3 et 4 de la directive 2019/790 du 17 avril 2019 dite « droit d'auteur dans le marché unique numérique »

Présidente : Mme Alexandra BENSAMOUN, professeure de droit à l'Université Paris-Saclay

Rapporteur : M. Yohann BOUQUEREL, maître des requêtes en service extraordinaire au Conseil d'Etat

La directive 2019/790 instaure **deux dispositifs en faveur de la fouille de textes et de données** (*text and data mining* ou TDM en anglais) : l'une à des **fins de recherche** exclusivement (art. 3), à laquelle ne peuvent pas s'opposer les titulaires de droit concernés, et l'autre, dite « exception ou limitation », **très générale car sans finalité prédéfinie**, qui peut faire l'objet d'une opposition des titulaires de droit (mécanisme d'*opt-out* ou option de retrait).

Le CSPLA a souhaité être éclairé sur les enjeux de la transposition en droit interne de ces textes, dans la continuité du rapport de J. Martin, président, et L. de Carvalho, rapporteure, sur l'exploration de données¹ (CSPLA, juill. 2014) et du rapport de A. Bensamoun et J. Farchy, présidentes, et P.-F. Schira, rapporteur, sur l'intelligence artificielle et la culture² (CSPLA, janv. 2020).

Ce questionnaire, à but informatif, a pour objet de simplifier votre contribution au rapport ou de guider votre audition par la mission du CSPLA. Des questions complémentaires pourront être posées lors de l'audition, le cas échéant. Dans la mesure du possible, des éléments de réponse écrits sont souhaités et peuvent être remis avant le 1^{er} novembre 2020. Ces éléments de réponse pourront rester confidentiels si vous le souhaitez. Vous pouvez évidemment faire le choix de ne répondre qu'à certaines questions ou de faire une réponse groupée.

Contributeur : BNF
Secteur concerné :

¹ <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

² <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-sur-les-enjeux-juridiques-et-economiques-de-l-intelligence-artificielle-dans-les-secteurs-de-la-creation-culturelle>

I. Questions générales

- 1) La fouille – ou exploration – de textes et de données, entendue comme l'ensemble des techniques automatisées visant à « *analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations* »³, occupe-t-elle une place importante dans votre secteur d'activité ? Pouvez-vous quantifier cette activité et donner des exemples récents ? Les activités de fouille de textes et de données dans votre secteur sont-elles toujours en lien avec l'intelligence artificielle ?

La BnF dispose d'importantes collections numériques issues de la numérisation de livres, périodiques, manuscrits, images, objets et documents audiovisuels (6 millions de documents) et de l'archivage de l'internet au titre du dépôt légal (1 Pétaoctet de données). Ces gisements documentaires font depuis 2014 environ régulièrement l'objet de demande d'accès pour de la fouille de textes et de données de la part d'équipes de recherche, principalement françaises ou européennes. La demande s'est intensifiée ces dernières années et plusieurs dizaines de projets sont actuellement en cours, essentiellement dans des disciplines rattachées aux humanités (littérature, histoire, sociologie, sciences politiques...) ou dans la recherche en informatique.

Exemples de projets :

- Un laboratoire de Traitement Automatique des Langues (TAL) d'une université parisienne utilise un corpus de poésie numérisée libre de droits pour entraîner une IA de génération de textes
- Un laboratoire d'informatique rattaché à Inria expérimente des outils de recherche sémantiques dans les images à partir d'un corpus de manuscrits médiévaux enluminés numérisés
- Un centre d'humanités numériques basé dans une école d'ingénieur en Suisse utilise des technologies d'IA pour géoréférencer d'anciens plans de Paris et reconstituer l'histoire des rues, quartiers et parcelles
- Un programme de recherche européen financé pour 5 ans, impliquant une université strasbourgeoise et une université allemande, étudie des vidéos collectées dans le cadre du dépôt légal de l'internet pour étudier l'histoire du corps (nutrition, santé, sexualité, sport...) au 20^e siècle
- Etc...

Tous ces projets n'utilisent pas nécessairement l'IA, néanmoins le recours à des techniques de type machine learning ou deep learning est de plus en plus fréquent dans ces disciplines.

Dans le cadre de ces projets, soit les œuvres concernées sont dans le domaine public, soit l'opération de TDM est couverte par l'exception dépôt légal : les corpus sont entrés dans les collections dans le cadre du dépôt légal, seule la BnF fait les actes de reproduction nécessaires au TDM (comme le lui permet l'article L 132-4 du Code du patrimoine), seuls des chercheurs accrédités mènent les études, des copies numériques des œuvres ne peuvent pas être récupérées par les chercheurs.

- 2) Pensez-vous que la fouille de textes et de données se développera dans les années à venir ? Pensez-vous que les acteurs de la fouille de textes et de données seront majoritairement européens ?

³ Directive 2019/790, article 2, 2).

Dans le domaine de la recherche scientifique et plus spécifiquement des humanités numériques, cette activité s'est déjà considérablement augmentée ces 5 dernières années du fait de la disponibilité croissante de masses de données. Les universités se sont dotées de centres d'expertise transverse à portée internationale (ex. le SCAI à Sorbonne Université) et les bibliothèques comme la BnF créent des services dédiés pour encourager ces recherches et les développer encore davantage. A la BnF, le futur Data Lab a vocation à accueillir ce type de projets, les développer et les faire connaître. Il ouvrira en 2021. Cet espace permettra de développer le TDM dans les emprises de la BnF. Les projets se développent souvent à une échelle européenne car il y a des financements, néanmoins il y a beaucoup d'acteurs français.

- 3) Quels sont les avantages et les inconvénients attendus du développement de la fouille de textes et de données, en termes de protection des droits d'auteur et assimilés, de croissance économique, de bienfaits pour les usagers ?

L'inconvénient principal pour les titulaires de droits sera qu'ils pourront s'estimer spoliés de leurs droits d'exploitation afférents à leurs œuvres (en particulier par le dispositif de l'article 4), sans qu'ils puissent en tirer une rémunération. Cela pourra les conduire à exercer l'opt-out. L'avantage économique qu'ils pourraient en retirer serait d'une part le développement des usages de leurs contenus (accroissement de la visibilité) voire la commercialisation de nouveaux services produits à partir de leurs données, et d'autre part l'accès à des techniques leur permettant de réaliser du ciblage marketing, de mieux connaître leur public, de produire le cas échéant des contenus personnalisés ou ciblés, de générer des données ou métadonnées automatiquement pour réduire leurs coûts.

S'agissant des usagers de la BnF, la fouille de textes et de données permet de développer de nouvelles approches scientifiques, d'analyser de grandes masses de données ce qui était impossible avec des méthodes manuelles, mais aussi de faire progresser les outils de type IA qui auront plus tard des applications industrielles. On voit ainsi des start-ups s'impliquer dans des projets de recherche portant sur des contenus patrimoniaux pour améliorer des algorithmes qu'ils mettront ensuite sur le marché.

- 4) Avez-vous été associés, directement ou indirectement, à l'élaboration de l'article 38 de la loi n°2016-1321 du 7 octobre 2016 pour une République numérique, qui a instauré, en faveur de la fouille de textes et de données, pour les besoins exclusifs de la recherche scientifique, une exception au monopole des auteurs pour la reproduction de leurs œuvres, ainsi qu'aux droits des producteurs de bases de données⁴ ? Ces dispositions vous semblent-elle adaptées⁵ ?

⁴ CPI, art. L. 122-5 : « Lorsque l'œuvre a été divulguée, l'auteur ne peut interdire : (...) 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques 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 ». – V. aussi art. L. 342-3, 5°.

⁵ Pour l'heure, l'exception française est doublement limitée : par son champ d'application, réservé à l'écrit, excluant ainsi la fouille sur des œuvres audiovisuelles, plastiques ou musicales ; par ses bénéficiaires, puisque seule une partie de la recherche publique est concernée, celle qui ne repose pas sur des partenariats publics-privés.

La BnF a été associée au projet de décret d'application, abandonné dans l'attente de la directive.

L'article 38 présente un certain nombre d'inconvénients :

- Pas de définition certaine de la notion d'écrits scientifiques (les sciences humaines sont-elles comprises ?)
- Caractère restrictif de l'objectif : la recherche publique (et non toute la recherche) et absence de finalité commerciale.
- Par qui sont assurées la conservation et la communication des fichiers produits ? Un tiers de confiance ?

5) Avez-vous été associés, directement ou indirectement, à l'élaboration des articles 3 et 4 de la directive 2019/790 ?

Le service juridique de la BnF a été associé dans le cadre du copyright working Group de la CENL (Conference of European National Librarians), qui fait connaître ses positions à la Commission européenne. En l'occurrence, le groupe était favorable à la création d'une exception TDM au profit des bibliothèques.

6) Quelle appréciation générale portez-vous sur ces deux articles ?

Article 3 : le dispositif manque de clarté sur les personnes pouvant, d'une part, réaliser les actes de reproduction/extraction des œuvres et données et, d'autre part, réaliser l'acte de fouille.

L'article 3 vise « *la reproduction / l'extraction effectuées par des organismes de recherche / institutions du patrimoine culturel en vue de procéder, à des fins de recherche scientifique, à une fouille de textes ou de données* ». Il y a donc l'idée que le processus peut s'exécuter en deux voire trois étapes : (1) la reproduction/l'extraction d'œuvres/données, (2) la fouille proprement dite sur ce corpus, et éventuellement (3) l'extraction/reproduction d'œuvres/données dans le cadre de la production des résultats de la fouille.

On peut imaginer soit que le processus est immédiat (le lancement du logiciel ou script par une personne entraîne l'enchaînement des 3 étapes – donc une seule personne met en œuvre l'exception), soit qu'il puisse être exécuté en plusieurs temps et donc théoriquement réalisé par différentes personnes (ex : la BnF produit un corpus thématique, et un chercheur réalise la fouille).

La question se pose donc de savoir qui, dans le respect de la directive, est fondé à réaliser les actes de reproduction/extraction et qui peut réaliser la fouille, puisque ces deux étapes peuvent être distinguées (l'éventuelle troisième étape rejoignant la première).

Concernant l'acte de reproduction/extraction :

A priori, les reproductions / extractions d'œuvres/données doivent être réalisées par des « organismes de recherche et institutions du patrimoine culturel » (l'article 3 de la directive est plutôt clair sur ce point dans sa formulation).

Toutefois, on pourrait considérer que des tiers pourraient directement réaliser les actes de reproduction/extraction à la lecture du considérant 14 qui associe les « personnes rattachées » à l'organisme de recherche / institution du patrimoine culturel à l'exception relative à la fouille (donc dans son intégralité), et non seulement à l'acte de fouille proprement dit (cf. infra sur la notion de « personne rattachée »). Cela signifierait que dès que la source des données/œuvres est licite, toute personne qui justifie d'un *rattachement* à l'organisme de recherche peut invoquer l'exception de l'article 3. Cette approche large n'est pas anodine car cela est susceptible de couvrir des corpus d'œuvres très importants, par exemple au regard de

l'exception bibliothèque qui permet à la BnF de numériser et rendre accessible à tous dans ses emprises des œuvres encore protégées provenant de ses collections et à des fins recherche (les finalités des deux exceptions se rejoignant). Cela étant, cette « étape » de reproduction/extraction peut être évitée dans le cas où le logiciel de fouille peut-être lancé sans nécessité d'actes de reproduction/extraction, la reproduction ayant été au préalable faite par la BnF par la seule mise à disposition de la base de données dans son ensemble.

Concernant l'acte de fouille :

La reproduction/extraction d'œuvres et données dans le cadre d'une fouille est nécessairement un acte compris dans le monopole d'exploitation, ce qui explique la création d'une nouvelle exception ; en revanche, on peut s'interroger sur l'acte de fouille qui est un *procédé d'analyse*, et donc pas un acte d'exploitation en tant que tel.

Le considérant 9 de la directive rappelle à ce titre que « la fouille (...) peut également porter sur de simples éléments factuels ou données non protégés par le droit d'auteur et dans ce cas, aucune autorisation au titre de la législation sur le droit d'auteur n'est nécessaire. Il peut également exister des situations dans lesquelles la fouille (...) n'entraîne pas d'actes de reproduction (...) ».

Dans la mesure où l'acte de fouille est matériellement détachable de l'acte de reproduction/extraction, on peut s'interroger si celui-ci est un acte régi par le monopole. S'il n'est pas régi par le monopole, toute personne autre que l'organisme de recherche / institution du patrimoine doit pouvoir le réaliser. S'il l'est (ou si l'on considère que bien que distinguable matériellement, il ne l'est pas intellectuellement), la question des personnes fondées à réaliser la fouille autres que l'organisme/l'institution se pose.

Si l'on se place dans la deuxième hypothèse, l'article 3 §1 peut être lu de deux manières quant à l'identification de la personne qui réalise la fouille : (1) une lecture stricte où seul l'organisme de recherche / institution du patrimoine culturel est autorisé à le faire (la personne qui reproduit/extrait est la même qui fouille) ; (2) une lecture plus souple qui autoriserait que la personne qui fouille soit différente de celle qui a préalablement reproduit/extrait les œuvres. La lecture stricte semble plus conforme à la lettre de l'article 3, mais une lecture plus souple semble soutenue par les considérants de la directive.

Le considérant 14 de la directive précise : « l'exception (...) devrait s'appliquer aux organismes de recherche et aux institutions du patrimoine culturel, y compris aux personnes qui y sont rattachées, pour ce qui est des contenus auxquels ils ont accès de manière licite ».

Quelles sont ces « personnes » « rattachées » aux organismes et institutions ? Sont-elles leur personnel – mais la précision aurait peu de sens, une personne morale ne pouvant elle-même réaliser de la fouille de données ?

Sont-elles simplement leurs filiales ?

Sont-elles « leurs » chercheurs, ceux-ci pouvant avoir différents statuts ?

- Chercheurs salariés : inclus dans l'exception.
- Chercheurs subventionnés : exclus de l'exception ?
- Chercheurs salariés de partenaires publics : inclus dans l'exception, si le partenaire est un organisme de recherche / institution du patrimoine culturel ?
- Chercheurs indépendants (ex : personnes habilitées pour accéder aux collections du dépôt légal au titre de l'article L132-4 du code du patrimoine) : exclus de l'exception ?

Sont-elles plus largement leurs usagers, chercheurs ou simples lecteurs ? Le considérant 14 ajoute, à l'appui de cette hypothèse, à propos de l'accès licite aux œuvres, que « les personnes rattachées à un organisme de recherche (...) qui a souscrit des abonnements à certains contenus sont réputées bénéficiées de l'accès licite à ces abonnements ». Les abonnements à des contenus peuvent être faits pour le personnel mais aussi pour les usagers (sur place ou à distance). Les « personnes rattachées » pourraient donc bien être les usagers de l'organisme.

Dans le cas où serait considéré que l'acte de fouille est en tant que tel compris dans le monopole des ayants droit, l'exception devrait être transposée en tenant compte de cette notion de « personnes rattachées » afin que tant les agents (qui sont nécessairement inclus dans le dispositif) que les usagers de l'organisme/institution puissent réaliser des fouilles.

Sont-elles, enfin, leurs partenaires (et par extension leurs agents) ? Auquel cas il faudrait faire un lien avec le considérant 11 relatif aux partenariats public-privé : les organismes et institutions « devraient pouvoir s'appuyer sur leurs partenaires privés pour effectuer des fouilles (...) ».

Etant précisé qu'on ne définit pas ce qu'est un « *partenariat* public-privé » : on pourrait envisager une interprétation large pour couvrir tout type de situation du moment que la finalité de recherche est respectée, et intégrer à la fois des personnes morales et des personnes physiques. Une interprétation trop large du « partenariat public-privé » pourrait toutefois être mal perçue des ayants droit (tout comme une interprétation trop large de la notion de « personnes rattachées » pour la réalisation de l'acte de reproduction et/ou de la fouille).

Par ailleurs, la constitution des corpus en vue de procéder à de la fouille a un coût, qui peut se révéler d'autant plus important s'il se révélait que la BnF est la seule fondée juridiquement à procéder aux reproductions/extractions. La BnF pourra-t-elle tarifier ce service ?

L'article 4 pose encore davantage de questions. Le dispositif est ouvert à tous (mais la BnF pourra servir de source pour les documents) et les actes de reproduction pourront être réalisés par chacun, sans avoir à passer par une institution culturelle. La façon dont cela sera fait à la BnF reste à déterminer. A ce titre, les usagers de la BnF souhaitant faire du TDM devraient pouvoir le faire en salle de lecture ou sur Gallica (la bibliothèque numérique de la BnF). Qui sera par ailleurs en charge de la conservation des reproductions et des extractions (en particulier quand les documents proviennent de la BnF) ? La BnF aura-t-elle en outre à signaler les documents ne pouvant donner lieu à du TDM (par exemple dans son catalogue général) et comment ces signalements lui seront-ils transmis ?

Sachant par ailleurs que la BnF conserve une quantité très importante et variée d'œuvres protégées susceptibles de faire l'objet d'une fouille : un ayant droit pourra-t-il autoriser ou refuser le dispositif œuvre par œuvre (et souhaite-t-on ce que soit possible) ?

Pour les collections du dépôt légal du web (à supposer que le dispositif de l'article 3 ne s'applique pas aux chercheurs habilités, et qui souhaiteraient utiliser le dispositif de l'article 4) : n'y-a-t-il pas un risque que cela devienne très difficilement gérable, au regard du fait que les sites Internet sont des œuvres massivement composites ?

Enfin, pour ces deux articles, la responsabilité de la BnF pourrait-elle se trouver engagée en cas d'application inadéquate de ces dispositifs par ses usagers ? S'exposerait-elle à une complicité de contrefaçon par fourniture des moyens ?

Dans le cadre de l'article 4, pour que la BnF ne puisse voir sa responsabilité engagée, sera-t-il suffisant de simplement signaler aux usagers les œuvres exclues du dispositif par les ayants droit, ou devra-t-elle prendre des mesures techniques pour empêcher le TDM sur les œuvres retirées du dispositif ?

En outre, si l'on se situe dans le deuxième cas et que l'on prend l'exemple de Gallica Intramuros (dont une partie des documents encore protégés sont numérisés et rendus accessibles dans le cadre de l'exception bibliothèque), il conviendrait que le système permette aux logiciels de fouille de tourner sur toutes les œuvres (y compris les œuvres ayant fait l'objet d'un *opt out*) à l'exception de ceux qui supposent un acte de reproduction ou une extraction à un moment du procédé, pour les œuvres ayant fait l'objet d'un *opt out* seulement. Cela résulte du considérant 9 qui souligne qu'il faut distinguer les actes de fouille nécessitant

une autorisation préalable des ayants droit (car nécessitant un acte de reproduction pour réaliser la fouille), de ceux ne nécessitant pas d'autorisation préalable.

Eu égard à ces interrogations juridiques et techniques, la BnF pourrait-elle décider, en dernier recours et pour tout ou partie de ses bases, temporairement ou définitivement, de ne pas rendre techniquement possible le dispositif de l'article 4 si cela se révèle trop complexe à mettre en place ?

- 7) Avez-vous connaissance d'une transposition des articles 3 et 4 de la directive 2019/790 dans d'autres Etats membres ?

Nos collègues juristes des bibliothèques nationales de Finlande, de République tchèque et des Pays-Bas nous ont indiqué que ces articles sont en cours de transposition dans leurs lois nationales respectives.

- 8) Disposez-vous d'éléments de comparaison entre le cadre juridique européen en matière de fouilles de textes et de données et les règles applicables en dehors de l'Union européenne (par exemple aux Etats-Unis et au Japon) ?

Non

- 9) Pensez-vous que, dans le cadre juridique français actuel, les fouilles de textes et de données peuvent être réalisées, en tout ou partie :
- en dehors des règles du droit d'auteur et assimilés, pour les données brutes et données publiques notamment ?
 - sur le fondement de l'exception de citation ?
 - sur le fondement de l'exception au profit de la recherche ?
 - sur le fondement de l'exception de reproduction technique provisoire ?

Nous ne pensons pas que d'autres fondements juridiques pourraient permettre le TDM :

- Données brutes et données publiques :
Le droit des informations publiques n'interdit pas le TDM (hors présence de secrets protégés ou des droits de propriété intellectuelle de tiers sur les documents – mais on retombe alors dans la problématique droits de propriété intellectuelle nécessitant une exception).
Pour les données brutes, s'il n'y a pas de droits de propriété intellectuelle, il est difficile d'imaginer un fondement juridique interdisant le TDM à partir du moment où les données sont rendues librement accessibles par la personne qui en est à l'origine.
- Courte citation : le TDM n'implique pas nécessairement qu'une œuvre soit citée et il n'y a pas systématiquement intégration dans une œuvre nouvelle.
- Exception recherche : non (en particulier pour l'article 4).
- Exception reproduction technique provisoire : non, la finalité n'est pas la même et la copie n'est pas forcément provisoire.

Toutefois, la BnF réalise des opérations de TDM sur le fondement de l'exception du dépôt légal (article L 132-4 du Code du patrimoine) avec des chercheurs (cf réponse au I-1 pour les modalités).

- 10) Si vous êtes éditeurs de presse, comment imaginez-vous l'articulation entre le nouveau régime des droits voisins, mentionné à l'article 15 de la directive 2019/790, et les dispositifs prévus aux articles 3 et 4 du même texte ? Souhaitez-vous attirer l'attention de la mission sur des problématiques spécifiques ?

II. Questions sur l'article 3 de la directive 2019/790

Article 3

Fouille de textes et de données à des fins de recherche scientifique

1. *Les États membres prévoient une exception aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel, en vue de procéder, à des fins de recherche scientifique, à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont accès de manière licite.*
2. *Les copies des œuvres ou autres objets protégés effectuées dans le respect du paragraphe 1 sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche.*
3. *Les titulaires de droits sont autorisés à appliquer des mesures destinées à assurer la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés. Ces mesures n'excèdent pas ce qui est nécessaire pour atteindre cet objectif.*
4. *Les États membres encouragent les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir d'un commun accord des bonnes pratiques concernant l'application de l'obligation et des mesures visées aux paragraphes 2 et 3, respectivement.*

Pour précision, cette exception est **obligatoire** et on ne peut y déroger par le contrat (dir. 2019/790, art. 7.1⁶).

- 1) Le cadre posé et les conditions de mise en œuvre de l'exception (bénéficiaires identifiés, fins de recherche scientifique exclusivement, accès licite, absence de compensation du fait du préjudice minime⁷) vous semblent-ils satisfaisants ?

Voir la réponse sous la question I-6

Il est étonnant que l'exception prévue à l'article 3 ne vise pas la directive 2009/24/CE relative aux logiciels, contrairement à l'exception de l'article 4 qui vise bien cette directive : est-ce une erreur ou volontaire ? Y-a-t-il des conséquences à en tirer ?

- 2) Les partenariats public-privé peuvent bénéficier de cette exception « académique »⁸, ce qui en étend considérablement la portée. On pourrait par exemple imaginer, grâce à un tel

⁶ « Toute disposition contractuelle contraire aux exceptions prévues aux articles 3, 5 et 6 est non exécutoire. »

⁷ V. cons. 17.

⁸ Dir. 2019/790, cons. 11 : « (...) En concordance avec l'actuelle politique de la recherche de l'Union, qui encourage les universités et les instituts de recherche à collaborer avec le secteur privé, les organismes de recherche devraient également bénéficier d'une telle exception lorsque leurs activités de recherche sont menées

partenariat, de lancer une solution d'IA créative musicale sur une base de données d'archives comme celle de l'INA, ou d'IA créative graphique et plastique sur un fonds muséal. Comment transposer cette ouverture ?

La transposition claire et sans ambiguïté dans le droit national de la possibilité d'utiliser l'exception de l'article 3 dans le cadre de partenariats public-privé semble essentielle.

On peut également s'interroger sur les partenariats public-public dont l'un des partenaires n'entre pas dans la catégorie d'organisme de recherche ou d'institution du patrimoine culturel.

- 3) Afin de favoriser les activités de recherche et l'interopérabilité entre les œuvres, seriez-vous favorables à l'utilisation d'un format standardisé dans votre secteur d'activité (par exemple XML pour les textes et MIDI pour la musique) ?

C'est toujours mieux d'avoir des formats normalisés et d'une façon générale nous œuvrons à leur usage le plus répandu lorsqu'ils existent, mais il ne semble pas souhaitable d'aller jusqu'à ce niveau de détail dans la transposition de la directive. Certains domaines n'ont pas encore normalisé de format d'échange, cela paraît prématuré.

- 4) Comment vérifier que la durée de conservation des copies des œuvres et des autres objets protégés, mentionnée à l'alinéa 2, n'est pas excessive ?

L'article 3 ne vise pas à proprement parler de durée de conservation, mais la conservation « à des fins recherche ». Tant que cette finalité est respectée, les copies des œuvres devraient pouvoir être conservées sans limitation de durée.

Les corpus devraient aussi pouvoir être conservés / archivés une fois la recherche initiale terminée pour des recherches ultérieures.

- 5) Quelles modalités souhaiteriez-vous voir mises en place pour la conservation des copies⁹ ?

Les organismes et institutions ne devraient être tenus (à supposer que cela soit pertinent du point de vue de la recherche entreprise) que de conserver les corpus créés et fouillés par eux.

Ils ne devraient être « tiers de confiance » que dans le cadre d'accords conclus à titre gratuit ou onéreux.

La conservation de ces copies, du point de vue de la BnF, peut s'entendre de 2 façons :

- Lorsque la BnF réalise l'extraction/reproduction, elle peut souhaiter la conserver si elle a représenté un coût ou un effort important et qu'elle pourrait être réutilisée par d'autres projets de recherche. Mais, elle ne devrait pas non plus être tenue de le faire car cette conservation représente également un coût. S'il y a conservation, elle se fait a

dans le cadre de partenariats public-privé. Les organismes de recherche et les institutions du patrimoine culturel devraient rester les bénéficiaires de cette exception, mais ils devraient pouvoir s'appuyer sur leurs partenaires privés pour effectuer des fouilles de textes et de données, y compris en utilisant leurs outils technologiques. »

⁹ V. cons. 15 : « Dans certains cas, par exemple aux fins de la vérification a posteriori des résultats de la recherche scientifique, les organismes de recherche et les institutions du patrimoine culturel pourraient avoir besoin de conserver des copies faites dans le cadre de l'exception aux fins d'effectuer de la fouille de textes et de données. En pareils cas, les copies devraient être stockées en lieu sûr. Les États membres devraient être libres de décider, au niveau national et après concertation avec les acteurs concernés, des modalités plus précises de conservation des copies, notamment la capacité de désigner des organismes de confiance aux fins du stockage de ces copies. Afin que l'application de cette exception ne soit pas restreinte indûment, ces modalités devraient être proportionnées et limitées à ce qui est nécessaire à la conservation des copies en toute sécurité et à la prévention de leurs utilisations non autorisées. »

priori dans les mêmes conditions que la conservation des œuvres elles-mêmes (exemple typique : à partir des fichiers WARC des archives web, on génère un fichier WAT qui correspond à un standard préparé pour la fouille. Cette génération est lourde à réaliser, on pourrait vouloir reverser ces WAT dans SPAR, qui est le dispositif de conservation de la BnF, et les garder.)

- Lorsque l'extraction/reproduction est remise au chercheur pour réaliser la fouille, celui-ci pourrait vouloir la conserver au même titre que les résultats de la fouille à des fins de vérification scientifique. Dans ce cas, il pourrait se tourner vers la BnF pour demander cette conservation (qui serait sans doute payante) ou utiliser une infrastructure adéquate dans le domaine de la recherche (par ex. le CINES). La 2^e option est typiquement plus conforme au processus actuel de la BnF car si ce sont les chercheurs qui veulent conserver un corpus (qu'il s'agisse d'un corpus source ou des résultats de la fouille) aujourd'hui on considère que ce n'est pas à la BnF de fournir ce service. La BnF pourrait être rapidement dépassée par l'ampleur de la demande et en outre, cela pourrait menacer l'équilibre de ses relations avec l'ESR et la répartition des responsabilités de chacun. NB : de toute façon, actuellement, lorsqu'il y a fouille d'un corpus sous droits, a priori l'extraction/reproduction n'est pas conservée par la BnF et elle ne la remet pas non plus au chercheur qui est censé l'utiliser dans les emprises et ne pas en faire lui-même de copie.

- 6) Quelles pourraient être les mesures mises en œuvre par les titulaires de droits pour assurer la sécurité et l'intégrité des réseaux, structures et bases de données sollicités, en permettant le TDM ?

Question qui relève plutôt des titulaires de droits

Mais certains titulaires de droits (typiquement les éditeurs de presse) seraient assez satisfaits que le TDM ne se fasse pas sur leurs infrastructures, car cela représente un coût important pour eux (sécurisation + performance machine). S'ils pouvaient renvoyer les personnes qui souhaitent fouiller vers un tiers de confiance comme la BnF ils n'auraient pas à supporter ces coûts. Certains ont rapporté que la charge des usagers qui aspirent leur site plus ou moins licitement à des fins de fouille est de plus en plus considérable.

- 7) Comment le gouvernement pourrait-il encourager les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir conjointement des bonnes pratiques pour l'application de cet article 3 ? Seriez-vous favorable à l'élaboration rapide d'une charte des bonnes pratiques entre les acteurs concernés ? Souhaiteriez-vous que ce processus soit mené dans le cadre du CSPLA ?

Une charte des bonnes pratiques serait en effet utile et pourrait être élaborée par le CSPLA.

III. Questions sur l'article 4 de la directive 2019/790

Article 4

Exception ou limitation pour la fouille de textes et de données

1. Les États membres prévoient une exception ou une limitation aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE, à l'article 4, paragraphe 1, points a) et b), de la directive 2009/24/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions d'œuvres et d'autres objets protégés accessibles de

manière licite aux fins de la fouille de textes et de données.

2. Les reproductions et extractions effectuées en vertu du paragraphe 1 peuvent être conservées aussi longtemps que nécessaire aux fins de la fouille de textes et de données.

3. L'exception ou la limitation prévue au paragraphe 1 s'applique à condition que l'utilisation des œuvres et autres objets protégés visés audit paragraphe n'ait pas été expressément réservée par leurs titulaires de droits de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne.

4. Le présent article n'affecte pas l'application de l'article 3 de la présente directive.

- 1) La directive semble donner aux Etats membres une certaine marge de manœuvre lors de sa transposition en droit interne, renvoyant à une approche fondée sur une « exception » ou sur une « limitation » des droits d'auteurs et assimilés. Quelles conséquences en tirez-vous ?

Les titulaires de droits accepteraient beaucoup plus difficilement le principe d'une exception. Une limitation serait peut-être davantage envisageable, et elle le serait encore plus si une compensation au profit des titulaires de droits était prévue (mais est-ce possible et cela ne risque-t-il pas de dissuader les acteurs du TDM ?).

Si une compensation est mise en place, le mode de calcul et de versement devrait être simplifié au regard de la masse d'œuvres et d'ayants droit potentiels.

- 2) Le cadre et les conditions de mise en œuvre (tous bénéficiaires, tous usages, y compris commerciaux, accès licite) vous paraissent-ils satisfaisants ?

Voir la réponse sous la question I-6

- 3) Comment comprenez-vous le droit d'option ouvert au titulaire et qui permet de se retirer du dispositif ?

C'est la faculté pour les titulaires de droits d'interdire que leurs œuvres fassent l'objet de TDM. Il reste toutefois à déterminer comment ils signaleront cette intention.

- 4) Quels seraient, selon vous, les moyens techniques appropriés (« procédés lisibles par machine ») que les titulaires de droits pourraient utiliser pour s'opposer (« *opt-out* », *i.e.* option de retrait) à une fouille de textes et de données réalisée par des organismes autres que ceux mentionnés à l'article 3 (métadonnées, conditions générales d'utilisation, accords contractuels, déclarations unilatérales, fichier centralisé, mesures techniques de protection, ou autres moyens) ? Faudrait-il standardiser ces moyens ? *Quid* en cas de doute sur l'exercice de l'*opt-out* ?

La solution d'un fichier centralisé serait préférable (comme la base ReLIRE pour les indisponibles, gérée par la BnF).

Il serait pertinent que la déclaration d'*opt out* soit standardisée à la fois dans la formulation et dans la transmission.

Techniquement ce serait quand même très lourd à gérer pour la BnF.

5) Pensez-vous exercer cette option de retrait ?

Sans objet

6) Pensez-vous que les titulaires de droits dans votre secteur d'activité vont s'opposer massivement à la fouille de textes et de données pour des besoins autres que ceux de la recherche scientifique ?

C'est très possible. En particulier les éditeurs.

7) En cas d'opposition des titulaires de droit à la fouille de textes et de données, les organismes qui souhaitent réaliser ces fouilles pourront-ils se fonder sur les dispositions citées à la question 9 du I) du présent questionnaire (autres exceptions/limitations) ?

Nous estimons que ces autres fondements sont inapplicables en l'espèce (sous réserve de l'exception dépôt légal, précédemment évoquée).

8) Comment apprécier la durée de conservation des reproductions et extractions de données ? Faudrait-il prévoir une sanction en cas de durée de conservation excessive ?

La durée de conservation devrait être la même que celle des droits d'accès obtenus par la BnF. En cas de conservation excessive, les auteurs ou ayants droit devraient pouvoir intenter une action en contrefaçon, puisque les conditions de l'exception ne seraient pas respectées si bien que cela constituerait une atteinte à un droit de propriété intellectuelle. Une sanction spécifique ne semble pas nécessaire.

9) Pensez-vous que cette exception est susceptible de limiter très fortement l'intérêt à investir dans des bases de données (par exemple, les bases de données d'images ou de textes élaborées par les éditeurs) ?

Le système d'*opt out* devrait être suffisamment protecteur pour des bases de données totalement originale / nouvelle, car la question « d'être en *opt out* » pourra être traitée en amont de la diffusion au public.

10) A quelles conditions les titulaires de droit pourraient-ils accepter les fouilles de textes et de données mentionnées à l'article 4 de la directive ? Le recours à une licence prévoyant des contreparties financières vous semble-t-il adapté ?

Une compensation serait en effet peut-être une bonne motivation.

11) Pensez-vous que la fouille de textes et de données dans un but commercial pourrait paradoxalement être freinée en cas de nombreuses options de retrait de la part des titulaires de droit ?

En effet. De trop nombreuses oppositions rendraient le dispositif inefficace et peu utilisable.

12) D'un point de vue technique, comment peut-on mettre en œuvre les dispositions de cet article 4 sans faire obstacle aux dispositions de l'article 3 de la directive¹⁰ ?

¹⁰ L'article 4.4 de la directive l'impose : « Le présent article n'affecte pas l'application de l'article 3 de la présente directive. »

A priori, les modalités de l'article 4 ne s'opposent pas à celles de l'article 3 : les bénéficiaires ne sont pas les mêmes, l'objet du TDM est différent (la recherche pour l'article 3), l'opt-out ne devrait pas pouvoir empêcher la recherche.

- 13)** Quelles pourraient être les règles communes aux dispositifs prévus aux articles 3 et 4 de la directive (par exemple, durée de conservation des reproductions et extractions) ? Une approche sectorielle vous paraît-elle utile ?

En effet, les durées de conservation pourraient être les mêmes. Mais il pourrait également être considéré que la recherche justifierait que les durées de conservation soient plus longues.

IV. Questions diverses

- 1)** Quelles seraient les précisions relatives à la fouille de textes et de données que vous souhaiteriez voir inscrites dans les textes législatifs et réglementaires de transposition des articles 3 et 4 de la directive 2019/790, sous réserve qu'elles ne soient pas contraires à la lettre et à l'esprit de ces dispositions ?

Cf réponses précédentes

- 2)** En cas d'expertise technique sur le TDM, quels formats utilisez-vous ? Avez-vous connaissance d'un document scientifique récapitulatif sur ce point ?

Pas de réponses à apporter

- 3)** Quelles autres remarques, générales ou spécifiques, souhaitez-vous formuler auprès de la mission ?

Pas de remarques supplémentaires

CONSEIL SUPÉRIEUR DE LA PROPRIÉTÉ LITTÉRAIRE ET ARTISTIQUE (CSPLA)

Mission sur l'exception de fouille de textes et de données en vue de la transposition en droit interne des articles 3 et 4 de la directive 2019/790 du 17 avril 2019 dite « droit d'auteur dans le marché unique numérique »

Présidente : Mme Alexandra BENSAMOUN, professeure de droit à l'Université Paris-Saclay

Rapporteur : M. Yohann BOUQUEREL, maître des requêtes en service extraordinaire au Conseil d'Etat

La directive 2019/790 instaure **deux dispositifs en faveur de la fouille de textes et de données** (*text and data mining* ou TDM en anglais) : l'une à des **fins de recherche** exclusivement (art. 3), à laquelle ne peuvent pas s'opposer les titulaires de droit concernés, et l'autre, dite « exception ou limitation », **très générale car sans finalité prédéfinie**, qui peut faire l'objet d'une opposition des titulaires de droit (mécanisme d'*opt-out* ou option de retrait).

Le CSPLA a souhaité être éclairé sur les enjeux de la transposition en droit interne de ces textes, dans la continuité du rapport de J. Martin, président, et L. de Carvalho, rapporteure, sur l'exploration de données¹ (CSPLA, juill. 2014) et du rapport de A. Bensamoun et J. Farchy, présidentes, et P.-F. Schira, rapporteur, sur l'intelligence artificielle et la culture² (CSPLA, janv. 2020).

Ce questionnaire, à but informatif, a pour objet de simplifier votre contribution au rapport ou de guider votre audition par la mission du CSPLA. Des questions complémentaires pourront être posées lors de l'audition, le cas échéant. Dans la mesure du possible, des éléments de réponse écrits sont souhaités et peuvent être remis avant le 1^{er} novembre 2020. Ces éléments de réponse pourront rester confidentiels si vous le souhaitez. Vous pouvez évidemment faire le choix de ne répondre qu'à certaines questions ou de faire une réponse groupée.

Contributeur : [Conseil national du numérique](#)

Secteur concerné : [secteur numérique - commission consultative indépendante](#)

¹ <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

² <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-sur-les-enjeux-juridiques-et-economiques-de-l-intelligence-artificielle-dans-les-secteurs-de-la-creation-culturelle>

I. Questions générales

- 1) La fouille – ou exploration – de textes et de données, entendue comme l'ensemble des techniques automatisées visant à « *analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations* »³, occupe-t-elle une place importante dans votre secteur d'activité ? Pouvez-vous quantifier cette activité et donner des exemples récents ? Les activités de fouille de textes et de données dans votre secteur sont-elles toujours en lien avec l'intelligence artificielle ?

Le Conseil national du numérique s'est positionné à plusieurs reprises en faveur de la fouille – ou exploration – de textes et de données.

En premier lieu, le Conseil national du numérique a proposé d'instaurer une exception au droit d'auteur autorisant la fouille de texte et de données pour la recherche en 2015, notamment dans le cadre de son avis relatif au projet de loi pour une République numérique⁴. Cet avis précisait que : « *la fouille de textes et de données désigne un ensemble de traitements informatiques consistant à extraire de connaissances selon un critère de nouveauté ou de similarité dans des textes ou des bases de données. Elle permet par exemple la recherche de « signaux faibles » difficiles à appréhender par la lecture cursive, le repérage ou l'analyse de comptes rendus d'expérimentations ratées. Elle a été considérée comme porteuse de nombreux potentiels pour la découverte scientifique et le développement de nouvelles connaissances. Elle doit permettre à la recherche de profiter des avancées en matière de l'analyse des mégadonnées ("big data"), qui a vocation à devenir un enjeu majeur de la compétitivité internationale. L'Irlande, le Royaume Uni mais aussi les Etats-Unis ou encore le Japon le permettent aujourd'hui.*

Considérant que :

- *La fouille automatisée de textes et de données, en tant qu'activité de lecture et d'extraction d'informations, est une pratique qui ne se distingue pas fondamentalement du relevé manuel des informations qui a toujours été effectué par la recherche ;*
- *Le droit d'auteur, qui protège la forme d'expression et non les idées, permet aujourd'hui de lire et de réutiliser des informations ou données incluses dans un texte sur lequel on a obtenu un droit d'accès ;*

Le CNNum estime qu'il n'y a pas de raison légitime à restreindre ce droit dans le cadre d'un traitement automatisé. Les grands éditeurs qui détiennent la majeure partie des publications scientifiques, peuvent aujourd'hui proscrire, par des solutions contractuelles, la fouille de textes et de données –notamment la copie provisoire, techniquement nécessaire afin de la réaliser –aux chercheurs, même lorsque ces derniers disposent d'un accès légal à l'ensemble des publications scientifiques comprises dans les bases de données fouillées. Cette interdiction s'appuie notamment sur le droit sui generis des bases de données. Cette pratique nécessite donc la création d'une exception au droit d'auteur, sur la base d'une réinterprétation de l'exception pour la recherche, à l'image de l'interprétation du Royaume-Uni. Compte tenu des limites et contraintes imposées par les solutions contractuelles, le CNNum recommande donc d'instaurer une véritable exception au droit d'auteur autorisant la fouille de textes et de données. »

³ Directive 2019/790, article 2, 2).

⁴ CONSEIL NATIONAL DU NUMÉRIQUE, Avis n°2015-3 relatif au projet de loi pour une République numérique, novembre 2015, p. 15-16. Disponible en ligne : <https://cnnumerique.fr/files/2017-10/Avis-du-CNNum-sur-le-projet-de-loi-numerique.pdf>.

En second lieu, **cette position a été approfondie dans le cadre du rapport du député Cédric Villani intitulé « Donner un sens à l'intelligence artificielle »⁵ de mars 2018, auquel le secrétariat général du Conseil a contribué, et dans la feuille de route du Conseil sur l'environnement et le numérique de juillet 2020⁶.**

Dans le cadre du rapport de Cédric Villani, il était recommandé de favoriser sans attendre les pratiques de fouille de texte et de données étant précisé que *« le cadre juridique européen doit favoriser les nouveaux usages sur la donnée. Dans cet objectif, la réforme en cours du cadre juridique relatif au droit d'auteur et à la protection des bases de données est l'occasion d'atteindre un équilibre plus favorable à la circulation des données et l'accès aux données de certains publics. Parmi les éléments de cette réforme, il en est un qui intéresse particulièrement le développement de l'IA au niveau national et européen : une exception aux droits d'auteur et des producteurs de bases de données pour la fouille de texte et de données (text and data mining). La fouille de textes et de données désigne un ensemble de traitements informatiques consistant à extraire des connaissances selon un critère de nouveauté ou de similarité dans des textes ou des bases de données. Elle permet par exemple la recherche de « signaux faibles » difficiles à appréhender par la lecture cursive, le repérage ou l'analyse de comptes rendus d'expérimentations ratées. La fouille de texte et de données est porteuse de potentiels énormes pour la découverte scientifique et le développement de nouvelles connaissances. Aujourd'hui, la reproduction des bases de données nécessaire à la mise en place de systèmes de fouille suppose un accord explicite du propriétaire de l'œuvre ou du détenteur des bases de données concernées – y compris lorsque l'accès à ces données est licite, par exemple lorsqu'un chercheur s'est acquitté des droits d'accès pour lire les articles présents dans la base d'un éditeur d'articles scientifiques. (...) Par l'absence de cadre légal clair, l'Europe accuse un retard important pour la compétitivité de sa recherche et, donc, pour ses capacités d'innovation. Associé aux nouvelles méthodes d'apprentissage de l'intelligence artificielle, l'accès aux données ainsi autorisé représente un potentiel important pour de nombreux projets scientifiques, en particulier dans le cadre de recherches interdisciplinaires. La question de savoir si une telle exception devrait être limitée à un cadre scientifique et à des fins non commerciales reste à trancher. Aussi notre mission est partisane d'une concertation large sur cette question. En effet, de nombreux acteurs – journalistes, associations, entreprises – pourraient largement bénéficier de cette exception, notamment pour le traitement automatisé d'informations accessibles en ligne. Aujourd'hui, un journaliste d'investigation qui souhaiterait utiliser les techniques de fouille de textes et de données pour analyser les contenus de sites – pour lesquels il dispose pourtant d'un accès licite – doit impérativement se conformer aux licences d'accès prévus par ces différents sites ou négocier individuellement avec chaque site l'autorisation d'opérer ce traitement automatisé. La loi pour une République numérique a déjà consacré une telle exception pour la recherche publique. Ces dispositions sont encore en attente d'un décret d'application. Les chercheurs doivent pouvoir bénéficier de cette exception sans avoir à accuser un retard supplémentaire. D'autant qu'une fois entrés en vigueur, les*

⁵ VILLANI Cédric, *Donner un sens à l'intelligence artificielle - Pour une approche nationale et européenne*, rapport remis en présence du Président de la République, mars 2018. Disponible en ligne sur : https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf.

⁶ CONSEIL NATIONAL DU NUMÉRIQUE, *Feuille de route sur l'environnement et le numérique, 50 mesures pour un agenda national et européen sur un numérique responsable c'est-à-dire sobre et au service de la transition écologique et solidaire et des objectifs de développement durable*, juillet 2020. Disponible en ligne sur : https://cnnumerique.fr/environnement_numerique_2020.

textes européens en question devront faire l'objet d'une transposition en droit national, ce qui pourrait encore allonger ce délai. »⁷

Dans le cadre de la feuille de route sur l'environnement et le numérique, le CNNum a réitéré sa position en faveur d'une science ouverte et la recherche autour des données au service de la transition écologique et solidaire. Partant, il recommandait d'une part de transposer les deux exceptions « text and data mining » de la directive sur le droit d'auteur afin de pouvoir effectuer, à des fins de recherche, des traitements de masse sur des informations qui y sont stockées⁸, et d'autre part de rendre disponible les données d'intérêt général pour la recherche⁹.

- 2) Pensez-vous que la fouille de textes et de données se développera dans les années à venir ? Pensez-vous que les acteurs de la fouille de textes et de données seront majoritairement européens ?

Cf. Supra réponse à la question 1.

- 3) Quels sont les avantages et les inconvénients attendus du développement de la fouille de textes et de données, en termes de protection des droits d'auteur et assimilés, de croissance économique, de bienfaits pour les usagers ?

En 2015, une étude empirique a démontré l'impact négatif d'une absence d'exception pour le TDM sur l'innovation en Europe continentale, à la fois pour les STM (sciences, technique et médecine) et les SHS (sciences humaines et sociales)¹⁰. L'argument suivant lequel une exception risquerait d'entraîner un risque pour les droits de propriété intellectuelle notamment du fait des potentielles divulgations des bases de données, est fallacieux.

- 4) Avez-vous été associés, directement ou indirectement, à l'élaboration de l'article 38 de la loi n°2016-1321 du 7 octobre 2016 pour une République numérique, qui a instauré, en faveur de la fouille de textes et de données, pour les besoins exclusifs de la recherche scientifique, une exception au monopole des auteurs pour la reproduction de leurs œuvres, ainsi qu'aux droits des producteurs de bases de données¹¹ ? Ces dispositions vous semblent-elle adaptées¹² ?

⁷ VILLANI Cédric, *op. cit.*, p. 35-36.

⁸ CONSEIL NATIONAL DU NUMÉRIQUE, Feuille de route précitée, mesure n°27, p. 80.

⁹ Sur la notion de données environnementales d'intérêt général, v.: CONSEIL NATIONAL DU NUMÉRIQUE, *Faire des données environnementales des données d'intérêt général*, avis de juillet 2020. Disponible en ligne sur : https://cnnumerique.fr/environnement_numerique_2020.

¹⁰ Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research, 2015. Disponible en ligne sur : http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2608513.

¹¹ CPI, art. L. 122-5 : « Lorsque l'œuvre a été divulguée, l'auteur ne peut interdire : (...) 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques 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 ». – V. aussi art. L. 342-3, 5°.

¹² Pour l'heure, l'exception française est doublement limitée : par son champ d'application, réservé à l'écrit, excluant ainsi la fouille sur des œuvres audiovisuelles, plastiques ou musicales ; par ses bénéficiaires, puisque seule une partie de la recherche publique est concernée, celle qui ne repose pas sur des partenariats publics-privés.

Le CNNum a été auditionné par le Parlement dans le cadre de l'article 38 de la loi pour une République numérique. Il avait produit une note sur l'opportunité d'une exception au droit d'auteur pour la fouille de textes et de données (*text and data mining* - TDM) à des fins de recherche publique et non-commerciale et la faisabilité juridique d'une exception au droit d'auteur et au droit *sui generis* sur les bases de données pour le TDM.

- 5) Avez-vous été associés, directement ou indirectement, à l'élaboration des articles 3 et 4 de la directive 2019/790 ?

Non.

- 6) Quelle appréciation générale portez-vous sur ces deux articles ?

Se référer à l'audition par le CSPLA d'Annie Blandin, ancienne membre du CNNum.

- 7) Avez-vous connaissance d'une transposition des articles 3 et 4 de la directive 2019/790 dans d'autres Etats membres ?

Non.

- 8) Disposez-vous d'éléments de comparaison entre le cadre juridique européen en matière de fouilles de textes et de données et les règles applicables en dehors de l'Union européenne (par exemple aux Etats-Unis et au Japon) ?

L'Irlande, le Royaume-Uni, les États-Unis, la Chine, le Japon, et plus récemment l'Allemagne et l'Estonie, ont adopté des dispositions législatives qui permettent aux chercheurs de réaliser des reproductions numériques d'une base de données à partir d'une source licite.

- 9) Pensez-vous que, dans le cadre juridique français actuel, les fouilles de textes et de données peuvent être réalisées, en tout ou partie :
- en dehors des règles du droit d'auteur et assimilés, pour les données brutes et données publiques notamment ?
 - sur le fondement de l'exception de citation ?
 - sur le fondement de l'exception au profit de la recherche ?
 - sur le fondement de l'exception de reproduction technique provisoire ?

Cf. *Supra* réponse à la question 1 : le CNNum s'est positionné en faveur de la fouille de textes et de données sur le fondement de l'exception de citation et de l'exception au profit de la recherche. Pour les données publiques, il n'y pas besoin d'exception puisque la loi pour une République numérique a consacré l'ouverture de celles-ci par défaut.

- 10) Si vous êtes éditeurs de presse, comment imaginez-vous l'articulation entre le nouveau régime des droits voisins, mentionné à l'article 15 de la directive 2019/790, et les dispositifs prévus aux articles 3 et 4 du même texte ? Souhaitez-vous attirer l'attention de la mission sur des problématiques spécifiques ?

RAS.

II. Questions sur l'article 3 de la directive 2019/790

Article 3

Fouille de textes et de données à des fins de recherche scientifique

- 1. Les États membres prévoient une exception aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel, en vue de procéder, à des fins de recherche scientifique, à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont accès de manière licite.*
- 2. Les copies des œuvres ou autres objets protégés effectuées dans le respect du paragraphe 1 sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche.*
- 3. Les titulaires de droits sont autorisés à appliquer des mesures destinées à assurer la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés. Ces mesures n'excèdent pas ce qui est nécessaire pour atteindre cet objectif.*
- 4. Les États membres encouragent les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir d'un commun accord des bonnes pratiques concernant l'application de l'obligation et des mesures visées aux paragraphes 2 et 3, respectivement.*

Pour précision, cette exception est **obligatoire** et on ne peut y déroger par le contrat (dir. 2019/790, art. 7.1¹³).

- 1) Le cadre posé et les conditions de mise en œuvre de l'exception (bénéficiaires identifiés, fins de recherche scientifique exclusivement, accès licite, absence de compensation du fait du préjudice minime¹⁴) vous semblent-ils satisfaisants ?

Le champ d'application est relativement large, mais mérite de s'interroger sur d'éventuels conflits d'intérêt en cas de partenariats de recherche publics/privés. Dès lors que le financement engage les deniers publics, la publication de la recherche devrait être publique et les acteurs privés ne sauraient se l'approprier.

- 2) Les partenariats public-privé peuvent bénéficier de cette exception « académique »¹⁵, ce qui en étend considérablement la portée. On pourrait par exemple imaginer, grâce à un tel

¹³ « Toute disposition contractuelle contraire aux exceptions prévues aux articles 3, 5 et 6 est non exécutoire. »

¹⁴ V. cons. 17.

¹⁵ Dir. 2019/790, cons. 11 : « (...) En concordance avec l'actuelle politique de la recherche de l'Union, qui encourage les universités et les instituts de recherche à collaborer avec le secteur privé, les organismes de recherche devraient également bénéficier d'une telle exception lorsque leurs activités de recherche sont menées dans le cadre de partenariats public-privé. Les organismes de recherche et les institutions du patrimoine culturel devraient rester les bénéficiaires de cette exception, mais ils devraient pouvoir s'appuyer sur leurs partenaires privés pour effectuer des fouilles de textes et de données, y compris en utilisant leurs outils technologiques. »

partenariat, de lancer une solution d'IA créative musicale sur une base de données d'archives comme celle de l'INA, ou d'IA créative graphique et plastique sur un fonds muséal. Comment transposer cette ouverture ?

- 3) Afin de favoriser les activités de recherche et l'interopérabilité entre les œuvres, seriez-vous favorables à l'utilisation d'un format standardisé dans votre secteur d'activité (par exemple XML pour les textes et MIDI pour la musique) ?

Le recours à un format standardisé pour un secteur donné peut être un moyen intéressant permettant de garantir l'interopérabilité des services concernés. Dans son étude de cas sur l'interopérabilité des réseaux sociaux¹⁶, le Conseil a identifié plusieurs risques liés à la mise en œuvre de l'interopérabilité et notamment à la mise en place de formats standardisés. L'un de ces risques concerne l'impact sur l'innovation des acteurs qui seraient concernés. En particulier, il résulte des auditions menées par le Conseil que la mise en place de normes communes pourrait donner lieu à la création de standard de fait. Aussi, afin de préserver la spécificité des modèles économiques, l'interopérabilité devrait se limiter à un standard commun a minima, si besoin complété au cas par cas. En effet, il ne faudrait pas limiter la capacité des services à lancer de nouvelles fonctionnalités ou à mettre à jour des produits.

Par ailleurs, afin d'améliorer l'accessibilité numérique, qui constitue une obligation légale pour le secteur audiovisuel, le Conseil est en faveur d'une interopérabilité des contenus audiovisuels accessibles. Dans le cadre de notre rapport sur "l'accessibilité numérique, entre nécessité et opportunité"¹⁷, il ressort des auditions que le manque d'accessibilité des contenus audiovisuels serait notamment lié à la multiplicité des supports pour un même contenu audiovisuel. En effet, les normes d'accessibilité d'un contenu varient avec le support utilisé pour y accéder (site web, application mobile...). Or la compatibilité des systèmes faciliterait la portabilité des contenus et permettrait une expérience utilisateur équivalente en termes d'accessibilité sur l'ensemble des plateformes (téléviseur, set-top box, smartphone, tablette). Aussi, le Conseil serait d'avis à encourager la mise en place de normes d'accessibilité communes pour l'ensemble des supports de contenus audiovisuels (recommandation n°17 dudit rapport).

- 4) Comment vérifier que la durée de conservation des copies des œuvres et des autres objets protégés, mentionnée à l'alinéa 2, n'est pas excessive ?

Il serait possible d'instaurer un tiers de confiance (comme la Bibliothèque nationale de France) pour héberger les copies nécessaires à l'activité de fouille ou encore d'instaurer une notion de délai raisonnable de conservation de données à des fins de recherche qui sera laissée à l'appréciation du juge.

- 5) Quelles modalités souhaiteriez-vous voir mises en place pour la conservation des copies¹⁸ ?

¹⁶ CONSEIL NATIONAL DU NUMÉRIQUE, *Concurrence et régulation des plateformes : Etude de cas sur l'interopérabilité des réseaux sociaux*, avis de juillet 2020. Disponible en ligne sur : https://cnnumerique.fr/Interoperabilite_Concurrence_Etude.

¹⁷ CONSEIL NATIONAL DU NUMÉRIQUE, *Accessibilité numérique : entre nécessité et opportunité*, rapport de février 2020. Disponible en ligne sur : https://cnnumerique.fr/publication_rapport_accessibilite_numerique.

¹⁸ V. cons. 15 : « Dans certains cas, par exemple aux fins de la vérification a posteriori des résultats de la recherche scientifique, les organismes de recherche et les institutions du patrimoine culturel pourraient avoir besoin de conserver des copies faites dans le cadre de l'exception aux fins d'effectuer de la fouille de textes et de données. En pareils cas, les copies devraient être stockées en lieu sûr. Les États membres devraient être libres de décider, au niveau national et après concertation avec les acteurs concernés, des modalités plus précises de conservation des copies, notamment la capacité de désigner des organismes de confiance aux fins du stockage de ces copies. Afin

Le Conseil national du numérique soutient le chiffrement¹⁹: celui-ci pourrait être utilisé pour la conservation des données et il serait dès lors opportun de consulter l'ANSSI afin qu'il rappelle les meilleurs règles de l'art en matière de cybersécurité.

- 6) Quelles pourraient être les mesures mises en œuvre par les titulaires de droits pour assurer la sécurité et l'intégrité des réseaux, structures et bases de données sollicités, en permettant le TDM ?

RAS.

- 7) Comment le gouvernement pourrait-il encourager les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir conjointement des bonnes pratiques pour l'application de cet article 3 ? Seriez-vous favorable à l'élaboration rapide d'une charte des bonnes pratiques entre les acteurs concernés ? Souhaiteriez-vous que ce processus soit mené dans le cadre du CSPLA ?

Mettre le ministère de la Culture et le ministère de la Recherche comme tiers de confiance pourrait être une solution idoine. La Bibliothèque nationale de France bénéficie également d'une expertise en termes de traitements de données.

III. Questions sur l'article 4 de la directive 2019/790

Article 4

Exception ou limitation pour la fouille de textes et de données

1. *Les États membres prévoient une exception ou une limitation aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE, à l'article 4, paragraphe 1, points a) et b), de la directive 2009/24/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions d'œuvres et d'autres objets protégés accessibles de manière licite aux fins de la fouille de textes et de données.*
2. *Les reproductions et extractions effectuées en vertu du paragraphe 1 peuvent être conservées aussi longtemps que nécessaire aux fins de la fouille de textes et de données.*
3. *L'exception ou la limitation prévue au paragraphe 1 s'applique à condition que l'utilisation des œuvres et autres objets protégés visés audit paragraphe n'ait pas été expressément réservée par leurs titulaires de droits de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne.*

que l'application de cette exception ne soit pas restreinte indûment, ces modalités devraient être proportionnées et limitées à ce qui est nécessaire à la conservation des copies en toute sécurité et à la prévention de leurs utilisations non autorisées. »

¹⁹ CONSEIL NATIONAL DU NUMÉRIQUE, Avis Prédiction, chiffrement & libertés, septembre 2017. Disponible en ligne sur : <https://cnnumerique.fr/nos-travaux/chiffrement>.

4. Le présent article n'affecte pas l'application de l'article 3 de la présente directive.

- 1) La directive semble donner aux Etats membres une certaine marge de manœuvre lors de sa transposition en droit interne, renvoyant à une approche fondée sur une « exception » ou sur une « limitation » des droits d'auteurs et assimilés. Quelles conséquences en tirez-vous ?

RAS.

- 2) Le cadre et les conditions de mise en œuvre (tous bénéficiaires, tous usages, y compris commerciaux, accès licite) vous paraissent-ils satisfaisants ?

Il s'agit d'un cadre relativement large qui mériterait de mener des études d'impacts avec des cas d'usages pour déterminer si le cadre et les conditions de mises en œuvre sont proportionnés (exemple : usage commercial de données de recherche ou encore justice prédictive avec les données Légifrance).

- 3) Comment comprenez-vous le droit d'option ouvert au titulaire et qui permet de se retirer du dispositif ?

RAS.

- 4) Quels seraient, selon vous, les moyens techniques appropriés (« procédés lisibles par machine ») que les titulaires de droits pourraient utiliser pour s'opposer (« *opt-out* », *i.e.* option de retrait) à une fouille de textes et de données réalisée par des organismes autres que ceux mentionnés à l'article 3 (métadonnées, conditions générales d'utilisation, accords contractuels, déclarations unilatérales, fichier centralisé, mesures techniques de protection, ou autres moyens) ? Faudrait-il standardiser ces moyens ? *Quid* en cas de doute sur l'exercice de l'*opt-out* ?

RAS.

- 5) Pensez-vous exercer cette option de retrait ?

RAS.

- 6) Pensez-vous que les titulaires de droits dans votre secteur d'activité vont s'opposer massivement à la fouille de textes et de données pour des besoins autres que ceux de la recherche scientifique ?

D'après les nombreuses auditions menées par le CNNum dans le cadre du partage de données d'intérêt général, la plupart des contributeurs économiques sont en faveur d'une approche non contraignante et considèrent que les entreprises doivent être encouragées à mettre à disposition les données qu'elles estiment être d'intérêt général sur la base du volontariat²⁰.

²⁰ CONSEIL NATIONAL DU NUMÉRIQUE, Synthèse des états généraux du numériques - Les données d'intérêt général, mai 2020. Disponible en ligne sur : <https://cnnumerique.fr/files/uploads/2020/CNNum%20-%20EGNum%20-%20DIG.pdf>.

- 7) En cas d'opposition des titulaires de droit à la fouille de textes et de données, les organismes qui souhaitent réaliser ces fouilles pourront-ils se fonder sur les dispositions citées à la question 9 du I) du présent questionnaire (autres exceptions/limitations) ?

RAS.

- 8) Comment apprécier la durée de conservation des reproductions et extractions de données ? Faudrait-il prévoir une sanction en cas de durée de conservation excessive ?

RAS.

- 9) Pensez-vous que cette exception est susceptible de limiter très fortement l'intérêt à investir dans des bases de données (par exemple, les bases de données d'images ou de textes élaborées par les éditeurs) ?

RAS.

- 10) A quelles conditions les titulaires de droit pourraient-ils accepter les fouilles de textes et de données mentionnées à l'article 4 de la directive ? Le recours à une licence prévoyant des contreparties financières vous semble-t-il adapté ?

RAS.

- 11) Pensez-vous que la fouille de textes et de données dans un but commercial pourrait paradoxalement être freinée en cas de nombreuses options de retrait de la part des titulaires de droit ?

Cf. supra réponse à la question 6.

- 12) D'un point de vue technique, comment peut-on mettre en œuvre les dispositions de cet article 4 sans faire obstacle aux dispositions de l'article 3 de la directive²¹ ?

L'élaboration d'un guide pratique commun technique (auquel pourraient être associées, en raison de leur expertise technique, la Commission nationale de l'informatique et des libertés, la Commission d'accès aux documents administratifs et l'Agence nationale de la sécurité des systèmes d'information) pourrait permettre d'apporter des réponses techniques à la mise en œuvre des dispositions de l'article 4 sans faire obstacle aux dispositions de l'article 3 de la directive.

- 13) Quelles pourraient être les règles communes aux dispositifs prévus aux articles 3 et 4 de la directive (par exemple, durée de conservation des reproductions et extractions) ? Une approche sectorielle vous paraît-elle utile ?

L'expérience du Conseil national du numérique démontre que l'organisation de concertations avec les acteurs privés et publics, avec au besoin des spécifications sectorielles permet de dégager des solutions pour décider de la meilleure gouvernance commune.

²¹ L'article 4.4 de la directive l'impose : « Le présent article n'affecte pas l'application de l'article 3 de la présente directive. »

IV. Questions diverses

- 1) Quelles seraient les précisions relatives à la fouille de textes et de données que vous souhaiteriez voir inscrites dans les textes législatifs et réglementaires de transposition des articles 3 et 4 de la directive 2019/790, sous réserve qu'elles ne soient pas contraires à la lettre et à l'esprit de ces dispositions ?

RAS.

- 2) En cas d'expertise technique sur le TDM, quels formats utilisez-vous ? Avez-vous connaissance d'un document scientifique récapitulatif sur ce point ?

RAS.

- 3) Quelles autres remarques, générales ou spécifiques, souhaitez-vous formuler auprès de la mission ?

RAS.



CONTRIBUTION A LA MISSION DU CSPLA SUR L'INTRODUCTION D'EXCEPTIONS DE FOUILLE DE TEXTES ET DONNEES

Hadopi - octobre 2020

Table des matières

1. LE BESOIN DE SECURISATION DU CADRE JURIDIQUE EN MATIERE DE FOUILLE DE TEXTES ET DE DONNEES.....	4
1.1 Le recueil et le traitement des données, un outil central pour une nouvelle forme d'économie, de recherche et pour la régulation du numérique.....	4
1.2 Un enjeu international	5
1.3 Les apports de la Directive	6
1.3.1 Les limites de la logique contractuelle	6
1.3.2 Un relatif consensus sur une exception aux fins de recherche.....	6
1.3.3 Une exception sans limites de finalités ni de bénéficiaires mais soumise à une option de retrait des titulaires de droits.....	7
2. LA MISE EN ŒUVRE EFFECTIVE DES EXCEPTIONS : LES SOLUTIONS ATTENDUES.....	9
2.1 Des standards uniformisés pour un meilleur traitement des données	9
2.2 Les enjeux de conservations des copies.....	10
2.2.1 Les modalités de conservation des copies	10
2.2.2 Les durées de conservation.....	11
2.3 Les modalités de protection des œuvres et objet protégés	12
LEGISLATIONS ETRANGERES RELATIVES A LA FOUILLE DE TEXTES ET DE DONNEES.....	16
I – Pays de l'Union européenne	16
A. Tableau de synthèse des réponses reçues.....	16
B. Présentation détaillée pays par pays	18
II – Pays hors Union européenne	23
A. Tableau de synthèse des réponses reçues.....	23
B. Présentation détaillée pays par pays	24

« Fouiller dans l'obscur pour trouver la clarté »¹

Cette citation pourrait expliquer une grande part des activités artistiques, philosophiques et scientifiques.

Désormais, la fouille est également informatique et cybernétique. Cette nouvelle dimension ouvre de nouvelles perspectives vers de nouvelles connaissances et services dans une course internationale à l'innovation et à la croissance. Elle soulève, par ailleurs, des interrogations eu égard à sa matière première inédite, les « *data* » et « *big data* », leurs sources et leur ampleur inégalées.

La fouille de textes et de données dite en anglais « *text and data mining* » ou « TDM » est un processus de recherche informatique automatisée permettant d'extraire des volumes importants de données telles que des textes, des sons, des images afin de les analyser en vue d'acquérir de nouvelles connaissances.

Sur le plan pratique, la fouille de textes et de données implique deux étapes successives totalement automatisées : la consultation des données sources et leur analyse (reconnaissance des mots, phrases, images et relations) ; l'interprétation de cette analyse, permettant de sélectionner les informations pertinentes et d'en extraire la valeur informationnelle.

L'encadrement juridique de cette pratique soulève de nombreuses interrogations dès lors qu'elle se situe au croisement de nombreuses pratiques et de technologies telles que l'intelligence artificielle, dont surtout le *deep learning*, et le *web scraping*. Pour les personnes publiques - et au premier chef pour les autorités de régulation - ces questions s'inscrivent de surcroît dans le contexte de déploiement de la science des données « *Data Science* » et de l'ouverture des données publiques « *Open Data* ».

Ces interrogations alimentent de plus fort la réflexion à conduire sur cette nouvelle économie de la donnée au regard des droits et libertés en présence, notamment sur la nature des données analysées, la profondeur de l'intrusion nécessaire à l'aspiration de ces données, la nature des traitements automatisés effectués, la licéité de ces pratiques par rapport aux monopoles et autres dispositions légales auxquelles elles sont confrontées : droit d'auteur, secret industriel, droit des bases de données, atteinte au système de traitement automatisé de données, protection des données à caractère personnel, etc.

L'introduction d'exceptions de fouille de textes et de données au niveau européen est principalement motivée par la compétitivité des acteurs européens de l'économie numérique, la recherche scientifique et l'innovation, après avoir constaté les limites des exceptions existantes à sécuriser ce nouveau type d'activités. Aux Etats Unis ou au Japon, le cadre juridique encourage d'ores et déjà vivement les accords de licence liés à la fouille de textes et de données.

Si la France, a déjà reconnu une exception en faveur de la fouille de textes dans la loi n°2016-1321 du 7 octobre 2016 pour une République numérique², son champ d'application restait beaucoup plus circonscrit que les exceptions européennes consacrées par les articles 3 et 4 de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE (ci-après la « Directive »).

Au-delà d'une simple exception dite « académique » au bénéfice des organismes de recherche et des institutions du patrimoine culturel, à des fins de recherche scientifique (article 3), la Directive fixe un second niveau d'exception dont la portée est très générale et sans finalité prédéfinie (article 4). Le président du Conseil supérieur de la propriété littéraire et artistique (CSPLA) a confié le 10 juin 2020 au professeur Alexandra Bensamoun une mission visant à préparer l'introduction en droit français de ces dispositions de la Directive.

¹ Les Écureuils de Central Park sont tristes le lundi, Katherine Pancol,

² Voir les articles [L. 122-5 10° du code de la propriété intellectuelle](#) et [L. 342-3 5° du même code](#)

Nous explorerons donc, aux fins de contribuer à cette mission, les enjeux liés à la reconnaissance de cette exception au regard de l'émergence de cette activité, en Europe et dans le monde, tant pour les acteurs privés que publics (I), ainsi que les enjeux liés à sa mise en œuvre effective (II).

1. LE BESOIN DE SECURISATION DU CADRE JURIDIQUE EN MATIERE DE FOUILLE DE TEXTES ET DE DONNEES

A l'instar du développement de l'intelligence artificielle, la fouille de textes et de données pourrait être amenée à jouer un rôle important pour favoriser l'innovation et la compétitivité (1.1) tant française qu'européenne ou internationale (1.2).

La Directive offre un encadrement juridique de cette pratique, nécessaire pour répondre au besoin de sécurité juridique et de souplesse des parties prenantes (1.3).

1.1 Le recueil et le traitement des données, un outil central pour une nouvelle forme d'économie, de recherche et pour la régulation du numérique

Il n'est pas contesté, en matière de recherche scientifique (sciences naturelles et sociales) que l'automatisation de la recherche dans une immensité de documents permet d'obtenir des résultats inaccessibles par des méthodes manuelles. Par exemple, pour le projet Text2genome, c'est une technologie de fouille de textes et de données qui a permis de cartographier le génome humain en compilant automatiquement trois millions de publications.

La consécration par l'article 3 de la Directive, au bénéfice de la recherche, d'une véritable exception au monopole des titulaires de droits de propriété intellectuelle, a vocation à renforcer non seulement la légitimité de leurs pratiques mais également leur efficacité en particulier face aux USA pour lesquels la pratique a été admise par la jurisprudence sur le fondement du *fair use*³.

Allant plus loin dans l'analyse, le Conseil national du numérique, qui s'est positionné en faveur d'une telle exception dès 2016, lors des travaux parlementaires de la loi pour une République numérique, soulignait « *au-delà des apports intellectuels, des bienfaits sociaux ou des progrès en matière de santé publique, il s'agit d'un enjeu crucial pour la compétitivité de la recherche en France, moteur de l'innovation et de la transformation de l'ensemble de notre tissu économique, créatrice de richesse et d'emplois* »⁴.

La fouille – ou exploration – de textes et de données, entendue comme l'ensemble des techniques automatisées visant à « analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations », occupe-t-elle une place importante dans votre secteur d'activité ?

A n'en pas douter, l'appétence pour la fouille de textes et de données dépasse le seul cadre de la sphère marchande et celui de la recherche scientifique, comme le souligne le considérant 18 de la Directive, et intéresse également les acteurs publics, au premier rang desquels les régulateurs dans le secteur numérique.

Si à ce jour, l'Hadopi n'a pas développé de telles activités en raison de leur absence d'encadrement juridique spécifique, l'introduction de nouvelles dispositions en droit interne pourrait offrir un cadre plus propice de mise en œuvre de cette exception dans le cadre de ses missions d'observation, d'encouragement à l'offre légale notamment.

³ Aux États-Unis : Voir : Cour suprême, 18 avril 2016, U.S. Court of appeal, décision du 16 octobre 2015 et United States Circuit judge, 14 novembre 2013

⁴ https://cnnumerique.fr/files/2017-10/CNNum_Fiche_TDM.pdf

Les régulateurs français semblent s'être emparés de ces questions de l'efficacité de leurs outils à l'ère du développement de l'intelligence artificielle, des « *data science* » et du « *big data* ». Comme le soulignait le Président du CSA, Roch Olivier Maistre, le 6 février 2019 :

« En réalité, on le voit bien : ce qui se dessine pas à pas, à travers ces différentes évolutions, c'est une forme de nouvelle régulation, un nouveau mode de régulation des contenus. Evidemment, il y a encore beaucoup d'incertitudes, ne serait-ce que parce que les modèles d'affaires des grandes plateformes de contenus évoluent. Mais une chose, au moins, est claire : les outils de la régulation ne seront plus tout à fait les mêmes. La régulation des grandes plateformes de contenus a vocation à combiner les méthodes classiques de la régulation, comme l'observation, avec de nouvelles, comme la *data science* »⁵.

L'ARCEP, sous la présidence de Sébastien Soriano, a fait de la « *régulation par la data* » l'un de ses principaux objectifs. Elle loue ainsi « *un nouveau mode d'action qui complète les outils traditionnels du régulateur. Son principe : utiliser la puissance de l'information afin d'orienter le marché dans la bonne direction* »⁶.

1.2 Un enjeu international

Avez-vous connaissance d'une transposition des articles 3 et 4 de la directive 2019/790 dans d'autres Etats membres ?

La direction des affaires juridiques et internationales de l'Hadopi s'est rapprochée de ses interlocuteurs au sein des pays membres l'Union européenne pour connaître l'état d'avancement de la transposition de la Directive et, le cas échéant, leur approche de la transposition.

Les réponses obtenues sont disponibles en annexe de cette contribution (Annexe 1). Les pays ayant répondu et indiqué être à un stade de réflexion plus abouti sont recensés dans le tableau de synthèse ci-dessous.

	Législation préexistante en la matière	Travaux de transposition en cours	Consultation en cours
Allemagne	X	X	
Autriche			X
Belgique		X	X
Espagne		X	
Grèce		X	
Hongrie		X	X
Italie		X	
Irlande	X	X	
Lettonie			X
Luxembourg		X	
République Tchèque		X	
Slovaquie		X	X
Slovénie		X	

Disposez-vous d'éléments de comparaison entre le cadre juridique européen en matière de fouilles de textes et de données et les règles applicables en dehors de l'Union européenne (par exemple aux Etats-Unis et au Japon) ?

Des éléments de réponses relatifs aux législations américaine, australienne, britannique, canadienne, japonaise, néozélandaise, suisse et turque sont également disponibles en annexe de cette contribution (Annexe 1).

⁵ Discours de Roch-Olivier Maistre à la 6e Conférence des présidents du Réseau francophone des régulateurs des médias (REFRAM) à Tunis le 6 septembre 2019

⁶<https://www.arcep.fr/la-regulation/grands-dossiers-thematiques-transverses/la-regulation-par-la-data.html>

1.3 Les apports de la Directive

1.3.1 Les limites de la logique contractuelle

Les premiers travaux sur le sujet⁷ faisaient déjà état des nombreuses promesses du *data mining*. Face aux potentiels conflits avec le droit d'auteur et le droit des bases de données, l'autorégulation via des systèmes de licences avec les titulaires de droits avait été privilégiée, tant en France qu'au sein de l'Union européenne⁸.

Quels sont les avantages et les inconvénients attendus du développement de la fouille de textes et de données, en termes de protection des droits d'auteur et assimilés, de croissance économique, de bienfaits pour les usagers ?

Les solutions contractuelles, ont cependant montré leurs limites en matière de recherche en raison d'une part, des difficultés rencontrées en l'absence d'identification des titulaires de droits et, d'autre part, des complexités inhérentes à la nécessité de multiplier les licences dans les projets de recherches impliquant la consultation de plusieurs corpus éditoriaux, outre la durée des négociations des accords de licence.

Il est assez largement admis, en matière de recherche, que les avantages de l'exception introduite par l'article 3 de la Directive sont supérieurs aux préjudices qu'encourent les titulaires de droits du fait de ces usages (voir notamment le considérant 17 de la Directive) et qu'en tout état de cause on ne peut pas se satisfaire d'une communauté scientifique qui ne pourrait se saisir de certains sujets ou résultats de recherche qu'avec l'aval des titulaires de droits.

A noter de surcroît, qu'un certain nombre de procédés ou méthodes de fouille de textes et de données ne réaliserait pas de reproduction des données fouillées, du moins pas au sens juridique du terme, mais se contenterait d'en restituer la « connaissance » c'est-à-dire le sens informationnel, comme l'évoque le considérant 9 de la Directive.

Les équilibres à rechercher apparaissent, en revanche, plus fragiles dans le cadre de la transposition de l'article 4 de la Directive entre les enjeux de sécurité juridique à apporter aux utilisateurs sur les reproductions et les extractions effectuées et la faculté de réserver ces droits, laissée aux ayants droit et ce, compte tenu de la méfiance résultant de leurs craintes liées à la sécurité et à l'intégrité de leurs réseaux et bases de données.

1.3.2 Un relatif consensus sur une exception aux fins de recherche

La loi n° 2016-1321 du 7 octobre 2016 pour une République numérique a introduit une exception de fouille de textes et de données au bénéfice du monde de la recherche scientifique publique, dans un périmètre très circonscrit puisqu'elle ne s'applique qu'aux écrits scientifiques (incluant les textes et données qu'ils contiennent) et non aux autres œuvres comme les images, les œuvres musicales ou audiovisuelles⁹.

L'exception prévue à l'article 3 de la Directive répond plus largement aux attentes du monde scientifique puisqu'elle vise plus largement les organismes de recherche et les institutions du patrimoine culturel œuvrant pour la recherche en sciences naturelles mais également en sciences humaines (cons. 12 de la Directive) et ce, dans un cadre public mais également dans le cadre de partenariats public-privé (cons. 11 et 12 de la Directive), qu'ils agissent à but non lucratif ou dans le cadre d'une mission d'intérêt public reconnu.

⁷ Voir notamment le rapport de la mission du CSPLA en 2014 confiée à Maître Jean Martin et les rapports rédigés sous la direction du Professeur Hargreaves

⁸ Voir notamment le projet ISTE

⁹ Articles L. 122-5 10° et L. 342-3, 5° du code de la propriété intellectuelle

Le cadre posé et les conditions de mise en œuvre de l'exception (bénéficiaires identifiés, fins de recherche scientifique exclusivement, accès licite, absence de compensation du fait du préjudice minime) sur l'article 3 de la directive 2019/790 vous semblent-ils satisfaisants ?

Le cadre posé par l'article 3 de la Directive cherche à trouver un juste équilibre permettant de répondre aux attentes des différents acteurs du monde de la recherche¹⁰ tout en préservant les intérêts des titulaires de droits en leur apportant des garanties visant à la protection de leurs droits.

Les interrogations portent davantage sur les modalités opérationnelles de mises en œuvre de ces garanties.

La condition d'accès licite imposée par le paragraphe 1 de l'article 3 de la Directive et précisée par le considérant 14 implique que les organismes de recherche et les institutions du patrimoine culturel prennent des précautions quant aux contenus auxquels les personnes qui leur sont rattachées auront accès. Le considérant 14 précise qu'il peut s'agir d'accès à des contenus fondés sur « une politique de libre accès » ou selon des « arrangements contractuels » avec les titulaires de droits.

Cette condition de licéité d'accès à la source nécessite néanmoins des clarifications concernant les contenus disponibles selon une politique de libre accès afin d'aider les chercheurs à identifier les accès et les contenus licites de ceux qui ne le sont pas.

Les copies nécessaires à la fouille de textes et de données devront être « stockées avec un niveau de sécurité approprié » (article 3 paragraphe 2 de la Directive), et selon des modalités de conservation des copies plus précises, par exemple en désignant des « organismes de confiance » (cons. 15 de la Directive).

Les titulaires de droits seront enfin autorisés à **appliquer des mesures permettant d'« assurer la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés »**, à condition qu'elles n'excèdent pas « ce qui est nécessaire pour atteindre cet objectif » (article 3 paragraphe 3 de la Directive). Des exemples concrets sont données par le considérant 16 de la Directive.

Le considérant 16 de la Directive prend soin de préciser que ces mesures décidées par les titulaires de droits devraient être « proportionnées aux risques inhérents » et « ne devraient pas aller au-delà de ce qui est nécessaire pour atteindre l'objectif visant à assurer la sécurité et l'intégrité du système ».

Compte tenu des incertitudes entourant encore les mesures de sécurité et du principe de proportionnalité qui doit guider le choix de ces mesures, il pourrait être utile, comme le préconise le paragraphe 4 de l'article 3 de la Directive, de définir d'un commun accord des bonnes pratiques entre les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel concernant l'application l'obligation de sécurisation des copies des contenus exploités et des mesures techniques de sécurité des bases et des réseaux d'hébergement.

Contrairement à l'exception obligatoire prévue par l'article 3 de la Directive, l'exception ou la limitation issue de l'article 4 de la Directive est plus largement controversée.

1.3.3 Une exception sans limites de finalités ni de bénéficiaires mais soumise à une option de retrait des titulaires de droits

Le cadre et les conditions de mise en œuvre (tous bénéficiaires, tous usages, y compris commerciaux, accès licite) de l'article 4 de la

Un climat de méfiance entoure davantage l'introduction de l'article 4 de la Directive dont le champ d'application est très large et dont les conditions de mise en œuvre sont floues.

¹⁰ A noter à ce titre que la communauté de recherche française avait indiqué dans un communiqué commun du 11 octobre 2017 souhaiter « un geste fort de la part de l'Union afin que cette dernière soutienne sans réserve un usage harmonisé et le plus ouvert possible de cette nouvelle exception, sans restrictions inutiles » : https://www.couperin.org/images/stories/Organisation/Communiqu%C3%A9_soutien_Sparc_Europe-2017-10-11.pdf

directive 2019/790 vous paraissent-ils satisfaisants ?

Les dispositions de l'article 4 de la Directive fixent un second niveau d'exception dont la portée est très générale et sans finalité prédéfinie, mais qui peut faire l'objet d'une opposition des titulaires de droits (mécanisme *d'opt-out* ou option de retrait).

Cette exception s'applique à toute entité privée ou publique qui pratique la fouille de textes et de données, y compris à des fins commerciales, pourvu qu'elle ait un accès licite aux données et à condition que les titulaires de droits n'aient pas interdit la fouille de textes et de données sur les contenus leur appartenant ou sur les réseaux et bases de données hébergeant ces contenus.

A ce jour, l'exception de l'article 4 de la Directive fait ainsi l'objet de réserves plus prononcées, notamment de la part des titulaires de droits puisqu'il y a encore, beaucoup d'incertitudes sur les usages qui seront faits de leurs contenus et les garanties qui seront apportées pour minimiser leurs craintes.

Cette exception reste pourtant réclamée par de nombreux acteurs y compris dans le monde de la recherche scientifique à des fins commerciales et pourrait être mise en œuvre par des acteurs publics n'ayant pas d'activité de recherche, à des fins non commerciales. Par ailleurs, l'introduction d'une telle exception ou limitation a pu être saluée par les entreprises du secteur privé qui envisagent de réaliser de la fouille de textes et de données pour des finalités autres que la recherche scientifique non commerciale.

Il n'en demeure pas moins que les conditions de mise en œuvre de l'exception doivent être clarifiées et ce, notamment afin d'éviter que les titulaires de droits décident massivement d'exercer leur option de retrait. En effet, compte tenu du climat de méfiance généralisé concernant cette exception, il est à craindre que *l'opt-out* soit adopté globalement par nombre d'ayants droit. Cela aurait pour effet de réduire voire d'anéantir la portée de cette exception.

Comment comprenez-vous le droit d'option ouvert au titulaire et qui permet de se retirer du dispositif ?

Un ayant droit exerçant largement son option de retrait contraint les potentiels bénéficiaires de l'exception à devoir individuellement obtenir un accord pour pouvoir pratiquer de la fouille de textes et de données. Cela implique donc un retour aux droits exclusifs et aux systèmes de contractualisation par voie de licence d'exploitation.

Cependant, la réservation de droits ne doit pas affecter « d'autres utilisations » ni faire échec à l'exception de fouille de textes et de données au profit de la recherche scientifique, ni à l'exception de reproduction provisoire instituée par l'article 5 §1 de la directive 2001/29/CE du 22 mai 2001.

Une série de questions est alors soulevée par ce texte : s'il existe un risque *d'opt-out* global, ne peut-on pas considérer que l'option de retrait ouverte aux titulaires de droit puisse être exercée partiellement ? Les ayants droit pourront-ils décider d'exercer leur option de retrait en fonction des projets de fouilles de textes et de données et/ou en fonction des œuvres et contenus fouillés ?

La lecture du considérant 18 de la Directive semble indiquer qu'il soit possible de réserver seulement une partie des droits par exemple dans le cadre d'un accord contractuel, d'une déclaration unilatérale ou de conditions générales d'utilisation.

Indépendamment du débat concernant le caractère synthétique ou analytique des droits d'exploitation de l'auteur, l'auteur peut céder partiellement ou totalement ses droits pour lui conférer une emprise maximale sur les différentes utilités économiques de son œuvre. Le droit positif français permet d'ailleurs ces démembrements contractuels puisque seuls certains droits peuvent être cédés et d'autres réservés.

A l'inverse, une approche stricte et généralisée de l'option de retrait prévue par l'article 4 de la Directive pourrait être contraire à l'esprit de la Directive qui vise à chercher un équilibre entre les bénéfices des utilisations de la fouille de textes et de données et le fait de maintenir un pouvoir de contrôle des titulaires

de droits. La difficulté va donc résider dans la recherche des moyens permettant de mettre en œuvre cet équilibre subtil.

La réservation de droits pourra ainsi être faite selon les moyens suivants :

- Concernant les contenus mis à la disposition du public en ligne, au moyen de procédés lisibles par machine. Dans ce cas, le considérant 18 de la Directive précise que « *la réservation de droits ne devrait être jugée appropriée que si elle est effectuée au moyen de procédés lisibles par machine, y compris les métadonnées et les conditions générales d'utilisation d'un site internet ou d'un service* ».
- Dans les autres cas, par d'autres moyens comme des accords contractuels ou une déclaration unilatérale.

L'*opt-out* pourrait se matérialiser à différents niveaux : au niveau de l'œuvre elle-même (pour indiquer qu'une œuvre ne peut pas faire l'objet de *text and data mining*) ou au niveau d'un « contenant » tel qu'une base de données, un site Internet ou une plateforme numérique (pour indiquer si cette structure ainsi que les données ou contenus qu'elle renferme ne peuvent pas être parcourus et exploités par un outil de fouille de textes et de données).

Ces deux options pourraient d'ailleurs être cumulatives, ou éventuellement hiérarchisées. Dans ce cas, il convient de s'interroger sur la règle à appliquer si une plateforme autorisant *le text and data mining* contient une œuvre pour laquelle les titulaires de droits se sont opposés à un tel processus ou inversement (risque de conflit au niveau de l'*opt-out*).

Si une option de retrait est exercée par un titulaire de droits au niveau du contenant pour s'appliquer à tout le contenant et donc à tous ses contenus, cela signifie-t-il que tout accès en violation de cet *opt-out* serait considéré non autorisé et ne remplirait donc plus la condition d'accès licite imposée par le paragraphe 1 de l'article 4 de la Directive ?

2. LA MISE EN ŒUVRE EFFECTIVE DES EXCEPTIONS : LES SOLUTIONS ATTENDUES

2.1 Des standards uniformisés pour un meilleur traitement des données

Afin de favoriser les activités de recherche et l'interopérabilité entre les œuvres, seriez-vous favorables à l'utilisation d'un format standardisé dans votre secteur d'activité (par exemple XML pour les textes et MIDI pour la musique)?

La multiplication des formats de données et de structures de données n'est pas de nature à en faciliter la fouille ou le traitement par des outils automatisés.

La coexistence de nombreux standards relatifs aux œuvres (les contenus) et aux bases de données (les contenants) contrarie en pratique la fouille de textes et de données à grande échelle.

Si l'absence d'harmonisation généralisée n'empêche pas cette fouille, elle peut rendre plus complexe les travaux en ce qu'elle requiert d'importantes opérations de retraitement et d'ajustement des données avant analyse.

Des formats génériques existent par exemple MP4 ou H.264 pour la vidéo, MP3 ou FLAC pour l'audio, JPEG ou PNG pour l'image, EPUB ou PDF pour les documents.

Mais les structures de données et en particulier les métadonnées associées à ces contenus – afin de les caractériser et de les identifier au mieux – ne font pas encore l'objet de consensus ou ne sont pas systématiquement respectées. Il existe pourtant des modèles de données multi-formats, tels que l'IMM (IPTC Information Interchange Model) ou le XMP (Extensible Metadata Platform qui peuvent être placés dans l'entête des fichiers contenant les œuvres numérisées, ou dans des fichiers descriptifs

accompagnant ces œuvres. Cependant leur utilisation reste isolée. L'encouragement à l'utilisation de ces standards serait par conséquent souhaitable pour favoriser les activités de fouille.

Ce point reste néanmoins un objectif à long terme et se situe presque en périphérie du *text and data mining*. Il s'agit avant tout de faciliter globalement les traitements de données, en lien avec la fouille mais aussi dans le cadre de nombreux autres usages.

2.2 Les enjeux de conservations des copies

2.2.1 Les modalités de conservation des copies

Il est nécessaire de s'interroger sur la portée de l'obligation imposée par l'article 3 paragraphe 2 de la Directive aux opérateurs de fouille de textes et de stocker les données collectées « avec un niveau de sécurité approprié ».

Un tel niveau de sécurité est rendu nécessaire compte tenu du fait que les copies des œuvres ou contenus réalisées pour les besoins de la fouille de textes et de données peuvent être conservées sur une longue durée, comme l'autorise l'article 3 paragraphe 2 de la Directive qui indique que les copies des contenus « *peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche* ».

Des fuites liées à des négligences de la part des opérateurs de fouille pourraient en effet être fort dommageables pour les ayants droit. Un défaut de sécurisation des bases générées pourrait en effet entraîner de la contrefaçon massive, comme le craignent les ayants droit d'une manière générale concernant la fouille de textes et de données. On serait donc logiquement en droit d'attendre que les opérateurs de fouille fassent preuve des meilleurs efforts à ce niveau (sachant que cette notion de meilleurs efforts est de nature évolutive en fonction des progrès techniques mais aussi des menaces).

Il serait par conséquent logique de chercher à responsabiliser les exploitants de solutions impliquant de la fouille de texte et de données. Les questions ne manquent pas sur ce point :

- Faut-il prévoir des sanctions spécifiques en cas de manquement ? Des mécanismes de responsabilité, des modalités d'indemnisation ?
- Qui devrait pouvoir statuer sur le caractère satisfaisant ou insuffisant des mesures de protection prises, en cas de litige ?
- Les opérateurs de fouille devraient-ils être soumis à des audits sur la sécurité de leurs systèmes ?

A l'instar de l'obligation de sécurité qui existe en matière de traitement des données à caractère personnel, et même si ces données revêtent un plus haut degré de sensibilité, il pourrait être utile de se référer à certaines recommandations de la CNIL, de l'ANSSI ou du TPN (*Trusted Partner Network*)¹¹, voire même que des recommandations spécifiques ou des bonnes pratiques soient rédigées pour assurer un niveau de sécurité approprié pendant la durée de conservation des copies des œuvres et des contenus nécessaires à la fouille de textes et de données.

C'est d'ailleurs ce que préconise le considérant 15 de la Directive qui invite les Etats membres à « *décider, au niveau national et après concertation avec les acteurs concernés, des modalités plus précises de conservation des copies, notamment la capacité de désigner des organismes de confiance aux fins du stockage de ces copies* ».

L'intervention d'un régulateur pourrait d'ailleurs se justifier dans ce cadre, en cas de différend entre acteurs portant sur les moyens de sécurisation des données conservées, ainsi que sur d'autres modalités pratiques (respect par les opérateurs de fouille de l'option de retrait des ayants droit,

¹¹ Le *Trusted Partner Network* (TPN) est une initiative mondiale de protection du contenu audiovisuel réunissant plusieurs associations professionnelles internationales. A travers un programme d'audit et de sensibilisation, le TPN aide les entreprises à améliorer la protection des contenus et à prévenir les fuites et les incidents de sécurité.

proportionnalité des mesures de protection prises par ces derniers, absence de conservation des données abusive, etc.).

2.2.2 Les durées de conservation

Comment apprécier la durée de conservation des reproductions et extractions de données réalisées dans le cadre des exceptions des articles 3 et 4 ? Comment vérifier qu'elle n'est pas excessive ? Faudrait-il prévoir une sanction en cas de durée de conservation excessive ?

Aucune durée de conservation standard ne semble proposée ou définie de manière précise ou globale par la Directive. Il n'est pas davantage précisé qui, y compris éventuellement des ayants droit ou des opérateurs de fouille, pourrait fixer ce critère.

Il s'agit d'une modalité d'exercice du bénéfice de la conservation de la copie dont l'appréciation ou la fixation pourrait être confiée à un régulateur. L'hypothèse d'une sanction pourrait trouver à s'inscrire, le cas échéant, dans le cadre d'une procédure de règlement de différends.

L'article 3 de la Directive se contente d'indiquer que les copies des œuvres *« peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche »* et le considérant 15 précise par ailleurs que *« Les utilisations des copies aux fins d'activités de recherche scientifique autres que la fouille de textes et de données, comme l'examen scientifique par des pairs ou la recherche commune, devraient continuer à relever, le cas échéant, de l'exception ou de la limitation prévue à l'article 5, paragraphe 3, point a), de la directive 2001/29/CE. »*

L'article 4 de la Directive évoque une conservation *« aussi longtemps que nécessaire »* dont on comprend qu'elle pourrait dépendre de la nature de la fouille de textes et de données et qu'elle pourrait, dans l'absolu, être indéterminée.

Pour autant, au vu des craintes des titulaires de droits et afin d'éviter qu'elle puisse être jugée excessive, cette durée de conservation pourrait être appréciée au regard de certains plafonds et pourrait d'ailleurs varier en fonction des types d'œuvres ou de données concernées : une base de données, un film ou un article scientifique n'obéissent en effet pas aux mêmes règles en ce qui concerne les usages licites (types de droits distincts, éventuels impacts d'ordre chronologique, etc.).

Quelles pourraient être les règles communes aux dispositifs prévus aux articles 3 et 4 de la directive (par exemple, durée de conservation des reproductions et extractions) ? Une approche sectorielle vous paraît-elle utile ?

D'un point de vue pratique et par souci de cohérence, il pourrait apparaître logique que nombre de règles en termes de fouille de textes et de données soit commune aux fouilles menées à des fins de recherche scientifique et à celles menées à d'autres fins et ce d'autant plus que les moyens de protection envisagée pourraient s'avérer être les mêmes (cf tableau ci-après).

L'ensemble des mécanismes prévus (à l'exception de l'*opt-out*) et les principales règles d'usage paraissent pouvoir s'appliquer de manière générique, ce qui devrait contribuer à faciliter la mise en œuvre du *text and data mining*. L'adoption de règles de fonctionnement trop distinctes d'un secteur à l'autre pourrait en outre s'avérer dissuasif car complexe à mettre en œuvre et pourrait freiner inutilement les projets en matière de fouille de données.

Pour autant, un distinguo pourrait être appliqué dans la mesure où certains travaux scientifiques ne peuvent être menés que dans la durée. Dans ce cas, une durée de conservation de plusieurs années ne serait pas abusive. En revanche dans le cadre de l'exception prévue par l'article 4 de Directive, dès lors qu'il existe une pluralité de finalités autorisées, la durée de conservation doit s'adapter pour ne pas être excessive dans le cas d'une fouille de données menée sur du court terme à des fins commerciales.

Une approche sectorielle peut également faire sens en vue d'adopter des standards communs par média ou pour faciliter l'évaluation du temps de conservation nécessaire aux fins de fouille de textes et de données qui peut varier d'un type d'œuvre à un autre.

2.3 Les modalités de protection des œuvres et objet protégés

Quelles pourraient être les mesures mises en œuvre par les titulaires de droits pour assurer la sécurité et l'intégrité des réseaux, structures et bases de données sollicités, en permettant le TDM [de l'article 3 de la Directive] ?

Le considérant 16 de la Directive concernant l'exception de fouille académique, précise que des mesures techniques peuvent être prises par les titulaires de droits pour « *par exemple servir à garantir que seules les personnes ayant accès de manière licite à leurs données puissent y accéder, notamment sur la base de la validation de leur adresse IP ou de l'authentification de l'utilisateur* ».

Plusieurs solutions applicables peuvent être paramétrées à différents niveaux de contrainte.

Si les mesures techniques de protection (MTP ou DRM) sont le plus souvent utilisées afin de gérer et de contrôler les accès licites aux contenus, les « dômes » de protection et API servent davantage, respectivement, à limiter les accès abusifs et à encadrer des accès privilégiés ou particuliers au service. Le recours à ces outils pourrait être justifié dans la mesure où il est important de se prémunir contre certains excès tels que le pillage massif de données ou le chargement intensif de données, susceptible de saturer les systèmes.

Pour autant, ces mesures sont considérées par la Directive « *lorsqu'il existe un risque que la sécurité et l'intégrité de leurs systèmes ou bases de données soient compromises* » et à la condition qu'elles soient proportionnées à l'objectif de sécurité et d'intégrité du système.

Il convient de veiller à ce que ces mesures de protection ne soient pas poussées à l'extrême sans justification dans l'unique but de gêner les opérations de fouilles. En cas de différend persistant, le recours à un médiateur ou régulateur pourrait être envisagé, sur saisine, afin de garantir le bénéfice de l'exception de fouille.

Quels seraient, selon vous, les moyens techniques appropriés (« procédés lisibles par machine ») que les titulaires de droits pourraient utiliser pour s'opposer (« opt-out », i.e. option de retrait) à une fouille de textes et de données réalisée par des organismes autres que ceux mentionnés à l'article 3 (métadonnées, conditions générales d'utilisation, accords contractuels, déclarations unilatérales, fichier centralisé, mesures techniques de protection, ou autres moyens) ? Faudrait-il standardiser ces moyens ? Quid en cas de doute sur l'exercice de l'opt-out ?

L'article 4 de la Directive introduit également avec la logique d'*opt-out*, un dispositif qui permet également aux ayants droit d'envisager des solutions de contrôle voire de blocage de certains accès au contenant comme aux contenus en fonction des restrictions qu'ils semblent imposer.

Les solutions techniques existantes sont recensées ci-après.

Il est possible d'imaginer également des solutions plus structurées et éventuellement hybrides, c'est-à-dire permettant de gérer l'*opt-out* aussi bien pour les contenants (bases de données, plateformes, sources de contenus, etc.) que pour les œuvres et les contenus. La « construction » de ces solutions, plus universelles et plus structurées que les précédentes, implique toutefois la mise en œuvre d'efforts coordonnés – et un niveau d'acceptabilité partagé par l'ensemble des acteurs – afin d'éviter une multiplication incohérente d'outils trop disparates.

Certaines des solutions envisagées ci-après permettent de contrôler voire de bloquer certains accès en fonction des restrictions imposées par les titulaires de droits. Elles filtrent donc les accès.

D'autres, au contraire, sont peu contraignantes et n'assurent pas que les déclarations faites par l'utilisateur sur sa qualité, son statut et ses motivations sont faites de bonne foi, en l'absence de toutes vérifications à ce titre.

↳ Les modalités techniques de protection des « contenants » par les ayants droit dans le cadre des articles 3 et 4 de la Directive

Solutions techniques existantes	Avantages	Inconvénients
Déclaration (au moyen d'un paramétrage technique) Il peut s'agir d'un paramètre standard intégré à une base de données, ou d'un « meta tag » placé sur un site Internet, tel que la balise NOINDEX qui permet à l'administrateur d'un site Internet de signaler aux moteurs de recherche qu'il ne faut pas parcourir ou archiver une page ou un site tout entier.	Technique simple et déclarative, peu onéreuse à mettre en place. Cette mesure d'information a vocation à être respectée par les acteurs soumis à l'article 4 de la Directive. Les acteurs agissant à des fins de recherche scientifique peuvent l'ignorer.	Cette solution est relativement fragile et se base avant tout sur l'espoir que la consigne sera respectée. Elle ne permet pas de bloquer l'accès.
« Dôme » de protection contre les intrusions Un mécanisme de contrôle d'accès ou de gestion des droits (DRM ou Digital Rights Management) bloque tout accès intempestif ou répétitif au système, empêchant ainsi les opérations de fouille non autorisées.	Cette solution bloque efficacement toute intrusion non autorisée et permet de fixer des seuils correspondant à une utilisation légitime des ressources. Elle permet en outre de s'assurer que la fouille de texte et de données ne vient pas saturer le système en le sur-sollicitant.	Le système de protection est actif par défaut et pourrait par conséquent faire obstacle au bénéfice de l'exception à des fins de recherche scientifique (à moins qu'un accès privilégié puisse être accordé par ailleurs de plein droit).
Accès pour la fouille via une API (interface de programmation applicative) L'API permet d'identifier le demandeur et de gérer les accès aux contenus du système (exemple : l'API de YouTube)	Cette solution permet d'optimiser et de moduler l'accès aux données dans le cadre de la fouille, en accordant des accès spécifiques. La fouille se fait de manière transparente et selon des fonctions dédiées.	Cela revient à conditionner ce qui relève en théorie d'une exception (notamment pour la recherche scientifique).

↳ Les modalités techniques de protection des « contenus » dans le cadre du mécanisme d'*opt-out* de l'article 4 de la Directive

Solutions techniques existantes	Avantages	Inconvénients
DRM, au niveau du contenu Un système de gestion des droits empêche la fouille ou indique que cette pratique n'est pas autorisée pour le contenu en question.	Gestion fine par contenu.	En pratique, les MTP ou DRM sont aujourd'hui de plus en plus implémentées au niveau des plateformes et non individuellement au niveau de chaque fichier contenant une œuvre. Par ailleurs, le système de protection est actif par défaut et pourrait par conséquent faire obstacle au bénéfice de l'exception à des fins de recherche scientifique.
Métadonnées, au niveau du contenu Un marqueur situé au sein des métadonnées d'un contenu, indique si l' <i>opt-out</i> est sollicité ou non pour le contenu et la finalité en question.	Ce système est assez simple et léger à mettre en place.	Il faut compter sur la bonne foi des utilisateurs en espérant qu'ils tiendront compte de cette simple indication. Par ailleurs, notamment en cas d'évolution dans le temps des souhaits exprimés par les ayants droit, il est possible que certaines copies d'un contenu autorisent la fouille tandis que d'autres mentionnent l' <i>opt-out</i> . Il existe enfin un risque d'effacement de ces métadonnées.
Marquage en dur (watermarking), au niveau du contenu Un marqueur intégré au sein même d'un contenu. Un outil de détection de marquage (ou de tatouage) numérique permet de lire cette indication.	Ce système est relativement proche de celui des métadonnées, mais est plus résistant.	Il faut là encore compter sur la bonne foi des acteurs procédant à la fouille de textes et de données, en espérant qu'ils respecteront l'indication donnée par le marquage. Par ailleurs, il est possible que certaines copies d'un contenu autorisent la fouille tandis que d'autres mentionnent l' <i>opt-out</i> , de manière contradictoire.

- **Solutions de protection universelle et centralisée**

La multiplication de ce genre de fichiers serait contreproductive et contraire à l'idée de guichet unique. Les conditions de mise en œuvre restent par ailleurs à déterminer. A qui devrait en revenir la gestion ? Faudrait-il en rendre l'usage obligatoire et en déduire que tout ayant droit qui n'y a pas déclaré son opt-out a renoncé à ce dernier ? En outre, ce type de solution n'a de sens que si les acteurs de la fouille de textes et de données s'en servent et respectent les consignes qui se trouvent dans le fichier.

Solutions techniques envisagée	Avantages	Inconvénients
Fichier centralisé Les titulaires de droits désireux d'utiliser l' <i>opt-out</i> le font savoir en déclarant ce choix dans un fichier centralisé et formalisé, faisant autorité.	Il s'agit d'un outil structuré et facilement identifiable par tous. Les titulaires de droits désireux d'utiliser l' <i>opt-out</i> peuvent s'y référencer. Les acteurs désireux de faire de la fouille de textes et de données (pour des raisons autres que la recherche scientifique) peuvent également consulter ce fichier, qui joue en quelque sorte un rôle de « guichet unique ». Cette solution devra également être évolutive et permettre aux titulaires de droits de mettre à jour leurs instructions, si nécessaire.	La solution du fichier centralisé seul est complexe à mettre en œuvre car il peut s'avérer difficile d'établir un lien entre un contenu collecté dans le cadre d'une fouille et sa référence dans le fichier centralisé. Il n'est clairement pas envisageable ni réaliste de référencer des parties entières d'Internet et de préciser pour chaque occurrence de chaque contenu si l'option de retrait s'applique ou non. La solution d'un fichier centralisé utilisé en lien avec un outil de reconnaissance de contenu résout quant à elle ce problème.
Fichier centralisé utilisé en lien avec les outils de reconnaissance de contenus Les titulaires de droits désireux d'utiliser l' <i>opt-out</i> renseignent ce choix dans des bases de références qui contiennent déjà (ou qui contiennent par ailleurs) d'autres données relatives aux œuvres, notamment des données liées aux technologies de reconnaissance de contenus	Cette solution repose potentiellement sur des fichiers centraux préexistants (certaines bases de référence d'empreintes numériques contiennent déjà aujourd'hui des millions d'empreintes d'œuvres protégées ¹²). Il faudrait y ajouter des informations concernant un éventuel <i>opt-out</i> . Ce modèle simplifierait l'identification des œuvres et des titulaires de droits, en combinant le stockage d'informations liées aux contenus (dont l' <i>opt-out</i>) avec les éléments permettant justement de reconnaître efficacement ces contenus. Il devient alors possible d'identifier et de reconnaître un contenu collecté dans le cadre d'une fouille, et de vérifier la règle qui s'applique pour l'œuvre en question. Peu importe que l'on dispose de dizaines ou de centaines de versions d'une même œuvre (même sans titre ou métadonnées).	La solution du fichier centralisé en lien avec les outils de reconnaissance de contenus concerne surtout les œuvres et les contenus, plutôt que les contenant. Technologiquement parlant, cette dernière solution est la plus avancée et la plus complète. Elle s'inscrit d'ailleurs dans l'évolution globale souhaitée par certains titulaires de droits internationaux en matière de protection des contenus. Mais c'est aussi sans doute celle qui demanderait le plus de développements et le plus de moyens techniques pour être mise en œuvre.

¹² On peut notamment citer les bases INA-Signature de l'Ina, Content-ID de YouTube ou Rights Manager de Facebook. Ces initiatives peuvent être portées par des acteurs privés ou publics, par organismes de gestion collective ou par des consortiums.

D'un point de vue technique, comment peut-on mettre en œuvre les dispositions de cet article 4 sans faire obstacle aux dispositions de l'article 3 de la directive¹³ ?

Il ressort des tableaux ci avant que le choix de certains procédés *d'opt-out* seraient susceptibles de faire obstacle à la fouille de textes et de données à des fins de recherches scientifiques. On peut ainsi classer ces différentes technologies en deux catégories :

- Les procédés « actifs » : le recours aux mesures techniques de protection (MTP ou DRM), aux « dômes » de protection sur les plateformes, voire aux interfaces de programmation d'applications (API), pourrait poser problème dans l'hypothèse où ces outils s'appliquent indistinctement et ne sont pas capables par défaut de distinguer spontanément si une opération de fouille est menée à des fins scientifiques ou à d'autres fins. Au mieux, certains indicateurs techniques peuvent suggérer qu'une fouille relève de la recherche scientifique (si par exemple une opération de fouille est menée depuis un serveur dont l'adresse IP est attribuée à une université) mais ces indications restent partielles et sont rarement infaillibles. Une solution, toutefois contraignante, consisterait à identifier au préalable les utilisateurs et à appliquer différentes règles selon qu'ils sont de simples utilisateurs ou des acteurs réalisant du *text and data mining* à des fins scientifiques ou non (sous réserve qu'un tel fonctionnement contraignant soit compatible avec la notion d'exception).
- Les procédés « passifs » : il s'agit ici des méthodes informatives utilisées pour signaler l'*opt-out* : fichier central, marquage numérique, déclaration, métadonnées, etc. Dans ces cas, la situation serait relativement simple et ne gênerait pas la fouille à des fins scientifiques. Il suffirait en effet que les acteurs agissant dans le cadre de l'article 3 de la Directive fassent abstraction de l'indication d'*opt-out* et ignorent même tout simplement cette dernière, tandis que les acteurs intervenant dans le cadre de l'article 4 devraient, en théorie, respecter cette information.

¹³ L'article 4.4 de la directive l'impose : « Le présent article n'affecte pas l'application de l'article 3 de la présente directive. »

ANNEXE 1

LEGISLATIONS ETRANGERES RELATIVES A LA FOUILLE DE TEXTES ET DE DONNEES

I – Pays de l'Union européenne¹⁴

A. Tableau de synthèse des réponses reçues

Pays	Législation spécifique au TDM préexistante	Projet de transposition de la Directive rendu public	Mission ou consultation sur la transposition
<u>ALLEMAGNE</u> Ministère fédéral de la justice et de la protection du consommateur Direction « Copyright and publishing law » [16 septembre 2020]	OUI Articles 60d et 60h de la loi allemande sur le droit d'auteur du 9 septembre 1965 (modifiée le 28 novembre 2018) [lien cliquable pour accéder à la version en anglais]	OUI Premier projet de transposition de la Directive du 15 janvier 2020 proposé par le ministère fédéral de la justice et de la protection des consommateurs [lien cliquable pour accéder à la version en allemand] + Article en allemand du professeur Benjamin Raue sur le projet de transposition de la Directive [lien cliquable pour accéder à la version en allemand]	NON
<u>AUTRICHE</u> Ministère fédéral de la justice [14 septembre 2020]	NON	NON	Une consultation informelle sur la mise en œuvre des articles 3 et 4 de la Directive a été réalisée jusqu'en juillet 2020. Les réponses sont encore à l'étude.
<u>BELGIQUE</u> Direction générale de la Réglementation économique Office de la propriété intellectuelle [22 octobre 2020]	NON	NON [Projet en cours d'élaboration]	La Belgique dispose d'un Conseil de la propriété intellectuelle qui a pour mission de conseiller le Ministre de la culture notamment sur la transposition de la Directive.

¹⁴ Les pays suivants membres de l'Union européenne n'ont pour le moment pas répondu à nos interrogations : Bulgarie, Chypre, Croatie, Danemark, Estonie, Finlande, Lituanie, Malte, Pays-Bas, Suède.

<u>ESPAGNE</u> Ministère de la culture et des sports [24 septembre et 1 ^{er} octobre 2020]	NON	NON [Projet en cours d'élaboration]	Une consultation a été lancée à la fin de l'année 2019 au sujet du projet de loi sur le droit d'auteur dans le marché unique numérique.
<u>GRECE</u> Hellenic Copyright Organization (HCO) [18 septembre 2020]	NON	NON [Projet en cours d'élaboration]	Une consultation sur le sujet a été lancée et est achevée. La HCO assure un rôle de conseil auprès du ministère de la culture et des sports pour toutes problématiques liées au droit d'auteur. Elle est compétente en matière de règlement alternatif des litiges.
<u>HONGRIE</u> Office de la propriété intellectuelle hongrois [9 octobre 2020]	NON	NON [Projet en cours d'élaboration]	Une consultation publique est en cours au sujet du projet de transposition de la Directive.
<u>IRLANDE</u> Unité « Propriété intellectuelle » du ministère « Business, Enterprise and Innovation » [7 octobre 2020]	OUI Articles 14 et 31 de la loi sur le droit d'auteur et les autres droits de propriété intellectuelle entrée en vigueur le 2 décembre 2019 [lien cliquable] => modifications de la loi de 2000 sur le droit d'auteur et les droits connexes (CRRA) [lien cliquable]	NON [Projet en cours d'élaboration]	Une consultation sur le sujet est en cours.
<u>ITALIE</u> AGCOM [10 septembre 2020]	NON	NON [Projet en cours d'élaboration]	Aucun organe (ni l'AGCOM, ni le CCPDA) n'est pour le moment associé à la transposition de la Directive.
<u>LETTONIE</u> Ministère de la culture [24 septembre 2020]	NON	NON	Le ministère de la culture a créé un groupe de travail sur l'amélioration de la législation sur le droit d'auteur qui étudie la transposition de la Directive. Le mécanisme de recours extrajudiciaire est régi par la loi lettone sur le droit d'auteur (Chapitre X.1 Médiateurs) [lien cliquable pour accéder à la version en anglais]

<u>LUXEMBOURG</u> Ministère de l'Economie du Grand- Duché de Luxembourg [15 septembre et 1 ^{er} octobre 2020]	NON	NON [Projet en cours d'élaboration]	NON
<u>POLOGNE</u> Ministère de la Culture [13 octobre 2020]	NON	NON	Une consultation publique a été lancée et vient de se terminer à la fin du mois de septembre.
<u>PORTUGAL</u> <u>Visapress</u> [10 septembre 2020]	NON	NON [Projet en cours d'élaboration]	NON
<u>REPUBLIQUE TCHEQUE</u> Ministère de la culture [18 septembre 2020]	NON	NON [Projet en cours d'élaboration]	NON
<u>ROUMANIE</u> Office de la propriété intellectuelle [13 octobre 2020]	NON	NON [Projet en cours d'élaboration]	NON
<u>SLOVAQUIE</u> Ministère de la Culture [14 octobre 2020]	NON	NON [Projet en cours d'élaboration]	Une consultation sur le projet de texte de transposition est en cours.
<u>SLOVENIE</u> Ministère du développement économique et de la technologie [24 septembre 2020]	NON	NON [Projet en cours d'élaboration]	NON

B. Présentation détaillée pays par pays

Cette présentation retranscrit fidèlement les informations qui nous ont été communiquées (en anglais) par nos interlocuteurs étrangers.

Il ne s'agit pas d'une interprétation juridique des textes cités.

1) Allemagne

Selon les informations qui nous ont été communiquées le 16 septembre 2020 par la Division « *Copyright and publishing law* » du Ministère fédéral de la justice et de la protection du consommateur, l'Allemagne fait partie des quelques pays européens dont le **droit positif dispose déjà d'une exception de fouille de textes et de données**, prise sur le fondement de l'article 5 (3) de la directive 2001/29/CE qui autorise les Etats membres à prévoir des exceptions ou limitations pour des finalités strictement limitées, notamment « *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 s'avère impossible, la source, y compris le nom de l'auteur dans la mesure justifiée par le but non commercial poursuivi* ».

Cette exception est prévue à l'**article 60d de la loi allemande sur le droit d'auteur** (cliquer [ICI](#) pour accéder à cet article) au bénéfice de la recherche scientifique non commerciale.

Le bénéfice de l'exception est assez large puisqu'il ne se limite pas à des personnes ou à des institutions spécifiques : toute personne qui procède à l'analyse automatique d'un grand nombre d'œuvres à des fins de recherche scientifique peut invoquer l'exception pour autant qu'elle ne poursuive que des objectifs non commerciaux.

L'article 60 d s'applique à tous les types d'œuvres, y compris les bases de données qui sont protégées par le droit *sui generis* (article 87c). Dans le cas des bases de données, certaines précisions sont apportées à l'article 60 d, section 2, afin de garantir qu'elles puissent également être exploitées.

L'utilisation de l'exception de fouille de textes et de données nécessite un accès licite aux contenus protégés.

Les titulaires de droits peuvent mettre en place des mesures techniques de protection. Toutefois, lorsqu'un utilisateur a un accès licite à une œuvre, il peut faire valoir son droit à exécuter la fouille de textes et de données (article 95b (1) 11 de la loi allemande sur le droit d'auteur).

Conformément à l'**article 60 h de la loi allemande sur le droit d'auteur**, les auteurs ont droit à une rémunération équitable en contrepartie de la fouille de textes et de données.

Par ailleurs, s'agissant de la transposition de la Directive, un premier projet de loi a été publié en janvier 2020 proposant une transposition assez fidèle des articles 3 et 4 (cliquer [ICI](#) pour y accéder) L'exception existante telle que prévue aux articles 60 d et 60 h va être révisée et adaptées aux exigences de l'article 3 de la Directive, notamment en ce qui concerne la rémunération. L'article sera transposé par un nouvel article 44 b.

Un article du Professeur Benjamin Raue (cliquer [ICI](#) pour y accéder) nous a été adressé par notre correspondant en Allemagne. Il traite spécifiquement de la question de la transposition de la Directive en droit allemand.

2) Autriche

Selon les informations qui nous ont été communiquées, le 14 septembre 2020, par le Ministère fédéral de la justice, l'Autriche a procédé jusqu'à la fin du mois de juillet 2020 à une consultation informelle sur la transposition des articles 3 et 4 de la Directive (Voir les consultations en allemand [ICI](#) et [ICI](#)). Les réponses étaient encore à l'étude en septembre 2020.

Notre interlocuteur nous indique par ailleurs que l'Autriche envisageait une transposition fidèle des dispositions européennes.

3) Belgique

Selon les informations qui nous ont été communiquées par l'Office de la Propriété intellectuelle le 22 octobre dernier, le projet de transposition de la Directive est en cours de finalisation après qu'un avis ait été rendu par le Conseil de la propriété intellectuelle.

L'avis est disponible en cliquant [ICI](#). Notre interlocuteur nous précise que les projets de dispositions et l'exposé des motifs tels qu'ils figurent dans cet avis ne sont pas nécessairement les dispositions définitives.

En ce qui concerne la transposition des articles 3 et 4 de la Directive, le document de travail prévoit une transposition assez « cohérente » (sic) (voir page 13 et suivante et pages 134 et 137 de l'avis du Conseil).

4) Espagne

D'après les informations qui nous ont été communiquées le 24 septembre et le 1^{er} octobre 2020 par le Ministère de la culture et des sports, en l'absence de textes particuliers applicables à la fouille de texte et de données, l'Espagne a lancé à la fin de l'année 2019 une consultation au sujet du projet de loi sur le droit d'auteur dans le cadre du marché unique numérique visant à transposer la Directive.

Le projet de loi de transposition de la Directive est en cours. Malheureusement, le texte, qui contient en effet des dispositions relatives à la fouille de texte et de données, est à ce jour confidentiel.

5) Grèce

D'après les informations qui nous ont été communiquées par la Hellenic Copyright Organization le 18 septembre 2020, la transposition de la Directive fait actuellement l'objet de travaux préparatoires après qu'une consultation publique ait été lancée auprès des différentes parties prenantes.

La Hellenic Copyright Organization (HCO) joue un rôle de conseil similaire à celui du CSPLA. Il s'agit de la seule autorité compétente en charge des problématiques liées aux droits d'auteur et aux droits voisins en Grèce. Par ailleurs, la HCO fournit des recommandations au ministère de la culture et des sports. Elle est également compétente en matière de règlement alternatif des litiges (médiation et arbitrage) en matière de droit d'auteur et de droits voisins et s'agissant de la gestion collective des droits.

6) Hongrie

Selon les informations qui nous ont été communiquées par l'Office hongrois de la propriété intellectuelle le 9 octobre 2020, la Hongrie dispose d'un projet de transposition de la Directive qui est soumis à une consultation publique depuis l'été dernier. Le projet reprend strictement les dispositions de la Directive.

L'Office hongrois de la propriété intellectuelle collabore avec le ministère de la justice pour transposer la Directive.

7) Irlande

Comme l'Allemagne, l'Irlande fait partie des quelques rares pays européens à avoir déjà adopté des dispositions spéciales sur la fouille de textes et de données.

Selon les informations qui nous ont été communiquées le 7 octobre 2020 par l'unité « *Propriété intellectuelle et copyright* » du ministère du « *Business, Enterprise and Innovation* » du gouvernement irlandais, les articles 14 et 31 de la loi irlandaise sur le droit d'auteur de 2019 prévoient une telle exception à des fins non commerciales (cliquer [ICI](#) pour y accéder). Ces dispositions sont entrées en vigueur le 2 décembre 2019. L'article 14 prévoit une exception pour la réalisation de copies d'une œuvre et l'article 31 est une disposition miroir prévoyant la réalisation d'une copie, d'une représentation ou d'un enregistrement à condition que l'utilisation soit faite uniquement à des fins de recherche non commerciale et que l'utilisation soit accompagnée d'une mention suffisante de la source de l'œuvre originale.

L'article 14 de la loi de 2019 insère un nouvel article 53A dans la loi de 2000 sur le droit d'auteur et les droits connexes (CRRA) et l'article 31 de la loi de 2019 insère un nouvel article 225A (voir la version consolidée non officielle de la CRRA en cliquant [ICI](#)).

L'exception s'applique à tous les contenus protégés par le droit d'auteur et les droits connexes sous réserve que les critères posés par les articles 14 et 31 soient respectés. Ces dispositions ne précisent pas les types de recherches auxquelles s'applique l'exception. Les chercheurs doivent avoir un accès légal aux contenus fouillés (voir articles 53A et 225A de la CRRA).

Par ailleurs, aucune compensation financière n'est prévue pour les titulaires de droits, dès lors que l'exception ne peut être utilisée qu'à des fins non commerciales. Le test des trois étapes de la Convention de Berne est pris en compte.

Concernant les mesures de protection qui peuvent être prises pour les réseaux et bases de données des contenus fouillés, il n'y a pas de texte spécifique applicable à la fouille de textes et de données mais l'article 374, partie VII de la CRRA prévoit des dispositions applicables aux mesures techniques de protection.

Enfin, concernant l'état d'avancement de la transposition de la Directive, l'unité « *Propriété intellectuelle et copyright* » du ministère du « *Business, Enterprise and Innovation* » précise que le travail de transposition est en cours et fait suite à un processus de consultation publique qui a reçu beaucoup de contributions, notamment concernant l'interaction de certaines dispositions contenues dans la Directive avec la législation irlandaise existante en matière de droit d'auteur. L'examen du projet de transposition n'étant pas encore terminé, il demeure confidentiel.

8) Italie

D'après les informations qui nous ont été communiquées le 10 septembre 2020 par l'AGCOM, un projet de loi de transposition est en cours d'approbation au Parlement et le ministère du patrimoine et des activités culturelles italien sera chargé du suivi du dossier. Une attention particulière sera portée sur la sécurité des données du fait des opérations d'extraction. Aucun organe tiers (ni l'AGCOM, ni le CCPDA) n'est pour le moment associé à la transposition de la Directive.

9) Lettonie

D'après les informations qui nous ont été communiquées le 23 septembre 2020 par le Ministère de la Culture de Lettonie, aucun projet de loi de transposition de la Directive n'est à ce jour disponible mais un groupe de discussion créé par le Ministre de la Culture à ce sujet traite actuellement du sujet.

10) Luxembourg

D'après les informations qui nous ont été communiquées le 15 septembre et le 1^{er} octobre 2020 par le Ministère de l'Economie du Grand-Duché de Luxembourg, le projet de loi qui transposera la Directive est en cours d'élaboration. Une transposition littérale de ces exceptions est plutôt la voie envisagée.

11) Pologne

Selon les informations qui nous ont été communiquées par le Ministère polonais de la culture le 13 octobre dernier, il n'y a, à ce jour, pas encore de projet de loi de transposition préparé. La Pologne vient seulement de terminer le processus de consultation publique. La Pologne semble s'orienter vers une transposition littérale. La Pologne a voté contre la directive et a introduit un recours devant la CJUE.

12) Portugal

D'après les informations, le Portugal ne dispose pas de législation spécifique liée à la fouille de textes et de données et le Gouvernement n'a pas encore fait part de sa position officielle concernant la transposition de la Directive.

13) République Tchèque

D'après les informations qui nous ont été communiquées par le Ministère de la Culture le 18 septembre 2020, la République Tchèque finalise actuellement un projet de modification de la loi sur le droit d'auteur comportant également quelques modifications mineures du code civil dans le cadre de la transposition

de la Directive. Le Ministère espère que le projet de loi entrera dans peu de temps dans la prochaine phase de la procédure législative, c'est à dire la procédure interministérielle.

Il n'existe pas de version officielle de ce projet de loi pour le moment mais le Ministère annonce une transposition littérale de la Directive.

Il n'y a pas de mission similaire à celle du CSPLA en République Tchèque, ni d'organisme public ou d'autorité en charge de la réglementation et/ou du règlement extrajudiciaire des litiges pour les questions spécifiques liées au droit d'auteur. Il existe en revanche des règles générales en matière de médiation et d'arbitrage (y compris la Cour d'arbitrage rattachée à la Chambre de commerce tchèque et à la Chambre d'agriculture de la République tchèque) et des règles spécifiques liées à la médiation des litiges liés à la gestion collective. Ils n'ont pas l'intention de créer un nouvel organisme de règlement extrajudiciaire des litiges.

14) Roumanie

Notre interlocuteur au sein de l'office roumain de la propriété intellectuelle nous a indiqué le 13 octobre dernier qu'à ce jour, le projet de texte de transposition de la Directive était en cours de finalisation et qu'il rentrerait prochainement en phase de consultation publique avant d'être soumis au Parlement.

Dans la version de travail, il est envisagé une transposition littérale de la Directive.

15) Slovaquie

D'après les informations qui nous ont été transmises par le Ministère de la culture de la République Slovaque le 14 octobre 2020, le projet de texte de transposition des exceptions prévues par la Directive est en cours de préparation et une consultation a lieu simultanément avec les parties prenantes. La Slovaquie semble essayer de s'en tenir au sens littéral et il n'est pas envisagé d'élargir la portée des exceptions.

16) Slovénie

D'après les informations qui nous ont été communiquées par le Ministère du développement économique et de la technologie le 24 septembre dernier, un projet de loi de transposition de la directive est en cours d'élaboration par le Ministère du développement économique et de la technologie et l'Office slovène de la propriété intellectuelle (SIPO). Il est à ce jour trop tôt pour avoir une idée précise de la teneur de la transposition.

II – Pays hors Union européenne

A. Tableau de synthèse des réponses reçues

PAYS	Législation spécifique au TDM	Commentaire
<u>AUSTRALIE</u> Australian Screen Association [15 septembre 2020]	NON	
<u>CANADA</u> Institut de la Propriété intellectuelle (IPIC) [18 et 21 septembre 2020]	NON mais existence d'une exception de « <i>fair dealing</i> »	La question de l'introduction d'une exception de fouille de textes et de données a été soulevée lors de consultations publiques et plusieurs pistes ont été envisagées. Ces pistes sont reprises dans le rapport du Comité permanent de l'industrie, des sciences et de la technologie de juin 2019 (cf. pages 107 et s.) [lien cliquable pour accéder à la version en français]
<u>ETATS-UNIS</u> Office américain de la propriété intellectuelle [23 octobre 2020]	NON mais existence d'une exception de « <i>fair use</i> »	
<u>JAPON</u> Office du droit d'auteur [11 et 15 septembre 2020]	OUI Articles 30-4, 47-4 et 47-5 de la loi sur le droit d'auteur modifiée en 2018 [lien cliquable pour accéder à la version en anglais]	
<u>NOUVELLE-ZELANDE</u> Ministère de la culture [23 septembre 2020]	NON	La loi sur le droit d'auteur de 1994 est actuellement en cours de révision sur de nombreux aspects, y compris la fouille de textes et de données. En novembre 2018, un document de réflexion a été publié pour la révision de la loi sur le droit d'auteur . [lien cliquable pour accéder à la version en anglais] Un processus de consultation est en cours au sujet de la révision de la législation sur le droit d'auteur . [lien cliquable pour accéder à la version en anglais]
<u>PEROU</u> INDECOP (Institut national pour la défense de la concurrence et la protection de la propriété intellectuelle) [29 septembre 2020]	NON	Pour le moment, il n'est pas prévu à court terme d'intégrer la fouille de textes et de données en droit péruvien mais pour autant, le sujet n'est pas exclu à plus long terme.

<u>ROYAUME-UNI</u> Office de la propriété intellectuelle [24 septembre 2020]	OUI Article 29A de la loi britannique sur le copyright [lien cliquable pour accéder à la version en français]	Un guide sur les exceptions aux fins de recherche nous a été communiqué par l'Office de la propriété intellectuelle anglais. [lien cliquable pour accéder à la version en anglais] L' examen post-mise en œuvre (IPR, « post implementation revue ») des exceptions au droit d'auteur de 2014 vient d'être achevé et publié. [lien cliquable pour accéder à la version en anglais]
<u>SUISSE</u> Institut Fédéral de la Propriété intellectuelle [14 septembre 2020]	OUI Article 24d de la loi sur le droit d'auteur modifiée le 1^{er} avril 2020 [lien cliquable pour accéder à la version en français]	
<u>TURQUIE</u> Ministère de la culture et du tourisme Direction générale du droit d'auteur [17 septembre 2020]	NON	Les questions liées au droit d'auteur sont suivies par le Ministère dans le cadre d'un groupe de travail entre la Commission européenne et la Turquie sur les droits de propriété intellectuelle mais notre interlocuteur précise qu'il n'y a pas, à l'ordre du jour, un projet de réglementation ou de modification de la loi à cet égard.

B. Présentation détaillée pays par pays

Cette présentation retranscrit les informations qui nous ont été communiquées par nos interlocuteurs étrangers. Il ne s'agit pas d'une interprétation juridique des textes cités.

1) Australie

Selon les informations qui nous ont été communiquées par l'Australian Screen Association le 15 septembre 2020, l'Australie ne dispose pas d'une exception au droit d'auteur pour la fouille de textes et de données. Le sujet avait été étudié en 2018 lors d'une révision de la législation sur le droit d'auteur mais les propositions d'amendements publiés par le Gouvernement n'incluaient pas la fouille de textes et de données.

2) Canada

D'après les informations qui nous ont été communiquées par l'Institut de la Propriété intellectuelle du Canada (IPIC¹⁵) les 18 et 21 septembre 2020, il n'existe pas actuellement d'exception spécifique permettant la fouille de textes et de données au Canada.

Néanmoins, la question a été soulevée lors de consultations publiques et plusieurs pistes ont été envisagées. Ces pistes sont reprises dans le rapport du Comité permanent de l'industrie, des sciences et de la technologie de juin 2019, (accessible [ICI](#)) et sont les suivantes :

- La création d'une exception aux fins d'analyse informationnelle,
- L'ajout d'une disposition comparable à celle fixée par l'article 29A de la CDPA du Royaume-Uni (cliquer [ICI](#) pour y accéder),

¹⁵ <https://ipic.ca/>

En tout état de cause, selon l'IPIC, « l'exception serait limitée à l'utilisation de contenu protégé auquel on a accédé légalement¹⁶ (...) [et] le Parlement devra songer aux répercussions des licences et des MTP sur l'efficacité de la nouvelle exception lors de la modification de la Loi¹⁷ ».

Le Comité a donc formulé les observations et la recommandation suivante : « Les témoignages présentés ont persuadé le Comité que le fait de faciliter l'analyse informationnelle du contenu protégé acquis légalement pourrait aider à concrétiser l'avenir prometteur du secteur canadien de l'intelligence artificielle. Le Comité recommande donc :

Recommandation 23

Que le gouvernement du Canada envisage de modifier la Loi sur le droit d'auteur afin de faciliter l'utilisation d'une œuvre ou d'un autre objet protégé à des fins d'analyse informationnelle. »¹⁸.

Début 2020, il était prévu que le Gouvernement donne suite à ces recommandations mais malheureusement le calendrier de travail a été bouleversé par l'épidémie de Covid-19.

En conséquence, il existe seulement à ce jour une exception de « *fair dealing* », similaire au « *fair use* » américain, qui s'applique aux reproductions qui sont faites à des fins de recherche ou d'études privées.

L'utilisation doit être jugée « équitable » en fonction de différents critères.

Cette exception peut s'appliquer à la fouille de textes et de données selon les tribunaux canadiens qui ont fait une interprétation large de l'exception de « *fair dealing* » initialement destinée à couvrir des moyens de recherche plus traditionnels. Néanmoins, l'analyse de l'« équité » peut être assez élaborée, ce qui introduit un certain niveau d'incertitude.

3) **Etats Unis**

Notre interlocuteur à l'Office américain de la propriété intellectuelle (**United States Patent and Trademark Office (USPTO)**) le 23 octobre dernier que nous avons interrogé en particulier sur la jurisprudence postérieure à celle de *Hathitrust*, nous invite à prendre connaissance :

- du rapport public publié en octobre 2020 intitulé « Public Views on Artificial Intelligence and Intellectual Property Policy » qui traite de la question de la fouille de texte et de données aux pages 23 et suivante. Ce rapport est disponible en cliquant [ICI](#);
- de la jurisprudence *Fox News v. TVEyes* du 27 février 2018, accessible en cliquant [ICI](#).

En dehors du « *fair use* » et des autres exceptions existantes au sein de l'article 17 « Droit d'auteur » du Code des Etats Unis d'Amérique, il n'est pas prévu d'introduire une exception spécifique pour la fouille de textes et de données compte tenu du recours aux licences contractuelles.

Les litiges ne seraient pas très fréquents.

¹⁶ INDU (2018), [Témoignages](#), 1445 (Stewart et Bourne-Tyson, CBUA); INDU (2018), [Témoignages](#), 1720 (Gagnon, Element AI); INDU (2018), [Témoignages](#), 1640 (Merkley, CrCC); INDU (2018), [Témoignages](#), 1715 (Tribe et Aspiazu, OpenMedia); INDU (2018), [Témoignages](#), 1530 (Kerr-Wilson, BCBC); INDU (2018), [Témoignages](#), 1550 (Fewer, CIPPIC); INDU (2018), [Témoignages](#), 1715 (Chapdelaine); INDU (2018), [Témoignages](#), 1600, 1630 (Tarantino et Lovrics, IPIC); INDU (2018), [Témoignages](#), 1540 (Geist); Microsoft Canada, [Mémoire présenté à INDU](#), 13 septembre 2018; ABRC, [Mémoire présenté à INDU](#), 28 septembre 2018; CBUA, [Mémoire présenté à INDU](#), 28 septembre 2018; RCDR, [Mémoire présenté à INDU](#), 28 septembre 2018; BSA The Software Alliance, [Mémoire présenté à INDU](#), 12 octobre 2018; Element AI, [Mémoire présenté à INDU](#), 12 octobre 2018; FIAIB, [Mémoire présenté à INDU](#), 12 octobre 2018; Internet Association, [Mémoire présenté à INDU](#), 20 novembre 2018; IPIC, [Mémoire présenté à INDU](#), 4 décembre 2018; Dessa, [Mémoire présenté à INDU](#), 14 décembre 2018; Geist, [Mémoire présenté à INDU](#), 14 décembre 2018; AMII et coll., [Mémoire présenté à INDU](#), 14 décembre 2018; Réseau Portage, [Mémoire présenté à INDU](#), 7 janvier 2019; Tawfik et coll., [Mémoire présenté à INDU](#), 18 janvier 2019. Voir aussi Université MacEwan, [Mémoire présenté à INDU](#), 13 septembre 2018.

¹⁷ IPIC, [Mémoire présenté à INDU](#), 4 décembre 2018.

¹⁸ <https://www.noscommunes.ca/DocumentViewer/fr/42-1/INDU/rapport-16/page-144#34>

4) Japon

D'après les informations qui nous ont été communiquées par l'Office japonais du droit d'auteur le 11 et 15 septembre 2020, la loi sur le droit d'auteur au Japon a été modifiée en 2018, y compris concernant les dispositions liées à la fouille de textes et de données (ancien article 47-7 de la loi japonaise sur le droit d'auteur).

La loi modifiée sur le droit d'auteur prévoit des dispositions souples de limitation des droits permettant l'utilisation d'œuvres protégées sans autorisation pour des services utilisant des données volumineuses qui ne nuisent pas au marché des contenus protégés par le droit d'auteur.

Les dispositions pertinentes incluses dans la section des « limitations au droit d'auteur » dans la loi sur le droit d'auteur de 2018 sont les suivantes :

- Article 30-4 : Exploitation d'une œuvre sans objectif de profiter des pensées ou des sentiments exprimés dans cette œuvre ;
 - Article 47-4 : Exploitation accessoire à l'exploitation des œuvres sur un ordinateur ;
 - Article 47-5 : Exploitation mineure accessoire au traitement informatisé des données et à la fourniture des résultats de celui-ci.
- (Cliquer [ICI](#) pour accéder à ces dispositions)

5) Royaume-Uni

Le Royaume-Uni fait partie des premiers pays à avoir introduit une exception de fouille de textes et de données au sein de l'Union européenne en 2014.

Selon les informations qui nous ont été communiquées par l'Office de la propriété intellectuelle le 24 septembre dernier, cette exception a été introduite, dans le cadre de la transposition de l'article 5 paragraphe 3 a) de la directive 2001/29/CE, à l'article 29A de la loi sur le copyright de 1988 modifiée (cliquer [ICI](#) pour accéder à ce texte).

Un guide sur les exceptions aux fins de recherche nous a été communiqué par l'Office de la propriété intellectuelle anglais (cliquer [ICI](#) pour y accéder).

L'examen post-mise en œuvre (IPR, « *post implementation revue* ») des exceptions au droit d'auteur introduites en 2014 vient d'être achevé et peut être consulté sur Internet en cliquant [ICI](#). La conclusion de cet examen semble être que l'exception de fouille de textes et de données fonctionne bien. Cependant, certains commentaires ont été formulés concernant l'interaction avec le régime des mesures techniques de protection (MTP).

Cette exception s'applique à tout type de contenus et à tout type de recherche, il n'y a pas de différence faite entre la recherche publique et privée. Des conditions sont imposées par l'article 29A de la loi sur le copyright et notamment, la condition d'accès licite aux œuvres, la finalité non commerciale de la recherche, l'ajout d'un crédit des œuvres et des contenus, dans la mesure du possible.

De nombreuses questions se posent concernant la nature commerciale ou non commerciale de la recherche (étant précisé que ces notions ne font pas l'objet de définition en droit du copyright anglais), particulièrement lorsqu'elle implique :

- une variété d'acteurs, y compris ceux du secteur privé, des organisations philanthropiques, des entreprises privées créées par les universités ;
- des résultats qui ne sont disponibles que par un accès payant (« *paywall* »), etc.

L'exception est obligatoire et ne peut pas être limitée par contrat. Néanmoins, la condition d'accès légal (par exemple, au moyen d'un abonnement ou un accès libre) s'applique toujours et des orientations ont été prises pour indiquer que les ayants droit pouvaient mettre en place des mesures pour garantir la sécurité et l'intégrité de leurs systèmes.

Il n'existe aucune compensation financière pour les titulaires de droits. L'impact initial de cette exception sur la mise en œuvre était prévu à « £0 », ce qui semble être toujours le cas. Aucun élément sur d'éventuelles conséquences négatives de la fouille de textes et de données n'a été fourni par les titulaires de droits dans le cadre du récent processus d'examen post-mise en œuvre de l'exception.

Les éditeurs et les fournisseurs de contenus sont en mesure d'appliquer des mesures raisonnables pour maintenir la sécurité ou la stabilité de leur réseau. Ces mesures ne doivent pas empêcher ou restreindre de manière excessive la capacité des chercheurs à extraire des textes et des données. Ils sont notamment libres de prévoir des mesures nécessaires pour garantir que seules les personnes ayant un accès légal à une œuvre puissent y accéder, comme par exemple une zone protégée par un mot de passe d'un site destiné aux abonnés.

L'IPO précise qu'il n'est pas considéré comme un régulateur dans ce domaine mais qu'il est l'agence gouvernementale britannique responsable de la politique du droit d'auteur en général. L'IPO propose par ailleurs un service de médiation volontaire dont les décisions ne sont pas contraignantes.

Les principales organisations du Royaume-Uni dans ce domaine sont le National Centre for Text Mining (NaCTem) et le Jisc. Les documents suivants sont consultables :

- [La fiche d'information commune publiées par le Jisc et le NaCTem](#) [lien cliquable pour accéder à la version en anglais]
- [Les directives techniques sur le TDM émises par le NaCTem](#) [lien cliquable pour accéder à la version en anglais]
- [Le guide de l'exception de TDM publié par le Jisc](#) [lien cliquable pour accéder à la version en anglais]
- [Diverses pages de directives pour les praticiens publiées par FutureTDM.](#) [lien cliquable pour accéder à la version en anglais]

L'Office de la propriété intellectuelle précise qu'il ne promeut aucune de ces organisations ou publications mais qu'elles jouent un rôle essentiel pour comprendre le paysage de la fouille de textes et de données au Royaume-Uni.

6) Suisse

Selon les informations qui nous ont été transmises par un membre de l'Institut Fédéral de la Propriété Intellectuelle suisse le 14 septembre dernier, la Suisse a adopté des modifications de la loi sur le droit d'auteur le 1^{er} avril 2020, notamment concernant une restriction au droit d'auteur en faveur de la recherche scientifique (cf. article 24d « Utilisation d'œuvres à des fins de recherche scientifique¹⁹).

La confection automatique des copies nécessaires à l'analyse de gros volumes de données (text and data mining) ne peut plus être interdit par les auteurs.

Cependant, l'invocation de la restriction en faveur de la recherche présuppose :

- que les copies des œuvres s'opèrent principalement à des fins de recherche et qu'elles soient techniquement nécessaires ;
- que les chercheurs aient eu accès licitement aux œuvres, soit parce qu'ils les ont p. ex. achetées eux-mêmes, soit parce qu'ils ont un accès légal à des revues électroniques.

La reproduction des logiciels est exclue de la restriction.

Toutes les formes de recherche scientifique (y compris la recherche commerciale) sont concernées.

Les reproductions confectionnées peuvent être conservées à des fins d'archivage et de sauvegarde une fois la recherche scientifique terminée pour assurer la possibilité de la vérification du résultat.

¹⁹ <https://www.admin.ch/opc/fr/classified-compilation/19920251/index.html#a24d>

Les auteurs n'ont pas droit à une rémunération au motif que l'œuvre n'est pas utilisée en tant que telle, mais est l'objet d'une utilisation automatisée des données qu'elle contient, ce qui ne représente pas un préjudice financier pour les auteurs.

7) Turquie

D'après les informations qui nous ont été communiquées par la Direction générale du droit d'auteur au Ministère de la culture et du tourisme de Turquie le 17 septembre 2020, il n'existe pas dans le droit national turc d'exception relative à la fouille de textes et de données mais notre interlocuteur précise qu'il existe néanmoins plusieurs exceptions dans le domaine de l'éducation et de la liberté de citation qui sont régies par la loi sur le droit d'auteur n°5846 : l'exception pour les établissements d'enseignement (article 33), l'exception à des fins éducative et pédagogiques (article 34), l'exception de libre citation (article 35).

Ces questions sont suivies par le Ministère dans le cadre d'un groupe de travail entre la Commission européenne et la Turquie sur les droits de propriété intellectuelle sans qu'il n'y ait pour autant, à l'ordre du jour, un projet de réglementation ou de modification de la loi à cet égard.

CONSEIL SUPERIEUR DE LA PROPRIETE LITTERAIRE ET ARTISTIQUE (CSPLA)

Mission sur l'exception de fouille de textes et de données en vue de la transposition en droit interne des articles 3 et 4 de la directive 2019/790 du 17 avril 2019 dite « droit d'auteur dans le marché unique numérique »

Présidente : Mme Alexandra BENSAMOUN, professeure de droit à l'Université Paris-Saclay

Rapporteur : M. Yohann BOUQUEREL, maître des requêtes en service extraordinaire au Conseil d'Etat

La directive 2019/790 instaure **deux dispositifs en faveur de la fouille de textes et de données** (*text and data mining* ou TDM en anglais) : l'une à des **fins de recherche** exclusivement (art. 3), à laquelle ne peuvent pas s'opposer les titulaires de droit concernés, et l'autre, dite « exception ou limitation », **très générale car sans finalité prédéfinie**, qui peut faire l'objet d'une opposition des titulaires de droit (mécanisme d'*opt-out* ou option de retrait).

Le CSPLA a souhaité être éclairé sur les enjeux de la transposition en droit interne de ces textes, dans la continuité du rapport de J. Martin, président, et L. de Carvalho, rapporteure, sur l'exploration de données¹ (CSPLA, juill. 2014) et du rapport de A. Bensamoun et J. Farchy, présidentes, et P.-F. Schira, rapporteur, sur l'intelligence artificielle et la culture² (CSPLA, janv. 2020).

Ce questionnaire, à but informatif, a pour objet de simplifier votre contribution au rapport ou de guider votre audition par la mission du CSPLA. Des questions complémentaires pourront être posées lors de l'audition, le cas échéant. Dans la mesure du possible, des éléments de réponse écrits sont souhaités et peuvent être remis avant le 1^{er} novembre 2020. Ces éléments de réponse pourront rester confidentiels si vous le souhaitez. Vous pouvez évidemment faire le choix de ne répondre qu'à certaines questions ou de faire une réponse groupée.

Contributeur : INA

Secteur concerné : audiovisuel

¹ <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

² <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-sur-les-enjeux-juridiques-et-economiques-de-l-intelligence-artificielle-dans-les-secteurs-de-la-creation-culturelle>

I. Questions générales

- 1) La fouille – ou exploration – de textes et de données, entendue comme l’ensemble des techniques automatisées visant à « analyser des textes et des données sous une forme numérique afin d’en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations »³, occupe-t-elle une place importante dans votre secteur d’activité ? Pouvez-vous quantifier cette activité et donner des exemples récents ? Les activités de fouille de textes et de données dans votre secteur sont-elles toujours en lien avec l’intelligence artificielle ?**

L’INA est susceptible de fouiller dans ses propres collections mais aussi dans les contenus de tiers, à des fins de recherche scientifique mais aussi à des fins commerciales. Ses contenus peuvent aussi être “fouillés” par des tiers et l’INA est alors susceptible d’intervenir en qualité de “titulaire de droits”. Le TDM n’est pas l’activité principale de l’Ina mais fait partie des outils qui sont régulièrement utilisés pour son activité de valorisation des fonds audiovisuels dont il a la charge ainsi que, plus largement, d’analyse des médias. Son utilisation commence également à être mise en œuvre dans le cadre de la description automatique de certains flux. L’intelligence artificielle est un domaine très large au contour assez flou. Le TDM peut avoir recours à des approches statistiques classiques ou des modélisations mathématiques assez complexes sans que pour autant elles puissent être qualifiées d’intelligence artificielle. Une approche souvent utilisée est l’apprentissage automatique (*machine learning*, famille des algorithmes dans laquelle se trouve le *deep learning*) qui nécessite des bases de données afin d’entraîner les algorithmes. Ces approches font également partie de la palette d’outils mis en œuvre à l’Ina, mais pas exclusivement. À titre d’exemple, outre les publications scientifiques, on peut lire des articles et ouvrages de vulgarisation présentant ces travaux :

- *Pendant le confinement, le temps de parole des femmes a baissé à la télévision et à la radio*, R. Uro et D. Doukhan, La revue des médias, 09 septembre 2020
- *Information à la télé et coronavirus : l’INA a mesuré le temps d’antenne historique consacré au Covid-19*, A. Bayet et N. Hervé, La revue des médias, 24 mars 2020
- *Les « gilets jaunes », trou noir médiatique*, R. Poirot et N. Hervé, La revue des médias, 30 juillet 2019
- *L’information à tout prix*, J. Cagé, N. Hervé et M-L. Viaud, Médias et Humanités, Ina Editions, Mars 2017

- 2) Pensez-vous que la fouille de textes et de données se développera dans les années à venir ? Pensez-vous que les acteurs de la fouille de textes et de données seront majoritairement européens ?**

Le développement de ces approches dans les années à venir est une évidence. Les principaux acteurs, d’un point de vue technologique, sont aujourd’hui les GAFAM. Prédire ce que sera l’avenir est assez difficile dans le contexte actuel, mais il est peu probable que l’Europe devienne majoritaire.

Quelle application dans l’espace ? Est-ce que les pays hors Union Européenne peuvent bénéficier de ces exceptions ?

³ Directive 2019/790, article 2, 2).

3) Quels sont les avantages et les inconvénients attendus du développement de la fouille de textes et de données, en termes de protection des droits d'auteur et assimilés, de croissance économique, de bienfaits pour les usagers ?

- Avantages en matière de bienfaits pour les usagers :
 - Permettre le développement d'études innovantes sur les collections audiovisuelles, notamment des approches relevant des humanités numériques et du big data.
 - Permettre une meilleure recherche dans les fonds, dans les contenus eux-mêmes et non dans les notices descriptives ou données associées, par ou pour les usagers du fonds INA et donc un accès plus pertinent aux ressources

Avantages en matière de croissance économique : La fouille de textes et de données permet pour l'INA d'avoir une meilleure indexation des contenus et donc d'améliorer ses services ou d'en développer de nouveaux (les données sont plus « découvrables » et peuvent être de plus en plus croisées). En facilitant le traitement de recherches « visuelles » notamment pour ses clients sur InaMediapro (par exemple, je recherche tel logo, tel monument, tel objet, telle personnalité...), cela pourrait permettre à l'INA de « mieux » vendre et d'avoir ainsi des retombées économiques bénéfiques.

Le TDM peut également permettre à l'INA de proposer de nouveaux produits ou services commerciaux, tels que des études ou expertises menées grâce à la fouille de ses collections.

Avantages en matière de protection des droits d'auteur et assimilés : développement d'outils permettant de mieux identifier les œuvres d'art par exemple ou les ayants droits (artistes-interprètes par exemple) pour mieux les rémunérer.

Inconvénients : questions d'ordre social sur l'évolution des pratiques de description, et donc des métiers

- 4) Avez-vous été associés, directement ou indirectement, à l'élaboration de l'article 38 de la loi n°2016-1321 du 7 octobre 2016 pour une République numérique, qui a instauré, en faveur de la fouille de textes et de données, pour les besoins exclusifs de la recherche scientifique, une exception au monopole des auteurs pour la reproduction de leurs œuvres, ainsi qu'aux droits des producteurs de bases de données⁴ ? Ces dispositions vous semblent-elle adaptées⁵ ?**

- 5) Avez-vous été associés, directement ou indirectement, à l'élaboration des articles 3 et 4 de la directive 2019/790 ?**

⁴ CPI, art. L. 122-5 : « Lorsque l'œuvre a été divulguée, l'auteur ne peut interdire : (...) 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques 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 ». – V. aussi art. L. 342-3, 5°.

⁵ Pour l'heure, l'exception française est doublement limitée : par son champ d'application, réservé à l'écrit, excluant ainsi la fouille sur des œuvres audiovisuelles, plastiques ou musicales ; par ses bénéficiaires, puisque seule une partie de la recherche publique est concernée, celle qui ne repose pas sur des partenariats publics-privés.

Rencontre en juin 2015 à Bruxelles de :

- Monsieur Philipp Runge, représentant de Madame Maria Martin Prat, en charge du Copyright au sein de l'Unité F5 de la Direction « Digital Economy and Coordination » de la DG Connect (Commission Européenne), <https://ec.europa.eu/dgs/connect/who-we-are> ;
- Monsieur Florian Blazy, conseiller juridique de la représentation permanente de la France auprès de l'UE, et son adjointe, Madame Brigitte Favarel ;
- Mme Anna Herold, Membre du Cabinet du Commissaire Oettinger, https://ec.europa.eu/commission/2014-2019/oettinger/team_en ;
- Mme Pauline Rouch, Conseiller politique auprès du Président Juncker, https://ec.europa.eu/commission/2014-2019/president/team_en.

Réunion téléphonique, en juin 2015, avec Madame Colette Bouckaert, représentante du Député européen Jean-Marie Cavada.

- 6) Quelle appréciation générale portez-vous sur ces deux articles ?
- 7) Avez-vous connaissance d'une transposition des articles 3 et 4 de la directive 2019/790 dans d'autres Etats membres ?
- 8) Disposez-vous d'éléments de comparaison entre le cadre juridique européen en matière de fouilles de textes et de données et les règles applicables en dehors de l'Union européenne (par exemple aux Etats-Unis et au Japon) ?
- 9) Pensez-vous que, dans le cadre juridique français actuel, les fouilles de textes et de données peuvent être réalisées, en tout ou partie :
 - en dehors des règles du droit d'auteur et assimilés, pour les données brutes et données publiques notamment ?
 - sur le fondement de l'exception de citation ?
 - sur le fondement de l'exception au profit de la recherche ?
 - sur le fondement de l'exception de reproduction technique provisoire ?
- 10) Si vous êtes éditeurs de presse, comment imaginez-vous l'articulation entre le nouveau régime des droits voisins, mentionné à l'article 15 de la directive 2019/790, et les dispositifs prévus aux articles 3 et 4 du même texte ? Souhaitez-vous attirer l'attention de la mission sur des problématiques spécifiques ?

II. Questions sur l'article 3 de la directive 2019/790

Article 3

Fouille de textes et de données à des fins de recherche scientifique

1. *Les États membres prévoient une exception aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel, en vue de procéder, à des fins de recherche scientifique, à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont accès de manière licite.*
2. *Les copies des œuvres ou autres objets protégés effectuées dans le respect du paragraphe 1 sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche.*
3. *Les titulaires de droits sont autorisés à appliquer des mesures destinées à assurer la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés. Ces mesures n'excèdent pas ce qui est nécessaire pour atteindre cet objectif.*
4. *Les États membres encouragent les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir d'un commun accord des bonnes pratiques concernant l'application de l'obligation et des mesures visées aux paragraphes 2 et 3, respectivement.*

Pour précision, cette exception est **obligatoire** et on ne peut y déroger par le contrat (dir. 2019/790, art. 7.1⁶).

1) Le cadre posé et les conditions de mise en œuvre de l'exception (bénéficiaires identifiés, fins de recherche scientifique exclusivement, accès licite, absence de compensation du fait du préjudice minime⁷) vous semblent-ils satisfaisants ?

Sur les bénéficiaires : Qu'entend-on par "organisme de recherche" ? Comment distinguer un chercheur qui relève d'un organisme de recherche (quel lien de droit doit-il avoir ?) d'une personne qui effectue des recherches scientifiques « hors organisme de recherche » ?

Sur la finalité : l'INA effectue de la fouille dans ses collections pour générer des résultats (tels que annotations automatiques, transcription, segmentation automatique de contenus diffusés selon différentes granularités (journées, émission) ...) dans le but de les mettre à disposition :

- des chercheurs de l'Inathèque (pour leurs propres recherches)
- des documentalistes de l'INA, dans le cadre de sa mission de dépôt légal et/ou de conservation et valorisation du patrimoine (mieux décrire, mieux indexer, etc.)

Cette fouille devrait entrer dans le champ de l'article 3 (précision à apporter dans un décret d'application ?). A défaut, les titulaires de droits pourraient s'opposer à l'exercice des missions de service public (mieux conserver, mieux décrire, mieux indexer, mieux valoriser...) de l'INA.

⁶ « Toute disposition contractuelle contraire aux exceptions prévues aux articles 3, 5 et 6 est non exécutoire. »

⁷ V. cons. 17.

Sur l'accès licite : quel accès licite des organismes de recherche aux contenus du DL ?

Un organisme de recherche souhaite effectuer du TDM dans les collections de l'INA : L'Ina doit-il / peut-il faire droit à ces demandes de fouille ? les conditions de l'article 3 sont-elles réunies ?

L'INA n'est pas « autorisé » en application du Code du patrimoine à remettre des copies du Dépôt Légal à un tiers (y compris s'il est chercheur au sens du Code du patrimoine);

« Une exception n'est pas un droit » : l'INA ne devrait pas non plus être « obligé » de fournir des copies du Dépôt légal à un tiers (en vue notamment de leur fouille par ce tiers).

Ainsi, nous interrogeons-nous sur l'articulation entre le Code du Patrimoine et l'article 3 de la Directive tel qu'il sera transposé.

Partant du postulat que l'INA n'est pas autorisé à fournir des copies du DL aux organismes de recherche, nous pourrions envisager que :

- l'INA effectue ces copies pour le compte des organismes de recherche, dans ses emprises, avec ses propres technologies ou avec celles des organismes de recherche
- ou
- les organismes de recherche effectuent ces copies, dans les emprises de l'INA, sans possibilité de les conserver en dehors des emprises de l'INA.

L'INA agirait comme un « tiers de confiance » vis-à-vis des titulaires de droits, garantissant d'ores et déjà dans le cadre de ses missions légales la conservation des collections objets de la fouille et en garantissant en outre la sécurité des copies, dans ses emprises. Cela nécessiterait toutefois des moyens et du temps (en matière de serveurs, de technologies et de moyens humains) dont l'INA ne dispose pas aujourd'hui.

2) Les partenariats public-privé peuvent bénéficier de cette exception « académique »⁸, ce qui en étend considérablement la portée. On pourrait par exemple imaginer, grâce à un tel partenariat, de lancer une solution d'IA créative musicale sur une base de données d'archives comme celle de l'INA, ou d'IA créative graphique et plastique sur un fonds muséal. Comment transposer cette ouverture ?

Quels sont les partenariats éligibles ? Quels sont les critères pour définir les partenariats public-privé ? Financements publics, subvention, mission d'intérêt général, projet reconnu par la communauté scientifique (partenariats européens de type H2020 ou nationaux de type ANR), etc.?

L'exception peut être invoquée pour la fouille effectuée par l'INA, à des fins de recherche scientifique, dans ses collections, avec ses propres outils.

Mais qu'en est-il lorsque l'INA fouille en utilisant les technologies de tiers dans le cadre de partenariat public-privé (pas de reproduction/copie des contenus par le partenaire mais partage

⁸ Dir. 2019/790, cons. 11 : « (...) En concordance avec l'actuelle politique de la recherche de l'Union, qui encourage les universités et les instituts de recherche à collaborer avec le secteur privé, les organismes de recherche devraient également bénéficier d'une telle exception lorsque leurs activités de recherche sont menées dans le cadre de partenariats public-privé. Les organismes de recherche et les institutions du patrimoine culturel devraient rester les bénéficiaires de cette exception, mais ils devraient pouvoir s'appuyer sur leurs partenaires privés pour effectuer des fouilles de textes et de données, y compris en utilisant leurs outils technologiques. »

des résultats avec lui). Se pose ici la question des modèles d'entraînement et de leurs utilisations par les entreprises privées.

3) Afin de favoriser les activités de recherche et l'interopérabilité entre les œuvres, seriez-vous favorables à l'utilisation d'un format standardisé dans votre secteur d'activité (par exemple XML pour les textes et MIDI pour la musique) ?

Il s'agit d'une bonne pratique et non d'un préalable, il vaut mieux des formats ouverts et bien documentés plutôt que strictement standardisés.

4) Comment vérifier que la durée de conservation des copies des œuvres et des autres objets protégés, mentionnée à l'alinéa 2, n'est pas excessive ?

Cela nécessite de définir des bonnes pratiques avec les chercheurs, qui peuvent apporter une vision opérationnelle des besoins.

5) Quelles modalités souhaiteriez-vous voir mises en place pour la conservation des copies⁹ ?

En l'absence de possibilité de contrôler la destruction des copies, il serait utile d'instaurer des "Data management plans" tels que ceux que l'on trouve dans les projets ANR, définissant les conditions d'accès, de sécurité, de conservation de respect de normes (telles que le RGPD).

6) Quelles pourraient être les mesures mises en œuvre par les titulaires de droits pour assurer la sécurité et l'intégrité des réseaux, structures et bases de données sollicités, en permettant le TDM ?

Un audit des bénéficiaires de l'exception, demandé par les titulaires de droits, sur la base de « data mangement plans » et effectué par un organisme de contrôle ou tiers de confiance à déterminer (à l'instar de la CNIL pour le respect du RGPD ou de la Business Software Agency (BSA) pour les licences de logiciels) pourrait être envisagé.

7) Comment le gouvernement pourrait-il encourager les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir conjointement des bonnes pratiques pour l'application de cet article 3 ? Seriez-vous favorable à l'élaboration rapide d'une charte des bonnes pratiques entre les acteurs concernés ? Souhaiteriez-vous que ce processus soit mené dans le cadre du CSPLA ?

Nous sommes favorables à l'élaboration d'une charte : la charte des bonnes pratiques est nécessaire pour encadrer et préciser l'application de l'exception de fouille de textes et de données. Demeure la question des sanctions en cas de violation de cette charte.

⁹ V. cons. 15 : « Dans certains cas, par exemple aux fins de la vérification a posteriori des résultats de la recherche scientifique, les organismes de recherche et les institutions du patrimoine culturel pourraient avoir besoin de conserver des copies faites dans le cadre de l'exception aux fins d'effectuer de la fouille de textes et de données. En pareils cas, les copies devraient être stockées en lieu sûr. Les États membres devraient être libres de décider, au niveau national et après concertation avec les acteurs concernés, des modalités plus précises de conservation des copies, notamment la capacité de désigner des organismes de confiance aux fins du stockage de ces copies. Afin que l'application de cette exception ne soit pas restreinte indûment, ces modalités devraient être proportionnées et limitées à ce qui est nécessaire à la conservation des copies en toute sécurité et à la prévention de leurs utilisations non autorisées. »

Le CSPLA pourrait être une bonne option.

III. Questions sur l'article 4 de la directive 2019/790

Article 4

Exception ou limitation pour la fouille de textes et de données

- 1. Les États membres prévoient une exception ou une limitation aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE, à l'article 4, paragraphe 1, points a) et b), de la directive 2009/24/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions d'œuvres et d'autres objets protégés accessibles de manière licite aux fins de la fouille de textes et de données.*
- 2. Les reproductions et extractions effectuées en vertu du paragraphe 1 peuvent être conservées aussi longtemps que nécessaire aux fins de la fouille de textes et de données.*
- 3. L'exception ou la limitation prévue au paragraphe 1 s'applique à condition que l'utilisation des œuvres et autres objets protégés visés audit paragraphe n'ait pas été expressément réservée par leurs titulaires de droits de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne.*
- 4. Le présent article n'affecte pas l'application de l'article 3 de la présente directive.*

- 1) La directive semble donner aux Etats membres une certaine marge de manœuvre lors de sa transposition en droit interne, renvoyant à une approche fondée sur une « exception » ou sur une « limitation » des droits d'auteurs et assimilés. Quelles conséquences en tirez-vous ?
- 2) **Le cadre et les conditions de mise en œuvre (tous bénéficiaires, tous usages, y compris commerciaux, accès licite) vous paraissent-ils satisfaisants ?**

Sur l'accès licite : la question de l'accès licite est à préciser. Par exemple, sur notre site inamediapro, l'accès n'est donné sur accréditation qu'aux professionnels de l'audiovisuel : l'accès par une entité n'entrant pas dans ce cadre ne constitue-t-il donc pas un accès licite au contenu ? De même, sur ina.fr et madelen.fr, les CGU réservent l'accès aux contenus aux personnes physiques pour une utilisation "dans le cadre du cercle de famille". Dans ce cas, faut-il considérer que l'accès aux contenus pour une utilisation commerciale ne constitue pas un accès licite ? Ces CGU limitatives définissent-elles l'étendue de l'accès licite et/ou sont-elles des modalités d'opt-out ?

- 3) **Comment comprenez-vous le droit d'option ouvert au titulaire et qui permet de se retirer du dispositif ?**

C'est au titulaire de droit de démontrer qu'il a exercé son droit de retrait.

- 4) **Quels seraient, selon vous, les moyens techniques appropriés (« procédés lisibles par machine ») que les titulaires de droits pourraient utiliser pour s'opposer (« *opt-out* », *i.e.* option de retrait) à une fouille de textes et de données réalisée par des organismes autres que ceux mentionnés à l'article 3 (métadonnées, conditions générales d'utilisation, accords contractuels, déclarations unilatérales, fichier centralisé, mesures techniques de protection, ou autres moyens) ? Faudrait-il standardiser ces moyens ? *Quid* en cas de doute sur l'exercice de l'*opt-out* ?**

Cf. 12) ci-dessous

Les clauses du type « utilisation des contenus à des fins privées, dans le cadre du cercle de famille » (dans les CGU) ne sont par exemple pas assez explicites.

5) Pensez-vous exercer cette option de retrait ?

L'INA détient un fonds composé d'archives soumises aux droits de différents et nombreux titulaires de droits : auteurs, producteurs, coproducteurs, artistes-interprètes, etc. Chacune de ces catégories d'ayants droits pourrait (individuellement ou collectivement) s'opposer à la fouille de tout ou partie de ses œuvres/objets protégés par la Propriété Intellectuelle à des fins commerciales. Si tel est le cas, l'INA sera conduit à exercer un droit de retrait sur l'ensemble de son catalogue pour ne pas avoir à gérer les demandes d'*opt-out* des différents ayants-droit sur chaque contenu, ce qui, compte-tenu de la nature du fonds de l'INA, ne serait pas envisageable.

- 6) Pensez-vous que les titulaires de droits dans votre secteur d'activité vont s'opposer massivement à la fouille de textes et de données pour des besoins autres que ceux de la recherche scientifique ?

7) En cas d'opposition des titulaires de droit à la fouille de textes et de données, les organismes qui souhaitent réaliser ces fouilles pourront-ils se fonder sur les dispositions citées à la question 9 du I) du présent questionnaire (autres exceptions/limitations) ?

L'opération de fouille de données sur des œuvres ou objets protégés implique la copie et la reproduction desdites œuvres, et - dans le cas de l'application de l'article 4 - pour une finalité commerciale (ou autre qu'une simple finalité de recherche scientifique). Les exceptions actuelles au droit d'auteur, citées à la question 9 du I) ne sont pas adaptées ou trop restrictives et donc inutilisables.

L'exception de courte citation n'est également pas adaptée dans la mesure où Il ne s'agit pas de « citer » des œuvres/objets mais bien d'utiliser le contenu en lui-même pour générer des résultats, sans rapport direct avec l'œuvre source utilisée.

L'exception de reproduction technique provisoire n'est pas plus adaptée en ce qu'elle est trop restrictive. Cette exception implique que la copie soit provisoire, transitoire et constitue une partie intégrante et essentielle d'un procédé technique. Or, si l'œuvre source n'est utilisée que pour permettre la fouille et en tirer des résultats, les copies des œuvres et objets protégés doivent pouvoir être conservées pendant une durée plus longue et a minima le temps de comparer, de vérifier et de tester les résultats avec leurs sources.

8) Comment apprécier la durée de conservation des reproductions et extractions de données ? Faudrait-il prévoir une sanction en cas de durée de conservation excessive ?

L'appréciation doit être faite en fonction de l'usage envisagé des résultats de la fouille: bonnes pratiques selon une approche sectorielle?

9) Pensez-vous que cette exception est susceptible de limiter très fortement l'intérêt à investir dans des bases de données (par exemple, les bases de données d'images ou de textes élaborées par les éditeurs) ?

10) A quelles conditions les titulaires de droit pourraient-ils accepter les fouilles de textes et de données mentionnées à l'article 4 de la directive ? Le recours à une licence prévoyant des contreparties financières vous semble-t-il adapté ?

Outre la difficulté de « gestion des opt out » (cf. point 5 ci-dessus), s'ajoute la difficulté de gérer les éventuelles rémunérations des différents titulaires de droits : qui perçoit ? comment répartir ? quid pour les ayants droits non représentés par une société de gestion collective (comme les producteurs titulaires de droits voisins) ?

11) Pensez-vous que la fouille de textes et de données dans un but commercial pourrait paradoxalement être freinée en cas de nombreuses options de retrait de la part des titulaires de droit ?

Oui puisque l'exercice massif de l'opt-out empêcherait toute fouille en dehors du cas de recherche scientifique. On sortirait de l'exception pour se retrouver sur l'exercice plein et entier du monopole des titulaires de droit.

12) D'un point de vue technique, comment peut-on mettre en œuvre les dispositions de cet article 4 sans faire obstacle aux dispositions de l'article 3 de la directive¹⁰ ?

1. Pour les données disponibles sur des sites web, mise en place d'un fichier du type robots.txt, standardisé, indiquant quels sont les contenus et/ou les entités exclus de l'exception TDM. Cette mesure doit donc pouvoir être ignorée par les bénéficiaires mentionnés à l'article 3 pour des fins de recherche académique.

2. Si toutefois une solution technique venait à être mise en place pour bloquer automatiquement certains robots aspirant les sites web, il faut prévoir une exception sous forme de liste blanche pour les projets de recherche. La procédure pour être inscrit sur cette liste blanche ne doit pas conduire à l'exclusion de fait des chercheurs pour cause de complexité abusive.

13) Quelles pourraient être les règles communes aux dispositifs prévus aux articles 3 et 4 de la directive (par exemple, durée de conservation des reproductions et extractions) ? Une approche sectorielle vous paraît-elle utile ?

Les règles de stockage et de conservation pourraient être communes aux deux articles.

¹⁰ L'article 4.4 de la directive l'impose : « Le présent article n'affecte pas l'application de l'article 3 de la présente directive. »

Néanmoins, les durées de conservations dépendent des utilisations que l'on veut faire des résultats de la fouille, une approche sectorielle peut donc être envisagée.

IV. Questions diverses

- 1) Quelles seraient les précisions relatives à la fouille de textes et de données que vous souhaiteriez voir inscrites dans les textes législatifs et réglementaires de transposition des articles 3 et 4 de la directive 2019/790, sous réserve qu'elles ne soient pas contraires à la lettre et à l'esprit de ces dispositions ?
- 2) En cas d'expertise technique sur le TDM, quels formats utilisez-vous ? Avez-vous connaissance d'un document scientifique récapitulatif sur ce point ?
- 3) Quelles autres remarques, générales ou spécifiques, souhaitez-vous formuler auprès de la mission ?

CONSEIL SUPERIEUR DE LA PROPRIETE LITTERAIRE ET ARTISTIQUE
(CSPLA)

Mission sur l'exception de fouille de textes et de données en vue de la transposition en droit interne des articles 3 et 4 de la directive 2019/790 du 17 avril 2019 dite « droit d'auteur dans le marché unique numérique »

Présidente : Mme Alexandra BENSAMOUN, professeure de droit à l'Université Paris-Saclay

Rapporteur : M. Yohann BOUQUEREL, maître des requêtes en service extraordinaire au Conseil d'Etat

La directive 2019/790 instaure **deux dispositifs en faveur de la fouille de textes et de données** (*text and data mining* ou TDM en anglais) : l'une à des **fins de recherche** exclusivement (art. 3), à laquelle ne peuvent pas s'opposer les titulaires de droit concernés, et l'autre, dite « exception ou limitation », **très générale car sans finalité prédéfinie**, qui peut faire l'objet d'une opposition des titulaires de droit (mécanisme d'*opt-out* ou option de retrait).

Le CSPLA a souhaité être éclairé sur les enjeux de la transposition en droit interne de ces textes, dans la continuité du rapport de J. Martin, président, et L. de Carvalho, rapporteure, sur l'exploration de données[1] (CSPLA, juill. 2014) et du rapport de A. Bensamoun et J. Farchy, présidentes, et P.-F. Schira, rapporteur, sur l'intelligence artificielle et la culture[2] (CSPLA, janv. 2020).

Ce questionnaire, à but informatif, a pour objet de simplifier votre contribution au rapport ou de guider votre audition par la mission du CSPLA. Des questions complémentaires pourront être posées lors de l'audition, le cas échéant. Dans la mesure du possible, des éléments de réponse écrits sont souhaités et peuvent être remis avant le 1^{er} novembre 2020. Ces éléments de réponse pourront rester confidentiels si vous le souhaitez. Vous pouvez évidemment faire le choix de ne répondre qu'à certaines questions ou de faire une réponse groupée.

Contributeur : Comité pour la Science Ouverte

Secteur concerné : Recherche

I. Questions générales

- 1) La fouille – ou exploration – de textes et de données, entendue comme l'ensemble des techniques automatisées visant à « *analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations* »[3], occupe-t-elle une place importante dans votre secteur d'activité ? Pouvez-vous quantifier cette activité et donner des exemples récents ? Les activités de fouille de textes et de données dans votre secteur sont-elles toujours en lien avec l'intelligence artificielle ?

Le TDM est une activité majeure et en croissance dans le domaine de la recherche. Cela concerne bien sûr l'intelligence artificielle, mais va bien au-delà. Un grand nombre de fronts pionniers sont ouverts grâce à cette approche. Pour donner quelques exemples : lors des Journées nationales de la science ouverte, nous avons organisé une session dédiée au TDM.

- *Des microbes dans mon fromage ?!*
- *Apprendre à lire aux ordinateurs pour améliorer le blé*
- *Cartographier le cerveau grâce aux publications en neurosciences*
- *Le projet politoscope*
- *Fouille de texte pour l'enrichissement et l'analyse sémantique de la presse ancienne européenne*

<https://jnso2018.sciencesconf.org/resource/page/id/4>

- 2) Pensez-vous que la fouille de textes et de données se développera dans les années à venir ? Pensez-vous que les acteurs de la fouille de textes et de données seront majoritairement européens ?

Il existe de fortes attentes dans la recherche publique en France vis-à-vis du TDM et ces pratiques sont appelées à se développer, mais à condition de pouvoir disposer d'un cadre juridique praticable et adapté.

La France a de réels atouts dans le domaine du TDM et de l'IA, côté recherche, mais elle est restée limitée jusqu'à présent par un régime juridique contraint. Rappelons-nous que Google fait du TDM depuis 20 ans et que la Chine dispose aussi d'un cadre juridique permettant la pratique du TDM.

L'exception va redonner une respiration aux acteurs européens et leur donner une chance d'aller plus loin. Certains pays européens, comme l'Angleterre ou l'Allemagne, ont mis en place des exceptions pour favoriser la pratique du TDM, mais l'absence d'uniformité à l'échelle de l'Union européenne n'est pas de nature à favoriser l'émergence de projets transnationaux, seule échelle qui permettrait de rivaliser avec des pays comme les Etats-Unis.

- 3) Quels sont les avantages et les inconvénients attendus du développement de la fouille de textes et de données, en termes de protection des droits d'auteur et assimilés, de croissance économique, de bienfaits pour les usagers ?

La ligne directrice suivie par les acteurs qui sont intervenus pour demander l'adoption de l'exception au niveau européen est la suivante : le droit de lire doit être le droit de fouiller. *"The right to read is the right to mine"*. C'est-à-dire que le TDM devrait être possible sur tous les contenus en accès licite. L'extraction automatisée d'informations n'est que le prolongement logique des opérations réalisées auparavant manuellement par les chercheurs.

Les bénéfices scientifiques, économiques et culturels sont très supérieurs aux inconvénients. Les ayants droit n'ont par ailleurs jamais réellement réussi à mettre en place une offre payante sur une base contractuelle pour ce genre de services. On ne peut citer aucun exemple de plateforme qui permettrait aujourd'hui de réaliser des opérations de TDM dans des conditions satisfaisantes pour la recherche à partir d'une infrastructure fournie directement par un

éditeur. Tout au plus a-t-on vu ces dernières années des clauses apparaître dans des licences nationales qui permettent la récupération d'archives de revues scientifiques pour les intégrer à la [plateforme ISTEEX](#) qui permet de réaliser des opérations de TDM. Ce dispositif ISTEEX est d'ailleurs la preuve que la recherche publique est en mesure de mettre en place des infrastructures sécurisées pour le TDM.

Il n'y a donc pas de manque à gagner pour les ayant-droits, parce qu'aucun d'eux n'est parvenu à mettre en place de manière convaincante de services payants de TDM.

Par ailleurs, la diversité des objets étudiés par la recherche via le TDM dépasse très largement toute conception amont de corpus. C'est tout le web qui est visé, dans son extraordinaire diversité et hétérogénéité, pas le seul secteur de l'édition.

4) Avez-vous été associés, directement ou indirectement, à l'élaboration de l'article 38 de la loi n°2016-1321 du 7 octobre 2016 pour une République numérique, qui a instauré, en faveur de la fouille de textes et de données, pour les besoins exclusifs de la recherche scientifique, une exception au monopole des auteurs pour la reproduction de leurs œuvres, ainsi qu'aux droits des producteurs de bases de données[4] ? Ces dispositions vous semblent-elle adaptées[5] ?

Oui, nous avons été associés à la loi pour une République numérique.

La France avait eu le mérite d'essayer de faire évoluer son cadre juridique pour faire progresser les pratiques de TDM. Mais les dispositions de la loi pour une République numérique comportaient un certain nombre de lacunes et de faiblesses, que l'on ne retrouve heureusement pas dans la directive européenne :

- Elles ne concernaient que les articles scientifiques (peut-être 1% des corpus ciblés) : seulement des contenus textuels, et seulement certains contenus textuels. Le TDM doit pouvoir être pratiqué sur tous les contenus protégés, du moment qu'existe un accès licite (images, musiques, oeuvres audiovisuelles, contenus en ligne, bases de données) ;

- Elles ne définissaient pas le TDM.

- Elles introduisaient une distinction entre les fichiers source et les "copies techniques" qui aurait sans doute rendu extrêmement complexe l'application de l'exception et que l'on ne retrouve heureusement pas dans la directive ;

- Elles ne formulaient pas exactement de la même manière l'exception au droit d'auteur et l'exception au droit des bases de données, ce qui constituait une complexité inutile.

- Elles n'ont pas pu être validées par le Conseil d'Etat.

- En revanche, elles indiquent que les fichiers issus des activités de TDM avaient la nature de “données de la recherche”, ce qui nous paraît positif.

Pour ces raisons, il importe de ne pas repartir du texte de la loi République numérique pour la transition, mais bien de la directive européenne elle-même.

5) Avez-vous été associés, directement ou indirectement, à l’élaboration des articles 3 et 4 de la directive 2019/790 ?

Oui.

6) Quelle appréciation générale portez-vous sur ces deux articles ?

Globalement, des deux articles sont satisfaisants et comportent des améliorations substantielles par rapport à l’exception de la loi République numérique :

- Ils comportent une définition claire du TDM, très utile.
- Ils ont une envergure européenne, ce qui est une excellente échelle, très supérieure à la portée de la loi pour une République numérique.
- La notion “d’accès licite” est très utile. Elle est la traduction directe du principe: “*The right to read is the right to mine*” ;
- Les articles s’appliquent à tous les types de documents : pas seulement textuels, et pas seulement articles scientifiques. Elle s’applique donc à tous les types de médias : oeuvres textuelles, musicales, audiovisuelles, graphiques, plastiques. On pourrait à ce titre lors de la transposition, renvoyer aux articles L 112-1 et L 112-2 de Code de la propriété intellectuelle pour couvrir tout le champ des oeuvres protégées ;
- Ils dépassent s’appliquent non seulement au droit d’auteur, mais aussi au droit des bases de données (ce qui est aussi le cas de l’exception de la loi République numérique).
- Ils définissent clairement les bénéficiaires de l’exception (voir ci après pour précisions) ;
- Le TDM est aussi ouvert aux institutions patrimoniales, ce que ne faisait pas la loi République numérique. BNF, INA, par exemple, en bénéficieront très largement. Il y a notamment un bel enjeu autour du dépôt légal du web issu de la loi DADVSI de 2006.

L’article 3 comporte néanmoins également quelques faiblesses.

Il s’agit notamment des “mesures de sécurisation” que les ayants droit sont autorisés à déployer. Les ayants droit peuvent en effet “*appliquer des mesures destinées à assurer la sécurité et l’intégrité des réseaux et des bases de données* » en précisant que ces mesures ne doivent pas excéder « *ce qui est nécessaire pour atteindre cet objectif* ». Bien qu’il repose sur des préoccupations légitimes, cet alinéa constitue aussi un des points qui peuvent fortement

nuire à l'effectivité de l'exception TDM. Il sera en effet difficile au législateur de trouver des formulations qui permettent de savoir précisément ce que les titulaires de droits peuvent déployer comme mesures de sécurisation et jusqu'où elles peuvent aller.

Ces aspects pourraient être détaillés dans un des décrets d'application ou dans un code des bonnes pratiques, comme y invite explicitement la directive. Néanmoins, étant donné le caractère mouvant des technologies de TDM et l'infinie variété des contenus mobilisables, il est assez improbable d'arriver à sécuriser complètement ce point, d'où l'importance de prévoir des mécanismes efficaces de règlement des conflits potentiels entre ayants droit et chercheurs (voir plus bas pour des propositions).

7) Avez-vous connaissance d'une transposition des articles 3 et 4 de la directive 2019/790 dans d'autres Etats membres ?

Pas à ce stade.

8) Disposez-vous d'éléments de comparaison entre le cadre juridique européen en matière de fouilles de textes et de données et les règles applicables en dehors de l'Union européenne (par exemple aux Etats-Unis et au Japon) ?

Aux USA, c'est la souplesse du système de "fair use" qui permet de faire du TDM (voir les suites de l'affaire Google Books tranchée par la Cour suprême des Etats-Unis qui a ouvert cette possibilité). Mais cette logique de "fair use" est éloignée de notre droit, qui raisonne en terme d'exceptions.

9) Pensez-vous que, dans le cadre juridique français actuel, les fouilles de textes et de données peuvent être réalisées, en tout ou partie :

- en dehors des règles du droit d'auteur et assimilés, pour les données brutes et données publiques notamment ?
- sur le fondement de l'exception de citation ?
- sur le fondement de l'exception au profit de la recherche ?
- sur le fondement de l'exception de reproduction technique provisoire ?

1) en dehors des règles du droit d'auteur et assimilés, pour les données brutes et données publiques notamment ?

Oui, pour les contenus appartenant au domaine public (mais à condition de pouvoir trouver des contenus certifiés comme appartenant au domaine public sans couche de droits rajoutées, ce qui est rare en France vu les pratiques des institutions culturelles). Oui également pour les données publiques à condition que celles-ci soient effectivement publiées par les administrations en application du principe d'ouverture par défaut prévu par la loi République

numérique (par exemple toutes les données ouvertes par les administrations sur le portail Data.gouv.fr et placées sous Licence Ouverte).

2) sur le fondement de l'exception de citation ?

Non, l'exception de citation n'est pas appropriée pour faire du TDM, du fait de la conception très restrictive des juges français. Elle peut néanmoins s'avérer être utile pour citer de courts extraits d'oeuvres pour illustrer les résultats du TDM.

3) sur le fondement de l'exception au profit de la recherche ?

Non. Le Conseil d'Etat a estimé que l'exception pédagogique et de recherche n'a pas vocation à être mobilisée pour faire du TDM et c'est ce qui a entraîné le blocage du décret d'application de la loi pour une République numérique. Maintenant que nous disposons d'une exception dédiée, il vaut mieux nettement distinguer les champs d'application des deux exceptions. Rappelons en outre que l'exception pédagogique et de recherche fait l'objet d'une compensation financière alors que c'est exclu par la directive pour l'exception TDM.

4) sur le fondement de l'exception de reproduction technique provisoire ?

Non. Le champ d'application de cette exception est très étroit. Elle joue un rôle utile pour les navigateurs web par exemple et gestion de leur mémoire cache, mais cela ne correspond pas aux besoins du TDM qui sont beaucoup plus larges.

10) Si vous êtes éditeurs de presse, comment imaginez-vous l'articulation entre le nouveau régime des droits voisins, mentionné à l'article 15 de la directive 2019/790, et les dispositifs prévus aux articles 3 et 4 du même texte ? Souhaitez-vous attirer l'attention de la mission sur des problématiques spécifiques ?

Nous ne sommes pas des éditeurs de presse.

II. Questions sur l'article 3 de la directive 2019/790

Article 3

Fouille de textes et de données à des fins de recherche scientifique

- 1. Les États membres prévoient une exception aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel, en vue de procéder, à des fins de recherche scientifique, à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont accès de manière licite.*
- 2. Les copies des œuvres ou autres objets protégés effectuées dans le respect du paragraphe 1 sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche.*
- 3. Les titulaires de droits sont autorisés à appliquer des mesures destinées à assurer la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés. Ces mesures n'excèdent pas ce qui est nécessaire pour atteindre cet objectif.*
- 4. Les États membres encouragent les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir d'un commun accord des bonnes pratiques concernant l'application de l'obligation et des mesures visées aux paragraphes 2 et 3, respectivement.*

Pour précision, cette exception est **obligatoire** et on ne peut y déroger par le contrat (dir. 2019/790, art. 7.1[6]).

- 1) Le cadre posé et les conditions de mise en œuvre de l'exception (bénéficiaires identifiés, fins de recherche scientifique exclusivement, accès licite, absence de compensation du fait du préjudice minime[7]) vous semblent-ils satisfaisants ?

L'article 3 délimite clairement le périmètre des bénéficiaires de l'exception. Il vise en effet « *les reproductions et les extractions effectuées par des organismes de recherche et des institutions du patrimoine culturel* ». On peut dès lors renvoyer au Code de la recherche (livre III notamment) pour identifier de manière très claire quels sont les organismes de recherche concernés et au Code du patrimoine pour les institutions du patrimoine culturel.

Les considérants de la directive ne ferment cependant pas la porte à ce que des entités privées puissent être considérés comme des organismes de recherche, à condition qu'elles aient un but non lucratif ou qu'elles exercent leurs activités dans le cadre d'une mission d'intérêt public reconnue par l'Etat (ex : Fondation Nationale des Sciences Politiques). On notera que c'est uniquement pour ces entités privées que la directive fait des distinctions sur le but lucratif, mais elle n'emploie pas ce genre de terme de manière générale. Il paraît donc peu judicieux d'insérer des formules du type « à l'exclusion de toute finalité commerciale » dans la loi, étant donné les difficultés d'interprétation de ce genre d'expression.

Concernant "l'accès licite", les considérants (14, 18) éclairent sur le double sens que l'on doit donner à cette expression : elle vise à la fois les contenus auxquels on accède sur la base d'une autorisation contractuelle (via un abonnement par exemple) et les contenus diffusés en libre accès sur Internet.

Ce dernier aspect est très important, car un des champs privilégiés du *Text and Data Mining* repose sur les analyses de contenus en ligne, type réseaux sociaux (LeBonCoin, Waze, Tinder, etc.).

Par ailleurs, il y a un réel intérêt à pouvoir utiliser l'exception sur des sources en Open Access, comme des archives ouvertes (HAL) ou des sites patrimoniaux (Persée) où les contenus sont certes en libre accès, mais pas systématiquement librement réutilisables.

Il est important de relever que l'exception dans la directive concerne aussi bien les oeuvres protégées par le droit d'auteur que les bases de données. La loi République numérique procédait d'ailleurs déjà ainsi et aurait conduit à l'insertion de deux exceptions : une au droit d'auteur (article L. 122-5) et une au droit des bases de données (article L343-3). Ce dernier point est essentiel, car de nombreux contenus en accès licite sur Internet, même s'ils ne constituent pas des oeuvres protégées, sont inclus dans des bases de données (exemple: réseaux sociaux type Facebook ou Twitter, sites comme LeBonCoin, *etc.*). Il faut donc bien que l'exception couvre tous les contenus en accès licite et pas simplement les oeuvres au sens du droit d'auteur.

- 2) Les partenariats public-privé peuvent bénéficier de cette exception «académique»[8], ce qui en étend considérablement la portée. On pourrait par exemple imaginer, grâce à un tel partenariat, de lancer une solution d'IA créative

musicale sur une base de données d'archives comme celle de l'INA, ou d'IA créative graphique et plastique sur un fonds muséal. Comment transposer cette ouverture ?

Le considérant 11 indique en effet explicitement que l'exception devrait être applicable dans le cadre de partenariats public-privé de recherche et ce point est essentiel, car ces partenariats jouent aujourd'hui un rôle très important dans la dynamique de l'innovation et de la recherche. La transposition devrait explicitement sécuriser cette possibilité.

Une manière de délimiter précisément le périmètre de ces partenariats inclus dans le champ de l'exception pourrait consister à reprendre le critère des recherches financées à au moins 50% par des crédits publics, déjà utilisé par la loi République numérique à son article 30 et qui permettrait une délimitation claire et simple à opérer parmi les projets.

3) Afin de favoriser les activités de recherche et l'interopérabilité entre les œuvres, seriez-vous favorables à l'utilisation d'un format standardisé dans votre secteur d'activité (par exemple XML pour les textes et MIDI pour la musique) ?

Non. Il existe des milliers de formats XML différents, il n'est pas raisonnable de rendre obligatoire l'adoption d'un standard. Il est quasiment impossible d'imposer un standard unique, qui impliquerait des coûts importants et, in fine, produirait sans doute des effets de déception. Les chercheurs rencontrent des difficultés lorsqu'ils doivent jongler avec des formats multiples, hétérogènes, médiocrement renseignés, etc., et ont l'habitude de gérer de telles contraintes.

En revanche :

- si l'éditeur dispose de plusieurs formats, pouvoir disposer du format le plus flexible et le plus sémantique (riche en informations) serait une excellente chose, mais seulement sur l'option "*best effort*" ; par exemple, du PDF en mode image n'est pas très exploitable.
- il faut s'assurer que le *scrapping* soit autorisé et possible, les API ne pouvant jamais couvrir l'ensemble des besoins ;
- il est tout-à-fait possible de produire des **bonnes pratiques, non contraignantes**, pour converger vers des formats communs ;
- il est nécessaire, cependant, que les formats soient lisibles et documentés quelque part : l'expression "*formats ouverts et lisibles par des machines*", qui figure par ailleurs déjà dans la loi République numérique, pourrait être reprise.

Définition du scrapping : "Le web scraping (parfois appelé harvesting) est une technique d'extraction du contenu de sites Web, via un script ou un programme, dans le but de le transformer pour permettre son utilisation dans un autre contexte, par exemple le référencement.". Source: https://fr.wikipedia.org/wiki/Web_scraping Pour plus de détails : https://en.wikipedia.org/wiki/Web_scraping

4) Comment vérifier que la durée de conservation des copies des œuvres et des autres objets protégés, mentionnée à l'alinéa 2, n'est pas excessive ?

L'exception de la loi République numérique s'appuyait sur la notion de « copies techniques » pour désigner les fichiers issus des traitements opérés à partir des sources mobilisées pour faire du TDM. Il en résultait une approche complexe, avec des fichiers ayant plusieurs statuts différents et une obligation de destruction des sources.

Ces distinctions auraient été en pratique une source majeure d'incertitude et d'insécurité. La nouvelle directive ne mentionne à aucun moment la nécessité d'opérer de telles distinctions et elle n'impose pas de détruire certains fichiers.

Bien au contraire, elle prévoit explicitement la possibilité de « stocker », mais aussi de « conserver » les copies, ce terme étant suffisamment général pour englober les copies techniques, mais aussi les fichiers source.

Ce point est crucial pour envisager la constitution « d'archives du TDM », qui permettront de garantir la vérifiabilité des recherches, mais aussi que de nouvelles recherches pourront être effectuées à partir de sources conservées.

A noter à ce sujet que même si la directive ne parle que de « reproductions » et pas de « communications », elle sous-entend une telle possibilité, puisqu'elle prévoit que les copies peuvent être *conservées à des fins de vérification et de recherche*. La vérification n'a de sens que si elle est effectuée par d'autres chercheurs que ceux qui ont réalisés l'analyse initiale, donc une communication des archives est bien possible et, si c'est le cas, on doit considérer que les copies peuvent aussi être conservées pour être transmises à d'autres chercheurs afin qu'ils puissent réaliser de nouvelles recherches.

Cette conservation doit néanmoins s'effectuer de manière sécurisée et la directive laisse une large marge de manœuvre pour définir au niveau national les conditions de cette conservation et les organismes habilités.

5) Quelles modalités souhaiteriez-vous voir mises en place pour la conservation des copies[9] ?

Cette discussion a déjà eu lieu en France sur le sujet lors de l'élaboration du décret d'application de la loi République numérique qui avait conduit à considérer que tous les organismes de recherche publique devraient être habilités à conserver les copies issues du TDM.

Il n'y a pas lieu aujourd'hui de revenir sur ce point, d'autant plus que les services d'archivage numérique se sont encore développés dans les universités et établissements de recherche depuis 2016. En particulier, l'ESRI dispose de 99 infrastructures nationales de recherche, et de nombreux data center et plateformes sécurisés adaptés à ces usages.

Pour les données les plus sensibles, l'ESRI dispose du CASD - Centre d'accès sécurisé aux données.

- 6) Quelles pourraient être les mesures mises en œuvre par les titulaires de droits pour assurer la sécurité et l'intégrité des réseaux, structures et bases de données sollicités, en permettant le TDM ?

Il n'y a pas lieu de rentrer dans le détail au niveau de la loi. La directive emploie l'expression *“mesures destinées à assurer la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés [...] n'excèdent pas ce qui est nécessaire pour atteindre cet objectif”* qui devrait être reprise telle quelle dans la loi lors de la transposition.

Néanmoins, comme signalé plus haut, cette question des mesures de sécurisation peut constituer un véritable point faible, susceptible de nuire à l'effectivité de l'exception si des mécanismes de régulation et de règlement des litiges potentiels ne sont pas mis en place.

On pourrait par exemple faire appel aux dispositions du Code de propriété intellectuelle qui confère à la Hadopi compétence pour examiner des demandes d'utilisateurs entravés dans l'exercice légitime d'une exception en raison de l'application de Mesures Techniques de Protection excessives.

La loi République numérique n'avait pas prévu une telle possibilité, mais il serait sans doute utile de prévoir un tel mécanisme d'arbitrage extra-judiciaire. Les technologies du TDM étant mouvantes et complexes, il paraît approprié de pouvoir s'en remettre à une autorité indépendante pour trancher d'éventuels litiges entre ayants droit et utilisateurs.

- 7) Comment le gouvernement pourrait-il encourager les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir conjointement des bonnes pratiques pour l'application de cet article 3 ? Seriez-vous favorable à l'élaboration rapide d'une charte des bonnes pratiques entre les acteurs concernés ? Souhaiteriez-vous que ce processus soit mené dans le cadre du CSPLA ?

Un instrument du type « code des bonnes pratiques du TDM » aurait indéniablement son intérêt, mais il importe là aussi que l'élaboration d'un tel document, qui peut s'avérer longue et complexe, ne bloque pas l'applicabilité de l'exception dès l'entrée en vigueur de la loi de transposition.

Il paraîtrait en outre judicieux de laisser d'abord les usages se mettre en place et ensuite seulement, entamer la réflexion sur les bonnes pratiques sur la base des retours de terrain. La loi pourrait même prévoir un rapport de bilan sur la pratique du TDM au terme d'une certaine durée (un an ?), faculté dont dispose le législateur français.

Il importera également de choisir avec soin les organes qui seront associés à l'élaboration de ce code des bonnes pratiques. Le Comité pour la Science Ouverte, qui rassemble de nombreux experts de ces questions, devrait en tant que tel être associés à ces travaux et on pourrait également imaginer que le Conseil National du Numérique supervise l'élaboration du code, cette instance ayant déjà travaillé sur la question du TDM, au moment de la préparation de la loi République numérique et à l'occasion du rapport « France Intelligence Artificielle » de Cédric Villani.

III. Questions sur l'article 4 de la directive 2019/790

Article 4

Exception ou limitation pour la fouille de textes et de données

- 1. Les États membres prévoient une exception ou une limitation aux droits prévus à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, à l'article 2 de la directive 2001/29/CE, à l'article 4, paragraphe 1, points a) et b), de la directive 2009/24/CE et à l'article 15, paragraphe 1, de la présente directive pour les reproductions et les extractions d'œuvres et d'autres objets protégés accessibles de manière licite aux fins de la fouille de textes et de données.*
- 2. Les reproductions et extractions effectuées en vertu du paragraphe 1 peuvent être conservées aussi longtemps que nécessaire aux fins de la fouille de textes et de données.*
- 3. L'exception ou la limitation prévue au paragraphe 1 s'applique à condition que l'utilisation des œuvres et autres objets protégés visés audit paragraphe n'ait pas été expressément réservée par leurs titulaires de droits de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne.*
- 4. Le présent article n'affecte pas l'application de l'article 3 de la présente directive.*

- 1) La directive semble donner aux Etats membres une certaine marge de manœuvre lors de sa transposition en droit interne, renvoyant à une approche fondée sur une « exception » ou sur une « limitation » des droits d’auteurs et assimilés. Quelles conséquences en tirez-vous ?

Nous plaidons pour une exception en bonne et due forme, pour des raisons de compétitivité et d’harmonisation européenne.

Notre principal souci est de permettre la création d’activités économiques issues de l’effort de recherche du secteur public, notamment via les SATT, vers le privé (création de *start-up* issues de projets de recherche, par exemple).

Il importe que le passage de l’article 3 à l’article 4 soit le plus harmonieux et fluide possible. Comme nous l’avons montré plus haut, il est simple de délimiter le champ d’application de l’article 3 en renvoyant à la définition des établissements et organismes de recherche qui figure au livre III du Code de la recherche. En revanche, si l’activité se déporte vers une entreprise, même liée à la recherche, alors c’est l’article 4 qui aurait vocation à s’appliquer.

- 2) Le cadre et les conditions de mise en œuvre (tous bénéficiaires, tous usages, y compris commerciaux, accès licite) vous paraissent-ils satisfaisants ?

Oui. Il paraît important que des acteurs privés puissent recourir aux techniques du TDM, y compris dans le cadre d’une activité commerciale. Le rapport “France Intelligence Artificielle” insistait sur le bénéfice qui pourrait en résulter pour le développement de l’intelligence artificielle en France. Par ailleurs, il s’agit d’un enjeu crucial de compétitivité au niveau international. Rappelons à ce titre que les entreprises aux Etats-Unis peuvent faire du TDM sur des contenus protégés depuis des années et que cela a contribué à ce que des acteurs comme les GAFAM prennent de l’avance considérable dans un domaine stratégique comme celui de l’intelligence artificielle.

- 3) Comment comprenez-vous le droit d’option ouvert au titulaire et qui permet de se retirer du dispositif ?

Un titulaire de droits a la possibilité d’exercer un “*opt-out*”, interdisant donc le TDM lorsqu’il excède le champ de l’article 3.

- 4) Quels seraient, selon vous, les moyens techniques appropriés (« procédés lisibles par machine ») que les titulaires de droits pourraient utiliser pour s’opposer (« *opt-out* »,

i.e. option de retrait) à une fouille de textes et de données réalisée par des organismes autres que ceux mentionnés à l'article 3 (métadonnées, conditions générales d'utilisation, accords contractuels, déclarations unilatérales, fichier centralisé, mesures techniques de protection, ou autres moyens) ? Faudrait-il standardiser ces moyens ? *Quid* en cas de doute sur l'exercice de l'*opt-out* ?

Les Mesures Techniques de Protection constitueraient une très mauvaise option. Elles n'arriveraient sans doute pas à faire la différence entre les acteurs publics et les acteurs privés. Dès lors, ces "protections" visant l'article 4 pourraient avoir des répercussions négatives sur l'exercice de l'article 3.

Nous plaillons pour une déclaration de l'*opt-out* via des métadonnées. qui a déjà fait ses preuves dans d'autres domaines. On peut par exemple penser au fichier web "robots.txt" qui permet déjà aux titulaires de droits de s'opposer à l'indexation de leurs contenus par les moteurs de recherche. Il faudrait arriver pour l'option de retrait de l'article 4 à un même niveau de standardisation et de simplicité d'usage.

Il faut par ailleurs absolument que la déclaration de retrait soit lisible par les machines. De simples CGU ne peuvent pas suffire, étant donné que les pratiques de TDM ont vocation à s'exercer à grande échelle.

Un fichier centralisé paraît inapproprié : il serait lourd, toujours incomplet et difficile à mettre à jour. Cela représenterait pour les ayants droit eux-mêmes une charge bien plus complexe à effectuer qu'une simple déclaration via un fichier simple à ajouter sur leur propre site web.

5) Pensez-vous exercer cette option de retrait ?

Nous ne sommes pas concernés par la possibilité d'activer l'option de retrait.

De manière générale, nous pensons important de trouver des voies consensuelles permettant de limiter les *opt-out*.

6) Pensez-vous que les titulaires de droits dans votre secteur d'activité vont s'opposer massivement à la fouille de textes et de données pour des besoins autres que ceux de la recherche scientifique ?

Il est fort possible que tous les éditeurs scientifiques type Elsevier activent l'option de retrait afin de se protéger des concurrents et pour éviter de favoriser des *outsiders*.

7) En cas d'opposition des titulaires de droit à la fouille de textes et de données, les organismes qui souhaitent réaliser ces fouilles pourront-ils se fonder sur les

dispositions citées à la question 9 du I) du présent questionnaire (autres exceptions/limitations) ?

Cela paraît très improbable. Certaines de ces exceptions ne sont ouvertes qu'aux acteurs publics (exception pédagogique et de recherche) ; les autres (citation, reproductions techniques temporaires, etc.) sont trop étroites.

- 8) Comment apprécier la durée de conservation des reproductions et extractions de données ? Faudrait-il prévoir une sanction en cas de durée de conservation excessive ?

Même réponse que pour l'article 3.

- 9) Pensez-vous que cette exception est susceptible de limiter très fortement l'intérêt à investir dans des bases de données (par exemple, les bases de données d'images ou de textes élaborées par les éditeurs) ?

Ce n'est pas l'intention de l'exception. L'exception TDM ne permet pas de dupliquer et communiquer à l'identique les contenus.

La Commission européenne a par ailleurs évalué plusieurs fois le droit des bases de données européen et était arrivé à la conclusion qu'il est en soi peu efficace sur le plan économique. Elle a même envisagé plusieurs fois de le supprimer (voir par exemple : <https://scoms.hypotheses.org/598>).

- 10) A quelles conditions les titulaires de droit pourraient-ils accepter les fouilles de textes et de données mentionnées à l'article 4 de la directive ? Le recours à une licence prévoyant des contreparties financières vous semble-t-il adapté ?

Le recours à des contreparties financières n'est pas l'esprit de l'exception. Les contreparties financières nous ramènent dans le cadre contractuel classique, qu'il s'agit de dépasser, car il n'est pas adapté dans de nombreuses hypothèses.

Des licences pourraient à la rigueur fonctionner lorsque le TDM s'effectue sur une seule source dont les titulaires de droits sont identifiables. Mais dès qu'il s'agit de constituer des corpus avec de nombreuses sources, les coûts de transaction explosent. Or, ce cas de figure est très courant.

- 11) Pensez-vous que la fouille de textes et de données dans un but commercial pourrait paradoxalement être freinée en cas de nombreuses options de retrait de la part des titulaires de droit ?

Oui, clairement.

On peut néanmoins imaginer que seuls des titulaires de droits particuliers, type éditeurs de presse, procéderont à cet *opt-out* systématique. Mais les titulaires de droits sont en réalité bien plus nombreux sur Internet que les seuls acteurs professionnels (ex: auteurs de blogs, etc.). Il est possible que les titulaires non-professionnels exercent l'*opt-out* et l'exception de l'article 4 gardera son intérêt pour ce type de contenus.

12) D'un point de vue technique, comment peut-on mettre en œuvre les dispositions de cet article 4 sans faire obstacle aux dispositions de l'article 3 de la directive[10] ?

Nous avons déjà donné des éléments sur ce point dans les réponses précédentes. Le moyen le plus sûr est de définir le plus clairement possible le cercle des bénéficiaires de l'article 3. On peut le faire sur la base d'un renvoi au Code de la Recherche (livre III) qui liste les organismes et établissements de recherche. Mais il faut éviter d'ajouter dans le texte de l'exception Article 3 une limitation du type "à l'exclusion de tout usage commercial" qui ne figure pas dans la directive et qui restreindrait indûment son périmètre (notamment parce qu'elle s'applique aussi aux partenariats public-privé). Pour délimiter les partenariats public-privé couverts, le critère des 50% de financements publics est pertinent.

13) Quelles pourraient être les règles communes aux dispositifs prévus aux articles 3 et 4 de la directive (par exemple, durée de conservation des reproductions et extractions) ? Une approche sectorielle vous paraît-elle utile ?

Non, il vaut mieux clairement séparer les choses et la loi n'a de toute façon pas à entrer dans des détails comme la durée de conservation.

IV. Questions diverses

1) Quelles seraient les précisions relatives à la fouille de textes et de données que vous souhaiteriez voir inscrites dans les textes législatifs et réglementaires de transposition des articles 3 et 4 de la directive 2019/790, sous réserve qu'elles ne soient pas contraires à la lettre et à l'esprit de ces dispositions ?

Ouvrir la possibilité de porter les éventuels litiges devant la Hadopi, au titre de son rôle de gardienne de l'effectivité des exceptions. Privilégier cette piste souple plutôt que de chercher à entrer trop dans les détails dans le texte.

- 2) En cas d'expertise technique sur le TDM, quels formats utilisez-vous ? Avez-vous connaissance d'un document scientifique récapitulatif sur ce point ?

La loi n'a pas pour nous vocation à entrer dans des détails comme la question des formats (question qui n'apparaît nulle part dans la directive). Une mention générique, comme la référence à des "formats ouverts et lisibles par les machines" pourrait suffire et elle présente l'avantage de déjà figurer dans la loi République numérique.

- 3) Quelles autres remarques, générales ou spécifiques, souhaitez-vous formuler auprès de la mission ?

Il nous paraît important de rester pour la transposition le plus proche possible des formulations même de la directive et ne pas chercher à rajouter des éléments dans la loi nationale. C'est le meilleur moyen pour laisser ouvertes les potentialités riches du texte européen, tout en évitant également des décalages par rapport aux législations des autres pays européens. L'homogénéité du régime juridique du TDM au niveau européen est un enjeu de premier plan si l'on veut pouvoir penser à cette échelle dans la compétition technologique internationale.

[1] <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-relative-au-text-and-data-mining-exploration-de-donnees>

[2] <https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique/Travaux/Missions/Mission-du-CSPLA-sur-les-enjeux-juridiques-et-economiques-de-l-intelligence-artificielle-dans-les-secteurs-de-la-creation-culturelle>

[3] Directive 2019/790, article 2, 2).

[4] CPI, art. L. 122-5 : « Lorsque l'œuvre a été divulguée, l'auteur ne peut interdire : (...) 10° Les copies ou reproductions numériques réalisées à partir d'une source licite, en vue de l'exploration de textes et de données incluses ou associées aux écrits scientifiques 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 ». – V. aussi art. L. 342-3, 5°.

[5] Pour l'heure, l'exception française est doublement limitée : par son champ d'application, réservé à l'écrit, excluant ainsi la fouille sur des œuvres audiovisuelles, plastiques ou musicales ; par ses bénéficiaires, puisque seule une partie de la recherche publique est concernée, celle qui ne repose pas sur des partenariats publics-privés.

[6] « Toute disposition contractuelle contraire aux exceptions prévues aux articles 3, 5 et 6 est non exécutoire. »

[7] V. cons. 17.

[8] Dir. 2019/790, cons. 11 : « (...) En concordance avec l'actuelle politique de la recherche de l'Union, qui encourage les universités et les instituts de recherche à collaborer avec le secteur privé, les organismes de recherche devraient également bénéficier d'une telle exception lorsque leurs activités de recherche sont menées dans le cadre de partenariats public-privé. Les organismes de recherche et les institutions du patrimoine culturel devraient rester les bénéficiaires de cette exception, mais ils devraient pouvoir s'appuyer sur leurs partenaires privés pour effectuer des fouilles de textes et de données, y compris en utilisant leurs outils technologiques. »

[9] V. cons. 15 : « Dans certains cas, par exemple aux fins de la vérification a posteriori des résultats de la recherche scientifique, les organismes de recherche et les institutions du patrimoine culturel pourraient avoir besoin de conserver des copies faites dans le cadre de l'exception aux fins d'effectuer de la fouille de textes et de données. En pareils cas, les copies devraient être stockées en lieu sûr. Les États membres devraient être libres de décider, au niveau national et après concertation avec les acteurs concernés, des modalités plus précises de conservation des copies, notamment la capacité de désigner des organismes de confiance aux fins du stockage de ces copies. Afin que l'application de cette exception ne soit pas restreinte indûment, ces modalités devraient être proportionnées et limitées à ce qui est nécessaire à la conservation des copies en toute sécurité et à la prévention de leurs utilisations non autorisées. »

[10] L'article 4.4 de la directive l'impose : « Le présent article n'affecte pas l'application de l'article 3 de la présente directive. »