



MINISTÈRE DE LA CULTURE

*Liberté
Égalité
Fraternité*

CONSEIL SUPERIEUR DE LA PROPRIETE LITTERAIRE ET ARTISTIQUE

RAPPORT DE MISSION

TRANSPOSITION DES EXCEPTIONS DE FOUILLE DE TEXTES ET DE DONNEES : ENJEUX ET PROPOSITIONS

Présidente de la mission :
Alexandra BENSAMOUN, professeure des universités, personnalité qualifiée au CSPLA

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

*Rapport présenté à la réunion plénière du CSPLA du 15 décembre 2020.
Son contenu n'engage que ses auteurs.*

Décembre 2020

Le contenu de ce rapport, fruit d'une réflexion menée à partir d'auditions, de contributions écrites, de contacts informels et de nombreuses lectures, n'engage que ses auteurs.

15 décembre 2020

Table des matières

Transposition des exceptions sur la fouille de textes et de données : enjeux et propositions

Synthèse.....	5
Introduction.....	9
I. La fouille de textes et de données : une activité en pleine expansion récemment appréhendée par le législateur.....	12
A. La fouille de textes et de données : une activité en pleine expansion	12
1) Une activité aux contours flous mais qui fait enfin l'objet d'une définition juridique au niveau européen	12
2) Une technique indispensable dans un environnement numérique en pleine expansion et favorable au développement de l'intelligence artificielle	14
3) Une activité prometteuse appelée à croître rapidement.....	17
B. La fouille de textes et de données concerne tous les secteurs protégés par le droit de la propriété littéraire et artistique, mais à des degrés divers	19
1) La fouille de textes et de données porte sur toutes les catégories de données numériques	19
2) La fouille de textes et de données est difficilement conciliable, dans le cadre juridique français, avec les exceptions aux règles du droit d'auteur	20
3) Les enjeux de la fouille sont divers pour les titulaires de droits.....	21
C. L'article 38 de la loi pour une République numérique : une tentative inaboutie pour promouvoir la fouille de textes et de données	22
1) Un dispositif réservé à la recherche publique et centré sur les écrits scientifiques	22
2) L'absence de décret d'application	24
D. Eléments de droit comparé.....	25
II. La directive du 17 avril 2019 prévoit deux dispositifs obligatoires en matière de fouille dérogeant au monopole des titulaires de droits	27
A. Les principes suivis par le législateur européen.....	27
1) Une approche globale et volontariste sur la fouille de textes et de données.....	27
2) La nécessité de concilier les intérêts des ayants droit et des acteurs de la fouille de textes et de données.....	28
3) La reconnaissance implicite que la fouille de textes et de données peut méconnaître les droits d'auteur et assimilés.....	29
4) Une directive qui respecte l'acquis européen.....	31
5) Mais qui s'articule parfois difficilement avec les exceptions actuelles	31
B. Article 3 : la traduction de la volonté de la Commission européenne de promouvoir la fouille de textes et de données en faveur de la recherche scientifique	33
1) La version initiale de l'article 3 proposée par la Commission européenne	33
2) Une exception qui déroge à de nombreuses prérogatives des titulaires de droits	34

3) Une exception dont ne peuvent bénéficier gratuitement que les organismes de recherche et les institutions du patrimoine culturel	35
4) Une exception obligatoire, qui ne peut être remise en cause ni par voie contractuelle, ni par l'exercice du droit de retrait mentionné à l'article 4 de la directive	37
5) Une exception qui ménage les intérêts des deux parties	37
6) Une exception qui invite les deux parties à négocier une charte de bonnes pratiques.....	38
C. Article 4 : un dispositif obligatoire introduit tardivement dans le projet de directive, qui explique son caractère atypique.....	39
1) Une disposition introduite tardivement et cependant obligatoire	39
2) Un dispositif qui déroge à de nombreuses prérogatives des titulaires de droits.....	40
3) Une très large ouverture de la dérogation	41
4) Une ouverture contrebalancée par l'option de retrait.....	42
5) Une disposition moins aboutie que l'article 3	42
D. Des règles communes aux articles 3 et 4 de la directive	43
1) Des dérogations encadrées.....	43
2) Des dispositifs soumis au test des trois étapes.....	44
3) Des dispositifs garantis.....	44
III. Présentation des différentes possibilités de transposition des articles 3 et 4 de la directive du 17 avril 2019	45
A. Le cadre juridique pour la transposition de cette directive	46
1) Des marges d'appréciation limitées lors de la transposition d'une directive d'harmonisation..	46
2) Le cas particulier des exceptions au droit d'auteur.....	47
3) Une transposition des articles 3 et 4 en cours dans les Etats membres.....	48
B. Les enjeux communs liés à la transposition des articles 3 et 4.....	49
C. Les options spécifiques liées à la transposition de l'article 3.....	56
D. Les options spécifiques liées à la transposition de l'article 4.....	73
IV. Propositions de la mission.....	87
A. Préconisations générales.....	87
B. Préconisations textuelles.....	89
Lettre de mission.....	93
Questionnaire élaboré par la mission	96
Liste des contributeurs.....	102

Les annexes, comprenant certaines des contributions écrites reçues, ont été réunies dans un document séparé.

Synthèse

Transposition des exceptions de fouille de textes et de données

Aux termes de l'article 1^{er} de la directive 2019/790 du 17 avril 2019¹, la fouille de textes et de données désigne « toute technique d'analyse automatisée 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 ». Cette définition recouvre un large spectre d'activités, qui vont de la veille médiatique à l'exploration des prix dans un secteur donné en passant par l'analyse de l'évolution d'une notion dans un journal, par exemple. L'idée commune de toutes les fouilles est de rendre explicites des liens dissimulés dans des masses de données afin de créer de nouvelles informations.

Cette activité appelée à croître rapidement est étroitement liée au développement de l'intelligence artificielle.

Considérant que les règles actuelles du droit d'auteur et assimilés étaient inadaptées à la fouille de textes et de données, le législateur européen a prévu deux dispositifs dans la directive pour encourager son développement :

- A l'article 3, une exception « académique » au bénéfice des organismes de recherche et des institutions du patrimoine culturel qui effectuent des fouilles à des fins de recherche scientifique, à laquelle les titulaires de droits ne peuvent s'opposer ;
- A l'article 4, une exception/limitation aux droits des titulaires au bénéfice de toute fouille, quelle que soit sa finalité, sous réserve toutefois que le titulaire n'ait pas exprimé son opposition (droit de retrait, également appelé *opt-out*).

L'article 7 de la même directive indique les dispositions communes à ces articles, notamment leur caractère obligatoire, l'applicabilité du test en trois étapes ou encore l'obligation de garantir le bénéfice effectif de ces dérogations.

L'article 34 de la loi portant diverses dispositions d'adaptation au droit de l'Union européenne en matière économique et financière (dite « loi DDADUE »)², du 3 décembre 2020, habilite le Gouvernement à légiférer par ordonnance pour transposer les articles 3, 4 et 7 dans un délai d'un an à compter de la promulgation de cette loi.

Si les Etats membres ne disposent pas de marge de manœuvre pour transposer les dispositions claires et précises de la directive d'harmonisation 2019/790, ils retrouvent une certaine latitude quand un thème est hors champ de la directive ou qu'il est traité par d'autres articles de la directive. Dans tous les cas, la mission rappelle la nécessité d'assurer un haut niveau de protection des droits et de préserver un équilibre entre les droits des titulaires de droits et les intérêts des utilisateurs.

Aussi, en vue de la transposition des deux dispositions concernées, la mission recommande les modifications suivantes.

¹ Directive (UE) 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.

² Loi n° 2020-1508.

Les propositions de modifications législatives

➤ Les dispositions communes

- 1) La mission propose de transposer les articles 3 et 4 aux articles L. 122-5 (droit d'auteur), L. 211-3 (droits voisins) et L. 342-3 (droit *sui generis* des bases de données) du code de la propriété intellectuelle (CPI).
- 2) La mission recommande de supprimer les dispositions issues de l'article 38 de la loi sur la République numérique qui portait sur l'exploration de textes et de données afin de ne pas maintenir un dispositif incompatible avec la directive.
- 3) La mission préconise de ne retenir dans la loi que les principes essentiels des articles 3 et 4 de la directive et de renvoyer au pouvoir réglementaire le soin de définir leurs conditions d'application, tandis que les parties prenantes sont invitées à élaborer une charte des bonnes pratiques sous la supervision du CSPLA.

➤ Article 3

- 1) La loi précisera que l'exception ne bénéficie qu'aux organismes de recherche et aux institutions du patrimoine culturel ayant un accès licite aux contenus et aux seules fins de la recherche scientifique.
- 2) Ces organismes et institutions ne pourront pas bénéficier de cette exception « académique » s'ils ont conclu une convention avec une entreprise lui accordant un accès privilégié aux résultats de la fouille.
- 3) Les copies et reproductions numériques effectuées lors d'une fouille de textes et de données sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins exclusives de recherche scientifique, y compris pour la vérification des résultats de la recherche.
- 4) Le titulaire de droits peut mettre en œuvre des mesures proportionnées et nécessaires afin d'assurer la sécurité et l'intégrité des réseaux et des bases de données dans lesquels les œuvres sont hébergées.
- 5) Une charte définit les bonnes pratiques relatives à l'application de cette exception « académique ».
- 6) Toute stipulation contraire aux dispositions précitées est nulle et non avenue.

➤ Article 4

- 1) Comme le CPI ne reconnaît pas la notion de limitation aux droits des auteurs et assimilés, la mission propose de transposer l'article 4 sous la forme d'une exception, même si le mécanisme semble davantage relever d'une limitation.
- 2) La mission propose de modifier les mêmes articles du CPI que pour l'article 3 de la directive, ainsi que l'article L. 122-6-1 du CPI relatif au droit d'auteur spécial portant sur les programmes d'ordinateur (cet ajout résulte clairement des termes de l'article 4 de la directive).
- 3) Toute personne peut effectuer des reproductions et des extractions d'œuvres et d'autres objets protégés accessibles de manière licite en vue de réaliser une fouille, quelle que soit sa finalité, sauf si le titulaire de droits s'y est opposé expressément au moyen de procédés appropriés.
- 4) Cette opposition ne peut avoir pour objet ou pour effet de remettre en cause l'exception « académique » mentionnée à l'article 3 de la directive.
- 5) Les copies et reproductions sont stockées avec un niveau de sécurité approprié puis détruites à l'issue de la fouille de textes et de données.

➤ **Hadopi**

La mission propose de confier trois nouvelles missions à la Haute autorité :

1) Une mission de régulation et de veille en matière de mesures techniques de protection (MTP) utilisées dans le cadre des fouilles.

Par la modification de l'article L. 331-31 du CPI, l'Hadopi pourra exercer une mission de conciliation sur le fondement de l'article L. 331-36 du même code pour toute question relative à la mise en œuvre effective des exceptions relatives à la fouille quand sont en jeu des MTP. Cette mission doit être prévue par les Etats membres compte tenu de l'article 7 de la directive qui oblige ceux-ci à garantir l'effectivité des exceptions.

2) Une mission de régulation pour arbitrer tous les différends en lien avec des fouilles relevant de l'article 3 de la directive, indépendamment des MTP précitées.

Cette nouvelle mission, inscrite directement dans le corps des articles L. 122-5, L. 221-3 et L. 342-3 du même code, portera notamment sur les mesures de protection du réseau hébergeant les œuvres protégées et sur les modalités de conservation des reproductions.

3) Une mission de suivi de la charte des bonnes pratiques en modifiant l'article L. 331-14 du CPI.

Les propositions de modifications réglementaires

➤ **Les règles transversales**

1) Dans la mesure où les deux nouvelles exceptions issues des articles 3 et 4 de la directive seront transposées à l'article L. 122-5 du CPI, un décret en Conseil d'Etat sera nécessaire pour définir leurs conditions d'application.

2) La mission recommande de définir par voie réglementaire la fouille de textes et de données en reprenant les termes de l'article 2 de la directive.

3) La mission préconise également de préciser la notion d'accès licite en raison de la spécificité de la fouille car la libre accessibilité de données en ligne semble présumer d'un accès licite. Les exceptions issues des articles 3 et 4 de la directive ne devraient pas s'appliquer aux fouilles réalisées à partir de sites ou de fournisseurs de services dont l'objectif manifeste est de se livrer à du piratage de droit d'auteur ou d'autres droits ou de faciliter cette activité.

➤ **Article 3**

La mission invite le pouvoir réglementaire à :

- Définir strictement le champ des bénéficiaires de l'exception académique, à savoir les organismes de recherche (universités, établissements publics à caractère scientifique, culturel et professionnel, organismes de recherche définis au livre III du code de la recherche, personnes morales de droit privé poursuivant une mission d'intérêt général en matière de recherche) et les institutions du patrimoine culturel (bibliothèques accessibles au public, musées, archives, institutions responsables du dépôt légal) ;
- Indiquer que la notion d'accès privilégié aux résultats de cette fouille recouvre notamment la communication des résultats provisoires en avant-première alors qu'une publication est prévue, la communication exclusive de tout ou partie des résultats pendant une période déterminée ou encore l'attribution initiale et exclusive d'un brevet éventuel ;

- Etendre l'application des règles encadrant les partenariats public-privé prévus pour les organismes de recherche aux institutions du patrimoine culturel ;
- Qualifier les résultats de la fouille « académique » de données de la recherche afin de souligner leur vocation à circuler parmi les chercheurs ;
- Clarifier l'articulation entre l'article 3 de la directive et l'exception posée par l'article L. 132-4 du code du patrimoine, dans la mesure où cet article crée un accès légal aux œuvres déposées mais que les chercheurs ne peuvent pas réaliser des copies de ces œuvres et les transporter en dehors des emprises des institutions depositaires (BNF, INA et CNC notamment) ;
- Autoriser ces institutions depositaires à instaurer si besoin des redevances pour couvrir tout ou partie des coûts liés à une fouille ;
- Préciser les règles protectrices de conservation des copies des œuvres et contenus protégés afin d'éviter tout risque de dissémination ;
- Mentionner les 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 ;
- Définir les thèmes que doit aborder *a minima* la charte des bonnes pratiques.

➤ **Article 4**

La mission invite le pouvoir réglementaire à :

- Préciser que le droit d'opposition est libre et qu'il n'a pas à être motivé ;
- Indiquer qu'un titulaire de droits est libre de changer d'avis (dans cette hypothèse, l'acteur de la fouille doit prendre en compte l'existence ou non d'un droit d'opposition à la date de la collecte des données) ;
- Prévoir que lorsque les données sont accessibles en ligne, les mesures appropriées pour exprimer l'opposition doivent être lisibles par une machine, ce qui autorise notamment le recours à des conditions générales d'utilisation (CGU) utilisant les formats les plus répandus, à des métadonnées, à des dômes de protection ou encore à des fichiers centralisés (y compris avec reconnaissance de contenus) ;
- Préciser qu'un titulaire de droits peut exprimer son opposition par de multiples canaux, à charge pour lui d'assurer la cohérence des informations délivrées ;
- Indiquer qu'il incombe au « fouilleur » de s'assurer qu'aucun droit d'opposition n'a été exprimé, ce qui implique qu'il paramètre le cas échéant ses robots lorsqu'il moissonne un site ;
- Inviter les acteurs de la fouille « hors recherche » à assister en tant qu'observateurs à la négociation de la charte de bonnes pratiques.

La mission a constaté que les solutions techniques envisagées pour exprimer l'opposition des titulaires de droits sont d'efficacité très variable. Un dispositif a retenu l'attention de la mission : dénommé ORRC (Outil de Régulation des Robots de Crawling) et développé par le Centre français d'exploitation du droit de copie (CFC), cet outil permet d'identifier les robots autorisés à collecter des données. Ce dispositif est riche de promesses car il permettrait de donner une grande efficacité au droit d'opposition sans bloquer les fouilles « académiques ». Il nécessite toutefois un investissement financier et humain qui est aujourd'hui hors de portée d'un très grand nombre d'ayants droit.

Introduction

De la société de l'information au marché unique numérique, l'Union européenne a entrepris d'appliquer à la propriété intellectuelle sa nouvelle stratégie, notamment fondée sur la **donnée**³ et sur les potentialités associées en termes d'**intelligence artificielle**⁴.

En matière de droit d'auteur, l'utilisation de l'intelligence artificielle (IA) pose des questions en amont de la création et en aval⁵.

D'une part, la réflexion européenne est en cours sur la question, en aval, des créations générées par une IA, comme l'indique notamment la Commission européenne dans sa Communication du 25 novembre 2020⁶.

D'autre part, se pose la question, en amont, du statut des données entrantes en cas d'exploration de contenus. Utilisant les techniques d'apprentissage-machine, les outils d'IA sont nourris par des masses de données. Aussi, pour permettre le développement de ces technologies, la **directive 2019/790 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 (dite plus loin « directive de 2019 ») a prévu deux dispositifs en faveur de la fouille de textes et de données (également appelée *text and data mining* ou TDM) aux articles 3 et 4 : une exception à des **fins de recherche** exclusivement (art. 3), à laquelle ne peuvent pas s'opposer les titulaires de droits concernés, et une « *exception ou limitation* », **très ouverte car sans finalité prédéfinie**, mais qui peut faire l'objet d'une opposition des titulaires de droit (mécanisme d'*opt-out* ou option de retrait).

Avec pour objectif « *d'adapter et de compléter le cadre actuel de l'Union en matière de droit d'auteur, tout en maintenant un niveau élevé de protection du droit d'auteur et des droits voisins* » (cons. 3), la directive énonce les règles suivantes :

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.

³ Communications de la Commission intitulées « *Une stratégie européenne pour les données* » (COM(2020)0066) et « *Une nouvelle stratégie industrielle pour l'Europe* » (COM(2020)0102).

⁴ Résolution du Parlement européen du 20 octobre 2020 sur les droits de propriété intellectuelle pour le développement des technologies liées à l'intelligence artificielle (2020/2015(INI)).

⁵ Ces deux aspects sont mis en exergue dans le rapport CSPLA de A. Bensamoun, J. Farchy, présidentes, et P.-F. Schira, rapporteur, sur « *Les enjeux juridiques et économiques de l'intelligence artificielle dans les secteurs de la création culturelle* », janv. 2020 : <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>

⁶ Communication de la Commission européenne, « *Exploiter au mieux le potentiel d'innovation de l'Union européenne. Un plan d'action en faveur de la propriété intellectuelle afin de soutenir la reprise et la résilience dans l'Union européenne* », 25 nov. 2020, COM(2020) 760 final.

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.*

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.*

Ces deux dérogations étant **obligatoires**, les États membres doivent les transposer en droit national. Aussi, dans la continuité du rapport sur l'intelligence artificielle et la culture⁷, le CSPLA a souhaité être éclairé sur **les enjeux de la transposition en droit interne** de ces textes⁸, et a donc confié, par une lettre du 10 juin 2020, une mission à Alexandra Bensamoun, professeure de droit à l'Université Paris-Saclay, présidente, rejointe par Yohann Bouquerel, maître des requêtes en service extraordinaire au Conseil d'État, rapporteur.

L'article 34 de la loi portant diverses dispositions d'adaptation au droit de l'Union européenne en matière économique et financière (dite « loi DDADUE »)⁹, du 3 décembre 2020, habilite le Gouvernement à légiférer par ordonnance pour transposer les articles 3, 4 et 7 dans un délai d'un an à compter de la promulgation de cette loi.

⁷ *Op. cit.*

⁸ Le CSPLA avait déjà, de manière précurseur, proposé un rapport en juill. 2014 sur le sujet de l'exploration de données, réalisé par J. Martin, président, et L. de Carvalho, rapporteure : <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>

⁹ Loi n° 2020-1508.

Prenant acte de la technicité du sujet¹⁰, la mission a d'abord diffusé un questionnaire indicatif identifiant les principaux enjeux posés par les articles 3 et 4 de la directive. Elle a ensuite procédé à de nombreuses auditions¹¹ et recueilli plusieurs contributions écrites (dont certaines ont été reproduites dans les annexes).

S'agissant du rapport lui-même, l'objet même de la mission a imposé une démarche pédagogique. La mission a donc d'abord expliqué l'activité de fouille et ses implications (I), avant de présenter les orientations de la directive et des dispositions concernées (II). Une fois ce cadre général posé, la mission a exposé les avantages et inconvénients des différentes options offertes, en justifiant ses choix (III). Enfin, des propositions textuelles clôturent ce rapport (IV).

¹⁰ Qui suscite d'ailleurs un intérêt relatif au sein de la doctrine. V. principalement N. Binctin, « TDM : un enjeu de l'intelligence artificielle », RIDA 262, oct. 2019, p. 5 ; E. Rosati, « *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity* », Asia Pacific Law Review, sept. 2019 : <https://ssrn.com/abstract=3452376> ; C. Geiger, G. Frosio, Giancarlo et O. Bulayenko, « *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU* », oct. 2019, <https://ssrn.com/abstract=3470653>. – Ces collègues ont contribué à la mission.

¹¹ Compte tenu du contexte sanitaire actuel, les auditions ont été réalisées à distance.

I. La fouille de textes et de données : une activité en pleine expansion récemment appréhendée par le législateur

L'activité de fouille de textes et de données est relativement récente, d'où sa prise en compte tardive par les législateurs, tant en France (C) qu'à l'étranger, notamment aux Etats-Unis (D). Pour autant, elle est en pleine expansion (A) et son développement implique inégalement les objets protégés par la propriété littéraire et artistique (B).

A. La fouille de textes et de données : une activité en pleine expansion

1) Une activité aux contours flous mais qui fait enfin l'objet d'une définition juridique au niveau européen

Dans son rapport du CSPLA de 2014, notre collègue Me Jean Martin, assisté de Mme Liliane de Carvalho, indiquait que l'expression même de fouille de textes et de données ne faisait pas l'unanimité au tournant des années 2010, certaines personnes préférant évoquer « l'exploration » ou « l'analyse » de données. Ce rapport définissait « *l'activité d'exploration de données* » comme un processus de recherche automatisé portant sur un ensemble de données numériques dans le but de découvrir de nouvelles connaissances ou de nouvelles idées¹².

L'article 38 de la loi du 7 octobre 2016 pour une République numérique, qui avait pour objet principal de créer une exception au droit d'auteur au profit de la recherche publique lorsqu'elle se livre à l'exploration de textes et de données, ne définissait pas cette dernière.

Dans son rapport intitulé « *Donner un sens à l'intelligence artificielle* » de mars 2018, le député Cédric Villani avait également relevé que la fouille de textes et de données « *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* » et qu'elle est à ce titre « *porteuse de potentiels énormes pour la découverte scientifique et le développement de nouvelles connaissances* »¹³.

L'article 1^{er} de la directive 2019/790 est venu mettre un terme aux querelles sémantiques en définissant ainsi la fouille de textes et de données : « *toute technique d'analyse automatisée 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*

¹² Rapport de la mission sur l'exploration de données, *text and data mining*, président M. Jean Martin, rapporteur Mme Liliane de Carvalho, juill. 2014, p. 9 : <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>

¹³ Rapport « *Donner un sens à l'intelligence artificielle - Pour une approche nationale et européenne* », de M. Cédric Villani, mars 2018, p. 35 : https://www.aiforhumanity.fr/pdfs/9782111457089_Rapport_Villani_accessible.pdf

corrélations ». La définition est ici **fonctionnelle**. Il s'agit en effet de décrire le processus technique et son objectif, mais sans qualifier juridiquement le résultat de la fouille notamment.

Comme l'a indiqué M. Dias, professeur à l'Université de Caen Normandie, spécialiste en traitement automatique des langues (TAL) et en fouille de données textuelles (*text mining*) lors de son audition par la mission, la fouille de textes et de données consiste à expliciter des liens, auparavant inconnus, parmi des données analysées en très grand nombre. La fouille de textes et de données ne doit donc pas être confondue avec une simple analyse de données, qui se borne par exemple à rechercher des mots dans un corpus.

Dans ce cadre, on peut remarquer que **la mention des « textes » dans l'expression « fouille de textes et de données » est sans grande portée**, car elle est couverte par l'expression de données. En effet, la notion de données numériques recouvre aussi bien le texte, la musique, l'image que la vidéo, et ce quelle que soit la qualification juridique. Aussi, les « données » visées peuvent être des œuvres de l'esprit, des contenus protégés par les droits voisins, des bases de données, ou plus largement des données brutes, non réservées par la propriété intellectuelle.

Cette extension du champ d'application de la fouille de textes et de données à toutes les données numériques est renforcée par le développement d'interfaces et de logiciels permettant de convertir plus ou moins facilement des formats au sein d'un même média (PDF vers des documents au format txt, ou format d'image JPEG vers un autre format, ou musique en format WAVE vers un format MP3), voire entre médias (des textes photographiés peuvent être convertis en textes). La facilité avec laquelle il est possible aujourd'hui de modifier le format des données rend pour l'essentiel secondaire la question de l'harmonisation des formats : selon les chercheurs auditionnés par la mission, la très forte diversité des standards actuels ne bloque pas la fouille de textes et de données.

Il convient toutefois de rappeler que le débat sur le périmètre exact de la notion de fouille de textes et de données n'est pas absolument clos. Comme l'a rappelé Pierre Zweigenbaum, directeur de recherche au LIMSI-CNRS (Université Paris-Saclay), le champ de l'extraction d'informations est plus large que celui de la fouille de textes et de données. Dans le premier cas, le chercheur sait ce qu'il cherche (l'objet de la recherche est déterminé), tandis que dans le second, il n'a pas encore définitivement arrêté l'objet de sa recherche. Dans cette dernière hypothèse, le chercheur peut tenter, dans une phase exploratoire, d'opérer des regroupements non supervisés entre données. La fouille de textes et de données ainsi définie apparaît comme beaucoup moins fréquente et plus complexe que l'extraction d'informations et comporte une visée heuristique : elle permet de faire émerger des questionnements ou hypothèses de recherche. Cette distinction entre extraction de données et fouilles de textes et de données est corroborée par un article de 2003 de Marti Hearst, actuellement professeur à la *School of Information* de l'Université de Californie à Berkeley¹⁴. Selon elle, dans une recherche classique sur internet, le chercheur est en quête d'un élément connu qui a déjà été exposé par quelqu'un d'autre, alors que dans la fouille de données, le but est de découvrir des informations jusqu'alors inconnues. Ainsi, en toute rigueur des termes, la plupart des chercheurs sur le langage font de l'extraction d'informations et non de la fouille de textes et de données. **Toutefois, du point de vue des opérateurs dans les entreprises, de la législation et de l'opinion publique, les chercheurs en extraction d'informations reconnaissent que leur activité est assimilable à de la fouille de textes et de données.** La distinction entre ces termes est devenue aujourd'hui

¹⁴ <https://people.ischool.berkeley.edu/~hearst/text-mining>

de pure sémantique. Ainsi, il apparaît clairement que l'expression « fouille de textes et de données » est préférée compte tenu de son acception large, y compris pour viser l'extraction de données, hypothèse que le législateur européen avait d'ailleurs sans doute en tête lors de l'élaboration de la directive 2019/790.

2) Une technique indispensable dans un environnement numérique en pleine expansion et favorable au développement de l'intelligence artificielle

L'explosion du nombre de données numériques ces dernières décennies a rendu nécessaire l'amélioration des techniques quantitatives pour les analyser. Il est indiqué dans un article de doctrine juridique relatif à la fouille et à son impact sur le développement de l'IA¹⁵ que :

- A l'échelle mondiale, les données numériques, qu'il s'agisse de données créées ou copiées, doublent tous les deux ans ;
- 90% des données numériques actuelles ont été créées ces deux dernières années ;
- L'univers digital devait compter en 2020 environ 44 zettabytes, soit 44 milliards de gigabytes.

Plus récemment, la Commission européenne a indiqué, dans son Livre blanc du 19 février 2020 sur l'intelligence artificielle, que « le volume des données produites dans le monde augmente rapidement et devrait passer de 33 zettaoctets en 2018 à 175 zettaoctets en 2025 »¹⁶.

Comme le remarquait la direction de l'information scientifique et technique du CNRS dans son livre blanc « Une Science ouverte dans une République numérique. Guide stratégique » d'octobre 2016 : « Chaque champ scientifique (astrophysique, biologie moléculaire, sociologie des organisations, géologie marine, linguistique...) a développé son propre arsenal de moyens et techniques de collecte de données : capteurs, sondes, satellites, séquenceurs, cameras, numérisation, simulation, analyses chimiques... qui conduisent à l'accumulation de vastes ensembles de données (big data) »¹⁷. De fait, il est fréquent qu'une extraction d'informations sur des textes concerne désormais des centaines de milliards de mots, contre un million de mots il y a 20 ans par exemple.

En outre, la recherche en intelligence artificielle¹⁸ a accompli des progrès considérables ces dernières années grâce notamment au *machine learning* (apprentissage-machine) et même au *deep learning* (apprentissage profond), qui fonctionne sur le modèle des réseaux de neurones. Depuis les années 2010, c'est l'approche dite « connexionniste », correspondant à un raisonnement inductif fondé sur l'expérience, qui l'a emporté (peut-être temporairement, serait-on tenté d'ajouter) sur l'approche dite « symbolique », qui privilégie le raisonnement

¹⁵ E. Rosati, « Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity », Asia Pacific Law Review, sept. 2019 : <https://ssrn.com/abstract=3452376>.

¹⁶ Commission européenne, Livre blanc « Intelligence artificielle : une approche européenne axée sur l'excellence et la confiance », COM(2020) 65 final, p. 6.

¹⁷ <https://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/07/Livre-Blanc-loi-numerique-October-2016.pdf>, p. 38-39.

¹⁸ Comme le souligne la Commission européenne dans son livre blanc du 19 février 2020 précité, l'intelligence artificielle « associe des technologies qui combinent données, algorithmes et puissance de calcul ».

déductif¹⁹. En effet, c'est l'ingestion de quantités faramineuses de données numériques qui permet à l'intelligence artificielle de mimer en partie l'intelligence humaine.

Pour illustration, les performances techniques du robot GPT-3 de l'entreprise américaine OpenAI ont récemment été présentées comme la démonstration des progrès de la recherche sur les réseaux de neurones profonds en intelligence artificielle. Ce robot, « *programmé sur l'un des dix plus puissants ordinateurs du monde* », « *écrit de la poésie, traduit, calcule, code des programmes informatiques, entretient des conversations en ligne, corrige des textes, rédige des tribunes* » après avoir été alimenté par plus de « *500 milliards de mots, l'équivalent de plus de 150 fois toute l'encyclopédie Wikipedia (dans toutes les langues), ou de plus de 2000 ans de lecture du quotidien Le Monde* »²⁰.

Mais l'IA ne recouvre pas seulement les techniques d'apprentissage. Aussi, pour certains chercheurs, le traitement automatique des langues est déjà une forme d'intelligence artificielle. Dès lors, certains considèrent que l'extraction de données (ou la fouille de textes et de données) est une technique d'intelligence artificielle.

Schématiquement, la fouille de textes et de données comprend **trois phases**, comme l'indique le schéma ci-dessous :

- La phase amont, de collecte des données, de leur nettoyage et de leur transformation ;
- La phase d'opérations de fouille à proprement parler grâce à un traitement informatique ou algorithme ;
- La phase aval, qui comprend l'interprétation, l'évaluation et la validation des résultats.

¹⁹ V. Rapport de la mission « *Intelligence artificielle et culture* », *op. cit.*, p. 7 : <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>

²⁰ D. Larousserie, « GPT-3, l'intelligence artificielle qui a appris presque toute seule à presque tout faire », *Le Monde*, 3 nov. 2020.

Source : Text Mining in Organizational Research, Vladimer B. Kobayashi, Stefan T. Mol, Hannah A. Berkers, Ga'bor Kismihok and Deanne N. Den Hartog, in Organizational Research Methods 2018, Vol. 21(3) 733-765

La première phase comprend elle-même **trois étapes** :

- La collecte de données et leur reproduction ;
- Leur « nettoyage » (qui prend dans certains cas jusqu'à 80% du temps d'une opération de fouille de textes et de données), afin notamment de supprimer les données incohérentes, non fiables ou redondantes ;
- Leur transformation dans un format spécifique, adapté à la recherche.

La mission considère que l'immense majorité des fouilles implique des copies et reproductions des œuvres collectées. Même si le considérant 9 de la directive du 17 avril 2019 indique que certaines fouilles peuvent ne pas entraîner de reproductions, il n'a pas été possible d'obtenir d'exemple concret en ce sens. D'après les recherches effectuées par la mission, ce cas de figure très théorique et donc rarissime pourrait se présenter si la personne qui effectue une fouille sait d'avance les éléments à collecter : elle pourrait alors aspirer uniquement ces éléments ciblés sans faire une copie du contenu protégé dans lequel ils s'inscrivent (par exemple, les locutions latines utilisées dans des articles de presse pendant une année donnée).

La deuxième phase est la plus rapide généralement, car elle consiste à appliquer un algorithme adapté à l'objet de la fouille de textes et de données.

La troisième phase est également chronophage et est potentiellement de durée illimitée en raison des règles de validation des travaux scientifiques (à tout moment, même trente ou quarante ans après la publication d'une avancée scientifique, ses auteurs doivent être en mesure de communiquer les données initiales, les données retraitées et les résultats de leur expérimentation, pour vérifier la pertinence de la recherche ou la prolonger en proposant une autre méthode).

Il convient à cet égard de rappeler que les projets de recherche subventionnés par des financements européens promeuvent le principe de *FAIR data* (*Findable, Accessible, Interoperable, Re-usable*) : les données de la science doivent être faciles à trouver, accessibles, interopérables et réutilisables.

Les droits d'auteur et assimilés sont évidemment susceptibles d'être méconnus lors de la première et de la deuxième phases de la fouille. Comme indiqué dans le rapport CSPLA précité sur l'intelligence artificielle et la culture : « *Les opérations de fouille supposent souvent en amont des reproductions intermédiaires des œuvres ou des bases de données concernées qu'il faut stocker, au moins temporairement, et parfois modifier (format, coupes, fusions, compilations...) pour les rendre analysables : chacune de ces opérations est alors susceptible d'être en contradiction avec les règles de réservation de la propriété intellectuelle, nécessitant un accord exprès de leurs titulaires* »²¹. De même, la conservation puis, le cas échéant, la communication de copies des données initiales à des tiers peuvent méconnaître frontalement des monopoles de propriété intellectuelle.

3) Une activité prometteuse appelée à croître rapidement

La fouille de textes et de données s'est d'abord développée dans le secteur du marketing et de la veille média.

Elle est également très présente sur les plateformes numériques de vidéos ou de musiques afin de faire des recommandations aux usagers, fondées sur les pages consultées et les tendances observées dans le public. Il en va de même pour les sites de vente en ligne.

En France, plusieurs entreprises se sont spécialisées dans la fouille de textes et de données en matière de veille économique et d'opinion sur les médias sociaux, les blogs, les forums, les sites de critique et d'information.

Mais il convient aussi de rappeler que des associations de défense des droits des consommateurs utilisent la fouille de textes et de données pour établir des corrélations dans les prix des biens vendus en supermarchés, ou encore que des associations de défense des droits des étrangers utilisent cette technique pour analyser les créneaux de rendez-vous proposés par les préfetures à ces derniers pour effectuer leurs démarches administratives.

Quant à la recherche scientifique, elle utilise de plus en plus la fouille de textes et de données dans le cadre de l'apprentissage machine. Ainsi, lors des journées nationales de la science ouverte en 2018²², plusieurs chercheurs ont présenté leurs projets utilisant la technique de la fouille de textes et de données, sur des thèmes aussi divers que :

- Apprendre à lire aux ordinateurs pour améliorer la qualité du blé et sa production ;
- Cartographier le cerveau grâce aux publications en neurosciences ;
- Améliorer l'analyse sémantique de la presse ancienne européenne.

²¹ *Op. cit.*, p. 56. V. <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>

²² <https://webcast.in2p3.fr/video/pour-quel-besoin-comment-quel-impact-enjeux-de-la-science-ouverte>

C'est aussi une technologie de fouille de textes et de données qui a permis de cartographier le génome humain en compilant automatiquement trois millions de publications dans le cadre du projet Text2genome.

Plusieurs autres exemples d'utilisation de fouille ont été directement donnés à la mission, à partir de solutions développées par des entreprises privées :

- La prédiction des risques de feux de forêt ;
- La surveillance de l'état de santé des troupeaux ;
- L'aide au diagnostic du cancer et de la cécité ;
- L'assistance des personnes aveugles ou malvoyantes pour se déplacer en leur fournissant des descriptions sonores d'objets grâce aux photographies prises à partir de leurs téléphones portables.

La fouille de textes est également très présente dans le domaine du droit : analyse de contrats, création automatisée de documents juridiques ou encore chatbots²³ juridiques.

Autre exemple d'actualité : une plateforme internet organisant des compétitions en sciences des données regroupe l'ensemble des données associées à la Covid. Initié par le Gouvernement américain et plusieurs groupes de recherche, le projet *COVID-19 Open Research Dataset* (CORD-19) regroupe plus de 57 000 articles universitaires, dont environ 45 000 en version intégrale, sur la COVID-19, le SARS-CoV-2 et les coronavirus connexes. Ces données sont mises librement à la disposition de la communauté mondiale des chercheurs, leur permettant d'appliquer les avancées récentes dans le traitement automatique du langage naturel et les autres techniques d'intelligence artificielle.

Comme le relevait la Commission européenne dans son étude d'impact du 14 septembre 2016 accompagnant la proposition de directive sur le droit d'auteur dans le marché unique numérique, il est difficile d'évaluer la place de la fouille de textes et de données en Europe. La Commission européenne estime toutefois que les Etats membres représentent 28,2% des articles scientifiques utilisant la fouille de textes et de données dans le monde, juste derrière l'Asie (32,4%) mais devant l'Amérique du Nord (20,9%). En outre, cette activité est appelée à se développer dans les années à venir car seulement 20% des chercheurs en Europe ont utilisé la fouille de textes et de données pour analyser la presse en 2016.

La majorité des personnes entendues par la mission considèrent que la fouille de textes et de données devrait croître considérablement dans les années à venir. Certaines personnes remettent toutefois en doute cette prévision en se fondant sur l'exemple du Royaume-Uni, qui a mis en place depuis 2014 une exception en faveur de la fouille de textes et de données. Un sondage effectué par la société de gestion collective britannique *Publishers Licensing Society* (PLS) a montré que les éditeurs britanniques avaient reçu à peu près le même nombre de demandes de fouille en 2015 (84) qu'en 2013 (79)²⁴.

²³ Il s'agit de logiciels programmés pour simuler une conversation en langage naturel avec un individu.

²⁴ <https://www.pls.org.uk/news-events/n-tdm-august-15/>

B. La fouille de textes et de données concerne tous les secteurs protégés par le droit de la propriété littéraire et artistique, mais à des degrés divers

1) La fouille de textes et de données porte sur toutes les catégories de données numériques

La fouille de textes et de données concerne **tous les types de données numériques, quelle que soit leur qualification juridique** :

- Données brutes ²⁵ ;
- Données publiques (en *open data*) ;
- Données protégées par le droit d'auteur ou les droits assimilés (droits voisins, droit des producteurs de bases de données) ;
- Données couvertes par un secret protégé par la loi ou par le contrat ;
- Données personnelles couvertes par le RGPD (règlement général sur la protection des données, 2016/679, 27 avril 2016) ;
- Données d'usage (données issues de la rencontre entre un utilisateur et un contenu) ;
- Métadonnées (données qui décrivent d'autres données) ²⁶.

Le rapport au CSPLA de notre collègue Valérie-Laure Benabou « *Droit de la propriété littéraire et artistique, données et contenus numériques* », d'octobre 2018, avait déjà souligné avec raison que « *la notion de données n'est pas davantage définie que celle de contenus numériques mais, en raison de la multiplicité des régimes qui s'attachent à ses acceptions variées, les règles applicables aux données sont susceptibles de perturber celles du droit de la propriété littéraire et artistique* » ²⁷.

Cette variabilité des données, de leur qualification et des régimes applicables est source de complexité juridique. En effet, quand un chercheur aspire des données numériques sur internet, il n'a pas toujours connaissance des différents régimes juridiques applicables. Certains chercheurs reconnaissent notamment que leurs données comprennent souvent des données protégées par le droit d'auteur et les droits assimilés.

D'autres chercheurs constituent au contraire leurs corpus auprès d'associations comme l'ELRA (*European Language Resources Association*), qui collecte et annote les données et signe les contrats nécessaires à leur reproduction avec les titulaires de droits. Il en va aussi de la légitimité

²⁵ Comme l'indique le considérant 9 de la directive 2019/790 : « *La fouille de textes et de données 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* ».

²⁶ Les métadonnées « *indiquent comment, quand, où et par qui les données ont été recueillies, mentionnent leur disponibilité et leur mode de distribution, le système de projection et de coordonnées qui les caractérisent, l'échelle de suivi, la résolution et la précision et la fiabilité à l'égard de certaines normes* » : <https://www.emse.fr/tice/uved/SIG/Glossaire/co/Metadonnees.html>

²⁷ Rapport « *Droit de la propriété littéraire et artistique, données et contenus numériques* », mission présidée par Valérie-Laure Benabou, avec la collaboration de Célia Zolynski, rapporteur Laurent Cytermann, CSPLA, oct. 2018, p. 5. V. <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-consequences-pour-la-propriete-litteraire-et-artistique-de-l-avenement-des-notions-de-donnees-et-de-contenus-numeriques>

de la recherche, notamment dans le milieu médical, qui impose une haute qualité et une parfaite traçabilité des données initiales.

2) La fouille de textes et de données est difficilement conciliable, dans le cadre juridique français, avec les exceptions aux règles du droit d’auteur

Sauf dans de très rares cas, la fouille de textes et de données implique une extraction et une reproduction de données. Or, comme le rapport du CSPLA sur l’intelligence artificielle et la culture de janvier 2020 l’a déjà montré, les acteurs de la fouille de textes et de données ne peuvent pas se fonder sur les exceptions actuelles au droit d’auteur prévues dans la législation française, et notamment ²⁸ :

- l’exception pour copie privée (L. 122-5, 2° du CPI ²⁹), dès lors que la fouille implique en règle générale une utilisation collective ;
- l’exception de reproduction technique provisoire présentant un caractère transitoire ou accessoire (L. 122-5, 6° du même code ³⁰), puisque la reproduction ne peut souvent être considérée comme telle ; en outre, cette exception exclut les bases de données ;
- l’exception de conservation (L. 122-5, 8° du même code ³¹) car cette exception a seulement pour finalité directe ou indirecte la préservation des collections d’œuvres.

L’inadaptation du cadre juridique actuel oblige les scientifiques à solliciter les ayants droit pour leur demander leur accord avant d’entreprendre leurs travaux de recherche impliquant une fouille. Or la fouille de textes et de données nécessitant souvent des millions de données pour un projet donné, les chercheurs n’ont ni le temps, ni la compétence juridique pour se conformer systématiquement aux règles du code de la propriété intellectuelle. Si la plupart des chercheurs entendus par la mission ont démontré leur souci constant de respecter les règles du droit d’auteur, d’autres ont reconnu qu’en pratique, certains corpus de données n’avaient pas été intégralement vérifiés ou qu’ils étaient parfois communiqués à d’autres chercheurs pour vérifier les résultats ou tester d’autres méthodes, alors même que l’autorisation initiale des ayants droit ne prévoyait pas explicitement une telle communication.

²⁸ *Op. cit.*, point 3.2.1.1, p. 53.

²⁹ « 2° Les copies ou reproductions réalisées à partir d'une source licite et strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, à l'exception des copies des œuvres d'art destinées à être utilisées pour des fins identiques à celles pour lesquelles l'œuvre originale a été créée et des copies d'un logiciel autres que la copie de sauvegarde établie dans les conditions prévues au II de l'article L. 122-6-1 ainsi que des copies ou des reproductions d'une base de données électronique ».

³⁰ « 6° La reproduction provisoire présentant un caractère transitoire ou accessoire, lorsqu'elle est une partie intégrante et essentielle d'un procédé technique et qu'elle a pour unique objet de permettre l'utilisation licite de l'œuvre ou sa transmission entre tiers par la voie d'un réseau faisant appel à un intermédiaire ; toutefois, cette reproduction provisoire qui ne peut porter que sur des œuvres autres que les logiciels et les bases de données ne doit pas avoir de valeur économique propre ».

³¹ « 8° La reproduction d'une œuvre et sa représentation effectuées à des fins de conservation ou destinées à préserver les conditions de sa consultation à des fins de recherche ou d'études privées par des particuliers, dans les locaux de l'établissement et sur des terminaux dédiés par des bibliothèques accessibles au public, par des musées ou par des services d'archives, sous réserve que ceux-ci ne recherchent aucun avantage économique ou commercial ».

3) Les enjeux de la fouille sont divers pour les titulaires de droits

La mission a identifié **cinq enjeux** s'agissant de la fouille de textes et de données pour les titulaires de droits :

1- Eviter une perturbation du fonctionnement normal de leur site par les robots qui aspirent leurs données.

Selon une étude concernant un quotidien français économique, 50% de la bande passante de son site est actuellement accaparée par des robots. Or les éditeurs en ligne attachent une grande importance à la fluidité de la consultation de leurs sites, afin que le temps de réponse soit le plus bref possible.

Même si elle apparaît moins forte qu'en 2014 lors de la rédaction du rapport de notre collègue M. Martin au nom du CSPLA, la crainte des ayants droit de voir des robots venus du monde entier, et en particulier de pays non européens, « piller » les sites français n'a pas disparu.

2- Eviter de mobiliser des ressources humaines et financières pour répondre aux demandes des chercheurs qui souhaiteront avoir des copies de contenus protégés sur le fondement de l'article 3 de la directive.

De nombreux ayants droit, en particulier les éditeurs de presse, craignent que le nouveau droit de fouille reconnu aux chercheurs par l'article 3 de la directive de 2019 ne se traduise par une obligation de devoir répondre à toutes les sollicitations tendant à obtenir des copies, ce qui impliquerait un coût non négligeable en personnel. En effet, une demande d'un chercheur entraîne parfois plusieurs heures de travail pour un salarié afin de préciser l'objet de la demande et préparer les données à transmettre, voire le cas échéant mettre en place ou adapter une API (*application programming interface*) pour transmettre ces données (l'installation d'une API peut coûter plusieurs milliers d'euros). Ce risque est d'autant plus fort que l'exception prévue à l'article 3 de la directive vise une utilisation transfrontière.

3- Empêcher la dissémination de contenus protégés en dehors du cadre posé par les articles 3 et 4 de la directive.

Plusieurs éditeurs craignent que les copies des contenus protégés obtenus par les acteurs de la fouille de données soient tôt ou tard utilisés à d'autres fins, voire diffusés sur internet, rendant inefficace la protection liée à la propriété intellectuelle.

4- Profiter de la valeur générée par l'activité de fouille de textes et de données commerciale.

Les données des journaux constituent une matière brute qui doit être valorisée. Les acteurs de la fouille de texte et de données ont besoin d'alimenter leurs algorithmes avec des données produites par des titulaires de droits. Il peut donc paraître logique que ces derniers revendiquent une part de la plus-value générée par la fouille de textes et de données leur appartenant.

Cet enjeu est particulièrement sensible pour les éditeurs et agences de presse en ligne, qui craignent que les nouveaux droits voisins qui leur ont été reconnus par l'article 15 de la directive 2019/790 restent lettre morte si les fouilles de textes et de données deviennent massives dans les années à venir.

Il ressort des auditions de la mission que la France possède un avantage comparatif certain en termes de propriété de données culturelles très convoitées, alors que les acteurs de la fouille qui visent à valoriser ces données ne sont pas européens, mais américains ou chinois. Eu égard à l'importance que revêt le secteur de la culture dans l'attractivité de notre territoire, plusieurs personnes auditionnées soulignent la nécessité de protéger les titulaires de droits contre une approche qui consisterait à encourager sans limite la fouille de textes et de données.

5- Conserver un droit de regard sur l'utilisation des résultats de la fouille de textes et de données.

Il peut paraître légitime que les ayants droit souhaitent être informés des finalités d'une fouille qui concernent leurs données dans certaines hypothèses. Par exemple, si des chercheurs en sciences humaines souhaitent enquêter sur le traitement réservé aux femmes en politique dans un journal depuis 1945, son propriétaire peut souhaiter avoir connaissance de la méthodologie retenue et des résultats obtenus, compte tenu des conséquences que cette étude pourrait entraîner pour l'image du journal.

La mission n'a pas eu connaissance du nombre de travaux de fouille de textes et de données par secteur d'activité protégé par la propriété intellectuelle.

Mais la mission a le sentiment que :

- le secteur de l'écrit est sans doute le plus concerné aujourd'hui par l'essor de la fouille de textes et de données ;
- les éditeurs de presse généraliste sont les plus touchés, du fait du modèle économique déployé (un grand nombre de leurs articles sont accessibles sur internet et sont massivement « moissonnés » par des robots), puis les éditeurs scientifiques ou juridiques (leurs articles sont souvent accessibles sur abonnement) et les éditeurs spécialisés ;
- les éditeurs littéraires, les titulaires de droits dans le secteur de la musique, de l'image et de l'audiovisuel apparaissent en comparaison moins touchés, et manifestent en conséquence moins d'inquiétude.

C. L'article 38 de la loi pour une République numérique : une tentative inaboutie pour promouvoir la fouille de textes et de données

1) Un dispositif réservé à la recherche publique et centré sur les écrits scientifiques

Dans son avis du novembre 2017, le Conseil national du numérique avait plaidé en faveur d'une disposition législative sur la fouille de textes et de données effectuée par des chercheurs. Il avait en effet rappelé 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 »³².

Le projet de loi pour une République numérique ne comportait pas initialement de disposition relative à la fouille de textes et de données. En effet, « *si la création d'une exception pour fouille de textes et de données figurait bien dans le projet de loi soumis à la consultation publique* », elle avait finalement disparu de la version adoptée en Conseil des ministres, compte tenu « *des réactions excessivement polémiques des parties concernées, en l'espèce des éditeurs et des organismes de recherche* », les premiers craignant un affaiblissement du droit d'auteur³³.

C'est par un amendement adopté en séance publique à l'Assemblée nationale, contre l'avis du Gouvernement et du rapporteur, que ces dispositions ont été introduites dans le texte.

Malgré l'opposition du Sénat à ce nouvel article 18 bis (la Commission de la culture, de l'éducation et de la communication avait adopté un amendement de réécriture de cet article afin de régir les fouilles non pas par une exception au droit d'auteur, mais par des contrats conclus entre les éditeurs et les organismes de recherche ou les bibliothèques³⁴, compte tenu de précédents encourageants³⁵), cet article a été maintenu dans le projet de loi.

L'article 38 de la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique a ainsi créé une exception au droit d'auteur (art. L. 122-5, 10° du code de la propriété intellectuelle³⁶) et au droit des producteurs de bases de données (art. L. 342-3, 5° du même code³⁷) en faveur de la fouille de textes et de données, à la double condition, d'une part, qu'elle soit exclusivement effectuée par des chercheurs qui ne poursuivent aucune finalité commerciale et, d'autre part, qu'elle porte sur des « *écrits scientifiques* ». Si ces deux conditions sont remplies, les

³² Avis n° 2015-3 relatif au projet de loi pour une République numérique, Conseil national du numérique, 30 nov. 2015, p. 15-16 :

<https://cnnumerique.fr/files/2017-10/Avis-du-CNNum-sur-le-projet-de-loi-numerique.pdf>

³³ Avis de la commission de la culture, de l'éducation et de la communication du Sénat, sur le projet de loi pour une République numérique, au rapport de Mme Colette Mélot, 5 avr. 2016, p. 45 :

<http://www.senat.fr/rap/a15-525/a15-5251.pdf>

³⁴ « *Ne jugeant pas opportun de créer une exception ad hoc au droit d'auteur qui serait contraire au droit européen, la commission de la culture a, en effet, préféré rendre obligatoire, dans les contrats conclus entre éditeurs et organismes de recherche ou bibliothèques, l'autorisation d'accès aux données et aux textes du corpus de publications scientifiques appartenant à l'éditeur, à des fins de fouille électronique, exclusivement pour la recherche publique et à l'exclusion de tout usage commercial* ». Voir le rapport de la commission des lois du Sénat sur le projet de loi pour une République numérique, au rapport de M. Frassa, p. 236 :

<http://www.senat.fr/rap/l15-534-1/l15-534-11.pdf>

³⁵ En effet, l'association STM, regroupant aujourd'hui 140 éditeurs scientifiques dans le monde, et le groupe PDR, regroupant des infrastructures de recherche du secteur pharmaceutique, ont publié une licence type en 2012 qui contient une clause pour les fouilles de textes et de données. Les éditeurs STM ont ensuite publié en novembre 2013 une déclaration engageant ses signataires à mettre en œuvre cette clause ou, à défaut, de permettre la fouille à des fins non commerciales sur des contenus sous abonnement sans frais supplémentaires.

³⁶ « *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* ».

³⁷ « *5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données incluses ou associées aux écrits scientifiques dans un cadre de recherche, à l'exclusion de toute finalité commerciale. La conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche pour lesquelles elles ont été produites, sont assurées par des organismes désignés par décret. Les autres copies ou reproductions sont détruites.* »

chercheurs peuvent réaliser des « *copies ou reproductions numériques* » à partir « *d'une source licite* », sans avoir à obtenir l'accord des titulaires de droits.

Toutefois, les règles de ces deux exceptions (au droit d'auteur et au droit *sui generis*) diffèrent sur trois points essentiels.

Tout d'abord, dans le cadre de l'exception au droit d'auteur, les bénéficiaires sont exclusivement les chercheurs publics, alors que l'exception aux droits des producteurs de bases de données bénéficie plus largement aux chercheurs, sans autre précision. Cependant, il s'agit vraisemblablement d'un oubli du législateur dès lors que les deux dispositifs excluent expressément la fouille de textes et de données à finalité commerciale.

Ensuite, un décret est prévu pour désigner les organismes chargés de conserver et communiquer les copies techniques issues des traitements quand est appliquée l'exception aux droits des producteurs de base de données, alors que le décret mentionné à l'article L. 122-5 du CPI ne prévoit pas expressément la désignation de tels organismes. Là encore, on peut imaginer qu'il s'agit d'un problème de cohérence.

Enfin, les fichiers produits au terme de la fouille de textes et de données effectuée dans le cadre de l'article L. 122-5 du CPI constituent des « *données de la recherche* », alors que cette qualification n'est pas reconnue au résultat de la fouille réalisée dans le cadre de l'article L. 342-3 du même code.

2) L'absence de décret d'application

Les articles L. 122-5 et L. 342-3 du CPI renvoient au pouvoir réglementaire le soin de préciser leurs conditions d'application, s'agissant notamment des modalités de conservation et de communication des fichiers produits à l'issue de la fouille de textes et de données.

Un groupe de travail a été mis en place, réunissant entre autres des chercheurs et des éditeurs. Ses travaux ont guidé le Gouvernement dans la rédaction du décret.

Compte tenu du dernier alinéa de l'article L. 122-5 du CPI qui renvoie à un décret en Conseil d'Etat le soin de définir les conditions d'application de toutes les exceptions prévues à cet article, le Gouvernement a soumis au Conseil d'Etat un projet de décret pour l'application de l'article 38 de la loi du 7 octobre 2016 pour une République numérique. Si la mission n'a pas eu connaissance de l'avis du Conseil d'Etat sur ce texte, compte tenu de l'absence de publicité des avis des sections administratives sur les projets de textes réglementaires, il ressort des auditions que la section administrative chargée d'examiner ce texte a émis un avis défavorable en raison, semble-t-il, de **l'incompatibilité entre les dispositions réglementaires en cause et le droit européen**, qui ne prévoyait pas à cette date de telles exceptions au droit d'auteur et aux droits des producteurs de bases de données³⁸.

³⁸ Notamment, l'article 5 de la directive 2001/29, dite « DADVSI », qui constitue le socle des exceptions possibles en droit de l'Union et reprend l'ensemble des traditions nationales, ne contient aucune disposition sur la fouille de textes et de données en matière de droit d'auteur.

D. Éléments de droit comparé

Aux **États-Unis**, la fouille de textes et de données semble principalement autorisée en vertu de la doctrine du « *fair use* » (ou usage loyal). La Constitution américaine autorise le Congrès à créer un cadre de protection de la propriété intellectuelle afin de « *favoriser le progrès de la science et des arts utiles, en assurant, pour un temps limité, aux auteurs et inventeurs le droit exclusif à leurs écrits et découvertes respectifs* » (article 1^{er}, section 8).

La doctrine du « *fair use* » ne porte pas spécifiquement sur la fouille de textes et de données car sa portée est plus large. Codifiées à la section 107 du *Copyright Act*³⁹, les règles du « *fair use* » confient au juge le soin de trouver un juste équilibre entre l'intérêt des auteurs, d'une part, et l'intérêt de la société pour la libre circulation des idées, de l'information et du commerce, d'autre part. Cet article dispose en effet que :

« Nonobstant les dispositions des articles 106⁴⁰ et 106 A⁴¹, l'utilisation loyale d'une œuvre protégée par le droit d'auteur, y compris son utilisation par reproduction dans des copies ou des phonogrammes ou par tout autre moyen spécifié par cet article, à des fins telles que la critique, les commentaires, les reportages, l'enseignement (y compris par l'intermédiaires de copies multiples pour une utilisation en classe), une bourse ou une recherche ne constitue pas une violation du droit d'auteur. Pour déterminer si l'utilisation faite d'une œuvre dans un cas particulier est une utilisation loyale, les facteurs à prendre en compte doivent inclure :

(1) le but et le caractère de l'utilisation, y compris si cette utilisation est de nature commerciale ou est à but non lucratif ;

(2) la nature de l'œuvre protégée par le droit d'auteur ;

(3) la quantité et l'importance de la partie utilisée par rapport à l'œuvre protégée dans son ensemble ;

et (4) l'effet de l'utilisation sur le marché potentiel ou la valeur de l'œuvre protégée par le droit d'auteur.

Le fait qu'une œuvre n'est pas publiée n'empêche pas par lui-même le constat d'une utilisation loyale si ce constat tient compte de tout ce qui précède ».

Il résulte de l'analyse de la jurisprudence américaine que :

- ces quatre critères ne sont pas limitatifs ;

- le premier critère consiste à vérifier si l'usage de l'œuvre protégée apporte une connaissance nouvelle et s'il en est fait un usage « transformatif », de sorte que plus une œuvre est transformée, moins les autres facteurs comme l'usage commercial sont importants ;

- le second critère joue rarement un rôle significatif en cas de litige ;

- le troisième critère consiste à évaluer l'importance des extraits en les comparant à l'œuvre initiale ;

³⁹ <https://www.copyright.gov/title17/title17.pdf>

⁴⁰ Cet article décrit le monopole d'exploitation.

⁴¹ Il s'agit du droit de se revendiquer auteur d'une œuvre visuelle et de veiller à son intégrité.

- le quatrième critère est le plus important car le juge contrôle si la nouvelle œuvre est susceptible de se substituer à l'œuvre originale sur son marché ;
- le jugement final procède d'une mise en balance des intérêts des deux parties à la lumière de l'ensemble des facteurs examinés.

Une affaire est systématiquement mise en avant pour rattacher la fouille de données au « *fair use* ». Dans l'affaire *Google Books* du 16 octobre 2015 (*United States Court of Appeals, 2nd Circuit*), le juge a considéré que le fait de numériser plus de mille livres dans une bibliothèque sans l'autorisation des auteurs n'était pas contraire aux règles du « *fair use* ». La « *création* » par Google de courts entrefilets contribue au caractère « *hautement transformatif* » de la fouille, permettant à l'utilisateur de déterminer si l'ouvrage recherché correspond bien à ses recherches.

Interrogée par la mission, Mme Jane Ginsburg, professeure de droit de la propriété littéraire et artistique à la *Columbia Law School*, a rappelé que si la décision *Google Books* a eu un profond retentissement international, il serait hâtif d'en conclure qu'elle fixe le cadre définitif des activités de fouille de textes et de données⁴². Selon elle, cette décision porte sur des faits remontant à 2005, sur des livres physiques qu'il s'agissait de scanner. La décision serait peut-être différente aujourd'hui avec des livres déjà scannés, ou pour des œuvres visuelles ou musicales. Ainsi, la professeure Ginsburg considère que la décision est **très liée aux faits de l'espèce** et qu'elle doit être considérée comme « *sui generis* ».

La fouille de textes et de données n'a pas donné lieu à d'autres décisions majeures sur le sujet, même si le « *fair use* » est évidemment mobilisé régulièrement. Certaines décisions pourraient cependant conforter l'analyse qui fait douter de son applicabilité à la fouille.

Ainsi, dans une décision *Fox News Network, LLC v. TVEyes, Inc* du 7 mars 2017 (*United States Court of Appeals, 2nd circuit*), le juge a considéré que *TVEyes*, qui enregistre en continu les programmes de toutes les chaînes de télévision américaine, les indexe et propose à ses clients de visionner des extraits limités à dix minutes correspondant à leurs recherches, a méconnu les règles du « *fair use* ». Le troisième critère défini par la section 107 du *Copyright Act* penchait en faveur de la chaîne Fox parce que la fonction « *visionnage* » proposée par *TVEyes* permettait à ses clients de voir et entendre virtuellement tous les programmes de la Fox qu'ils souhaitent. Quant au quatrième critère, il a été jugé que *TVEyes* a usurpé un droit pour lequel la chaîne Fox était fondée à demander une compensation à travers un accord de licence.

La professeure Ginsburg a en outre indiqué à la mission que les grands projets de fouille aux Etats-Unis se règlent en général par un contrat⁴³ et que même si l'affaire *Google Books* ne tranche pas vraiment la question de la fouille de textes et de données, les acteurs économiques s'en contentent et les opérations ne suscitent pas de contentieux.

⁴² J. C. Ginsburg, « *Secure Failure Fair Use Analysis* », janv. 2016 :

<https://www.mediainstitute.org/2016/01/25/security-failure-fair-use-analysis/>

⁴³ On peut noter en ce sens que le *Copyright Clearance Center* (CCC) offre un service destiné aux entreprises des sciences de la vie : *RightFind XML for Mining* donne accès à environ 10 millions d'articles en contenu XML provenant de plus de 60 éditeurs STM avec des métadonnées normalisées et des conditions de licence cohérentes pour l'extraction du contenu à des fins de recherche interne. Le système réduit la nécessité de négociations de licence ponctuelles, ainsi que les coûts administratifs associés, tout en fournissant des revenus supplémentaires aux titulaires de droits lorsque leur contenu est utilisé pour l'exploration de textes. V. <http://www.copyright.com/business/xmlformining/> et <https://www.rightsdirect.com/solution/rightfind/>

Pour toutes ces raisons, la mission relativise fortement l'allégation, fortement répandue, selon laquelle la législation américaine serait très favorable aux activités de fouille de textes et de données.

D'autres législations s'appuient sur une disposition propre pour autoriser largement la fouille. C'est le cas notamment du **Japon**, qui dispose d'une exception permettant toute exploitation, avec ou sans but commercial, dès lors que l'« *analyse de l'information (c'est-à-dire l'extraction, la comparaison, la classification, ou autre analyse du langage, d'un son, d'une image, ou de tout autre élément dont un grand nombre d'œuvres ou une grande quantité de données sont composées)* » n'a pas pour objectif de « **jouir** » de l'œuvre⁴⁴.

Cette voie est très différente de celle empruntée par le législateur européen et on ne saurait s'en inspirer, comme certains le proposent⁴⁵.

II. La directive du 17 avril 2019 prévoit deux dispositifs obligatoires en matière de fouille dérogeant au monopole des titulaires de droits

Prenant acte de certains principes (A), la directive 2019/790 du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique impose deux dispositions distinctes : l'une, attendue, à l'article 3, au bénéfice de certains organismes à des fins de recherche scientifique (B), l'autre, à l'article 4, atypique et qui pose une dérogation très ouverte mais contrebalancée par une option de retrait au bénéfice des ayants droit (C). Pour autant, des règles communes régissent ces deux dispositifs (D).

A. Les principes suivis par le législateur européen

1) Une approche globale et volontariste sur la fouille de textes et de données

Selon l'association France Digitale, qui représente plus de 1 800 startups et 150 investisseurs français du numérique, « *la concurrence et la différenciation se font principalement sur l'accès aux données plus que sur les compétences algorithmiques* » et les entreprises européennes souffrent d'un « *grand déficit d'accès aux données* ».

Le législateur européen semble partager ce constat. Aussi, le débat sur la fouille de textes et de données s'est inscrit dans la politique européenne en faveur de l'intelligence artificielle, promesse de compétitivité.

Selon un rapport d'un groupe d'experts sollicité par la Commission européenne en 2014, la fouille de textes et de données représente une « *opportunité économique importante pour*

⁴⁴ « Jouir de l'œuvre » signifie « satisfaire le désir de savoir ou s'amuser intellectuellement en écoutant ou en voyant ces œuvres » : v. M. Nagatsuka, « Lettre du Japon. La réforme de la loi sur le droit d'auteur en 2018 et à venir », Propr. intell. oct. 2019, n° 73, p. 121.

⁴⁵ Pour une vision complète des systèmes hors UE, v. en annexes la contribution de l'Hadopi.

l'Europe. L'utilisation prolifique de la FTD ajouterait des dizaines de milliards d'euros au PIB global de l'UE, à travers essentiellement une plus forte productivité des chercheurs et des effets (externalités) liés à l'augmentation du niveau des recherches »⁴⁶.

Après avoir présenté en avril 2018 la stratégie européenne pour l'intelligence artificielle, la Commission européenne a rendu public le 19 février 2020 un livre blanc intitulé « *Intelligence artificielle : une approche européenne axée sur l'excellence et la confiance* »⁴⁷.

Au cours des trois dernières années, le financement de l'UE en faveur de la recherche et de l'innovation dans le domaine de l'intelligence artificielle a atteint selon la Commission européenne 1,5 milliard d'euros, soit une augmentation de 70% par rapport à la période précédente.

Dans le même sens, le considérant 8 de la directive de 2019 indique que la fouille de textes et de données, qui rend possible le « *traitement de grandes quantités d'informations en vue d'acquérir de nouvelles connaissances et de découvrir de nouvelles tendances* », peut être « *en particulier profitable à la communauté des chercheurs, et ainsi soutenir l'innovation* ». Le considérant 9 de la même directive rappelle encore que « *les exceptions et limitations au droit d'auteur ne sont pas entièrement adaptées à l'utilisation de certaines technologies dans le domaine de la recherche scientifique* », et que « *la compétitivité de l'Union en tant qu'espace de recherche risque d'en pâtir, à moins que des mesures ne soient prises pour remédier à l'insécurité juridique concernant la fouille de textes et de données* ».

Ces mesures s'expriment, en matière de fouille, par la consécration de deux dispositifs différents : l'un en faveur de la recherche scientifique et au bénéfice de certaines institutions et l'autre ouverte à tous les usages mais à la discrétion des titulaires de droits qui ont la possibilité de s'en extraire. Le volontarisme européen a trouvé pour traduction le caractère obligatoire de ces dispositifs, qui doivent donc être transposés dans tous les Etats membres.

2) La nécessité de concilier les intérêts des ayants droit et des acteurs de la fouille de textes et de données

Malgré l'intention initiale de promouvoir la recherche et de renforcer l'innovation et la compétitivité de l'Union européenne, le législateur européen n'a pas entendu faire primer les intérêts des acteurs de la fouille de textes et de données sur ceux des titulaires de droits. En effet, le considérant 6 de la directive énonce que « *les exceptions et limitations prévues dans la présente directive tendent vers un juste équilibre entre les droits et les intérêts des auteurs et autres titulaires de droits, d'une part, et des utilisateurs, d'autre part* ».

Ainsi, il n'est pas question de sacrifier les uns au profit des autres. En ce sens, on verra que l'exception bénéficie très largement aux chercheurs mais qu'elle prévoit en contrepartie des règles sur la conservation des copies et autorise le titulaire de droits à protéger l'intégrité du réseau qui héberge ses œuvres. De même, si la dérogation de l'article 4 est ouverte à tous les

⁴⁶ Commission européenne, « *Standardisation dans le domaine de l'innovation et du développement technologique, notamment dans le domaine de l'exploration de texte et de données* », rapport du groupe d'experts, 2014, p. 6-7 :

http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf

⁴⁷ https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_fr.pdf

usages, y compris commerciaux, elle trouve un point d'équilibre dans l'option de retrait qui peut être exercée par les titulaires de droits.

En outre, on le sait, la recherche d'un haut niveau de protection est un objectif en propriété intellectuelle qui découle directement de l'article 167, paragraphe 4, du Traité sur le fonctionnement de l'Union européenne⁴⁸, imposant à l'Union de tenir compte des aspects culturels dans son action. Autrement dit, si la directive poursuit comme objectif premier la constitution d'un marché unique numérique sans distorsion de concurrence, cet objectif n'est pas le seul assigné au législateur européen.

La mission considère de manière générale que **la volonté du législateur européen de poser un cadre juridique équilibré entre la protection des droits des ayants droit et des intérêts des acteurs de la fouille de textes et de données doit être respectée lors de la transposition des articles 3 et 4 de la directive de 2019.**

3) La reconnaissance implicite que la fouille de textes et de données peut méconnaître les droits d'auteur et assimilés

Avant l'adoption de la directive du 17 avril 2019, le débat juridique était nourri sur la question de savoir si la fouille de textes et de données était susceptible de méconnaître le droit d'auteur et les droits voisins tels que définis au niveau européen. Au nom du principe selon lequel le droit de lire implique le droit de fouiller (« *the right to read is the right to mine* »), beaucoup d'acteurs considéraient que la fouille de textes et de données était, par construction, licite dès lors que l'accès à l'œuvre était lui-même licite. Ces acteurs soutenaient que le droit d'auteur ne protège que l'originalité de l'œuvre initiale, qui se trouve totalement déconstruite dans un projet de fouille : le résultat de la fouille rend méconnaissables et indiscernables les données entrantes. Autrement dit, la « déconstruction » des données entrantes puis la « (re)construction » d'une solution grâce à un algorithme spécifique justifiaient selon eux une approche extensive, voire illimitée, du champ d'application des fouilles. En outre, certains avançaient que les contenus protégés n'étaient pas utilisés pour eux-mêmes, mais pour leur contenu informationnel. Les ayants droit quant à eux défendaient la position inverse, considérant qu'en pratique, l'immense majorité des fouilles impliquait des reproductions d'œuvres protégées qui rendaient obligatoire leur accord préalable.

Un consensus est toutefois apparu sur l'insécurité juridique du cadre juridique européen pour les acteurs de la fouille de textes et de données. En effet, comme nous l'avons déjà montré s'agissant de la réglementation française (I, B, 2), toutes les fouilles de textes et de données n'entrent pas dans le champ (cette fois européen) :

- de l'exception de reproduction provisoire (article 5, paragraphe 1, de la directive 2001/29/CE⁴⁹) ;

⁴⁸ « L'Union tient compte des aspects culturels dans son action au titre d'autres dispositions des traités, afin notamment de respecter et de promouvoir la diversité de ses cultures ».

⁴⁹ « Les actes de reproduction provisoires visés à l'article 2, qui sont transitoires ou accessoires et constituent une partie intégrante et essentielle d'un procédé technique et dont l'unique finalité est de permettre :

a) une transmission dans un réseau entre tiers par un intermédiaire, ou
b) une utilisation licite

- de l'exception de copie privée (article 5, paragraphe 2, point b, de la même directive⁵⁰) ;
- de l'exception de reproduction d'œuvres par les bibliothèques (article 5, paragraphe 2, point c, de la même directive⁵¹) ;
- de l'exception relative à l'illustration dans le cadre de l'enseignement ou de la recherche scientifique (article 5, paragraphe 3, point a de la même directive⁵²).

L'insécurité juridique était d'autant plus grande que seule la première de ces quatre exceptions s'impose dans tous les Etats membres, les trois autres exceptions demeurant à leur discrétion.

Ainsi, les acteurs de la fouille de textes et de données ne savaient pas si les données litigieuses étaient protégées par le droit d'auteur, le droit *sui generis* des bases de données, ou les deux (considérant 8 de la directive de 2019)⁵³.

Cette insécurité juridique était perçue par le législateur européen comme un frein au développement de la fouille de textes et de données, contraire au projet de soutenir l'intelligence artificielle en Europe.

Compte tenu de la rédaction des articles 3 et 4 de la directive du 17 avril 2019, le législateur européen considère que dès lors qu'une fouille de textes et de données entraîne des reproductions de données, la réglementation sur le droit d'auteur et assimilés est susceptible de s'appliquer. C'est précisément parce que la fouille de textes et de données peut méconnaître les droits d'auteur, les droits voisins et le droit *sui generis* que la directive a instauré des exceptions et limitations en vue de favoriser son développement.

Par ailleurs, certaines personnes entendues par la mission regrettent que la fouille de textes et de données ne soit abordée que d'une manière parcellaire et par l'entremise du droit d'auteur et assimilés, alors qu'elle concerne en pratique, comme on l'a vu plus haut, toutes les catégories de données et nécessiterait une réglementation transversale.

d'une œuvre ou d'un objet protégé, et qui n'ont pas de signification économique indépendante, sont exemptés du droit de reproduction prévu à l'article 2. »

⁵⁰ « 2. Les États membres ont la faculté de prévoir des exceptions ou limitations au droit de reproduction prévu à l'article 2 dans les cas suivants : (...) »

b) lorsqu'il s'agit de reproductions effectuées sur tout support par une personne physique pour un usage privé et à des fins non directement ou indirectement commerciales, à condition que les titulaires de droits reçoivent une compensation équitable qui prend en compte l'application ou la non application des mesures techniques visées à l'article 6 aux œuvres ou objets concernés ».

⁵¹ « lorsqu'il s'agit d'actes de reproduction spécifiques effectués par des bibliothèques accessibles au public, des établissements d'enseignement ou des musées ou par des archives, qui ne recherchent aucun avantage commercial ou économique direct ou indirect »;

⁵² « Les États membres ont la faculté de prévoir des exceptions ou limitations aux droits prévus aux articles 2 et 3 dans les cas suivants :

a) 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 ne s'avère impossible, la source, y compris le nom de l'auteur, dans la mesure justifiée par le but non commercial poursuivi ».

⁵³ « Dans certains cas, la fouille de textes et de données peut entraîner des actes protégés par le droit d'auteur, par le droit *sui generis* sur la base de données, ou par les deux, notamment en ce qui concerne la reproduction d'œuvres ou autres objets protégés, l'extraction de contenus d'une base de données, ou les deux, ce qui est par exemple le cas lorsque les données sont normalisées lors du processus de fouille de textes et de données ».

4) Une directive qui respecte l'acquis européen

De manière générale, le paragraphe 2 de l'article 1^{er} de la directive 2019/790 rappelle que sauf dans les cas mentionnés à son article 24 (cet article porte principalement sur les exceptions à des fins d'illustration de l'enseignement ou de recherche scientifique et en faveur des bibliothèques accessibles au public), les directives précédentes en matière de droit d'auteur et assimilés continuent de s'appliquer sans modification : « *Sauf dans les cas mentionnés à l'article 24, la présente directive laisse intactes et n'affecte en aucune façon les règles existantes fixées dans les directives actuellement en vigueur dans ce domaine, en particulier les directives 96/9/CE, 2000/31/CE, 2001/29/CE, 2006/115/CE, 2009/24/CE, 2012/28/UE et 2014/26/UE* ».

5) Mais qui s'articule parfois difficilement avec les exceptions actuelles

La directive de 2019 propose une articulation entre les articles 3 et 4 et le droit européen en vigueur.

De fait, le considérant 5 de la directive 2019/790 indique que les exceptions et limitations issues du droit en vigueur dans l'Union européenne doivent continuer à s'appliquer tant qu'elles ne remettent pas en cause les nouvelles exceptions et limitations prévues par cette nouvelle directive : « *Les exceptions et limitations en vigueur dans le droit de l'Union devraient continuer à s'appliquer, notamment à la fouille de textes et de données (...), tant qu'elles ne limitent pas le champ d'application des exceptions ou limitations obligatoires prévues dans la présente directive, que les États membres sont tenus de mettre en œuvre dans leur droit national* ».

Le considérant 9 de la directive rappelle que ses articles 3 et 4 n'ont pas lieu de s'appliquer quand la fouille de textes et de données n'entraîne pas de reproduction ou peut s'appuyer sur l'exception de reproduction provisoire. Il peut exister « *des situations dans lesquelles la fouille de textes et de données n'entraîne pas d'actes de reproduction ou dans lesquelles les reproductions effectuées relèvent de l'exception obligatoire concernant les actes de reproduction provisoires prévue à l'article 5, paragraphe 1, de la directive 2001/29/CE, qui devrait continuer à s'appliquer aux techniques de fouille de textes et de données qui ne conduisent pas à la réalisation de copies qui dépassent le champ d'application de cette exception* ». La mission a déjà montré que les fouilles sans reproductions d'œuvres protégées étaient sans doute rarissimes, pour ne pas dire impossibles.

La mission souhaite ici démontrer que les fouilles peuvent difficilement se fonder sur l'exception obligatoire concernant les actes de reproduction provisoires prévue à l'article 5, paragraphe 1, de la directive 2001/29/CE, compte tenu de la jurisprudence de la CJUE qui, bien que tâtonnante, est exigeante sur le caractère provisoire des reproductions. Aussi est-il probable que le législateur européen ait voulu ici sécuriser une situation incertaine du fait d'une jurisprudence qui manque de clarté, soulignant aussi la nécessité de ne pas obérer l'avenir en conservant ouvertes les deux voies.

L'article 5 paragraphe 1 de la directive 2001/29 exempte « *Les actes de reproduction provisoires visés à l'article 2, qui sont transitoires ou accessoires et constituent une partie*

intégrante et essentielle d'un procédé technique et dont l'unique finalité est de permettre : a) une transmission dans un réseau entre tiers par un intermédiaire, ou b) une utilisation licite d'une œuvre ou d'un objet protégé » dès lors que ces actes « *n'ont pas de signification économique indépendante* ». Cette exception, adoptée sous la pression des fournisseurs d'accès à internet, trouve son pendant à l'article 13 de la directive 2000/31 du 8 juin 2000 sur le commerce électronique. Elle se justifie par la définition extensive retenue de la reproduction. L'idée est d'admettre que « *l'autorisation de "circuler" sur les réseaux numériques emporte l'autorisation de reproduire aux fins de la circulation* »⁵⁴.

S'il est clair que les conditions posées par l'article 5 paragraphe 1 de la directive de 2001 sont cumulatives⁵⁵, **la définition des critères de l'exception reste douteuse**. S'agissant du **caractère transitoire** de la reproduction, la CJUE a précisé qu'un acte ne peut être qualifié comme tel « *que si sa durée de vie est limitée à ce qui est nécessaire pour le bon fonctionnement du procédé technique concerné, étant entendu que ce procédé doit être automatisé de sorte qu'il supprime cet acte d'une manière automatique, sans intervention humaine, dès que sa fonction visant à permettre la réalisation d'un tel procédé est achevée* »⁵⁶. Concernant le **caractère accessoire** de la reproduction, alternatif au caractère transitoire, la CJUE y voit un acte qui « *n'a ni existence ni finalité autonomes par rapport au procédé technique dont il fait partie* »⁵⁷, ce qui correspond en l'espèce aux copies « en cache » faites sur le disque dur d'un internaute qui consulte un site, dont la durée de vie peut être plus longue que celle d'une reproduction « provisoire ». L'exception est encore conditionnée par l'existence d'un **troisième critère**, à savoir au fait que la reproduction soit une « *partie intégrante et essentielle d'un procédé technique* »⁵⁸. En outre, de manière transversale, la **finalité de la reproduction est limitée** : les actes doivent avoir « *pour unique finalité de permettre a) une transmission dans un réseau entre tiers par un intermédiaire, ou b) une utilisation licite d'une œuvre ou d'un objet protégé* ». Si la première hypothèse visée semble claire, la notion d'utilisation licite, à rapprocher de celle de source licite, ou encore d'utilisateur légitime, pose plus de difficultés. Il semblerait ici que la directive de 2001 ait voulu renvoyer aux stipulations contractuelles pour délimiter le champ de l'exception. Enfin, les actes de reproduction exemptés ne doivent pas avoir de « *signification économique indépendante* ». Là encore, cette notion n'a pas révélé tous ses secrets et la casuistique semble être la voie retenue par la CJUE qui la définit au gré de ses décisions⁵⁹.

D'évidence, le flou entourant l'application de l'article 5 paragraphe 1 de la directive de 2001 rend incertaine son application aux fouilles de textes et de données. La mission n'a pas réussi à identifier, malgré les demandes, des hypothèses de fouilles qui rempliraient toutes

⁵⁴ A. Lucas A. Lucas-Schloetter et C. Bernault, *Traité de la propriété littéraire et artistique*, LexisNexis, 5^e éd., 2017, n° 384.

⁵⁵ CJUE, 16 juill. 2009, aff. C-5/08, *Infopaq I*.

⁵⁶ *Infopaq I*, préc., pt 64.

⁵⁷ CJUE, 5 juin 2014, aff. C-360/13, *Public Relations Consultants Association Ltd*, pt 43.

⁵⁸ La CJUE y voit deux conditions cumulatives : il faut, d'une part, que « *les actes de reproduction soient entièrement effectués dans le cadre de la mise en œuvre d'un procédé technique* » et, d'autre part, que « *la réalisation de ces actes de reproduction soit nécessaire, en ce sens que le procédé technique ne pourrait pas fonctionner de manière correcte et efficace sans ces actes* » (*Public Relations Consultants Association Ltd*, pt 28). Ce qui est le cas des copies « en cache ».

⁵⁹ CJUE, 4 oct. 2011, aff. C-403/08, *Football Association Premier League*, pt 175-176 (fixation dans la mémoire des décodeurs et sur les écrans de télévision des utilisateurs) ; CJUE, ord., 17 janv. 2012, aff. C-302/10, *Infopaq II* (procédé d'acquisition de données permettant d'établir des synthèses d'une sélection d'articles de presse en fonction de thèmes choisis par les clients).

les conditions prévues à cet article. Notamment, l'effacement des copies semble clairement relever de la volonté de l'acteur qui réalise la fouille, ce qui ne semble correspondre ni à la lettre de l'article 5 paragraphe 1 ni au caractère automatique de l'effacement exigé par la CJUE. Certains considèrent que le critère de distinction entre l'exception technique et la fouille devrait être le caractère provisoire : la copie de l'article 5 paragraphe 1 est par nature provisoire alors que celle réalisée dans le cadre de la fouille peut être conservée dans certaines conditions. La mission n'adhère pas à ce critère. En effet, certes, la conservation en vue de vérifications de la recherche est permise dans le cadre de l'exception à des fins de recherche. Mais l'article 4 de la directive de 2019 ne semble permettre la conservation que le temps de l'opération de fouille, ce qui rend la copie également provisoire. On relèvera cependant que la décision *Infopaq II*⁶⁰ peut semer le trouble, en semblant indiquer que l'article 5 paragraphe 1 pourrait s'appliquer aux opérations de fouille. Il faut cependant garder à l'esprit qu'elle ne statue pas sur toutes les conditions de l'article 5 paragraphe 1 et pose des problèmes d'interprétation.

La mission a ainsi le sentiment que des acteurs privés exerçant des activités de fouille, notamment en veille média, pourraient être tentés de s'appuyer indûment sur l'article 5 paragraphe 1 de la directive de 2001, moins contraignant que l'article 4 de la directive de 2019 qui prévoit un droit d'opposition à la fouille de la part du titulaire de droits. **En conséquence, si l'activité est clairement de la fouille de textes et de données, ces acteurs doivent appliquer les règles fixées par l'article 4 de la directive de 2019/790 et ne pas se prévaloir de l'exception prévue à l'article 5 paragraphe 1 de la directive 2001/29, qui relève à l'évidence d'une autre philosophie.**

B. Article 3 : la traduction de la volonté de la Commission européenne de promouvoir la fouille de textes et de données en faveur de la recherche scientifique

1) La version initiale de l'article 3 proposée par la Commission européenne

Dans sa proposition de directive du Parlement européen et du Conseil sur le droit d'auteur dans le marché unique numérique, présentée le 14 septembre 2016, la Commission européenne n'avait prévu que l'article 3, et non l'article 4 sur la fouille de textes et de données hors recherche :

« 1. Les États membres prévoient une exception aux droits visés à l'article 2 de la directive 2001/29/CE, à l'article 5, point a), et à l'article 7, paragraphe 1, de la directive 96/9/CE, et à l'article 11, paragraphe 1, de la présente directive pour les reproductions et extractions effectuées par des organismes de recherche, en vue de procéder à une fouille de textes et de données sur des œuvres ou autres objets protégés auxquels ils ont légitimement accès à des fins de recherche scientifique.

2. Toute disposition contractuelle contraire à l'exception prévue au paragraphe 1 est sans effet.

⁶⁰ *Infopaq II*, précit.

3. Les titulaires des droits sont autorisés à appliquer des mesures destinées à assurer la sécurité et l'intégrité des réseaux et 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 des droits et les organismes de recherche à définir d'un commun accord des bonnes pratiques concernant l'application des mesures visées au paragraphe 3. »

Si la structure de l'article 3 de la version finale de la directive reprend celle de la proposition présentée par la Commission européenne, elle s'en distingue sur un point essentiel : l'introduction de règles sur la conservation des copies.

2) Une exception qui déroge à de nombreuses prérogatives des titulaires de droits

Aux termes du premier paragraphe de l'article 3 de la directive 2019/790 :

« 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. »

L'exception ainsi créée porte sur les droits présentés dans le tableau suivant :

Champ de l'exception prévue à l'article 3 de la directive 2019/790

Fondement juridique	Synthèse	Contenu
Article 5, point a) de la directive 96/9/CE	Droit d'auteur des producteurs de bases de données	L'auteur d'une base de données bénéficie, en ce qui concerne l'expression de cette base pouvant faire l'objet d'une protection par le droit d'auteur, du droit exclusif de faire ou d'autoriser : a) la reproduction permanente ou provisoire, en tout ou en partie, par quelque moyen et sous quelque forme que ce soit ;
Article 7, paragraphe 1, de la directive 96/9/CE	Droit <i>sui generis</i> des producteurs de bases de données	1. Les États membres prévoient pour le fabricant d'une base de données le droit d'interdire l'extraction et/ou la réutilisation de la totalité ou d'une partie substantielle, évaluée de façon qualitative ou quantitative, du contenu de celle-ci, lorsque l'obtention, la vérification ou la présentation de ce contenu attestent un investissement substantiel du point de vue qualitatif ou quantitatif.
Article 2 de la directive 2001/29/CE	Droit d'auteur Droits voisins « classiques »	Droit de reproduction Les États membres prévoient le droit exclusif d'autoriser ou d'interdire la reproduction directe ou indirecte, provisoire ou permanente, par quelque moyen et sous quelque forme que ce soit, en tout ou en partie : a) pour les auteurs, de leurs œuvres ; b) pour les artistes interprètes ou exécutants, des fixations de leurs exécutions ; c) pour les producteurs de phonogrammes, de leurs phonogrammes ;

		<p>d) pour les producteurs des premières fixations de films, de l'original et de copies de leurs films ;</p> <p>e) pour les organismes de radiodiffusion, des fixations de leurs émissions, qu'elles soient diffusées par fil ou sans fil, y compris par câble ou par satellite.</p>
Article 15, paragraphe 1, de la directive 2019/790	Droits voisins étendus aux éditeurs de presse en ligne	<p>Protection des publications de presse en ce qui concerne les utilisations en ligne</p> <p>1. Les États membres confèrent aux éditeurs de publications de presse établis dans un État membre les droits prévus à l'article 2 et à l'article 3, paragraphe 2⁶¹, de la directive 2001/29/CE pour l'utilisation en ligne de leurs publications de presse par des fournisseurs de services de la société de l'information. Les droits prévus au premier alinéa ne s'appliquent pas aux utilisations, à titre privé ou non commercial, de publications de presse faites par des utilisateurs individuels. La protection accordée en vertu du premier alinéa ne s'applique pas aux actes d'hyperliens. Les droits prévus au premier alinéa ne s'appliquent pas en ce qui concerne l'utilisation de mots isolés ou de très courts extraits d'une publication de presse.</p>

3) Une exception dont ne peuvent bénéficier gratuitement que les organismes de recherche et les institutions du patrimoine culturel

Le paragraphe 1 de l'article 3 de la directive dispose que seuls les organismes de recherche et les institutions du patrimoine culturel (ces dernières n'étaient pas prévues dans la proposition initiale de la Commission européenne) pourront bénéficier de l'exception prévue à cet article et ce gratuitement, c'est-à-dire sans compensation financière possible⁶².

Ces deux entités sont définies à l'article 2 de la directive, dont l'objet est de fixer les notions juridiques structurantes relatives au droit d'auteur et aux droits voisins en matière de marché unique numérique.

❖ D'une part, les organismes de recherche regroupent les **universités**, y compris leurs **bibliothèques**, les **instituts de recherche** et toute autre entité « ayant pour objectif premier de mener des recherches scientifiques, ou d'exercer des activités éducatives comprenant également des travaux de recherche scientifique :

a) à titre non lucratif ou en réinvestissant tous les bénéfices dans ses recherches scientifiques ;
ou

b) dans le cadre d'une mission d'intérêt public reconnue par un État membre ;

⁶¹ « 2. Les États membres prévoient le droit exclusif d'autoriser ou d'interdire la mise à la disposition du public, par fil ou sans fil, de manière que chacun puisse y avoir accès de l'endroit et au moment qu'il choisit individuellement : / a) pour les artistes interprètes ou exécutants, des fixations de leurs exécutions ; / b) pour les producteurs de phonogrammes, de leurs phonogrammes ; / c) pour les producteurs des premières fixations de films, de l'original et de copies de leurs films ; / d) pour les organismes de radiodiffusion, des fixations de leurs émissions, qu'elles soient diffusées par fil ou sans fil, y compris par câble ou par satellite ».

⁶² Aucune compensation financière ne peut être imposée par les ayants droit. En effet, le considérant 17 de la directive indique que « Compte tenu de la nature et l'étendue de l'exception, qui est limitée aux entités qui font de la recherche scientifique, le préjudice potentiel que cette exception pourrait occasionner aux titulaires de droits serait minime. Les États membres ne devraient, dès lors, pas prévoir de compensation pour les titulaires de droits en ce qui concerne les utilisations relevant des exceptions en matière de fouille de textes et de données introduites par la présente directive ».

de telle manière qu'il ne soit pas possible pour une entreprise exerçant une influence déterminante sur cet organisme de bénéficier d'un accès privilégié aux résultats produits par ces recherches scientifiques » (article 2, point 1 de la directive de 2019).

La mission relève que le législateur européen doit tenir compte de l'**hétérogénéité des structures juridiques effectuant de la recherche dans l'Union européenne**. Ainsi, la directive n'a pas retenu un critère organique par exemple pour définir les organismes de recherche. Le législateur européen n'a pas utilisé non plus l'expression de « recherche publique », mais il est clair que les acteurs de la recherche publique *lato sensu* sont visés, ainsi que les associations, fondations à but non lucratif, de même que toute personne privée chargée d'une mission de service public, dès lors que toutes ces entités respectent les conditions posées à l'article 3 de la directive. Le considérant 12 de la directive mentionne aussi les hôpitaux qui font de la recherche parmi les bénéficiaires de l'article 3 de la directive, ce qui en France renvoie aux CHU, rattachés à des universités.

La question des partenariats public-privé est centrale selon le législateur européen, dans la mesure où l'actuelle politique de la recherche de l'Union « *encourage les universités et les instituts de recherche à collaborer avec le secteur privé* » (considérant 11 de la directive de 2019). L'article 2 pose **deux critères fonctionnels alternatifs** pour définir les organismes de recherche, et un **troisième critère, obligatoire et transversal** selon la mission, qui vise à encadrer les partenariats public-privé. En tout état de cause, l'organisme de recherche qui a noué un partenariat avec un acteur privé ne peut jamais bénéficier de l'exception définie à l'article 3 si cet acteur exerce une influence déterminante sur lui de sorte qu'il obtiendrait un accès privilégié aux résultats de la fouille. Le considérant 12 apporte un éclairage important sur cette limite aux partenariats public-privé : « *les organismes sur lesquels des entreprises commerciales ont une influence déterminante leur permettant d'exercer un contrôle en raison d'éléments structurels tels que leur qualité d'actionnaire ou d'associé, ce qui pourrait conduire à un accès préférentiel aux résultats des recherches, ne devraient pas être considérés comme des organismes de recherche aux fins de la présente directive* ».

Le législateur européen n'interdit pas cependant que les organismes de recherche, lorsqu'ils bénéficient de l'exception posée à l'article 3 de la directive, utilisent les compétences techniques de leurs partenaires privés. En effet, le considérant 11 dispose que les organismes de recherche devraient pouvoir s'appuyer par exemple « *sur leurs partenaires privés pour effectuer des fouilles de textes et de données, y compris en utilisant leurs outils technologiques* ».

Par ailleurs, si l'article 3 est muet sur la question de la propriété des résultats de la fouille, le considérant 11 dispose que les organismes de recherche et les institutions du patrimoine culturel « *devraient toutefois rester les bénéficiaires de cette exception* ». Cette idée est aussi conforme à l'affirmation selon laquelle le préjudice aux titulaires est « *minime* », ce qui conduit à ne pas compenser l'exception (cons. 17).

Enfin, la finalité de la fouille de textes et de données doit toujours être la recherche scientifique, qu'il s'agisse de sciences naturelles ou de sciences humaines.

- ❖ D'autre part, les **institutions du patrimoine culturel** comprennent **les bibliothèques accessibles au public, les musées, les archives ou institutions dépositaires d'un patrimoine cinématographique ou sonore** (article 2, point 3 de la directive de 2019).

La mission constate, à ce stade, que la directive est peu prolixe sur la définition de ces institutions du patrimoine culturel : ainsi, les règles relatives aux partenariats public-privé applicables aux organismes de recherche n'ont pas été reprises ou déclinées pour ces institutions. Le silence du législateur européen sur ce point n'entraîne cependant pas mécaniquement l'exclusion de ces règles.

4) Une exception obligatoire, qui ne peut être remise en cause ni par voie contractuelle, ni par l'exercice du droit de retrait mentionné à l'article 4 de la directive

D'abord, il faut souligner le caractère obligatoire de l'exception posée à l'article 3 de la directive de 2019 : les Etats membres ont tous l'obligation de transposer la dérogation.

En outre, contrairement à l'exception/limitation de l'article 4 de la directive, qui peut être refusée par le titulaire de droits par la voie du droit de retrait (également appelé droit d'opposition ou *opt-out*), l'exception mentionnée à l'article 3 de la directive est opposable aux titulaires de droits qui ne peuvent pas la contester si les conditions prévues au même article 3 sont remplies. D'ailleurs, toute disposition contractuelle tendant à remettre en cause l'exception n'est réputée nulle et non avenue selon le 1^{er} paragraphe de l'article 7 de la directive⁶³.

Enfin, le législateur européen a expressément prévu, au paragraphe 4 de l'article 4⁶⁴, que l'exercice du droit de retrait ne peut jamais conduire à empêcher les organismes de recherche et les institutions du patrimoine culturel d'effectuer de la fouille de textes et de données dans le cadre de l'article 3.

5) Une exception qui ménage les intérêts des deux parties

Le paragraphe 2 de l'article 3 de la directive confère aux chercheurs la faculté de stocker les copies des œuvres ou autres objets protégés nécessaires à leurs fouilles sous réserve, d'une part, de garantir un niveau de sécurité approprié et, d'autre part, de ne poursuivre qu'une « *finalité de recherche scientifique* », y compris pour la vérification des résultats de la recherche.

Le sujet de la durée de conservation des données n'est pas abordé en tant que tel, le législateur ayant fait le choix de mentionner plus globalement leur stockage. La plupart des chercheurs auditionnés par la mission plaident pour une durée de conservation des données sinon illimitée, du moins très étendue, car les canons de la recherche scientifique imposent la possibilité de tester la robustesse des travaux des autres collègues et d'éprouver la pertinence de méthodes

⁶³ « Toute disposition contractuelle contraire aux exceptions prévues aux articles 3, 5 et 6 est non exécutoire ».

⁶⁴ « Le présent article n'affecte pas l'application de l'article 3 de la présente directive ».

alternatives. En tout état de cause, la directive n'octroie pas aux titulaires le droit d'exiger la destruction immédiate du corpus de données initiales.

La directive est également muette sur le contrôle du niveau de sécurité retenu par les chercheurs pour stocker les données. Toutefois, elle n'interdit pas expressément aux ayants droit de vérifier directement ou indirectement que le niveau de sécurité est approprié.

Par ailleurs, le paragraphe 3 de l'article 3 de la directive autorise les titulaires de droits « 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 ne doivent toutefois pas excéder « ce qui est nécessaire pour atteindre cet objectif »⁶⁵. Autrement dit, ces mesures ne doivent pas être détournées de leur finalité et se muer en obstacle injustifié à la fouille de textes et de données. La directive ne prévoit pas explicitement que ces mesures puissent avoir pour objet de garantir le fonctionnement du site internet du titulaire de droits, mais cet objectif semble découler assez directement de la notion d'intégrité du réseau.

La directive ne prévoit pas non plus l'instauration d'un tiers chargé d'arbitrer les différends entre titulaires de droits et chercheurs sur le caractère proportionné des mesures précitées. Mais elle impose la garantie de l'exception, comme la mission le rappellera plus loin.

6) Une exception qui invite les deux parties à négocier une charte de bonnes pratiques

Au paragraphe 4 de l'article 3, la directive impose aux Etats membres d'encourager les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel à définir d'un commun accord de bonnes pratiques concernant, d'une part, l'obligation pour les chercheurs de stocker les données avec un niveau de sécurité approprié, d'autre part, la faculté pour les titulaires de droits de mettre en place des mesures techniques permettant d'assurer la sécurité et l'intégrité des réseaux.

Compte tenu du caractère éminemment technique et évolutif des questions de sécurité numérique, et des nombreux sujets restés en suspens, le législateur européen a décidé de recourir à la voie contractuelle entre les parties prenantes pour fixer les modalités d'application concrète des paragraphes 2 et 3 de l'article 3 de la directive.

⁶⁵ V. avertissement cons. 16 : « Eu égard au nombre potentiellement élevé de demandes d'accès et de téléchargements de leurs œuvres ou autres objets protégés, les titulaires de droits devraient être autorisés à appliquer des mesures lorsqu'il existe un risque que la sécurité et l'intégrité de leurs systèmes ou bases de données soient compromises. Ces mesures pourraient 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. Ces mesures devraient rester proportionnées aux risques inhérents, 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 et ne devraient pas compromettre l'application effective de l'exception. »

C. Article 4 : un dispositif obligatoire introduit tardivement dans le projet de directive, qui explique son caractère atypique

1) Une disposition introduite tardivement et cependant obligatoire

Rappelons que la proposition de directive de 2016 ne contenait pas de disposition sur une dérogation en faveur de la fouille de données pour tous les usages. Il faut dire que l'intérêt est récent. En ce sens, la réflexion initiée par la Commission européenne en 2012, *Licences for Europe*, n'envisageait que la fouille « *for scientific research purposes* » (à des fins de recherche scientifique ; groupe de travail n° 4).

On relèvera seulement, dans l'*Analyse d'impact sur la modernisation des règles de l'UE en matière de droit d'auteur* accompagnant notamment la proposition de directive sur le droit d'auteur dans le marché unique numérique⁶⁶, l'évocation de quatre options élaborées sur la base de l'article 3 (exception de fouille à des fins de recherche). Le critère est donc toujours la recherche scientifique, mais l'option 4 propose une voie assez proche de l'exception finalement retenue à l'article 4, puisqu'elle suggère une exception obligatoire applicable à toute personne ayant un accès licite (y compris les organismes de recherche d'intérêt public et les entreprises) couvrant la fouille de textes et de données à des fins de recherche scientifique, qu'elle soit donc à finalité commerciale ou non⁶⁷. L'étude d'impact rejette cette option en raison de ses conséquences négatives majeures pour les titulaires de droits, notamment les éditeurs, qui ne pourraient plus conclure des licences pour la fouille de données à des fins scientifiques avec les acteurs commerciaux⁶⁸. Cette analyse met en exergue par anticipation l'importance que revêt le droit de retrait dans l'économie de l'article 4 de la directive.

Le dispositif de l'article 4 a donc été ajouté tardivement, avec une réflexion beaucoup moins aboutie que pour l'article 3, devant le Conseil. En réalité, la discussion a essentiellement porté sur le caractère obligatoire ou facultatif de cette exception, le Trilogue ayant finalement opté pour une harmonisation poussée puisque le dispositif doit être transposé par les Etats membres.

⁶⁶ SWD (2016) 301 : <https://ec.europa.eu/digital-single-market/en/news/impact-assessment-modernisation-eu-copyright-rules> ; v. Partie 1, p. 102 s.

⁶⁷ Version anglaise originale : « *Mandatory exception applicable to anybody who has lawful access (including both public interest research organisations and businesses) covering text and data mining for any scientific research purposes* ».

⁶⁸ « *This option would have a more significant negative impact on rightholders. As a consequence of the broad scope of the exception, STM publishers would no longer be able to licence TDM for scientific research purposes to commercial players, which represent an essential market for them, notably in areas such as life science and pharmaceutical. Industry estimates the value of the commercial TDM market (in Europe) to be worth more than 56 million euros by 2019. Two major STM publishers alone currently have 302 existing TDM licences with life science companies, which is a significant figure given the characteristics of the market. Publishers indicate that the use of TDM is also increasing outside the life science and pharmaceutical industry, including in sectors such as financial services and chemical manufacturing. Similarly to the other legislative options, this option would in principle not remove rightholders' ability to generate revenues from selling access to their content* » (p. 117).

2) Un dispositif qui déroge à de nombreuses prérogatives des titulaires de droits

Aux termes du premier paragraphe de l'article 4 de la directive 2019/790 :

« 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. »

L'exception ou limitation ainsi créée porte sur les droits présentés dans le tableau suivant :

Champ de l'exception prévue à l'article 4 de la directive 2019/790

Fondement juridique	Synthèse	Contenu
Article 5, point a), de la directive 96/9/CE	Droit d'auteur des producteurs de bases de données	L'auteur d'une base de données bénéficie, en ce qui concerne l'expression de cette base pouvant faire l'objet d'une protection par le droit d'auteur, du droit exclusif de faire ou d'autoriser : a) la reproduction permanente ou provisoire, en tout ou en partie, par quelque moyen et sous quelque forme que ce soit ;
Article 7, paragraphe 1, de la directive 96/9/CE	Droit <i>sui generis</i> des producteurs de bases de données	1. Les États membres prévoient pour le fabricant d'une base de données le droit d'interdire l'extraction et/ou la réutilisation de la totalité ou d'une partie substantielle, évaluée de façon qualitative ou quantitative, du contenu de celle-ci, lorsque l'obtention, la vérification ou la présentation de ce contenu attestent un investissement substantiel du point de vue qualitatif ou quantitatif.
Article 2 de la directive 2001/29/CE	Droit d'auteur Droits voisins « classiques »	Droit de reproduction Les États membres prévoient le droit exclusif d'autoriser ou d'interdire la reproduction directe ou indirecte, provisoire ou permanente, par quelque moyen et sous quelque forme que ce soit, en tout ou en partie : a) pour les auteurs, de leurs œuvres ; b) pour les artistes interprètes ou exécutants, des fixations de leurs exécutions ; c) pour les producteurs de phonogrammes, de leurs phonogrammes ; d) pour les producteurs des premières fixations de films, de l'original et de copies de leurs films ; e) pour les organismes de radiodiffusion, des fixations de leurs émissions, qu'elles soient diffusées par fil ou sans fil, y compris par câble ou par satellite.
Article 4, paragraphe 1, points a) et b) de la directive 2009/24/CE	Droit d'auteur spécial applicable aux programmes d'ordinateur	Actes soumis à restrictions 1. Sous réserve des articles 5 et 6, les droits exclusifs du titulaire au sens de l'article 2 comportent le droit de faire ou d'autoriser : a) la reproduction permanente ou provisoire d'un programme d'ordinateur, en tout ou en partie, par quelque moyen et sous quelque forme que ce soit ; lorsque le chargement, l'affichage, l'exécution, la transmission ou le stockage d'un programme d'ordinateur nécessitent une telle reproduction du programme, ces actes de reproduction sont soumis à l'autorisation du titulaire du droit ;

		b) la traduction, l'adaptation, l'arrangement et toute autre transformation d'un programme d'ordinateur et la reproduction du programme en résultant, sans préjudice des droits de la personne qui transforme le programme d'ordinateur ;
Article 15, paragraphe 1, de la directive 2019/790	Droits voisins étendus aux éditeurs de presse en ligne	<p>Protection des publications de presse en ce qui concerne les utilisations en ligne</p> <p>1. Les États membres confèrent aux éditeurs de publications de presse établis dans un État membre les droits prévus à l'article 2 et à l'article 3, paragraphe 2⁶⁹, de la directive 2001/29/CE pour l'utilisation en ligne de leurs publications de presse par des fournisseurs de services de la société de l'information. Les droits prévus au premier alinéa ne s'appliquent pas aux utilisations, à titre privé ou non commercial, de publications de presse faites par des utilisateurs individuels. La protection accordée en vertu du premier alinéa ne s'applique pas aux actes d'hyperliens. Les droits prévus au premier alinéa ne s'appliquent pas en ce qui concerne l'utilisation de mots isolés ou de très courts extraits d'une publication de presse.</p>

On observe ici l'ajout de l'article 4, paragraphe 1, points a) et b) de la directive 2009/24/CE, qui autorise à réaliser de la fouille sur les **programmes d'ordinateur**, là où l'article 3 ne le permet pas.

En conséquence, les champs d'application des articles 3 et 4 ne se sont pas parfaitement identiques. Cette différence de traitement est surprenante. En quoi la fouille, hors recherche, pour tous les autres usages, impliquerait un besoin différent de celui des chercheurs ? Dans ces conditions, l'absence de mention à l'article 3 laisse perplexe et on peut imaginer qu'il s'agit là d'un problème de cohérence que le législateur n'a pas relevé.

3) Une très large ouverture de la dérogation

Comme pour l'article 3, la fouille permise par l'article 4 vise tous types de données, textes, sons, images fixes ou animées, données brutes ou protégées organisées en bases de données. A cela s'ajoutent les programmes d'ordinateur, qui peuvent être librement fouillés dans le cadre de l'article 4.

En revanche, l'exception de l'article 4 diffère par sa **très large portée**, puisqu'elle bénéficie à tous, acteurs privés, pour des besoins propres en interne comme pour des solutions commerciales, mais aussi acteurs publics (notamment l'Hadopi dans le cadre de ses missions d'observation des usages et d'encouragement à l'offre légale), quelle que soit donc la finalité poursuivie.

Les incertitudes pesant sur les usages qui seront faits des contenus collectés sur le fondement de l'article 4 de la directive suscitent plus d'inquiétudes que les données moissonnées dans le cadre de l'article 3. Les ayants droit, y compris ceux qui ne nourrissent aucune inquiétude du fait de la consécration de cette exception, soulignent par conséquent la nécessité de rendre

⁶⁹ « 2. Les États membres prévoient le droit exclusif d'autoriser ou d'interdire la mise à la disposition du public, par fil ou sans fil, de manière que chacun puisse y avoir accès de l'endroit et au moment qu'il choisit individuellement : / a) pour les artistes interprètes ou exécutants, des fixations de leurs exécutions ; / b) pour les producteurs de phonogrammes, de leurs phonogrammes ; / c) pour les producteurs des premières fixations de films, de l'original et de copies de leurs films ; / d) pour les organismes de radiodiffusion, des fixations de leurs émissions, qu'elles soient diffusées par fil ou sans fil, y compris par câble ou par satellite ».

effectif le droit d'opposition à la fouille effectuée hors recherche. En effet, même pour ceux qui n'envisagent pas de l'exercer pour le moment, n'y voyant qu'un préjudice minime, la **préservation d'un *opt-out* plein et entier** est fondamental pour pouvoir réagir si les conditions d'exploitation devaient changer et que la fouille devait leur causer un préjudice économique.

4) Une ouverture contrebalancée par l'option de retrait

L'option de retrait (ou *opt-out*) constitue le **point d'équilibre de cette exception**, très ouverte et qui neutralise très largement les droits des titulaires. Le retrait de l'ayant droit conduit à réserver les contenus concernés ; il s'agit donc d'un **retour à l'exclusivité** et, le cas échéant, au **système de contractualisation par voie de licence d'exploitation**.

Pour cette raison, il importe que les conditions de l'exception soient bien précisées, afin de ne pas créer une méfiance générale chez les titulaires de droits, aboutissant à exercer par défaut – par précaution – le droit de retrait, ce qui réduirait à néant la volonté européenne de promouvoir la fouille hors recherche.

Un tel mécanisme aboutit, en pratique, à une **inversion du paradigme** puisque le titulaire de droits doit se manifester pour bénéficier de la protection. On comprend dès lors que ce droit de retrait concentrera une part importante des enjeux de l'article 4.

Que penser d'une telle disposition qui contraint le titulaire à une démarche afin de bénéficier de la protection ?

D'évidence, la situation pourrait être analysée comme mettant en péril **l'interdiction des formalités** posée à l'article 5 (2) de la Convention de Berne, en vertu duquel « *la jouissance et l'exercice [des] droits ne sont subordonnés à aucune formalité* » en matière de droit d'auteur. Pourtant, il faut considérer qu'un tel schéma est possible puisqu'une hypothèse comparable de réservation est prévue dans cette convention. En effet, son article 10 *bis* (1) autorise les Etats membres à « *permettre la reproduction par la presse, ou la radiodiffusion ou la transmission par fil au public, des articles d'actualité de discussion économique, politique ou religieuse, publiés dans des journaux ou recueils périodiques, ou des œuvres radiodiffusées ayant le même caractère, dans les cas où la reproduction, la radiodiffusion ou ladite transmission n'en est pas expressément réservée* » (nous soulignons). Certes, cette disposition n'est pas obligatoire, les Etats ayant la possibilité de l'exclure, mais elle met clairement en place une disposition dérogatoire, sous réserve d'*opt-out*. **L'article 4 de la directive de 2019 peut donc se prévaloir d'un précédent issu de la Convention de Berne.**

Enfin, la mission démontrera plus loin que le titulaire de droits n'a pas à motiver son opposition à la fouille hors recherche.

5) Une disposition moins aboutie que l'article 3

La tardiveté de l'intégration de la fouille pour tous les usages dans la réflexion européenne explique sans doute le caractère moins abouti de l'article 4 de la directive.

D'abord, l'article 4 ne fait aucune mention des conditions de la conservation. Certes, on pourrait considérer que cela n'est pas utile puisque les données ne peuvent être conservées que le temps « *nécessaire aux fins de la fouille de textes et de données* » (art. 4, para. 2). Cependant, le texte européen suggère qu'une certaine durée de conservation peut être requise puisque les copies peuvent être conservées « *aussi longtemps que nécessaire* » à l'objectif poursuivi par la fouille.

Ensuite, l'article 4 ne comporte pas de disposition comparable à l'article 3 au sujet de la préservation de la sécurité et de l'intégrité des réseaux et des bases. Ce silence peut étonner. En effet, alors que la fouille est permise pour tous les usages, le besoin de préservation est accru car les fouilles pourraient être massives. Sans doute le législateur européen a-t-il considéré que l'option de retrait permettait, nécessairement, de mettre en place ce type de mesures.

Enfin, on notera encore qu'il n'est nullement renvoyé à une charte des bonnes pratiques en ce domaine.

A ce stade, on peut s'étonner de ces différences de traitement. En effet, il pèse sur les chercheurs, bien souvent vertueux compte tenu des règles intrinsèques à la recherche scientifique, beaucoup plus d'obligations que sur les bénéficiaires de l'article 4. Cette **asymétrie** est d'autant plus étonnante que les risques de méconnaître les règles de la propriété intellectuelle seront plus fréquents et graves pour les fouilles effectuées sur le fondement de l'article 4 que pour celles réalisées dans le cadre de l'article 3 de la directive.

D. Des règles communes aux articles 3 et 4 de la directive

1) Des dérogations encadrées

Outre les conditions propres à chaque cas d'ouverture, il faut souligner que **les hypothèses de fouille visées ne suspendent que le droit de reproduction** (et l'extraction pour le droit *sui generis*). Ainsi, **les dérogations des articles 3 et 4 ne permettent pas de neutraliser le droit de communication au public**. Les contenus protégés ne sauraient donc être communiqués au public en tant que tels, **même partiellement**, dans le cadre d'une fouille de textes et de données. L'assertion est liée à la nature même de l'hypothèse visée : le propre de la fouille de données consiste à déconstruire les données entrantes en vue de construire un résultat étranger à la forme initiale. De l'avis des experts auditionnés par la mission, il n'est d'ailleurs pas possible, à partir du seul résultat, de remonter aux sources.

En outre, les dispositifs de fouille ne créent **pas un droit d'accès général aux contenus protégés** ; au mieux, il s'agirait d'une nouvelle hypothèse spécifique ouvrant l'accès à ces contenus dans les hypothèses décrites⁷⁰. Le bénéfice de l'exception est d'ailleurs conditionné à un « accès licite » aux contenus protégés dans les deux hypothèses de fouille.

⁷⁰ En ce sens, J.-C. Galloux, « Libres propos sur le droit d'accès aux données », in *Circulation des données et propriétés intellectuelles*, Propr. Intell. n° 76, juill. 2020, p. 11.

2) Des dispositifs soumis au test des trois étapes

L'article 7 de la directive prévoit que :

« 2. *L'article 5, paragraphe 5⁷¹, de la directive 2001/29/CE s'applique aux exceptions et aux limitations prévues dans le présent titre (...) ».*

Il résulte de la référence à l'article 5, paragraphe 5 de la directive 2001/29/CE que les exceptions et limitations prévues aux articles 3 et 4 de la directive 2019/790, lesquelles constituent des « *cas spéciaux* », ne doivent pas porter « *atteinte à l'exploitation normale de l'œuvre ou autre objet protégé* » ni causer « *un préjudice injustifié aux intérêts légitimes du titulaire du droit* ».

Autrement dit, comme toutes les exceptions, les dispositifs de fouille doivent respecter le test en trois étapes qui constitue, depuis la directive 2001/29, un filtre supplémentaire à la mise en œuvre d'une exception/limitation légale.

3) Des dispositifs garantis

L'article 7 de la directive de 2019 prévoit encore que :

« 2. (...) *L'article 6, paragraphe 4, premier, troisième et cinquième alinéas, de la directive 2001/29/CE, s'applique aux articles 3 à 6 de la présente directive. »*

Il résulte de la référence à l'article 6, paragraphe 4, premier, troisième et cinquième alinéas, de la directive 2001/29/CE, relatif aux obligations relatives aux mesures techniques, que :

- « *en l'absence de mesures volontaires prises par les titulaires de droits, y compris les accords entre titulaires de droits et d'autres parties concernées* », les États membres doivent « *prendre des mesures appropriées* » pour assurer que les bénéficiaires des exceptions ou limitations prévues aux articles 3 et 4 de la directive 2019/790/CE puissent effectivement exercer des activités de fouilles de textes et de données, dans la « *mesure nécessaire* » prévues par ces articles, s'ils ont un « *accès licite à l'œuvre protégée ou à l'objet protégé en question* » (premier alinéa du paragraphe 4 de l'article 6 de la directive 2001/29/CE). Autrement dit, les États membres ont l'obligation de rendre effectives les exceptions et limitations prévues aux articles 3 et 4 de la directive de 2019 ; cette hypothèse de garantie de l'effectivité de certaines exceptions au monopole des titulaires de droits⁷² a déjà été confiée en France à l'Hadopi⁷³. Il conviendra en conséquence d'étendre à ce champ les missions de l'Hadopi ou de l'autorité qui la remplacera (il est envisagé de fusionner le CSA et l'Hadopi pour créer l'Autorité de régulation de la communication audiovisuelle et numérique ou ARCOM).
- « *Les mesures techniques appliquées volontairement par les titulaires de droits, y compris celles mises en œuvre en application d'accords volontaires, et les mesures*

⁷¹ « *Les exceptions et limitations prévues aux paragraphes 1, 2, 3 et 4 ne sont applicables que dans certains cas spéciaux qui ne portent pas atteinte à l'exploitation normale de l'œuvre ou autre objet protégé ni ne causent un préjudice injustifié aux intérêts légitimes du titulaire du droit.* »

⁷² Art. L. 331-6 CPI.

⁷³ Art. L. 331-31 CPI.

techniques mises en œuvre en application des mesures prises par les États membres, jouissent de la protection juridique prévue au paragraphe 1 » (troisième alinéa du paragraphe 4 de l'article 6 de la directive 2001/29/CE). Pour mémoire, le paragraphe 1 de l'article 6 de la directive de 2001 dispose que : « *Les États membres prévoient une protection juridique appropriée contre le contournement de toute mesure technique efficace, que la personne effectue en sachant, ou en ayant des raisons valables de penser, qu'elle poursuit cet objectif* ». En d'autres termes, l'Etat a l'obligation de prévoir une protection juridique appropriée des mesures techniques de protection ou d'identification des œuvres utilisées par les titulaires de droits dans le cadre des articles 3 et 4 de la directive du 17 avril 2019 ; des dispositions sont déjà prévues dans le CPI pour sanctionner le contournement des mesures techniques⁷⁴. Elles s'appliqueront donc en l'état.

- « *Lorsque le présent article est appliqué dans le cadre des directives 92/100/CEE et 96/9/CE, le présent paragraphe s'applique mutatis mutandis* » (cinquième alinéa du paragraphe 4 de l'article 6 de la directive 2001/29/CE). Cet alinéa signifie que les obligations posées à l'article 6 de la directive 2001/29/CE s'imposent aux directives de 1992 (qui reconnaît notamment des droits voisins aux artistes interprètes, producteurs de phonogrammes et de vidéogrammes et organismes de radiodiffusion) et de 1996 (qui définit les droits des producteurs de base de données). Ainsi, en dépit des articles 3 et 4 de la directive 2019/790, les mesures techniques sur une œuvre, un objet de droit voisin ou une base de données, restent protégées et obéissent au régime énoncé dans le cadre d'une fouille.

On remarquera ici qu'il n'est pas renvoyé à l'alinéa 4 de l'article 6 paragraphe 4, en vertu duquel « *Les dispositions des premier et deuxième alinéas ne s'appliquent pas aux œuvres ou autres objets protégés qui sont mis à la disposition du public à la demande selon les dispositions contractuelles convenues entre les parties de manière que chacun puisse y avoir accès de l'endroit et au moment qu'il choisit individuellement.* ». En d'autres termes, la directive rend applicable l'obligation de rendre effectives les exceptions concernant les contenus à la demande quand ils font l'objet d'une fouille.

III. Présentation des différentes possibilités de transposition des articles 3 et 4 de la directive du 17 avril 2019

La transposition des dérogations imposées aux articles 3 et 4 de la directive 2019/790 s'inscrit dans un cadre juridique précis (A). Elle pose cependant un certain nombre d'options, que celles-ci soient communes (B) ou spécifiques (C et D), et qu'il appartiendra à la mission de trancher.

⁷⁴ Art. L. 335-3-1 et L. 335-3-2 CPI.

A. Le cadre juridique pour la transposition de cette directive

1) Des marges d'appréciation limitées lors de la transposition d'une directive d'harmonisation

Si les règlements européens sont d'application directe dans les Etats membres, les directives ne posent en principe que des objectifs à atteindre, qui doivent être précisés lors de la transposition en droit interne. Les Etats membres disposent donc dans cette hypothèse d'une certaine latitude dans la manière d'atteindre ces objectifs.

Toutefois, quand il s'agit d'une directive d'harmonisation en lien avec la constitution d'un marché intérieur, les marges de manœuvre des Etats membres sont limitées à la portion congrue, voire inexistantes.

En effet, dans son avis du 27 septembre 2018 sur le projet de loi portant suppression des « sur-transpositions » des directives européennes en droit français, le Conseil d'Etat rappelle que :

« Dans certains domaines, le droit européen interdit toute différence dans le droit de chacun des États membres et, par suite, toute sur-transposition. C'est le cas pour les directives – fréquentes dans les secteurs liés à la liberté de circulation des biens et des services – qui procèdent, dans un secteur déterminé, à une harmonisation complète sans ouvrir de possibilités d'option ou de dérogation (CJCE, 5 mai 1998, National Farmer's Union e.a., aff. C-157/96), ainsi que pour celles qui, par exemple, énumèrent les conditions de mise en œuvre d'une règle, faisant ainsi obstacle à ce que les mesures nationales de transposition introduisent des conditions supplémentaires (CJCE, 23 novembre 1989, Kommanditgesellschaft in Firma Eau de Cologne & Parfumerie Fabrik/Provide, aff. C-150/88) »⁷⁵.

Cette position fait écho à l'étude du Conseil d'Etat « Directives européennes : anticiper pour mieux transposer » de 2015⁷⁶ :

« Lorsque la directive procède à ce que l'on appelle une harmonisation totale, qui consiste en l'impossibilité pour les Etats membres d'adopter ou de maintenir, dans le domaine que régit la directive, des dispositions différentes de celles de cette dernière, la "sur-transposition" est alors interdite par le droit européen lui-même ».

Il revient en conséquence aux juges nationaux de faire prévaloir les directives européennes sur le droit national. Un requérant peut ainsi exercer un recours pour excès de pouvoir devant le juge administratif contre un acte réglementaire qui transpose inexactement une directive⁷⁷. Un tel recours est également ouvert à l'encontre de toutes les mesures réglementaires ultérieures⁷⁸. Enfin, le juge administratif a accepté de déclarer illégaux les actes réglementaires pris sur le

⁷⁵ <https://www.conseil-etat.fr/ressources/avis-aux-pouvoirs-publics/derniers-avis-publics/projet-de-loi-relatif-a-la-suppression-des-surtranspositions-des-directives-europeennes-en-droit-francais>

⁷⁶ Conseil d'Etat, *Directives européennes : anticiper pour mieux transposer*, La Documentation française, Les études du Conseil d'Etat, nov. 2015, p. 21.

⁷⁷ CE, 28 sept. 1984, *Confédération nationale des sociétés de protection des animaux de France et des pays d'expression française et autres*, n° 28467, aux Tables.

⁷⁸ CE, 7 déc. 1984, *Fédération française des sociétés de protection de la nature*, n° 41971-41972, au Recueil.

fondement d'une disposition législative incompatible avec les objectifs d'une directive européenne ⁷⁹.

En l'espèce, la mission considère que la directive 2019/790 est précisément une **directive d'harmonisation** puisque son article 1^{er} dispose que :

« La présente directive fixe des règles visant à poursuivre l'harmonisation du droit de l'Union applicable au droit d'auteur et aux droits voisins dans le cadre du marché intérieur, en tenant compte, en particulier, des utilisations numériques et transfrontières des contenus protégés ».

Ces dispositions font écho au premier considérant de la directive :

« Le traité sur le fonctionnement de l'Union européenne prévoit l'établissement d'un marché intérieur et l'instauration d'un système propre à empêcher les distorsions de concurrence dans le marché intérieur. Poursuivre l'harmonisation des dispositions législatives des États membres sur le droit d'auteur et les droits voisins devrait contribuer à la réalisation de ces objectifs. »

Certes, la seconde phrase de l'article 1^{er} pourrait semer le trouble car elle ne mentionne pas explicitement l'objectif d'harmonisation des règles au sein du marché unique, alors même que cet article traite explicitement des exceptions et limitations au droit d'auteur et aux droits voisins ⁸⁰. Mais la mission considère que cette seconde phrase doit être lue à la lumière de la première phrase de l'article 1^{er} et du considérant 1 de la directive.

2) Le cas particulier des exceptions au droit d'auteur

La politique législative menée par l'Union européenne au sujet des exceptions est ambiguë. D'abord, la directive 2001/29 énonce en son article 5 une liste exhaustive de 21 exceptions, reflétant l'ensemble des traditions des États membres et dont une seule doit impérativement être transposée, l'exception technique provisoire (art. 5 paragraphe 1). Une telle méthode rend l'harmonisation très théorique.

La directive de 2019 met en place une **autre méthode** puisqu'elle impose, parmi les « *mesures visant à adapter les exceptions et limitations à l'environnement numérique et transfrontière* », les deux dispositifs fouille de données (art. 3 et 4), une exception/limitation propre à l'enseignement numérique (art. 5) et une exception de conservation à des fins patrimoniales (art. 6).

Cela étant, il est apparu que même lorsque l'exception est facultative, son contenu, dès lors que l'État membre choisit de la transposer, a tendance à « se communautariser », par l'effet de l'utilisation par la CJUE de méthodes d'interprétation unificatrices, notamment la consécration de notions autonomes de droit de l'Union. En ce sens, dans l'arrêt *ACI Adam*, la CJUE a pu affirmer « *que les États membres ont la faculté d'instaurer ou non les différentes exceptions prévues à l'article 5 de cette directive, et ce conformément à leurs traditions juridiques, mais que, une fois qu'ils ont opéré le choix d'introduire une certaine exception, celle-ci doit être*

⁷⁹ CE, 28 févr. 1992, *S.A. Rothmans International France et S.A. Philip Morris France*, n° 56776-56777, au Recueil.

⁸⁰ « *Elle fixe également des règles relatives aux exceptions et limitations au droit d'auteur et aux droits voisins, à la facilitation des licences, ainsi que des règles destinées à assurer le bon fonctionnement du marché pour l'exploitation des œuvres et autres objets protégés* ».

appliquée de façon cohérente, de sorte qu'elle ne puisse pas porter atteinte aux objectifs poursuivis par la directive 2001/29 visant à assurer le bon fonctionnement du marché intérieur »⁸¹.

De l'ensemble de ces éléments, il ressort que **la marge de manœuvre des législateurs nationaux dans le cadre de la transposition des articles 3 et 4 apparaît assez réduite.**

3) Une transposition des articles 3 et 4 en cours dans les Etats membres

A la suite de son audition par la mission, l'Hadopi a consulté son réseau européen afin d'établir un état de la transposition des articles 3 et 4 de la directive du 17 avril 2019.

Il en ressort que seules l'Allemagne⁸² et l'Irlande disposaient d'une législation préexistante en matière de fouille de textes et de données (étant rappelé que le Royaume-Uni dispose depuis 2014 d'une loi sur la fouille⁸³), et qu'aucun Etat membre ne semble avoir achevé cette transposition (la date butoir, fixée au 7 juin 2021 par l'article 29 de la directive, sera sans doute dépassée compte tenu de la crise sanitaire liée à la COVID-19).

Etat des lieux de la transposition des articles 3 et 4 de la directive du 17 avril 2019

	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	

Source : Hadopi

Le détail de cet état des lieux figure en annexes du présent rapport.

⁸¹ CJUE, 10 avr. 2014, aff. C-435/12, pt 34.

⁸² Il existe en effet une exception TDM en Allemagne dans le cadre de la recherche scientifique et pour une utilisation non commerciale (*UrhG*, § 60d).

⁸³ Art. 29 A du *Copyright, Designs and Patents Act* 1988.

B. Les enjeux communs liés à la transposition des articles 3 et 4

➤ Quelle grille de lecture pour la transposition de ces deux articles ?

Compte tenu des réflexions précédentes, **la mission a retenu la méthode de transposition suivante :**

- Dès lors que le texte est **clair**, aucune marge de manœuvre n'est reconnue à l'Etat membre chargé de transposer une directive d'harmonisation prévoyant des exceptions et limitations obligatoires ;
- Lorsqu'un sujet est **hors champ de la directive** (et sans qu'on puisse l'interpréter du fait de dispositions proches ou par comparaison entre les articles 3 et 4), comme par exemple la qualification juridique des résultats de la fouille, l'Etat membre retrouve une certaine latitude, sous réserve de respecter bien sûr le droit européen ;
- Lorsqu'un article est muet sur des **sujets qui sont traités par d'autres articles** de la directive, l'Etat membre peut combler ces lacunes dans une démarche d'harmonisation ou au contraire acter la différence entre les dispositifs si les considérants de la directive y invitent ou pour des motifs tirés de l'intérêt général.

De manière générale, la mission estime que l'intérêt général commande de prévoir le plus souvent possible des **règles communes** pour les fouilles exercées sur le fondement des articles 3 et 4 de la directive. Ainsi, la mission préconise de prévoir les mêmes règles sur les partenariats public-privé pour les institutions du patrimoine culturel que pour les organismes de recherche, bien que ces règles ne soient expressément prévues que pour ces organismes. Toutefois, ce principe d'harmonisation n'est pas intangible car la mission propose de rendre obligatoire la destruction des copies effectuées lors des fouilles réalisées dans le cadre de l'article 4, sans étendre les règles prévues à l'article 3.

➤ Faut-il inscrire dans la loi ou le règlement la définition de la fouille de textes et de données ?

Comme nous l'avons vu précédemment, l'article 38 de la loi du 7 octobre 2016 pour une République numérique n'a pas défini l'activité d'exploration de textes et de données, préférant confier cette tâche au pouvoir réglementaire.

Afin de ne pas alourdir la rédaction des articles du CPI relatifs aux exceptions, la mission préconise de confier au pouvoir **réglementaire** le soin de définir la fouille de textes et de données en reprenant la définition donnée à l'article 2 de la directive de 2019.

➤ **Faut-il définir la notion d'« accès licite aux œuvres et objets protégés » ?**

Les dérogations des articles 3 et 4 de la directive sont expressément conditionnées à un accès licite aux contenus protégés. On peut éclairer la notion par renvoi aux considérants de la directive.

S'agissant de l'exception « académique », le considérant 14 de la directive de 2019 indique pour sa part que : « *L'accès licite devrait s'entendre comme couvrant l'accès à des contenus fondé sur une politique de libre accès ou en vertu d'arrangements contractuels entre, d'une part, les titulaires de droits et, d'autre part, les organismes de recherche ou les institutions du patrimoine culturel, comme des abonnements, ou en vertu d'autres voies légales. Par exemple, les personnes rattachées à un organisme de recherche ou à une institution du patrimoine culturel qui a souscrit des abonnements à certains contenus sont réputées bénéficier de l'accès licite à ces abonnements. **L'accès licite devrait également désigner l'accès à des contenus librement accessibles en ligne*** » (nous soulignons ⁸⁴).

S'agissant de l'hypothèse de l'article 4, le considérant 18 précise que « *cette exception ou limitation ne devrait s'appliquer que lorsque le bénéficiaire a accès de manière licite aux œuvres ou autres objets protégés, y compris lorsque ceux-ci ont été mis à la disposition du public en ligne, et dans la mesure où les titulaires de droits n'ont pas réservé de manière appropriée les droits de reproduction et d'extraction pour la fouille de textes et de données* ».

Une première option consisterait à ne pas définir la notion d'accès licite dans la mesure où elle n'est pas ignorée du droit de la propriété intellectuelle. La notion, comparable, de licéité de la source a en effet été discutée en matière d'exception de copie privée. La condition a été intégrée à l'article dédié par la loi du 20 décembre 2011, qui a réservé le bénéfice de la dérogation aux copies « *réalisées à partir d'une source licite* »⁸⁵. Cette exigence a ensuite été consacrée par la CJUE dans l'arrêt *ACI Adam*, du 10 avril 2014⁸⁶, en se fondant, d'une part, sur le bon fonctionnement du marché intérieur (pt 35) et, d'autre part, sur le test en trois étapes (pts 38 s.). Elle a aussi été appliquée à l'exception de reprographie⁸⁷. Ce principe de licéité peut également être rapproché de la notion d'« utilisateurs légitimes », qui encadre l'exception au droit exclusif en matière de logiciels et de bases de données⁸⁸.

Cela étant, les contours de l'exigence d'un accès licite ne sont pas clairement définis. Ils sont d'ailleurs appelés à évoluer avec **l'introduction de l'article 17 de la directive de 2019, qui permettra de « liciter » de nombreux contenus protégés mis en ligne par les utilisateurs grâce aux licences conclues entre les plateformes contributives et les ayants droit.**

De son côté, la Cour de justice a affirmé que la source de la copie est illicite si la reproduction est effectuée « *à partir d'œuvres protégées qui sont mises à disposition du public sans l'autorisation du titulaire des droits* »⁸⁹.

⁸⁴ La version anglaise utilise l'expression de « *freely available online* ».

⁸⁵ Art. L. 122-5, 2° CPI.

⁸⁶ CJUE, 10 avr. 2014, aff. C-435/12, *ACI Adam*.

⁸⁷ CJUE, 12 nov. 2015, aff. C-572/13, *Reprobel*.

⁸⁸ V. art. L. 122-6-1 et L. 342-3 CPI. – Critique sur cette exigence : S. Dusollier, « L'utilisation légitime de l'œuvre : un nouveau sésame pour le bénéfice des exceptions au droit d'auteur ? », CCE 2005, étude 38.

⁸⁹ CJUE, 5 mars 2015, aff. C-463/12, *Copydan*, pt 74.

Dans le même sens, certains titulaires souhaitent que la notion d'accès licite soit traduite par celle de source licite, à partir de contenus mis à la disposition avec l'autorisation des ayants droit, et que la condition soit vérifiée à chaque acte de reproduction effectué à l'occasion d'une fouille (accès initial, traitement des données, conservation, vérification des résultats...).

La mission considère qu'au vu du texte et du considérant 14, cette position risque d'aller au-delà de la volonté du législateur européen.

Dans ces conditions, la seconde option impliquerait de préciser la notion d'accès licite en raison de la spécificité de l'activité de fouille. En effet, dans les deux cas d'ouverture, on constate que la libre accessibilité en ligne semble présumer d'un accès licite. Il est encore exigé pour l'article 4, en plus de l'accès licite, que le droit de retrait n'ait pas été exercé.

Ainsi, l'accès licite semble inclure les contenus librement accessibles en ligne. L'affirmation est empreinte de pragmatisme et rappelle la jurisprudence en matière de liens hypertextes où la connaissance de l'illicéité semble être requise⁹⁰. Comment la personne qui fouille un contenu en ligne pourrait-elle vérifier la licéité des contenus librement accessibles ? Une telle obligation de vérification ne réduirait-elle pas à néant l'objectif de sécurisation juridique de l'activité de fouille poursuivi par le législateur européen ?

Pour autant, il faudrait considérer qu'il ne s'agit que d'une présomption simple d'accès licite, qui pourrait donc être combattue par la preuve contraire **lorsque l'évidence dément la vraisemblance**. En effet, le législateur européen n'a évidemment pas entendu valider la fouille sur des sites de contrefaçon dont l'illégalité est notoire ou manifeste. De même, une certaine prudence s'impose pour celui qui réalise la fouille. La mission recommande dès lors d'explicitier la notion d'accès licite dans les mesures **réglementaires** de la transposition de la directive, car cette notion semble avoir une portée plus large que celle habituellement retenue en droit de la propriété intellectuelle. Il pourrait encore être utile, dans un but de réaffirmation de la protection des droits, d'intégrer à la suite dans le texte réglementaire, sur le modèle du considérant 62 *in fine* de la directive, relatif à l'article 17, la précision selon laquelle l'évidence de l'illicéité ne peut permettre le bénéfice de la situation dérogatoire. Ainsi, **les dérogations des articles 3 et 4 de la directive ne devraient pas s'appliquer aux fouilles réalisées à partir de sites ou de fournisseurs de services dont l'objectif principal ou manifeste est de se livrer à du piratage de droit d'auteur ou d'autres droits ou de le faciliter**.

Quoi qu'il en soit, d'un point de vue pratique, il ressort des auditions de la mission que les éditeurs scientifiques mettent parfois à disposition des chercheurs des articles sur des plateformes dédiées. Il en va ainsi de PubMed, qui est un moteur de recherche de données bibliographiques spécialisé en biologie et en médecine, hébergé par la Bibliothèque américaine de médecine des Instituts américains de la santé. La plateforme Pubmed Medline met en ligne tous les ans des résumés d'articles en format html ou txt, tandis que sur la plateforme Pubmed Central les articles sont mis en ligne intégralement. Tous ces articles sont accessibles à toute personne, sans qu'il soit besoin de présenter une adresse mail de chercheur public.

De même, ELDA (société commerciale attachée à l'association ELRA) met à disposition, sous licence, des contenus « licités », c'est-à-dire purgés de tout risque contentieux au regard des règles du droit d'auteur (par contrat). Les personnes qui réalisent de la fouille peuvent ainsi

⁹⁰ CJUE, 13 févr. 2014, aff. C-466/12, *Svensson* ; CJUE, 8 sept. 2016, aff. C-160/15, *GS Media*.

demander à la société la mise à disposition de corpus pré-traités et pour lesquels les droits ont été gérés.

La mission a constaté, en matière de recherche (qu'elle soit publique ou privée), que l'**impératif de traçabilité** des données conduit les chercheurs à accorder une grande importance au choix de leurs sources. En effet, s'il existe certains travaux mineurs qui ne donnent pas lieu à des campagnes d'évaluation, la mise à disposition des collègues scientifiques des données initiales et des résultats est de plus en plus la règle, y compris pour les travaux des doctorants. Le regard des pairs permet de détecter le cas échéant des incohérences dans les données initiales, des malfaçons dans les données retraitées ou encore des biais méthodologiques. Ces campagnes d'évaluation peuvent également aboutir à une « compétition entre méthodes » et à l'émergence d'un algorithme plus adapté que celui retenu dans la recherche en question. Lorsqu'un jeu de données n'a été accordé que pour une finalité précise de recherche, les chercheurs s'interdisent le plus souvent de le communiquer à leurs pairs. Par conséquent, les chercheurs sont très favorables à l'idée d'autoriser sans restriction la communication des jeux de données à leurs pairs, dès lors que cette communication de données ne sort pas du champ de la recherche et a pour objectif de vérifier les résultats.

Les auditions de la mission ont montré que même les « doctorants CIFRE » (Conventions industrielles de formation par la recherche), qui sont des salariés en cours de rédaction d'une thèse⁹¹, ne sont autorisés à consulter et enregistrer les données des éditeurs scientifiques auxquels le laboratoire est abonné, que dans les locaux du laboratoire de recherche, en aucun cas au sein de l'entreprise, dès lors que cette dernière n'est pas abonnée aux sites en question.

Ainsi, une certaine **vertu dans l'accès** aux données est souvent exigée dans une démarche **scientifique** sérieuse.

➤ Les droits concernés par les articles 3 et 4 sont-ils identiques ?

La mission a constaté au II du présent rapport que le champ d'application des deux articles est assez large puisque ceux-ci dérogent au droit d'auteur, aux droits voisins, y compris les nouveaux droits voisins des éditeurs de presse, et au droit d'auteur et au droit *sui generis* des bases de données. Il a également été relevé que la coïncidence entre les champs d'application de ces deux articles n'est pas parfaite puisque l'article 4 déroge encore au droit d'auteur spécial portant sur les programmes d'ordinateur.

Déjà, le législateur européen avait exclu les logiciels et les bases de données du champ de l'article 5 paragraphe 1, de la directive 2001/29, relatif à l'exception technique provisoire. L'exclusion résultait alors indirectement du paragraphe 2 de l'article 1^{er}, dédié au champ d'application de la directive 2001/29, qui affirmait que celle-ci « *laisse intactes et n'affecte en aucune façon* » notamment les dispositions relatives aux programmes d'ordinateur et aux bases de données⁹². La mission considère que ces dispositions témoignent d'une mauvaise

⁹¹ Comme l'indique le site du ministère de l'enseignement supérieur, de la recherche et de l'innovation, une entreprise peut recruter en CDD ou CDI un jeune diplômé en master et lui confier un projet de recherche objet de sa thèse. En contrepartie d'un salaire minimal annuel brut de 23 484 euros accordé au doctorant, l'entreprise reçoit une aide annuelle de 14 000 euros de la part de l'Association nationale de la recherche et de la technologie (ANRT). En 2016, 1 377 nouvelles conventions ont été conclues.

⁹² V. aussi dir. 2001/29, cons. 20.

compréhension du fonctionnement des réseaux. En effet, les contenus qui circulent sur les réseaux intègrent souvent de façon indissociable des éléments logiciels ou de structuration sous forme de bases de données, de sorte que la distinction est impossible à mettre en œuvre⁹³.

On pourrait dès lors être tenté, dans une première option, de considérer que l'exclusion des programmes d'ordinateur du champ ouvert à la fouille de textes et de données défini par l'article 3 est une maladresse et qu'elle doit être corrigée en assurant une correspondance stricte entre les périmètres des fouilles effectuées sur le fondement des articles 3 et 4 de la directive.

Mais il pourrait alors être reproché une mauvaise transposition d'un texte pourtant clair. D'ailleurs, dans le cas de l'exception technique de la directive 2001/29, la France avait fait le choix de transposer l'exception à l'article L. 122-5, 6° du CPI en précisant que la seule reproduction provisoire exemptée est celle qui porte sur « *des œuvres autres que les logiciels et les bases de données* ».

Une seconde option consisterait donc à **prendre acte de la différence de portée des deux exceptions**.

La mission propose de suivre la seconde option, compte tenu du **précédent** en droit interne et des termes clairs de la directive, et de modifier en conséquence l'article L. 122-6-1 du CPI.

➤ **Faut-il prévoir des règles spécifiques aux publications de presse ?**

La loi n° 2019-775 du 24 juillet 2019⁹⁴ tendant à créer un droit voisin au profit des agences de presse et des éditeurs de presse a transposé l'article 15 de la directive 2019/790, qui confère des droits voisins aux éditeurs dont les publications sont mises en ligne par des « *fournisseurs de services de la société de l'information* ».

Cet article 15 de la directive confère, par renvoi aux articles 2 et 3, paragraphe 2, de la directive 2001/29, deux prérogatives aux éditeurs : le droit de reproduction et le droit de mise à disposition du public.

Grâce à ces droits, les auteurs d'œuvres intégrées dans une publication de presse recevront « *une part appropriée des revenus que les éditeurs de presse perçoivent des fournisseurs de services de la société de l'information pour l'utilisation de leurs publications de presse* » (art. 15, paragraphe 5 de la directive de 2019).

Ce même article 15 fixe des limites propres aux droits voisins concernés, précisément s'agissant :

- des actes d'hyperliens ;
- de l'utilisation de mots isolés ou de très courts extraits d'une publication de presse.

C'est pourquoi l'article 4 de la loi du 24 juillet 2019 a introduit dans le CPI l'article L. 218-2 qui dispose que :

⁹³ V. en ce sens, A. Lucas *et alii*, Traité précité, n° 385.

⁹⁴ Sur les enjeux, il sera renvoyé aux deux rapports CSPLA de Mme Laurence Franceschini (juill. 2016 et janv. 2018).

« L'autorisation de l'éditeur de presse ou de l'agence de presse est requise avant toute reproduction ou communication au public totale ou partielle de ses publications de presse sous une forme numérique par un service de communication au public en ligne. »

Tandis que l'article 2 de la même loi a créé un article L. 211-3-1 du même code pour prévoir que :

« Les bénéficiaires des droits ouverts à l'article L. 218-2 ne peuvent interdire :

1° Les actes d'hyperlien ;

2° L'utilisation de mots isolés ou de très courts extraits d'une publication de presse. Cette exception ne peut affecter l'efficacité des droits ouverts au même article L. 218-2. Cette efficacité est notamment affectée lorsque l'utilisation de très courts extraits se substitue à la publication de presse elle-même ou dispense le lecteur de s'y référer. »

La mission constate que le législateur français a pris soin de strictement encadrer la limite aux droits voisins tirée des très courts extraits : en aucun cas, cette exception ne doit remettre en cause les droits voisins reconnus aux éditeurs de presse par la directive de 2019⁹⁵.

Dès lors, la question se pose de savoir si les dispositifs de fouille doivent de même être strictement limités pour empêcher une remise en cause des nouveaux droits voisins. Certains ayants droit considèrent en ce sens qu'une limite du même type devrait être intégrée en matière de nouveaux droits voisins sur les publications de presse. La mission n'adhère cependant pas à cette analyse.

Une première option serait de s'inspirer du précédent de la loi du 24 juillet 2019, d'autant que le milieu de l'édition fait état de fortes inquiétudes. Aussi, il pourrait être utile de rappeler directement dans le CPI que l'exercice des dérogations tirées des articles 3 et 4 de la directive ne doivent pas remettre en cause les droits voisins reconnus récemment aux éditeurs de presse par l'article L. 218-2 du même code. Toutefois, une telle option aurait comme inconvénient d'instaurer un traitement spécial des nouvelles exceptions issues de l'article 3 et 4 de la directive. Elle impliquerait sans doute de transposer les exceptions de fouille à l'article L. 211-3-1 du CPI, alors que cet article n'est pas dédié à des exceptions mais à des frontières spéciales du monopole, ce qui justifie d'ailleurs que le test en trois étapes ne s'y applique pas (alors que les nouvelles exceptions sont bien soumises à ce test).

Une seconde option pourrait être de ne pas préciser cette limite, en la tenant pour redondante. En effet, on doit considérer qu'une exception ou limitation qui remettrait en cause l'efficacité d'un droit serait **nécessairement contraire au triple test**. Or, comme on l'a vu plus haut, l'article 7 de la directive de 2019 prévoit que les exceptions reconnues aux articles 3 et 4 de cette directive ne doivent pas porter « atteinte à l'exploitation normale de l'œuvre ou autre objet protégé » ni causer « un préjudice injustifié aux intérêts légitimes du titulaire du droit », compte tenu du paragraphe 5, article 5 de la directive de 2001. Dans ces conditions, si les hypothèses de fouille sont bien transposées dans les articles dédiés aux exceptions, le triple test

⁹⁵ Comp. cons. 58, *in fine*, de la dir. : « Compte tenu de l'agrégation et de l'utilisation massives de publications de presse par les prestataires de services de la société de l'information, il importe que l'exclusion des très courts extraits soit interprétée de manière à ne pas affecter l'efficacité des droits prévus dans la présente directive. »

y est mentionné. Notamment, le dernier alinéa de l'article L. 211-3 prévoit déjà que les exceptions aux droits voisins (y compris les nouveaux droits voisins des éditeurs de presse) « *ne peuvent porter atteinte à l'exploitation normale de l'interprétation, du phonogramme, du vidéogramme, du programme ou de la publication de presse ni causer un préjudice injustifié aux intérêts légitimes de l'artiste-interprète, du producteur, de l'entreprise de communication audiovisuelle, de l'éditeur de presse ou de l'agence de presse* ».

Aussi, la mission recommande de suivre la deuxième option, en transposant les articles 3 et 4 à l'article L. 211-3 ; **la soumission des exceptions/limitations au triple test interdit que celles-ci remettent en cause l'efficacité des droits, notamment par un effet de substitution.**

➤ **Faut-il confier à l'Hadopi le soin de veiller à la bonne mise en œuvre des articles 3 et 4 ?**

Trois niveaux d'intervention peuvent être dégagés.

Comme indiqué plus haut, par renvoi à la directive 2001/29, l'article 7 de la directive de 2019 impose que les dérogations posées aux articles 3 et 4 soient garanties. Il apparaît à la mission que la France pourrait utilement, comme elle l'a déjà fait pour les exceptions concernées au titre de la directive de 2001, confier ce pouvoir à l'Hadopi. **C'est pourquoi la mission estime nécessaire de modifier l'article L. 331-31 du CPI afin d'inscrire les activités de fouille dans la mission de régulation et de veille en matière de mesures techniques de protection (MTP) et d'identification des œuvres et des objets protégés par un droit d'auteur ou un droit voisin.** Ce faisant, l'Hadopi pourra exercer une mission de conciliation sur le fondement de l'article L. 331-36 du même code pour toute question relative à la mise en œuvre effective des exceptions relatives à la fouille quand sont en jeu des MTP.

Par ailleurs, **indépendamment des questions liées à la présence de MTP**, un grand nombre de personnes auditionnées par la mission plaident pour **l'instauration d'une autorité de régulation chargée d'arbitrer plus globalement tous les différends en lien avec les activités de fouilles de textes et de données.** Cette procédure, qui pourrait, comme pour la copie privée, être précédée d'une tentative de conciliation, serait de nature à rassurer les ayants droit qui craignent notamment une dispersion des corpus comme les chercheurs qui redoutent de ne pas pouvoir y accéder aisément. Le règlement de différends porterait alors plus largement que dans le cadre de l'article L. 331-31 sur l'ensemble des questions liées aux modalités d'exercice de l'exception de fouille de textes et de données (par exemple, sur les mesures de protection du réseau hébergeant les œuvres protégées, ou les modalités de conservation des reproductions). Là encore, la mission propose de confier ce rôle à l'Hadopi, afin de fluidifier la mise en œuvre de l'exception. Il conviendrait alors de préciser par voie législative que le règlement de ces différends peut être porté devant l'Hadopi qui procéderait, dans ce cas, selon les dispositions de l'article L. 331-35 du même code. En revanche, il apparaît à la mission que ce nouveau rôle confié à l'Hadopi devrait être réservé à la **fouille académique**, qui nécessite sans doute une certaine souplesse et une rapidité de traitement en cas de différend. S'agissant de la fouille prévue à l'article 4, les conditions sont beaucoup moins nombreuses et moins susceptibles de créer des conflits de mise en œuvre, les titulaires de droits disposant toujours du droit de retrait.

Enfin, la directive impose aux Etats membres d'encourager les acteurs de l'article 3 à s'engager dans la signature d'une **charte des bonnes pratiques**. Un tel instrument, une fois adopté, nécessite un **suivi (incluant une actualisation)**, notamment parce que les technologies sont éminemment évolutives. A nouveau, la mission propose de confier cette mission à l'Hadopi, qui dispose des compétences techniques requises. Cette troisième mission conduirait à modifier l'article L. 331-14 du CPI ⁹⁶.

Parmi ces trois nouvelles missions confiées à l'Hadopi, seule la première est obligatoire car découlant du texte même de la directive.

C. Les options spécifiques liées à la transposition de l'article 3

➤ Quelle doit être la définition des organismes de recherche ?

La mission préconise d'inscrire dans la loi que l'exception prévue à l'article 3 n'est ouverte qu'aux organismes de recherche et aux institutions du patrimoine culturel, mais de renvoyer leur définition précise au pouvoir réglementaire.

- ❖ S'agissant des **universités au sens large**, la mission suggère de se référer à l'article L. 711-2 du code de l'éducation, tel qu'il résulte de la loi n° 2013-660 du 22 juillet 2013 relative à l'enseignement supérieur et à la recherche.

Cet article définit les types d'établissements publics à caractère scientifique, culturel et professionnel, qui sont :

- les universités auxquelles sont assimilés les instituts nationaux polytechniques ;
- les écoles et instituts extérieurs aux universités ;
- les écoles normales supérieures, les écoles françaises à l'étranger et les grands établissements ;
- les communautés d'universités et d'établissements.

- ❖ S'agissant des **organismes de recherche à proprement parler**, la mission propose de viser les personnes morales poursuivant une mission de recherche publique mentionnée au chapitre II du titre I^{er} du livre I^{er} de la partie législative du code de la recherche.

Ce chapitre consacré aux objectifs et moyens institutionnels de la recherche publique comprend notamment les articles suivants :

- L. 112-2 (la recherche publique est organisée dans les services publics, notamment les établissements publics d'enseignement supérieur, les établissements publics de recherche et les établissements de santé, et dans les entreprises publiques) ;
- L. 112-4 (il reproduit la définition de la recherche agronomique et vétérinaire issue de l'article L. 830-1 du code rural et de la pêche maritime) ;

⁹⁶ « La Haute Autorité remet chaque année au Gouvernement et au Parlement un rapport rendant compte du respect de leurs obligations et engagements par les professionnels des différents secteurs concernés. Ce rapport est rendu public ».

- L. 112-5 (il reproduit la définition de la recherche appliquée sur la forêt et le bois issue de l'article L. 152-1 du nouveau code forestier).

Il est également envisageable de viser les personnes morales mentionnées au livre III du code de la recherche, et d'y ajouter la référence aux articles L. 112-4 et L. 112-5 précités. En effet, la référence au livre III du code de la recherche offre l'avantage de présenter toutes les personnes morales en charge de la recherche publique en France, qu'il s'agisse par exemple :

- Des établissements publics à caractère administratif comme le Centre national de la recherche scientifique (CNRS), l'Institut national de la recherche agronomique (INRA), l'Institut national de la santé et de la recherche médicale (INSERM) ;
- Des établissements publics à caractère industriel et commercial comme le Commissariat à l'énergie atomique et aux énergies alternatives (CEA) ou l'Agence nationale pour la gestion des déchets radioactifs (ANDRA) ;
- Des fondations de coopération scientifique.

❖ S'agissant des **personnes morales de droit privé poursuivant une mission d'intérêt général en matière de recherche**, la mission n'a pas identifié d'article précis dans la législation actuelle auquel il pourrait être renvoyé.

C'est pourquoi la mission préconise d'inscrire dans le décret d'application de la loi :

- Soit « les personnes morales de droit privé chargée d'une mission d'intérêt général en matière de recherche » ;
- Soit « les associations ou fondations reconnues d'utilité publique poursuivant une activité de recherche scientifique ».

Une discussion est sans doute ici nécessaire avec les différents acteurs, mais quelle que soit la formule retenue, des structures comme l'Institut Pasteur, fondation reconnue d'utilité publique, ou la Fondation nationale des sciences politiques devront bénéficier de l'exception posée par l'article 3 de la directive.

➤ **Quelles sont les entités couvertes par la notion d'institutions du patrimoine culturel ?**

❖ S'agissant des **bibliothèques accessibles au public**, elles sont définies au livre III du code du patrimoine.

Le titre I^{er} de ce livre concerne les dispositions communes, le titre II porte sur les bibliothèques municipales et intercommunales, le titre III a trait aux bibliothèques départementales.

Le titre IV du livre III de la partie réglementaire du code du patrimoine traite des bibliothèques nationales, c'est-à-dire de la Bibliothèque nationale de France et de la bibliothèque publique d'information.

- ❖ S'agissant des **musées**, ils sont définis à l'article L. 410-1 du code du patrimoine, qui ouvre le livre IV de ce code.

Aux termes de cet article, « *Est considérée comme musée, au sens du présent livre, toute collection permanente composée de biens dont la conservation et la présentation revêtent un intérêt public et organisée en vue de la connaissance, de l'éducation et du plaisir du public.* »

Sont ainsi visés les musées relevant de la compétence des collectivités territoriales et ceux relevant de la compétence de l'Etat (« musées de France »).

- ❖ S'agissant des **archives**, elles sont définies à l'article L. 211-1 du code du patrimoine.

Cet article dispose que : « *Les archives sont l'ensemble des documents, y compris les données, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité* ».

Il convient de distinguer les **archives publiques** des **archives privées**.

Les premières sont définies à l'article L. 211-4 du même code et regroupent :

« *1° Les documents qui procèdent de l'activité de l'Etat, des collectivités territoriales, des établissements publics et des autres personnes morales de droit public. Les actes et documents des assemblées parlementaires sont régis par l'ordonnance n° 58-1100 du 17 novembre 1958 relative au fonctionnement des assemblées parlementaires ;*

2° Les documents qui procèdent de la gestion d'un service public ou de l'exercice d'une mission de service public par des personnes de droit privé ;

3° Les minutes et répertoires des officiers publics ou ministériels et les registres de conventions notariées de pacte civil de solidarité ».

Les secondes sont définies à l'article L. 211-5 du même code et regroupent par défaut toutes les archives qui ne sont pas qualifiées de publiques :

« *Les archives privées sont l'ensemble des documents définis à l'article L. 211-1 qui n'entrent pas dans le champ d'application de l'article L. 211-4.* »

Compte tenu des termes retenus par le législateur européen pour définir les institutions du patrimoine culturel, la mission considère que ce dernier a entendu viser les archives publiques et non les archives privées, même si ce point peut être débattu⁹⁷.

- ❖ S'agissant des **institutions responsables du dépôt légal**, elles sont actuellement définies à l'article L. 132-3 du code du patrimoine.

Cet article dispose que :

« *Sont responsables du dépôt légal, qu'ils gèrent pour le compte de l'Etat, dans des conditions déterminées par décret en Conseil d'Etat : la Bibliothèque nationale de France, le Centre national de la cinématographie, l'Institut national de l'audiovisuel et le service chargé du dépôt légal du ministère de l'intérieur.*

⁹⁷ Notamment pour les archives privées données à une institution publique.

Ce décret peut confier la responsabilité du dépôt légal à d'autres établissements ou services publics, nationaux ou locaux, à la condition qu'ils présentent les garanties statutaires et disposent des moyens, notamment scientifiques, propres à assurer le respect des objectifs définis à l'article L. 131-1. »

La mission considère qu'il convient de viser cet article L. 132-3 du code du patrimoine, même si l'article 2 de la directive ne mentionne que les institutions dépositaires d'un patrimoine « cinématographique » ou « sonore ».

➤ **Quelles doivent être les règles encadrant les partenariats public-privé (PPP) afin d'éviter un détournement de la philosophie de l'article 3 de la directive ?**

Selon la mission, le législateur européen ne s'est pas opposé à ce que des chercheurs privés effectuent des travaux de fouille de textes et de données en s'appuyant sur l'exception prévue à l'article 3 de la directive, laissant à chaque Etat le soin de définir les contours exacts de la recherche.

En outre, il existe, au moins en théorie, le risque de voir un acteur privé conclure un accord de financement avec des chercheurs publics sur le fondement de l'article 3 de la directive pour contourner l'opposition d'un titulaire de droits à la fouille de textes et de données hors recherche mentionnée à l'article 4 de la même directive.

C'est vraisemblablement pour cette raison que le législateur européen a interdit le bénéfice de l'exception posée par l'article 3 aux chercheurs ayant conclu un PPP (partenariat public-privé) lorsque l'entreprise cocontractante, qui exerce une influence déterminante sur l'organisme de recherche, peut bénéficier d'un accès privilégié aux résultats produits par ces recherches scientifiques.

Au préalable, la mission souligne que la compréhension du 1) de l'article 2 de la directive peut soulever des difficultés. Les trois critères (activité non lucrative/réinvestissement intégral des bénéficiaires dans la recherche scientifique ; mission d'intérêt public ; absence d'accès privilégié aux résultats de la fouille) se rapportent-ils uniquement à l'expression « *toute autre entité* » ou bien également aux « *universités, y compris leurs bibliothèques* » et aux « *instituts de recherche* » ?

En tout état de cause, **la mission considère que le point déterminant est que le troisième critère, relatif à l'absence d'accès privilégié aux résultats de la fouille, s'applique à tous les organismes de recherche, quelle que soit leur nature** (et non seulement à « *toute autre entité* »).

Cette interprétation est corroborée par la dernière phrase du considérant 12 de la directive qui dispose que « *les organismes sur lesquels des entreprises commerciales ont une influence déterminante leur permettant d'exercer un contrôle en raison d'éléments structurels tels que leur qualité d'actionnaire ou d'associé, ce qui pourrait conduire à un accès préférentiel aux résultats des recherches, ne devraient pas être considérés comme des organismes de recherche aux fins de la présente directive* ». **Autrement dit, la mission estime que dès lors qu'une université ou un laboratoire public conclut une convention avec une entreprise lui attribuant un accès privilégié aux résultats de la fouille, cet organisme de recherche ne peut pas bénéficier de l'exception posée par l'article 3 de la directive.**

La mission a exploré plusieurs options pour apprécier la portée du troisième critère (accès privilégié aux résultats de la fouille) et circonscrire ainsi le risque de détournement des PPP.

Option 1 : réserver l'exception posée par l'article 3 à la recherche fondamentale et exclure du champ de l'exception les recherches appliquées et le développement expérimental⁹⁸.

Une telle option serait toutefois contraire à l'esprit de la directive qui ne prévoit pas une telle limitation et réduirait considérablement sa portée.

Option 2 : fixer des critères extérieurs à la directive, comme des plafonds de participation financière des acteurs privés. Dans cette hypothèse, seuls les projets dans lesquels les entreprises financent par exemple moins de la moitié ou d'un tiers d'un projet de recherche pourraient justifier le bénéfice de l'exception posée par l'article 3 de la directive.

Cette option peut en effet se prévaloir d'un précédent récent, à titre de comparaison : l'article 30 de la loi pour une République numérique a autorisé même après la conclusion d'un contrat d'édition, sous conditions (précisément en instaurant une période d'embargo), la mise à disposition gratuite d'un écrit scientifique dans un format ouvert, dès lors que le financement du chercheur est issu pour au moins 50% de fonds publics⁹⁹.

Cette option doit toutefois être écartée car elle introduirait des critères étrangers à la directive, dont l'objet est d'harmoniser le marché unique numérique, d'autant que la directive comprend d'autres critères poursuivant le même objectif de limitation du champ des PPP.

Option 3 : définir un ensemble d'indices en s'inspirant des critères tirés des considérants de la directive. Cette option consisterait ainsi à autoriser les chercheurs à bénéficier de l'exception introduite par l'article 3 de la directive à la double condition suivante :

1- la fouille ne doit entraîner qu'un préjudice minime pour le titulaire de droits (considérant 17 de la directive) ;

2- l'organisme de recherche doit demeurer le bénéficiaire direct de la fouille (considérant 11 de la même directive).

L'inconvénient de cette option est qu'elle complexifierait assez sensiblement le cadre juridique et pourrait également remettre en cause l'effectivité de l'exception définie à l'article 3 de la directive.

Option 4 : retenir comme **seul critère l'absence d'accès préférentiel de l'entreprise qui finance une recherche aux résultats de la fouille de textes et de données.**

⁹⁸ Pour la définition des différentes catégories de recherche, v. le *Manuel de Frascati 2015, Lignes directrices pour le recueil et la communication des données sur la recherche et le développement expérimental*, OCDE, p. 53-61 : <https://www.oecd-ilibrary.org/docserver/9789264257252-fr.pdf?expires=1604090543&id=id&accname=guest&checksum=75956029033543756F0A5C10881FBD0D>

⁹⁹ Art. L. 533-4 du code de la recherche : « I. Lorsqu'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne est publié dans un périodique paraissant au moins une fois par an, son auteur dispose, même après avoir accordé des droits exclusifs à un éditeur, du droit de mettre à disposition gratuitement dans un format ouvert, par voie numérique, sous réserve de l'accord des éventuels coauteurs, la version finale de son manuscrit acceptée pour publication, dès lors que l'éditeur met lui-même celle-ci gratuitement à disposition par voie numérique ou, à défaut, à l'expiration d'un délai courant à compter de la date de la première publication ».

En effet, c'est le seul critère prévu dans le dispositif de la directive. Le critère de « *l'influence déterminante* » est pour ainsi dire transparent par rapport au critère de l'accès privilégié aux résultats de la fouille. Si une entreprise finance à hauteur de 80% un projet mené par une université mais que la convention ne prévoit pas un accès préférentiel de l'entreprise aux résultats de la fouille, l'université pourra bénéficier de l'exception posée à l'article 3 de la directive.

En conséquence, la mission propose de retenir l'interprétation suivante : **tout organisme de recherche défini par voie réglementaire (voir plus haut) peut bénéficier de l'exception posée à l'article 3 de la directive, sauf si le projet de fouille fait l'objet d'une convention (de financement ou de mise à disposition de compétences humaines ou techniques) avec une entreprise qui aurait ainsi accès de manière privilégiée aux résultats de cette fouille** (par exemple, selon la mission, communication des résultats provisoires en avant-première alors qu'une publication est prévue, communication exclusive de tout ou partie des résultats pendant une période déterminée, attribution initiale et exclusive d'un brevet éventuel...).

- **Faut-il appliquer aux institutions du patrimoine culturel les mêmes règles qu'aux organismes de recherche en matière de partenariat public-privé ?**

Comme la mission l'a indiqué plus haut, les institutions du patrimoine culturel n'étaient pas mentionnées, dans le projet initial de directive présenté par la Commission européenne, parmi les bénéficiaires de l'article 3.

C'est peut-être cette omission qui explique que les règles relatives au partenariat public-privé n'ont pas été étendues ou déclinées à ces institutions. Or, il n'existe aucune différence substantielle entre les chercheurs travaillant dans un organisme de recherche et ceux intervenant dans une institution du patrimoine culturel au sens de l'article 2 de la directive.

Compte tenu de ce contexte, la mission préconise d'**appliquer aux institutions du patrimoine culturel les mêmes règles en matière de PPP** que celles prévues pour les organismes de recherche.

- **Faut-il retenir une interprétation extensive ou restrictive de la notion de personnes rattachées aux organismes de recherche et aux institutions du patrimoine culturel ?**

La mission considère que cette notion doit être interprétée de manière **stricte**.

En effet, cette notion, par un jeu de renvoi en cascade, ne saurait avoir pour conséquence de permettre à des chercheurs entretenant un lien indirect avec ces entités de bénéficier de l'exception posée par l'article 3 de la directive. Soit le chercheur est fonctionnaire ou salarié de l'entité éligible à la fouille, et il bénéficie alors de l'exception posée par l'article 3 de la directive, soit il travaille pour le compte d'un tiers privé, auquel cas il convient d'appliquer les règles relatives au partenariat public-privé définies à l'article 2 ou celles relatives à la fouille hors recherche académique définie à l'article 4 de la directive.

➤ **Les bénéficiaires de l'article 3 ont-ils un accès direct et inconditionné aux contenus ?**

On pourrait considérer, dans une première option, que les bénéficiaires de l'article 3 disposent d'un « droit à » la fouille. Une telle analyse traduirait la volonté européenne de soutenir la recherche, notamment en lien avec l'IA. L'affirmation selon laquelle l'option de retrait de l'article 4 ne peut porter atteinte à la fouille fondée sur l'article 3 pourrait aller dans le même sens.

Une seconde option postulerait plutôt une analyse en termes d'exception à un droit. D'ailleurs, seuls certains bénéficiaires peuvent réaliser de la fouille académique, ce qui suppose donc de pouvoir identifier le « fouilleur ». En outre, l'exception n'est pas un blanc-seing : elle est finalisée (recherche scientifique), implique un accès licite aux contenus et elle n'interdit pas la mise en place de mesures permettant de protéger l'intégrité des infrastructures.

La mission considère que **l'article 3 n'a pas entendu poser un droit d'accès immédiat, inconditionnel et impersonnel aux contenus protégés**. L'accès pourrait donc parfaitement impliquer la mise en place d'une solution technique permettant par exemple de préserver les réseaux.

➤ **Faut-il qualifier juridiquement les résultats de la fouille de textes et de données ?**

La mission constate que l'article 3 de la directive est muet sur la qualification juridique des résultats de la recherche. En particulier, en cas de partenariat public-privé, qui est le propriétaire des résultats de la fouille, quelles sont les règles de diffusion des résultats, quelles sont les règles en cas de dépôt de brevet ?

La mission estime par conséquent que les Etats membres sont libres de déterminer les règles relatives aux résultats de la recherche.

Trois options peuvent être distinguées.

La première option consiste à ne pas qualifier dans la loi ou le règlement les résultats de la fouille, en faisant prévaloir le cadre juridique actuel tel que défini dans le code des relations entre le public et l'administration (CRPA) et le code de la recherche.

La deuxième option vise à qualifier les résultats de la fouille de « données publiques ».

La troisième option consiste à qualifier les résultats de la fouille de « données de la recherche », dans le prolongement de la loi pour une République numérique.

• S'agissant de la première option :

Comme l'a souligné Mme Agnès Robin, Maître de conférences à l'Université de Montpellier, entendue par la mission, « *La production et la diffusion des connaissances scientifiques se*

*trouvent au carrefour de deux politiques opposées : l'une orientée vers la valorisation des résultats de la recherche scientifique, l'autre orientée vers l'ouverture des données »*¹⁰⁰.

D'une part, la valorisation des résultats de la recherche fait partie, depuis la loi du 15 juillet 1982¹⁰¹, d'une des missions de la recherche publique. Par l'ordonnance du 17 février 2014¹⁰², un livre spécifique a même été inséré dans le code de la recherche sur « *la valorisation des résultats de la recherche et le transfert de technologie en direction du monde économique et des associations et fondations, reconnues d'utilité publique* ». La valorisation des résultats de la recherche fait partie des quatre objectifs que doit poursuivre la politique nationale de la recherche selon l'article L. 111-1 du code de la recherche¹⁰³. Ces dernières années, les sociétés d'accélération du transfert de technologies (SATT) se sont par conséquent multipliées. Créées par un ou plusieurs établissements (universités et organismes de recherche), les treize SATT actuelles ont pour mission principale de « *traduire les découvertes et compétences de la recherche publique en applications concrètes et répondre aux besoins des entreprises* » à travers des activités aussi diverses que les dépôts de brevets, les opérations de preuve de concept, la création de start-up ou encore l'attribution de licences à des entreprises¹⁰⁴.

D'autre part, les résultats de la recherche scientifique sont soumis depuis 2005 à la politique générale d'ouverture des données, qui s'imposent aux établissements réalisant de la recherche publique comme à toutes les administrations. En effet, les articles L. 311-1 à L. 327-1 du CRPA définissent le triptyque suivant : communication des documents administratifs, diffusion et réutilisation des informations publiques. Depuis la loi du 28 décembre 2015¹⁰⁵, la réutilisation des informations scientifiques se situe dans le champ des données publiques.

Selon Mme Robin, l'articulation de ces deux missions, valorisation des recherches et diffusion/réutilisation des données, est source de complexité pour les chercheurs, faute pour le législateur ne pas avoir tenu compte de toutes les spécificités de la recherche publique.

En tout état de cause, le principe de diffusion et de réutilisation des données de l'administration connaît de très nombreuses limites.

Elles sont de deux ordres.

D'une part, l'article L. 311-4 du CRPA pose comme principe que « *les documents administratifs sont communiqués ou publiés sous réserve des droits de propriété littéraire et artistique* ». L'article L. 321-2 du même code dispose que ne sont pas considérés comme des

¹⁰⁰ A. Robin, « L'ouverture des données publiques scientifiques : de l'examen de la règle "*open as possible, closed as necessary*" », CCE n° 9, sept. 2020, p. 5.

¹⁰¹ Loi n° 82-610 du 15 juill. 1982 d'orientation et de programmation pour la recherche et le développement technologique de la France.

¹⁰² Ordonnance n° 2014-135 du 17 févr. 2014 modifiant la partie législative du code de la recherche.

¹⁰³ « *La politique nationale de la recherche et du développement technologique vise à :*

1° Accroître les connaissances ;

2° Partager la culture scientifique, technique et industrielle ;

3° Valoriser les résultats de la recherche au service de la société. A cet effet, elle s'attache au développement de l'innovation, du transfert de technologie lorsque celui-ci est possible, de la capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique, et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;

4° Promouvoir la langue française comme langue scientifique. »

¹⁰⁴ <https://www.enseignementsup-recherche.gouv.fr/cid67054/les-satt-societes-d-acceleration-du-transfert-de-technologies.html>

¹⁰⁵ Loi n° 2015-1779 du 28 déc. 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public.

informations publiques les documents sur lesquels des tiers à l'administration détiennent des « *droits de propriété intellectuelle* ». L'article L. 321-3 du même code précise que l'administration ne peut alléguer que ses droits auraient été méconnus pour s'opposer à la réutilisation du contenu des bases de données qu'elle a rendu publiques, sauf si cette publication a méconnu les droits de propriété intellectuelle détenus par des tiers à l'administration ¹⁰⁶.

D'autre part, les articles L. 311-5 et L. 311-6 du même code énumèrent les longues listes de documents dont la communication à des tiers est interdite (par exemple, les documents couverts par la protection de la vie privée, le secret médical ou encore le secret en matière commerciale et industrielle).

Cette première option a pour avantage de ne pas remettre en cause le cadre juridique actuel sur les données de la recherche : dès lors que les activités de fouille n'occupent pas aujourd'hui une place prépondérante dans la recherche en France, il serait inopportun de bouleverser le cadre actuel, aussi insatisfaisant soit-il. Elle a en revanche l'inconvénient de ne pas répondre au risque de contournement de l'article 4 de la directive.

- S'agissant de la deuxième option :

Cette deuxième option présente l'avantage de répondre au risque de contournement de l'article 4 par certaines entreprises privées en cas d'*opt-out* exercé par des titulaires de droits propriétaires de données et d'être cohérente avec la volonté du Gouvernement de promouvoir la « science ouverte ».

La mission considère toutefois qu'il ne serait pas opportun de qualifier systématiquement les résultats des fouilles de « données publiques ».

Non seulement cette notion connaît déjà de très nombreuses limites, comme on vient de le voir, mais elle aurait aussi comme conséquence néfaste de décourager les acteurs privés de cofinancer des projets de fouille.

Comme l'indique Mme Robin ¹⁰⁷, les contrats de collaboration de recherche déterminent « *la part contributive des partenaires à la recherche-développement* », dont le résultat est exprimé à travers un document généralement intitulé « *Annexe financière* ». Ce document fixe en effet « *des quotes-parts de copropriété* » et calcule les « *créances de rémunération des partenaires sur les produits d'exploitation de la technologie innovante (savoir-faire ou inventions) ou des bases de données produites en commun* ».

Cette analyse a été corroborée par les entretiens menés par la mission avec de nombreux chercheurs, qui ont révélé que la très grande majorité des contrats de partenariat en matière de fouille prévoient actuellement une publication dans une revue avec comité de lecture (sauf s'il s'agit d'un projet de faible ampleur, pour prouver un concept, ou d'un contrat répondant à l'intérêt exclusif et ponctuel de l'acteur privé, auxquels cas la publication n'est pas prévue), avec copropriété des résultats de la recherche et parfois engagement de l'organisme public de céder ses droits en cas de dépôt de brevet.

¹⁰⁶ « *Sous réserve de droits de propriété intellectuelle détenus par des tiers, les droits des administrations mentionnées au premier alinéa de l'article L. 300-2 du présent code, au titre des articles L. 342-1 et L. 342-2 du code de la propriété intellectuelle, ne peuvent faire obstacle à la réutilisation du contenu des bases de données que ces administrations publient en application du 3° de l'article L. 312-1-1 du présent code* ».

¹⁰⁷ *Op. cit.*, p. 7.

- S'agissant de la troisième option :

Dans la continuité de l'article 38 de la loi pour une République numérique, la mission propose de qualifier les résultats de la fouille effectuée sur le fondement de l'article 3 de la directive de données de la recherche, quand bien même cette notion n'a été que très récemment définie¹⁰⁸.

D'une part, l'étude d'impact du projet de loi pour une République numérique proposait la définition suivante¹⁰⁹ :

*« les données de la recherche sont l'ensemble des données factuelles issues d'observations, d'enquêtes, de corpus, d'archives, d'expériences ou d'analyses computationnelles, enregistrées sous tout format et sur tout support, dans une forme brute ou après avoir été traitées ou combinées, et sur lesquelles se fondent les raisonnements du chercheur et qui sont jugées nécessaires à la validation des résultats de la recherche »*¹¹⁰.

D'autre part et surtout, la directive concernant les données ouvertes et la réutilisation des informations du secteur public (refonte)¹¹¹ propose de définir les données de la recherche comme *« des documents se présentant sous forme numérique, autres que des publications scientifiques, qui sont recueillis ou produits au cours d'activités de recherche scientifique et utilisés comme éléments probants dans le processus de recherche, ou dont la communauté scientifique admet communément qu'ils sont nécessaires pour valider des conclusions et résultats de la recherche »* (art. 2, point 9).

La mission recommande de retenir l'option 3 qualifiant le résultat de la fouille de « données de la recherche ». Cependant, la mission n'a pas proposé d'inscrire cette solution dans le texte législatif proposé en fin de rapport (partie IV) car cette question s'intègre à une réflexion plus large en lien avec la politique de la recherche.

Ainsi, les chercheurs, compte tenu de leur liberté d'expression et de communication (art. L. 952-2 du code de l'éducation), resteraient libres de fixer au cas par cas avec le cofinanceur privé les règles sur la publicité des résultats, la propriété des résultats et les règles applicables en cas de dépôt de brevet, à condition de respecter, bien évidemment, les conditions de l'article 3 permettant de bénéficier de l'exception en cas de PPP. La qualification de « données de la recherche » est certes moins contraignante que celle de « données publiques », mais elle a le mérite de rappeler implicitement que les résultats de la fouille ont vocation à circuler entre chercheurs.

La mission invite en outre les ministères de tutelle des organismes de recherche et des institutions du patrimoine culturel qui bénéficieront de l'article 3 de la directive à indiquer systématiquement dans les formulaires de convention de partenariat qu'en aucun cas l'acteur privé ne peut avoir un accès privilégié aux résultats de la fouille.

¹⁰⁸ Pour une perspective historique de la notion : V. Rebouillat. *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs*, CNAM, 2019, p. 26-28.

¹⁰⁹ P. 49, note 31 : https://www.legifrance.gouv.fr/contenu/Media/Files/autour-de-la-loi/legislatif-et-reglementaire/etudes-d-impact-des-lois/ei_art_39_2015/ei_republique_numerique_cm_09.12.2015.pdf.pdf

¹¹⁰ https://www.legifrance.gouv.fr/contenu/Media/Files/autour-de-la-loi/legislatif-et-reglementaire/etudes-d-impact-des-lois/ei_art_39_2015/ei_republique_numerique_cm_09.12.2015.pdf.pdf

¹¹¹ Dir. 2019/1024 du 20 juin 2019.

- **Comment articuler l'exception de fouille « académique » avec l'article L. 132-4 du code du patrimoine, qui crée une exception au droit d'auteur au profit des chercheurs qui effectuent leurs travaux dans les organismes responsables du dépôt légal ?**

Pour mémoire, l'article L. 132-4 du code du patrimoine dispose que :

« L'auteur ne peut interdire aux organismes dépositaires, pour l'application du présent titre :

1° La consultation de l'œuvre sur place par des chercheurs dûment accrédités par chaque organisme dépositaire sur des postes individuels de consultation dont l'usage est exclusivement réservé à ces chercheurs ;

2° La reproduction d'une œuvre, sur tout support et par tout procédé, lorsque cette reproduction est nécessaire à la collecte, à la conservation et à la consultation sur place dans les conditions prévues au 1° ».

Les représentants de la Bibliothèque nationale de France et de l'Institut national de l'audiovisuel que la mission a auditionnés se sont interrogés sur l'articulation entre les dispositions de l'article L. 132-4 du code du patrimoine et la nouvelle exception « académique » prévue à l'article 3 de la directive de 2019.

De manière générale, il serait sans doute préférable de traiter dans un article unique, ou du moins dans une partie identifiée du code de la propriété intellectuelle, toutes les exceptions au droit d'auteur et assimilés. En outre, il est exact que le dépôt légal concerne les œuvres protégées par le droit d'auteur, les bases de données, les logiciels¹¹², ainsi que les phonogrammes, les vidéogrammes et les documents multimédias¹¹³ sans oublier les services de communication au public par voie électronique¹¹⁴.

La mission considère qu'à ce stade, il n'est pas indispensable de modifier l'article L. 132-4 du code du patrimoine, dès lors que son objet est différent de celui de l'article 3 de la directive de 2019. En effet, l'article L. 132-4 crée un droit pour les chercheurs de consulter sur place, dans les murs des institutions responsables du dépôt légal, les œuvres protégées, quel que soit l'objet de leur recherche, alors que l'article 3 de la directive ne prévoit pas un droit de consultation sur place pour les fouilles de textes et de données.

La question qui se pose cependant est de savoir si l'article L. 132-4 du code du patrimoine crée ou pas une hypothèse alternative d'accès licite pour les chercheurs, leur permettant

¹¹² L'article R. 132-9 du code du patrimoine dispose que « *Les logiciels et les bases de données sont déposés à la Bibliothèque nationale de France dès lors qu'ils sont mis à la disposition du public, à titre onéreux ou gratuit, par diffusion en nombre d'un support matériel de quelque nature que ce soit* ».

¹¹³ V. par exemple l'article R. 132-20 du code du patrimoine : « *Le dépôt des phonogrammes, vidéogrammes et documents multimédias est effectué en deux exemplaires pour les documents édités sur le territoire national et en un exemplaire pour ceux qui sont importés, au plus tard le jour de la mise à la disposition du public destinataire, à la Bibliothèque nationale de France* ».

¹¹⁴ V. par exemple l'article R. 132-23 du code du patrimoine, qui porte notamment sur le dépôt légal des « *services de communication au public en ligne au sens du IV de l'article 1er de la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique, enregistrés sous le nom de domaine .fr ou tout autre nom de domaine enregistré auprès du ou des organismes français chargés de la gestion de ces noms, enregistrés par une personne domiciliée en France ou produits sur le territoire français* ».

dès lors de bénéficiaire de l'article 3 de la directive. Cette interprétation pourrait se prévaloir du considérant 14 de la directive qui indique que « *l'accès licite devrait s'entendre comme couvrant l'accès à des contenus fondé sur une politique de libre accès ou en vertu d'arrangements contractuels entre, d'une part, les titulaires de droits et, d'autre part, les organismes de recherche ou les institutions du patrimoine culturel, comme des abonnements, ou en vertu d'autres voies légales* » (nous soulignons).

Interrogée par la mission, les responsables de la BNF accréditent cette thèse : l'accès au dépôt légal des œuvres protégées par des chercheurs habilités, dans la mesure où il est autorisé par l'article L. 132-4 du code du patrimoine, constituerait bien un « accès licite » au sens de l'article 3 de la directive de 2019.

En effet, selon la BNF, l'accès aux œuvres protégées doit être qualifié de licite si les trois conditions suivantes sont remplies :

- Le chercheur est accrédité par la bibliothèque de recherche dans le cadre de sa demande de travaux réalisés sur le fondement de l'article L. 132-4 du code du patrimoine. A l'issue d'un entretien, un « pass recherche » est délivré au chercheur qu'il doit payer. Toutefois, dans le cadre de partenariats entre la BNF et des organismes de recherche, le chercheur relevant du partenaire peut être exonéré du paiement du pass pour toute la durée du partenariat. La durée de validité du pass peut être d'une journée, de cinq jours ou d'une année ;
- Les documents consultés par le chercheur, sous forme matérielle ou dématérialisée, doivent être entrés dans les collections du dépôt légal de la BNF ;
- Ces documents doivent être consultés dans les emprises de la BNF uniquement.

Cette « autre voie légale » permettant l'accès aux ressources, mentionnée par le considérant 14 de la directive de 2019, permettrait donc aux chercheurs de constituer directement leurs corpus thématiques de fouille auprès de la BNF par exemple (au lieu de s'adresser aux éditeurs, ce qui est d'ailleurs souhaité par certains d'entre eux). L'INA a ainsi indiqué, lors de son audition, qu'elle ne s'opposait pas à être désignée comme « tiers de confiance » par les titulaires de droits, afin d'effectuer notamment des copies des œuvres déposées légalement pour le compte des organismes de recherche, dans ses emprises, avec ses propres technologies ou avec celles de ces organismes (v. Annexes).

Cependant, il reste que **ce mécanisme de type « exceptions en cascade » est susceptible d'interroger sur sa conformité au triple test.** A la connaissance de la mission, la question n'a jamais été posée en jurisprudence. En outre, si l'exception de l'article L. 132-4 ne vise que la consultation sur place des fonds numérisés, c'est sans doute parce qu'une transmission en ligne organiserait une concurrence évidente au préjudice des titulaires de droits.

En tout état de cause, la mission considère qu'il ne sera pas possible de faire jouer, en cascade, les deux exceptions (celle du dépôt légal prévu par l'article L. 132-4 du code du patrimoine, puis celle de l'article 3 de la directive de 2019), pour demander *à distance* de fouiller les fonds concernés. Il ne semble **pas possible, en « chaînant » les exceptions, d'autoriser des copies hors des emprises des organismes concernés.**

En revanche, la question reste entière lorsque le chercheur se déplace et consulte sur place. Cette consultation lui donne-t-elle un accès licite, qui l'autoriserait alors à fouiller les données ? Cette hypothèse pourrait être retenue, mais elle présente peu d'avantages concrets. D'un point de vue pratique, il est difficile d'imaginer que le chercheur fasse toutes ses recherches dans

l'enceinte de l'institution responsable du dépôt légal. Surtout, cette hypothèse implique que le chercheur se déplace dans l'enceinte de l'institution responsable du dépôt légal, alors que l'exception de l'article 3 s'intègre dans une série de mesures en lien avec « l'environnement numérique et transfrontière » (Titre II de la directive).

Au-delà de cette réflexion, la mission s'interroge sur **l'extension des missions qui pourraient être confiées aux organismes dépositaires**. En effet, les titulaires de droits pourraient « déléguer » aux organismes dépositaires le soin de fournir les corpus sur le fondement de l'article 3 de la directive, comme certains titulaires le souhaitent. Cette possibilité permettrait de réaliser des corpus thématiques, concernant plusieurs éditeurs. Ainsi, le chercheur pourrait s'adresser à l'ayant droit pour obtenir les contenus souhaités et celui-ci aurait la possibilité de renvoyer vers les organismes de dépôt légal qui réaliseraient alors les copies pour le compte de l'ayant droit. Dans cette hypothèse, les organismes concernés pourraient tout à fait être autorisés à facturer des frais de gestion, conformément à ce qui suit. Les modalités de cette **délégation facultative** pourraient être discutées dans la **charte**.

➤ **La mise à disposition de copies peut-elle donner lieu à une redevance visant à couvrir uniquement son coût réel ?**

Le considérant 17 de la directive dispose que « *Compte tenu de la nature et l'étendue de l'exception, qui est limitée aux entités qui font de la recherche scientifique, le préjudice potentiel que cette exception pourrait occasionner aux titulaires de droits serait minime. Les États membres ne devraient, dès lors, pas prévoir de compensation pour les titulaires de droits en ce qui concerne les utilisations relevant des exceptions en matière de fouille de textes et de données introduites par la présente directive* ».

Ce considérant interdit donc en principe que les titulaires de droits demandent une compensation aux chercheurs en contrepartie de la mise à disposition des copies ou reproductions d'objets protégés.

La mission tire de la lecture de ce considérant **deux conclusions** :

- d'une part, si un **titulaire de droits** devait connaître un afflux massif et soudain de demandes de chercheurs, la mission considère qu'il pourrait instaurer une rémunération couvrant au maximum les frais de mise à disposition des copies et reproductions ;
- d'autre part, ce considérant 17 n'interdit pas le paiement par les chercheurs d'une redevance au profit des **organismes compétents en matière de dépôt légal** (BNF, INA notamment), qui n'agissent pas alors en tant que titulaires de droits.

Le Conseil d'Etat, dans un arrêt d'Assemblée du 21 novembre 1958 (*Syndicat national des transporteurs aériens*) a jugé qu'une redevance pouvait être exigée des usagers en vue de couvrir les charges d'un service public déterminé dès lors que cette redevance trouve sa contrepartie directe dans des prestations fournies par ce service. Si une imposition ne peut être prévue que par la loi en vertu de l'article 34 de la Constitution, une redevance peut être mise en place par voie réglementaire. Il résulte de l'analyse de la jurisprudence du Conseil d'Etat et du Conseil constitutionnel qu'une redevance peut être créée si le service rendu constitue la finalité essentielle du service public, que ce service ne remplit pas des missions qui incombent par

nature à l'Etat, que la personne qui acquitte cette redevance est bien le bénéficiaire principal de ce service et enfin que le montant de cette redevance est proportionnel au service rendu ¹¹⁵.

La mission rappelle d'ailleurs que même en matière de réutilisation des informations publiques, le CRPA autorise l'établissement de redevances. Ainsi l'article L. 324-1 de ce code dispose que l'administration peut instaurer « *une redevance de réutilisation* » lorsqu'elle est tenue de « *couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement* » de sa mission de service public. Toutefois, « *le produit total du montant de cette redevance, évalué sur une période comptable appropriée, ne dépasse pas le montant total des coûts liés à la collecte, à la production, à la mise à la disposition du public ou à la diffusion de leurs informations publiques* ». Cette autorisation s'étend aux bibliothèques, aux musées et aux archives (art. L. 324-2 du même code). Le montant de la redevance doit être « *fixé selon des critères objectifs, transparents, vérifiables et non discriminatoires* » et actualisé « *au moins tous les cinq ans* ».

La mission ne s'oppose donc pas à ce que **le pouvoir réglementaire instaure des redevances pour services rendus** pour les organismes mentionnés à l'article L. 132-4 du code du patrimoine, sous réserve que le montant de ces redevances ne dépasse pas le coût réel de la mise à disposition des contenus protégés et qu'il n'instaure aucune discrimination entre chercheurs européens.

En tout état de cause, la question d'une éventuelle redevance instaurée au bénéfice des organismes responsables du dépôt légal ne se posera pas avec acuité avant plusieurs années, car la grande majorité des fouilles aujourd'hui s'effectuent directement sur internet sans mobiliser ces organismes et cette situation devrait perdurer dans les années à venir.

➤ **Quelles doivent être les règles de conservation des copies ?**

Le paragraphe 2 de l'article 3 de la directive dispose que les copies des œuvres ou autres objets protégés 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. Il est éclairé par le considérant 15 ¹¹⁶.

Ce paragraphe ne précise pas la personne en charge de stocker ces copies, ni la durée de leur stockage, ni même la possibilité expresse de les communiquer à d'autres scientifiques.

¹¹⁵ V. les conclusions de Guillaume Odinet, rapporteur public, sur l'affaire CE, 28 nov. 2018, *SNCF Réseau*, n° 413 839, au Recueil.

¹¹⁶ « *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. 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.* »

Comme le remarque avec raison l'Hadopi dans sa contribution : « *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)* ».

S'agissant de la **personne en charge de stocker les copies des œuvres et objets protégés**, la mission rappelle que le considérant 15 de la directive dispose que les Etats membres doivent être libres de désigner des « *organismes de confiance aux fins du stockage de ces copies* ». Le stockage des copies peut donc être confié soit aux chercheurs, soit à des tiers, soit aux deux. En aucun cas les entreprises, dans le cadre d'un PPP, ne doivent avoir la mission de conserver ces copies ou reproductions, ni y avoir accès directement.

Lors de son audition, M. Dacos, représentant du Comité pour la science ouverte, a rappelé que le ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation possède de nombreuses **infrastructures de stockage**, y compris des *data centers*, des plateformes sécurisées et un Centre d'accès sécurisé aux données (CASD) pour les données les plus sensibles, et que l'objectif du ministère était de développer ces infrastructures pour augmenter le stockage des données de la recherche.

La mission aurait une préférence pour confier cette mission de conservation aux **infrastructures publiques de stockage des données**, car elles pourraient harmoniser les règles de sécurité et de communication des copies à d'autres chercheurs. En particulier, si ces infrastructures ont le monopole de la conservation des copies, elles pourraient leur ajouter des métadonnées, des marquages et des avertissements pour rappeler que ces copies ne peuvent être utilisées que dans le cadre de l'article 3 de la directive. La mission considère que les copies et reproductions pourraient être « chiffrées » selon des modalités établies en partenariat avec l'ANSSI (Agence nationale de la sécurité des systèmes d'information) qui rappellerait les meilleures solutions techniques en matière de cybersécurité. Par ailleurs, la mission préconise d'instaurer des « *Data management plans* » comme ceux mis en œuvre dans les projets de l'ANR, définissant les conditions d'accès, de sécurité, de conservation de données (par exemple les données personnelles). Des audits par des organismes extérieurs devraient pouvoir être réalisés pour évaluer la pertinence des mesures de protection mises en œuvre. Les *data centers* devraient également tenir la liste des chercheurs qui ont obtenu ces copies. Ainsi, les ayants droit qui le souhaitent pourraient à tout moment demander des comptes aux *data centers*, en obtenant des informations sur les mesures de sécurité adoptées, la liste des chercheurs qui ont demandé des copies et la liste de ceux qui les ont obtenues.

Le fait de confier le stockage des copies et reproductions à des centres spécialisés rendrait plus facile la preuve du non-respect des conditions de l'exception, étant précisé que la directive ne prévoit aucune sanction en cas de manquement aux règles de conservation de ces documents.

Certaines personnes doutent cependant de la capacité pour les infrastructures publiques de stockage des données de supporter toutes les copies et reproductions relatives à la fouille.

Concernant la **durée de conservation des copies**, les auditions menées par la mission ont montré que les canons de la recherche scientifique imposent que les auteurs d'une recherche

scientifique conservent sans limitation temporelle les données initiales (copies des données sources) et « nettoyées » (documents élaborés par les chercheurs) ainsi que le résultat de la fouille (application de l'algorithme aux données « nettoyées »), pour permettre à leurs pairs de pouvoir réitérer leurs expériences, et identifier si besoin des irrégularités dans les données utilisées, des biais méthodologiques, voire tester des méthodes alternatives. Il paraît difficile à la mission de prévoir des durées différentes de conservation des copies pour tenir compte des avancées de la recherche car, comme nous l'avons déjà indiqué, une étude scientifique effectuée il y a vingt ou trente ans peut à tout moment retrouver un intérêt scientifique.

S'agissant de la portée de la notion de **communication des copies à d'autres chercheurs**, la mission considère qu'une interprétation extensive doit être retenue car le législateur européen a utilisé l'expression de conservation des copies à des « *fins de recherche scientifique* » : il n'a pas exigé que la communication ne se fasse qu'entre chercheurs d'une même discipline ou de discipline connexe. Ainsi, si un chercheur en sociologie dispose d'un corpus d'un journal sur le fondement de l'article 3 de la directive, ce même article l'autorise à le communiquer à un chercheur en linguistique. En tout état de cause, ce débat sur le périmètre de la communication des copies n'a pas lieu d'être dans la mesure où tout chercheur a le droit de demander à l'ayant droit la communication d'une copie ou d'une reproduction. Pour reprendre notre exemple, si le chercheur en linguistique ne peut obtenir les copies de la part du chercheur en sociologie, il pourra les demander directement auprès du journal (ou du responsable du dépôt légal éventuellement).

La mission considère que la loi doit poser le principe de la conservation des copies mais renvoyer à un décret le soin de définir ses modalités pratiques, y compris pour la communication éventuelle du corpus (voir *infra*) et à la charte des bonnes pratiques le soin de préciser ces modalités. En outre, comme indiqué précédemment, la mission recommande que l'Hadopi puisse être saisie en cas de différend relatif à l'ensemble de ces questions (conditions et durée de la conservation, communication...).

- **La loi ou le règlement doivent-ils définir les mesures que peuvent mettre en place les titulaires de droits pour assurer l'intégrité des infrastructures ?**

Pour rappel, l'article 3 paragraphe 3 de la directive autorise les titulaires de droits « 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* ». Il s'agit ici de mesures de protection du « contenant », qui se rajoutent aux mesures techniques de protection que peut prendre tout titulaires de droits (article 7 de la directive de 2019).

La mission considère qu'il revient à la loi de seulement poser le principe des mesures que peuvent mettre en œuvre les titulaires de droits, ainsi que la règle selon laquelle ces mesures doivent être « nécessaires » à l'objectif de protéger la sécurité et l'intégrité des réseaux et des bases de données où les œuvres ou autres objets protégés sont hébergés.

Un décret devrait en préciser les modalités d'application. La mission estime par ailleurs que le contenu de ces mesures, nécessairement évolutives, pourrait être précisé, à titre illustratif, par la charte des bonnes pratiques, comme suggéré au paragraphe 4 de l'article 3 de la directive.

➤ **La loi ou le règlement doivent-ils prévoir l'existence d'une charte entre les parties prenantes de la fouille de textes et de données pour mettre en œuvre les règles relatives à la conservation des copies et les mesures précitées ?**

Il est peu commun de prévoir dans la loi ou un règlement l'existence d'une charte, qui relève normalement de ce que d'aucuns appellent le « droit mou ». La charte est souvent considérée comme un type de convention infra-normatif, qui ne crée pas, au moins directement, d'obligation juridique entre les parties signataires¹¹⁷. Il existe toutefois des chartes dont les stipulations sont opposables¹¹⁸.

Dans la mesure où la directive encourage l'élaboration d'une charte des bonnes pratiques, la mission préconise de prévoir son existence dans la partie législative du CPI. Son contenu devrait être précisé dans la partie réglementaire. Il pourrait aussi y être renvoyé dans la partie législative pour en confier le suivi à l'Hadopi.

La charte devrait *a minima* porter sur les thèmes suivants :

1° procédure de vérification par le titulaire de droits de la qualité du chercheur qui demande une copie ou une reproduction et engagement de ce dernier, par un contrat écrit, à respecter la présente charte ;

Le titulaire de droits doit pouvoir s'assurer de l'identité du demandeur, et qu'il agit bien pour le compte d'un organisme de recherche ou d'une institution du patrimoine culturel habilités à effectuer des fouilles à des fins exclusives de recherche scientifique.

2° obligations pour le chercheur de ne pas porter atteinte au fonctionnement du site internet et des infrastructures du titulaire de droits, à son intégrité et à sa sécurité ;

Le chercheur ne doit pas utiliser un robot qui effectuerait un moissonnage des données portant préjudice au site du titulaire de droits.

3° modalités de conservation par les chercheurs des copies et des données, et mesures de sécurité envisagées pour éviter leur dissémination ;

Le titulaire de droits doit connaître en amont les modalités de conservation des copies et reproductions (par exemple, le chercheur les conserve-t-il lui-même, ou bien confie-t-il cette

¹¹⁷ V. A. Bensamoun, « Droit souple et lutte contre la contrefaçon sur internet », *Dalloz IP/IT*, avr. 2016, p. 182.

¹¹⁸ Ainsi l'article L 1111-1-1 du code général des collectivités territoriales reproduit-il *in extenso* la charte de l'élu local qui expose les principes déontologiques que doivent respecter tous les élus locaux.

Autre exemple : l'article L. 1222-9 du code du travail prévoit que « *le télétravail est mis en place dans le cadre d'un accord collectif, à défaut, dans le cadre d'une charte élaborée par l'employeur après avis du comité social et économique, s'il existe* ». Cet article énumère les 5 thèmes que doivent contenir l'accord collectif ou la charte (mise en œuvre du télétravail, fin du télétravail, plages horaires, contrôle du temps de travail par exemple).

Dernier exemple : le législateur a autorisé, à l'article L. 7342-9 du code du travail, les plateformes de mise en relation par voie électronique (il s'agit des activités de transport avec chauffeur et de livraison de marchandises au moyen d'un véhicule à deux ou trois roues) à établir unilatéralement « *une charte déterminant les conditions et modalités d'exercice de [leur] responsabilité sociale* » et définissant leurs droits et obligations ainsi que ceux des travailleurs avec lesquels elles sont en relation. Cet article liste les huit thèmes que doit comporter *a minima* la charte. De manière très originale, l'administration peut être saisie par la plateforme pour « *homologuer* » la charte, après consultation obligatoire des travailleurs indépendants. La charte est publiée sur le site internet de la plateforme et annexée aux contrats ou aux conditions générales d'utilisation qui la lient aux travailleurs.

mission à un tiers... ?), les mesures de sécurité prises, les moyens d'éviter leur dissémination et les moyens de vérifier sur pièce ou sur place que ces engagements sont respectés.

4° modalités de transmission éventuelle à d'autres chercheurs.

En tout état de cause, le pouvoir réglementaire doit exercer sa compétence et prévoir notamment les règles sur la conservation des copies et les mesures de protection du réseau hébergeant les œuvres protégées car la négociation de la charte peut ne pas aboutir : des règles supplétives réglementaires sont donc indispensables.

Beaucoup de personnes entendues par la mission ont considéré que le CSPLA pourrait être le lieu d'élaboration de cette charte, en réunissant l'ensemble des parties prenantes, leurs représentants et les ministères intéressés (le MESRI et le ministère de la Culture).

D. Les options spécifiques liées à la transposition de l'article 4

A titre liminaire, la mission indique que la qualification des résultats de la fouille n'a pas ici à être encadrée. La personne qui fouille doit pouvoir utiliser librement ces résultats tant qu'ils ne portent pas atteinte aux droits des titulaires.

➤ Faut-il prévoir une « exception » aux droits d'auteur et assimilés ou une « limitation » ?

Il est difficile de faire la distinction entre une exception et une limitation. En réalité, la terminologie est fluctuante et les dérogations au droit exclusif sont dénommées « exceptions », « limites », « limitations » ou encore « actes autorisés ».

En France, on renvoie traditionnellement aux exceptions. Mais il existe des situations où la paralysie du droit exclusif n'est pas absolue et donne par exemple naissance à une compensation – on parle alors parfois de licence légale au sujet de l'exception de copie privée. Lorsque le droit exclusif ne s'efface pas sans contrepartie ou définitivement (par exemple parce qu'il existe une faculté de revenir à la réservation), le terme de limitation est sans doute plus approprié. La directive 2001/29 et les traités administrés par l'OMPI (Organisation mondiale de la propriété intellectuelle) visent ensemble les exceptions et limitations, sans distinguer au cas par cas.

La CJUE a cependant opéré une distinction, dans l'arrêt *VG Wort*¹¹⁹, considérant que l'exception suppose que le droit exclusif est « *totalemt exclu* ». La limitation impliquerait au contraire que le droit exclusif puisse être sauvegardé. Aussi, « *il n'est pas exclu qu'une telle limitation puisse comporter, partiellement, en fonction des différentes situations particulières qu'elle régit, une exclusion, une restriction, voire le maintien dudit droit* » (pt 34). Même si les conséquences à tirer de ces considérations sémantiques ne sont pas claires, on peut estimer que **les dispositions de l'article 4, de par la possibilité de retour à la réservation, s'apparentent davantage à une limitation.**

¹¹⁹ CJUE, 27 juin 2013, aff. C-457/11 et C-458/11.

Cela étant, cette qualification n'a aucune portée en droit français et n'empêche pas, dans les articles dédiés, l'ajout de régimes spéciaux.

La mission propose donc de ne pas trancher la question et d'utiliser le terme d'« exception » au sens large, dans la mesure où aucune différenciation sémantique n'apparaît dans les articles dédiés du CPI.

➤ **Faut-il obliger le titulaire des droits à justifier son *opt-out* ?**

L'article 4 paragraphe 3 de la directive met en place une possibilité de retrait ou d'opposition (*opt-out*) qui permet au titulaire de réserver les œuvres et autres objets protégés. L'exercice du retrait aboutit à un retour au droit exclusif et donc à une possible contractualisation de la relation.

D'aucuns s'insurgent contre ce droit accordé au titulaire et qui réduirait à néant l'exception. Il faut sans doute relativiser cette crainte. Outre qu'il s'agit là sans doute du point d'équilibre de cette exception particulièrement permissive, il ressort des auditions que de nombreux ayants droit n'ont pas l'intention d'exercer cette option, soit parce qu'ils en ignorent l'existence, soit parce qu'ils estiment que les fouilles ne constituent pas un danger. L'*opt-out* généralisé se concentrera sans doute sur les ayants droit les plus exposés, qui seront ensuite amenés le cas échéant à contractualiser avec les acteurs de la fouille hors recherche. **L'exercice du droit de retrait n'aura donc pas le plus souvent comme objet d'interdire définitivement la fouille, mais plutôt de permettre au titulaire de droits de négocier dans des conditions équitables les modalités de cette fouille et la compensation financière attendue.**

En effet, contrairement à l'article 3 qui repose sur une certaine vertu des chercheurs, l'article 4 autorisera la fouille par des acteurs privés, pour des motifs commerciaux ou pour leur compte propre, de toutes les données pour lesquelles les ayants droit n'ont pas exercé leur droit de retrait et auquel l'accès est licite. On peut alors imaginer que les moissonnages risquent d'être plus massifs et les irrégularités plus difficiles à détecter (sauf dans le cas, très rare, de dépôt subséquent d'un brevet, qui oblige à dévoiler l'origine des données dans la demande de titre). Sans parler de la difficulté pour un titulaire de démontrer devant le juge une violation éventuelle de ses droits et l'existence d'un préjudice...

Une option 1 consisterait, pour éviter de freiner la fouille de textes et de données hors recherche, à imposer que l'exercice de l'*opt-out* soit motivé par des « raisons légitimes », comme la mission a pu l'entendre à l'occasion d'une audition. On pourrait en conséquence ici imaginer qu'une autorité indépendante puisse vérifier la légitimité de l'*opt-out*. Cette proposition semble à la mission impossible à soutenir dans la mesure où elle modifierait profondément l'équilibre du texte, voulu par le législateur européen, et qu'elle ajouterait une condition, restrictive et non anticipée, au texte d'une directive qui se veut d'harmonisation.

L'option 2 invite en conséquence à une réponse négative. Outre les arguments déjà développés, on peut ajouter que l'article 4 de la directive constitue déjà une inversion de la logique même des droits protégés par le code de la propriété intellectuelle (le titulaire doit faire une démarche pour accéder à l'exclusivité).

La mission recommande en conséquence de suivre l'option 2. Il semble à la mission que l'*opt-out* s'apparente à un **droit discrétionnaire**, dont le titulaire n'a pas à justifier l'exercice, lequel ne peut donc être contrôlé.

Dans ce cadre, il importe de rappeler que cet exercice ne saurait entraîner de conséquences négatives pour le titulaire de droits qui l'exerce¹²⁰. Notamment, sauf à violer des règles de droit de la concurrence notamment, **le bénéfice de certains services ne saurait être conditionné au non-exercice du retrait**.

➤ **L'exercice du droit de retrait est-il encadré dans le temps ?**

Dans le prolongement de ce qui précède, il ne serait pas légitime d'encadrer dans le temps l'exercice de l'*opt-out*, qui doit rester ouvert à tout moment au titulaire de droits : celui-ci peut s'opposer à la fouille dès la transposition de l'article 4 de la directive, un an ou cinq ans plus tard.

En outre, il a le droit de changer d'avis. Une solution inverse risquerait à nouveau de crispier les volontés et d'engager vers un *opt-out* massif, par précaution.

Cependant, cette liberté de changer d'avis nécessite quelques éclaircissements. *Quid* en cas d'exercice de l'*opt-out* postérieurement à la collecte de données en vue d'effectuer une fouille ?

Une première option consiste à retenir le moment de la collecte pour apprécier la licéité de l'accès. Si, à un temps *t*, le titulaire ne s'est pas opposé à la fouille, alors l'entreprise peut collecter les données et utiliser le contenu pour l'opération projetée. En revanche, si le titulaire s'y est opposé de manière appropriée, alors l'entreprise ne peut pas collecter les données. Cette option impose donc la prise en compte de la date du retrait.

Une deuxième option consisterait à obliger l'entreprise à vérifier tout au long du processus de fouille si le titulaire n'a pas exercé son *opt-out* dans les semaines ou mois qui suivent la collecte et tant que dure l'opération. Cette option est sans doute plus protectrice des intérêts des titulaires de droits mais elle manque de pragmatisme et compliquerait sérieusement les opérations de fouille. En outre, elle impliquerait une certaine rétroactivité, gênante, de l'option de retrait.

La mission propose donc de retenir la première option : **la réservation induite par l'*opt-out* – sachant que celui-ci peut être exercé à tout moment – ne peut être prise en compte qu'à compter de son exercice**.

➤ **Que faire en cas de contradiction d'information au sujet de l'*opt-out* ?**

Il existe deux situations dans lesquelles les informations sur cet *opt-out* peuvent être incohérentes.

La première situation concerne l'**articulation entre le « contenant » et le « contenu »**. On peut en effet imaginer qu'un site interdise la fouille des contenus hébergés alors qu'un contenu est marqué comme pouvant faire l'objet d'une telle opération, ou l'inverse. Cette situation

¹²⁰ Relevant déjà ce risque, Rapport CSPLA, « *IA et culture* » précit., p. 58.

concerne notamment les sites qui hébergent des contributeurs indépendants juridiquement (écrivains, musiciens, photographes...). Dans cette hypothèse, il faut sans doute considérer que le retrait exercé sur le contenant vaut blocage et il doit s'appliquer à l'ensemble des contenus. En revanche, dans l'hypothèse où le contenant manifeste une autorisation globale, il faudrait prendre en compte l'éventuel retrait opéré sur un ou plusieurs objets protégés hébergés sur ce site.

Une seconde situation peut concerner **les contenus eux-mêmes**. On peut en effet imaginer que le même contenu soit disponible en ligne sur plusieurs sites et que certains fichiers comprennent une option de retrait pendant que d'autres l'ignorent. Ici, la mission considère que, sauf à ce que l'erreur provienne du titulaire lui-même, il est fort probable que les contenus « ouverts » à la fouille soient accessibles de manière illicite au sens de la directive.

➤ **Faut-il prévoir que les acteurs de la fouille fassent des recherches diligentes sur l'exercice éventuel de l'*opt-out* ?**

Certains titulaires demandent à ce que les acteurs de la fouille réalisent des « recherches diligentes » sur l'exercice ou non du retrait. L'expression est empruntée à la directive 2012/28/UE du 25 octobre 2012 sur certaines autorisations autorisées des œuvres orphelines. L'œuvre orpheline y est définie par l'absence d'identification du titulaire de droits en dépit des recherches diligentes.

Une option 1 accèderait à cette demande et imposerait donc que les bénéficiaires potentiels de l'exception réalisent des recherches diligentes au sujet de l'exercice du retrait. Cette option est très protectrice des intérêts des titulaires de droits mais elle serait très difficile à mettre en œuvre. On le sait, dans la directive de 2012, les modalités de ce qui constitue une recherche diligente sont clairement définies (art. 3). D'ailleurs, le CPI ¹²¹ prévoit à l'article L. 135-7 l'adoption d'un décret en Conseil d'Etat pour définir les modalités d'application du régime, et notamment lister les sources d'informations à consulter pour chaque catégorie d'œuvres. Il faudrait alors imaginer qu'un tel décret renvoie vers des sources pertinentes où l'information de l'*opt-out* serait consignée. Une telle option, lourde à mettre en place, permettrait cependant de normaliser l'exercice de l'*opt-out*.

Une option 2 consisterait à ne pas imposer de recherches diligentes aux acteurs de la fouille, en considérant que ce serait ajouter une règle *ex nihilo* à un dispositif obligatoire dans une directive d'harmonisation.

La mission considère qu'il est plus raisonnable de ne pas ajouter une telle condition au risque de sur-interpréter le texte et de réduire la portée de l'article 4 de la directive. Pour autant, il importe ici de rappeler, d'une part, que l'exercice de l'*opt-out* par un moyen approprié ne saurait être ignoré et, d'autre part, que l'approvisionnement sur des sites douteux ne permet pas de prétendre à un « accès licite ».

¹²¹ La directive a été transposée par la loi n° 2015-195 du 20 février 2015, portant diverses dispositions d'adaptation au droit de l'Union européenne dans les domaines de la propriété littéraire et artistique et du patrimoine culturel, aux articles L. 135-1 à -7 du CPI, et par renvoi à ces dispositions à l'article L. 211-7 pour les droits voisins.

➤ Comment exprimer le droit de retrait ?

L'article 4 paragraphe 3 apporte des précisions sur les modalités d'exercice de l'*opt-out*. Ainsi la réservation par les titulaires de droits doit être « *expresse* » (plus exactement se faire « *expressément* ») et s'opérer « *de manière appropriée, notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne.* » L'utilisation de l'adverbe « *notamment* » (*such as* dans la version anglaise) suggère une certaine ouverture. Est-ce à dire que la réservation pourrait s'opérer autrement que par des procédés lisibles par la machine ?

Le considérant 18 peut aider à comprendre l'objectif législatif sur cette question. Ainsi, lorsque le contenu est mis à la disposition en ligne, la réservation est « *appropriée* » seulement « *si elle est effectuée au moyen de procédés lisibles par machine, y compris des métadonnées et les conditions générales d'utilisation d'un site internet ou d'un service* ». Ainsi, seul un procédé lisible automatiquement permet une réservation efficace. Le considérant poursuit en évoquant « *d'autres cas* ». L'expression manque de clarté. La traduction correcte de l'expression anglaise « *in other cases* » aurait sans doute dû être « dans les autres cas », c'est-à-dire les cas où le contenu n'est pas mis à la disposition en ligne. Dans ces hypothèses donc, « *il peut être approprié de réserver les droits par d'autres moyens, comme des accords contractuels ou une déclaration unilatérale.* »

Une lecture attentive du texte permet donc de comprendre les modalités de retrait :

- Lorsque le contenu est accessible en ligne, le procédé, quel qu'il soit (métadonnées, CGU...), doit être lisible par la machine ;
- Lorsque la mise à disposition se fait autrement, l'exigence n'est pas requise.

S'agissant de la **lisibilité par la machine**, le risque existe cependant que les ayants droit, dans leurs CGU notamment, utilisent des formules très variées et peu précises juridiquement, ce qui les fragiliserait ensuite en cas de différend avec des fouilleurs privés qui pourraient plaider leur bonne foi. La mission estime que l'élaboration de stipulations standardisées au niveau européen serait le moyen le plus approprié pour exprimer l'opposition des ayants droit aux fouilles effectuées sur le fondement de l'article 4 de la directive. Toutefois, cette hypothèse paraît envisageable seulement à moyen ou long terme. A court terme, la mission invite le pouvoir réglementaire à renvoyer, dans son décret, à un simple **arrêté ministériel** le soin de **définir un certain nombre de formules standardisées** à disposition des ayants droit pour exprimer leur droit d'opposition. Ces formules pourraient être des stipulations rédactionnelles à insérer dans les CGU par exemple, ou des syntaxes simples de type « CGU_TDM : NO » aisément lisibles par des robots. Il conviendrait aussi pour l'ayant droit de préciser la portée de son opposition. Couvre-t-elle toutes ses œuvres ou seulement une partie ?

Le recours à de telles formules standardisées aurait le mérite de sécuriser juridiquement à la fois les ayants droit et les « fouilleurs » privés : il serait impossible pour ces derniers de plaider qu'ils n'avaient pas connaissance de l'opposition des titulaires de droit.

La mission ne propose pas de charger les parties prenantes de définir *ab initio* dans la charte ces formules standardisées, car il n'est pas certain que sa négociation aboutisse dans les meilleurs délais. En revanche, à la lumière des retours d'expérience, la charte pourrait proposer

d'adapter les formules standardisées mentionnées dans l'arrêté ministériel. En tout état de cause, la lisibilité par la machine doit, au sens de la directive, être entendue largement. Comme l'indique le texte européen, les **CGU en format lisible par machine** (le HTML, par exemple) peuvent constituer un **moyen approprié d'exprimer le droit d'opposition du titulaire de droits**.

Cette lisibilité, lorsque le contenu est mis à disposition en ligne, est la seule exigence imposée. Il incombe donc au « fouilleur » de s'assurer qu'aucun droit d'opposition n'a été exprimé, ce qui implique qu'il paramètre le cas échéant son robot lorsqu'il moissonne un site.

Le titulaire reste donc libre d'exercer ou pas son droit de retrait et des moyens mis en œuvre du moment qu'ils sont appropriés.

La mission recommande de préciser ces points dans la partie **réglementaire** de la transposition.

- **Faut-il énumérer les moyens techniques par lesquels les titulaires de droits peuvent exprimer leur droit de retrait ?**

Comme indiqué, le titulaire des droits est libre de choisir le moyen approprié pour exercer son *opt-out*. On pourrait cependant vouloir énumérer les moyens mis à sa disposition.

Une première option pourrait consister à ne pas mentionner les modalités de retrait. Au vu des exigences distinctes posées par le texte européen, la proposition pourrait représenter une insécurité juridique.

Une deuxième option viserait, tout en assurant une certaine sécurité juridique, à laisser de la souplesse, en procédant par l'exemple.

Une troisième et dernière option assumerait d'énoncer l'ensemble des dispositifs efficaces permettant de réaliser un retrait.

La mission recommande de privilégier la deuxième option, qui a **l'avantage de la pédagogie et aussi d'un certain niveau de sécurité juridique, tout en privilégiant l'évolutivité indispensable en lien avec la technique**. La mission propose en conséquence d'illustrer ces moyens dans la partie **réglementaire** et dans la **charte**.

- **Quels sont les avantages et inconvénients des différents moyens à la disposition des titulaires de droits pour exercer l'*opt-out* sur les contenus en ligne ?**

Pour rappel, s'agissant des contenus non mis à la disposition en ligne, tout moyen peut être utilisé. Le texte européen mentionne en ce sens les accords contractuels, voire une déclaration unilatérale. S'agissant des contenus mis à la disposition en ligne, l'*opt-out* doit s'exercer par des dispositifs techniques lisibles par la machine.

On peut distinguer ici trois types de technologies permettant d'exercer une opposition :

- Celles qui protègent les « contenants », type métadonnées, « dômes » de protection ou API ;

- Celles qui protègent les « contenus », type métadonnées, DRM ou marquages ;
- Celles qui sont centralisées dans un fichier.

Forte de son expertise technique, l'Hadopi a parfaitement répertorié ces solutions, en mettant en exergue les avantages et les inconvénients de chacune. La mission reproduit ci-après les tableaux récapitulatifs communiqués par la Haute autorité.

Modalités techniques de protection des contenants

Solutions techniques existantes	Avantages	Inconvénients
<p>Déclaration (au moyen d'un paramétrage technique)</p> <p>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.</p>	<p>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.</p>	<p>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.</p>
<p>« Dôme » de protection contre les intrusions</p> <p>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.</p>	<p>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.</p>	<p>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).</p>
<p>Accès pour la fouille via une API (interface de programmation applicative)</p> <p>L'API permet d'identifier le demandeur et de gérer les accès aux contenus du système (exemple : l'API de YouTube)</p>	<p>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.</p>	<p>Cela revient à conditionner ce qui relève en théorie d'une exception (notamment pour la recherche scientifique).</p>

Modalités techniques de protection des contenus

Solutions techniques existantes	Avantages	Inconvénients
<p>DRM, au niveau du contenu</p> <p>Un système de gestion des droits empêche la fouille ou indique que cette pratique n'est pas autorisée pour le contenu en question.</p>	<p>Gestion fine contenu par contenu.</p>	<p>En pratique, les 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.</p>
<p>Métadonnées, au niveau du contenu</p> <p>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.</p>	<p>Ce système est assez simple et léger à mettre en place.</p>	<p>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.</p>
<p>Marquage en dur (watermarking), au niveau du contenu</p> <p>Un marqueur intégré au sein même d'un contenu. Un outil de détection de marquage (ou de tatouage) numérique permet de lire cette indication.</p>	<p>Ce système est relativement proche de celui des métadonnées, mais est plus résistant.</p>	<p>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.</p>

Modalités de protection universelles et centralisées

Solutions techniques envisagée	Avantages	Inconvénients
<p>Fichier centralisé</p> <p>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.</p>	<p>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 référer. 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.</p>	<p>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.</p>
<p>Fichier centralisé utilisé en lien avec les outils de reconnaissance de contenus</p> <p>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.</p>	<p>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 protégées¹²²). Il faudrait y ajouter des informations concernant un éventuel <i>opt-out</i>.</p> <p>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.</p> <p>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).</p>	<p>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.</p> <p>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.</p>

¹²² 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 des organismes de gestion collective ou par des consortiums.

En plus de ces éléments, les auditions de la mission ont montré que :

- Les conditions générales d'utilisation (CGU) et les déclarations unilatérales des titulaires de droits sont rarement lues par les personnes réalisant de la fouille de textes et de données en ligne, y compris des chercheurs académiques ;
- Les métadonnées ne sont pas généralisées dans tous les secteurs et elles ne constituent pas un obstacle parfaitement efficace, en tout état de cause, à la fouille de textes et de données ; en outre, des spécialistes de fouille ont indiqué à la mission que beaucoup de formats n'incluent pas de métadonnées sur le droit d'auteur ;
- Le recours au contrat pourra être très utile pour que les titulaires de droit et les sociétés recourant à la fouille de textes et de données hors recherche se partagent la valeur ajoutée liée à cette activité.

La mission voudrait également porter à la connaissance du CSPLA, à titre d'exemple, une solution technique développée par le CFC permettant de contrôler les robots moissonneurs.

L'exemple de la solution ORRC

De nombreux éditeurs ont constaté ces dernières années le développement de robots qui « moissonnent » massivement leurs données sur leurs sites web. Selon les éditeurs, plus de la moitié de la bande passante du site internet de certains journaux français est actuellement accaparée par des robots. Selon le CFC, les robots représentent entre 30 et 90% des visites des sites éditeurs et entre 20 à 50% des hits serveur. Un grand nombre de ces robots agissent pour le compte de « prestataires de veille web » qui proposent à leurs clients des services comme des revues de presse, des études sur l'image de leurs clients dans les médias et les réseaux sociaux, ou encore des analyses de tendances commerciales, en fournissant des hyperliens consacrés à des domaines et thèmes présélectionnés. Face à cette situation, qui rompt radicalement avec les pratiques de revue de presse papier bien encadrée juridiquement, des éditeurs ont souhaité disposer d'une solution de gestion permettant l'exercice effectif des droits attachés à leurs publications et ils ont confié depuis la fin 2017 au CFC la gestion de ces droits dans le cadre de l'utilisation de ces publications par des tiers. Le CFC, agissant au nom des éditeurs qui l'ont mandaté, délivre ainsi par contrat aux « prestataires de veille web » les autorisations de reproduction et de représentation dont ils ont besoin, en application de l'article L. 122-4 du CPI. En octobre 2020, 226 éditeurs ont passé des contrats d'apport volontaire en gérance des droits pour la veille Web, couvrant ainsi 637 sites internet de presse français.

Afin de contrôler l'application de ces contrats, le CFC propose aux éditeurs signataires une prestation payante dénommée ORRC (Outil de Régulation des Robots de Crawling). Cet outil a pour but d'identifier tous les robots qui aspirent les données des sites internet des éditeurs qui ont mandaté le CFC pour gérer leurs droits en matière de veille web et de bloquer en conséquence ceux qui n'ont pas d'autorisation. Les prestataires de veille web sont ainsi fortement incités à conclure des contrats avec le CFC quand ils constatent que l'accès de leurs robots aux sites des éditeurs est bloqué par l'ORCC.

Concrètement, l'outil ORRC s'appuie sur 2 éléments du marché :

- Le premier élément est un bot mitigation, qui identifie tous les robots qui représentent un danger en terme de sécurité. Le bot mitigation choisi par le CFC est développé depuis 12 ans et il est utilisé par 200 000 sites à travers le monde, ce qui lui permet d'avoir une liste très étoffée et actualisée des robots dangereux. Selon le CFC, 97% des robots identifiés par le bot mitigation ne posent pas de difficultés en termes de sécurité informatique (« good robots ») ;
- Le second élément est un pare feu qui est paramétré sur mesure par le CFC pour tenir compte des contrats d'autorisation signés avec les prestataires de veille web et, le cas échéant, des demandes spécifiques des éditeurs. Ainsi, le CFC utilise actuellement 80 paramètres, comme l'adresse IP, le nom du *user agent*, la localisation du robot ou encore son analyse comportementale (par exemple, afin de déjouer l'attention des éditeurs, certains robots agissent la nuit, en privilégiant le moissonnage de quelques données peu sensibles, avant d'aspirer massivement les données recherchées). C'est ce pare-feu qui gère *in fine* le passage de robots, et le cas échéant les bloque.

- Le *bot mitigation* et le pare-feu retenus par le CFC sont installés sur un CDN (*content delivery network*.¹²³) de 40 « pops » dans le monde, ce qui permet de ne pas ralentir le trafic sur les sites des éditeurs.

À partir des éléments donnés par ce bot mitigation, l'outil ORRC permet de repérer immédiatement les robots déjà connus ou présentant une immatriculation *ad hoc*, ainsi que les robots non encore identifiés, d'étudier les caractéristiques et le comportement de ces derniers et d'enregistrer, le cas échéant, les données obtenues par des recherches complémentaires, en particulier pour identifier la personne morale qui exploite chaque robot (que le CFC appelle « infomédiaire »). Lorsque les infomédiaires sont identifiés et selon les informations trouvées, les équipes du CFC qualifient chacun selon les services qu'il propose ou selon l'usage constaté des données (analyse d'audience, veille web, flux RSS avec services payants, moteur de recherche, etc.). Cette qualification est partagée avec l'éditeur de chaque site sur lequel le robot est identifié (un robot identifié sur un premier site n'a plus besoin d'être analysé lorsqu'il est repéré sur un autre site de presse). L'outil ORRC permet enfin au CFC ou aux éditeurs directement, selon la qualification du robot identifié, de paramétrer le pare-feu pour autoriser le robot à passer, le gêner (lui autoriser le passage mais dans des limites strictes, par exemple en termes de fréquence ou de hits) ou le bloquer. Les robots non identifiés peuvent être dirigés sur une page web spécifique, afin d'obliger leur infomédiaire à s'identifier.

Les éditeurs peuvent confier au CFC la gestion complète et en direct de leur flux de robots. Les éditeurs ayant déjà un CDN peuvent recourir à ORRC en parallèle de leur gestion de flux (lorsqu'une page est demandée par un robot, un script est activé, et le trafic est analysé en direct) ou en différé (les logs sont régulièrement intégrés et analysés, des recommandations et propositions de blocage sont envoyées à l'éditeur).

D'un point de vue technique, l'outil ORRC semble remplir les missions qui lui ont été assignées. Le CFC peut techniquement bloquer l'accès à un site d'éditeur à n'importe quel robot, qu'il s'agisse d'un robot agissant pour le compte d'un moteur de recherche, d'une personne publique ou d'un acteur privé.

L'outil ORRC est toutefois confronté à deux difficultés :

- 1) sa masse critique, d'une part. Cet outil ne pourra durablement discipliner les entités qui utilisent des robots pour moissonner les sites des éditeurs que si ces derniers recourent majoritairement à la prestation ORRC.
- 2) l'absence d'obligation juridique pour les robots de s'identifier, d'autre part. Certains robots annoncent leurs identités, mais d'autres présentent de fausses identités, tandis qu'un grand nombre agissent *incognito*. C'est pourquoi le CFC attribue aux prestataires de veille web qui contractent avec lui un numéro d'immatriculation que doivent présenter les robots agissant pour leurs comptes.

La mission estime que parmi les différents canaux mis à disposition des titulaires de droits pour exercer leur opposition à la fouille de textes et de données hors recherche sur le fondement de l'article 4 de la directive de 2019, un outil inspiré de l'ORRC présente le plus de garanties pour rendre effectif ce droit d'opposition. Un tel outil serait en outre compatible avec l'exception académique posée par l'article 3 de la directive, dans la mesure où les bénéficiaires devraient prouver, avant la collecte de données, leur identité aux titulaires de droits.

➤ Les moyens pour exercer l'*opt-out* peuvent-ils être cumulatifs ?

Dans la mesure où aucun moyen n'est par lui-même infallible, on peut soutenir un possible cumul des moyens réalisant l'*opt-out* (sous réserve de cohérence). Cependant, cette ouverture,

¹²³ Il s'agit d'une infrastructure technique optimisée pour délivrer le plus rapidement possible les contenus internet : il comprend le serveur d'origine et des serveurs périphériques implantés dans plusieurs endroits différents dans le monde.

multipliant les canaux d'expression des titulaires, est aussi susceptible de multiplier les hypothèses de contradiction.

Selon la mission, la seule limite imposée par le texte européen à l'expression du retrait est sa lisibilité par la machine lorsque le contenu est en ligne. Aussi, les titulaires de droits peuvent cumuler les moyens, à charge pour eux d'assurer la cohérence de l'information.

➤ **Comment assurer l'articulation entre l'article 4, imposant une possibilité d'*opt-out*, et l'article 3 ?**

Le choix de certains procédés techniques d'*opt-out* pourrait faire obstacle à la fouille de données à des fins de recherche de l'article 3. Or, l'article 4, paragraphe 4, précise bien que : « *Le présent article n'affecte pas l'application de l'article 3 de la présente directive* ».

Dans ce cadre, l'Hadopi propose de distinguer entre les procédés dits « *actifs* » (dôme, MTP, API), qui s'appliquent indistinctement quelle que soit la finalité de la fouille, et les procédés dits « *passifs* » ou méthodes informatives (fichier centralisé, déclaration, métadonnées...), qui peuvent tout simplement être ignorées par le chercheur.

Est-ce à dire que les procédés actifs ne devraient pas être utilisés pour ne pas empêcher la fouille académique ? Le texte ne se prononce pas en ce sens. La mission considère même qu'une telle solution serait en définitive *contra legem* en ce qu'elle empêcherait la mise en place de toute mesure efficace de lutte contre la contrefaçon en ligne.

La mission considère donc que le texte européen n'interdit pas d'exiger du chercheur qu'il s'identifie auprès du titulaire de droits afin que celui-ci mette en place les conditions d'une fouille sécurisée.

➤ **Quid de la conservation des corpus de fouille ?**

S'agissant d'abord de la **durée de conservation** des copies, l'article 4, paragraphe 2, de la directive renvoie à une conservation « *aussi longtemps que nécessaire aux fins de la fouille de de textes et de données* » (v. aussi cons. 18, al. 1^{er}, *in fine*). Comment interpréter ces dispositions ? Là aussi, deux options peuvent être distinguées.

La première option consiste à avoir une interprétation large de ces dispositions, en considérant que les finalités de la fouille incluent *implicitement mais nécessairement* la vérification de ses résultats. Autrement dit, les chercheurs travaillant dans des entreprises pourraient conserver, comme les chercheurs « académiques », les reproductions des œuvres protégées pendant plusieurs années pour vérifier le résultat des fouilles.

La seconde option retient une interprétation stricte des dispositions en cause, en soulignant que le législateur européen a retenu à dessein une rédaction différente aux articles 3 et 4 de la directive s'agissant de la durée de conservation des reproductions. En effet, il n'est pas évoqué la nécessité de conservation au-delà de l'opération de fouille elle-même, alors que c'est le cas pour l'exception académique mentionnée à l'article 3 qui prévoit une conservation « *à des fins de recherche scientifique, y compris pour la vérification des résultats de la recherche* ». Sans

doute cette nécessité sera-t-elle également avérée en cas de recherche privée, notamment dans le secteur médical. Mais le caractère indéfini de cette durée fait craindre chez les titulaires de droits une dissémination des contenus. D'autant qu'en pratique, les chercheurs d'entreprises privées contractualisent la relation avec les éditeurs notamment et la licence permettra dès lors de régler cette question de la conservation (durée et modalités).

C'est pourquoi la mission préconise de retenir la seconde option : **les reproductions effectuées dans le cadre d'une fouille relevant de l'article 4 de la directive doivent être détruites au plus tard lors de la production des résultats de la fouille.**

A défaut, le « fouilleur » n'est plus excusé par l'exception et le corpus conservé est contrefaisant. Pour autant, la production du résultat peut nécessiter une certaine durée de conservation (par exemple, une fouille qui s'étendrait sur plusieurs mois voire années). Notamment, les systèmes de fouilles suivent parfois un cycle d'amélioration continu nécessitant de nouvelles analyses du contenu d'origine.

S'agissant ensuite des **conditions de la conservation**, alors que l'article 3, paragraphe 3, prévoit que les corpus doivent être stockés « *avec un niveau de sécurité approprié* » et peuvent être conservés, y compris pour vérifier les résultats (tout en précisant les conditions dans les considérants), l'article 4, paragraphe 2, se contente d'indiquer que les données reproduites ou extraites « *peuvent être conservées aussi longtemps que nécessaire aux fins de la fouille de textes et de données* ».

On pourrait, dans une option 1, estimer, eu égard au silence du législateur européen, que la conservation des reproductions n'a pas à être sécurisée. Cette approche n'est toutefois pas exempte de critiques. D'une part, elle contribuerait à un exercice systématique de l'*opt-out*, ce qui n'est pas souhaitable. D'autre part, on voit mal pourquoi les exigences seraient plus poussées pour les organismes de recherche que pour les acteurs privés.

C'est pourquoi une option 2 consisterait à rapprocher les deux régimes sur ce point. On pourrait objecter que l'exigence n'est pas imposée dans le texte. Pour autant, là encore, on comprendrait difficilement ce qui pourrait justifier une telle différence de traitement, alors que la directive garantit, comme toutes les directives en matière de droit d'auteur et de droits voisins, un « *niveau élevé de protection aux titulaires de droits* » (cons. 2).

Ainsi, même si la durée supposée de conservation est limitée à la production du résultat, il importe d'imposer une **sécurisation de la conservation des copies et reproductions des contenus protégés**.

- **Faut-il étendre à l'article 4 le paragraphe 3 de l'article 3, relatif aux 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 ?**

Une réponse positive pourrait sembler logique. En effet, il serait peu compréhensible que les titulaires de droits disposent de cette faculté dans le cadre de fouilles académiques, mais pas lors de fouilles hors recherche, alors même que les secondes risquent d'être plus massives et donc plus agressives pour les infrastructures que les premières.

On pourrait cependant considérer, à l'inverse, que l'omission de cette règle à l'article 4 caractérise une volonté claire et univoque du législateur européen, dont la méconnaissance pourrait être sanctionnée par le juge.

La mission considère toutefois que ce débat n'a pas lieu d'être, dès lors qu'en pratique, l'exercice du droit de retrait peut avoir les mêmes effets que les mesures de protection de l'intégrité du réseau mentionné à l'article 3 de la directive.

➤ **Faut-il étendre à l'article 4 de la directive la possibilité de négocier une charte des bonnes pratiques entre les titulaires de droits et les acteurs de la fouille de textes et de données commerciale ?**

La directive prévoit, dans le cadre de l'exception académique, l'élaboration d'une charte des bonnes pratiques au sujet de la mise en œuvre des mesures techniques de protection des réseaux et des bases hébergeant les contenus protégés ainsi que des conditions de stockage et de conservation des corpus. Aucune mesure équivalente n'est prévue dans le texte pour l'article 4.

On pourrait considérer, dans une première option, que le silence du texte européen vaut exclusion.

Pour autant, on peut aussi imaginer, dans une seconde option, qu'un tel instrument, du moment qu'il ne restreint pas les conditions de l'exception, pourrait au contraire utilement servir à rassurer les parties prenantes et ainsi à rendre l'exception de l'article 4 plus effective. En effet, l'absence d'encadrement pourrait entraîner une méfiance et un *opt-out* « par défaut ». Aussi, toute démarche prompte à faciliter une mise en œuvre responsable de l'exception devrait être encouragée.

Par ailleurs, il faut se souvenir que les deux exceptions de fouille doivent s'articuler puisque le paragraphe 4 de l'article 4 précise que cette disposition « *n'affecte pas l'application de l'article 3* ». Dans ces conditions, la charte pourrait prévoir des modalités pratiques d'articulation entre les deux exceptions.

Cependant, compte tenu de la diversité des fouilles, de l'absence de coïncidence entre les acteurs des fouilles relevant de l'article 3 et de ceux effectuant des fouilles dans le cadre de l'article 4, et de la difficulté à identifier clairement les représentants institutionnels de la fouille relevant de l'article 4, la mission considère qu'il n'est **pas pertinent d'étendre le champ des négociateurs de la charte**. En revanche, il serait intéressant d'inviter les acteurs de l'article 4 à s'associer, en tant qu'observateurs, aux travaux de la charte, ne serait-ce que pour établir les modalités d'articulation entre les deux exceptions.

Enfin, à titre informatif, la mission ajoute que l'exception vise, évidemment, la relation entre un bénéficiaire et le titulaire du droit de reproduction, qui sera souvent un cessionnaire. Aussi, il serait pertinent d'envisager la question de la rémunération des titulaires originaires dans pareille hypothèse, en fonction le cas échéant de l'exercice du droit de retrait et de l'instauration d'une licence. Cette question, qui sort du cadre de la présente mission, est en lien avec celle, plus générale et centrale, de la rémunération des titulaires originaires dans l'environnement numérique.

IV. Propositions de la mission

Avant de formuler des propositions textuelles (B) visant à transposer les articles 3 et 4 de la directive 2019/790, la mission soumet quelques préconisations générales (A).

A. Préconisations générales

Au vu de ce qui précède, la mission préconise :

- d'abroger les dispositions législatives issues de la loi sur la République numérique car elles sont contraires au droit européen, antérieur et postérieur à la directive de 2019 ;
- d'intégrer dans les articles dédiés aux exceptions (principalement) les textes de transposition des articles 3 et 4 ;
- d'encourager dès maintenant les titulaires de droits et les acteurs de la fouille de textes et de données, académique et hors recherche, à se rencontrer pour réfléchir, sur le fondement du présent rapport, à une charte des bonnes pratiques qui pourrait être rédigée sous la supervision du CSPLA et dont le suivi pourrait revenir à l'Hadopi ;
- de garder à l'esprit l'objectif de maintenir un équilibre entre la protection des titulaires de droits et les intérêts des acteurs de la fouille de textes et de données.

Les préconisations formulées par la mission ont vocation à être déclinées à **trois niveaux normatifs différents** :

- dans la **loi** pour la transposition des exceptions elles-mêmes (principe de l'exception et mention des bénéficiaires pour les articles 3 et 4 ; caractère d'ordre public, principe de la conservation des reproductions, faculté de prendre des mesures de protection du réseau pour l'article 3 ; absence d'incidence de l'*opt-out* sur le bénéficiaire de l'exception académique pour l'article 4) et pour les nouvelles missions confiées à l'Hadopi qui en découlent (régulation des mesures techniques de protection mises en œuvre par les titulaires de droits comme pour toutes les exceptions, procédure de règlement des différends après conciliation, suivi de la mise en œuvre de la charte) ;
- dans un **décret en Conseil d'Etat** pour préciser toutes les conditions d'application des dispositions législatives mentionnées à l'article L. 122-5 du CPI (par exemple, définition de la fouille de textes et de données et de l'accès licite ; définition précise des bénéficiaires de l'exception « académique », ainsi que modalités pratiques de conservation – et éventuellement de transmission à d'autres chercheurs – des copies ; précisions des modalités pour exprimer l'*opt-out* des titulaires de droits pour l'exception de l'article 4)¹²⁴ ;

¹²⁴ Les articles L. 122-6-1 du CPI sur les exceptions au droit d'auteur des logiciels, L. 211-3 sur les exceptions aux droits voisins et L. 342-3 sur les exceptions au droit *sui generis* des bases de données ne prévoient pas un décret en Conseil d'Etat. La mission n'a pas proposé de prévoir un décret en Conseil d'Etat pour préciser les conditions d'application des exceptions relatives à la fouille de textes et de données car une telle mesure aurait créé une inégalité de traitement entre les différentes exceptions actuelles. En tout état de cause, le Gouvernement devra

- dans une **charte des bonnes pratiques**, négociée par les parties intéressées sous la supervision du CSPLA, pour illustrer les conditions pratiques et techniques des fouilles.

Pour des raisons évidentes de délai, la mission a concentré son attention sur la transposition de la directive dans la partie législative du CPI.

➤ Quels sont les principaux articles du CPI qu'il convient de modifier ?

Malgré le fait que l'article **L. 122-5** est déjà très long et que sa lecture est parfois mal aisée, la mission propose d'y inscrire les nouvelles exceptions issues des articles 3 et 4, dans une nouvelle rédaction de son 10°. Le premier avantage de cette option est de supprimer les dispositions issues de la loi pour une République numérique, qui sont incompatibles avec la directive de 2019. Un autre avantage est que les deux derniers alinéas de L. 122-5 du CPI s'appliqueraient automatiquement à la nouvelle exception :

1) ces nouvelles exceptions ne doivent pas « *porter atteinte à l'exploitation normale de l'œuvre ni causer un préjudice injustifié aux intérêts légitimes de l'auteur* ». Ce faisant, la transposition respecterait l'article 7 de la directive de 2019, qui renvoie à l'article 5, paragraphe 5, de la directive de 2001 prévoyant justement l'exigence du respect du triple test ;

2) un décret en Conseil d'Etat, qui implique une vigilance renforcée du ministère par rapport à un décret simple, est obligatoire¹²⁵.

Un dernier avantage est d'offrir à l'utilisateur du code toutes les exceptions au droit d'auteur au sein d'un seul et même article.

Les deux exceptions devraient encore être ajoutées à l'article **L. 211-3**, qui concentrent les exceptions communes à tous les droit voisins, dont les droits des éditeurs de presse.

Dans le même ordre d'idée, la mission suggère de décliner ces deux exceptions en proposant une nouvelle rédaction du 5° de l'article **L. 342-3**, relatif aux exceptions aux droits des producteurs de bases de données, afin de supprimer les dispositions issues de la loi pour la République numérique¹²⁶ et les remplacer par les hypothèses nouvelles.

La mission recommande encore de modifier l'article **L. 122-6-1** en matière de logiciels, pour y intégrer uniquement l'exception ouverte à l'article 4 de la directive.

Enfin, dans la mesure où le bénéfice de l'exception doit être garanti, la mission préconise d'étendre le champ de compétence de l'Hadopi sur ce point et donc de modifier l'article **L. 331-31** du CPI.

décliner pour tous les articles précités les principes retenus dans le décret en Conseil d'Etat relatif à l'article L. 122-5 du CPI.

¹²⁵ L'article L. 122-5 *in fine* prévoit en effet : « *Les modalités d'application du présent article, notamment les caractéristiques et les conditions de distribution des documents mentionnés au d du 3°, sont précisées par décret en Conseil d'Etat.* »

¹²⁶ Comme à l'article L. 122-5, le dernier alinéa de l'article L. 342-3 du CPI dispose que : « *Les exceptions énumérées par le présent article ne peuvent porter atteinte à l'exploitation normale de la base de données ni causer un préjudice injustifié aux intérêts légitimes du producteur de la base* ».

Articles du CPI à modifier :

- L. 122-5 (droit d'auteur),
- L. 122-6-1 (droit spécial du logiciel, uniquement pour l'exception « non académique »),
- L. 211-3 (droits voisins),
- L. 331-31 (sur les compétences de l'Hadopi pour la garantie des exceptions),
- L. 342-3 (droit *sui generis* des bases de données).

Il pourrait également être ajouté des **dispositions complétant les missions de l'Hadopi** dans le cadre de l'exception de fouille relevant de l'article 3 :

- règlement des différends indépendamment de la question des mesures techniques de protection et
- suivi de la charte.

B. Préconisations textuelles

Le code de la propriété intellectuelle est ainsi modifié :

I – Le 10° de l'article L. 122-5 est ainsi rédigé :

« 10° Les copies ou reproductions numériques d'une œuvre accessible de manière licite en vue de la fouille de textes et de données réalisées :

a) par des organismes de recherche ou des institutions du patrimoine culturel, aux seules fins de la recherche scientifique.

L'exception mentionnée à l'alinéa précédent n'est pas applicable lorsque l'organisme ou l'institution mentionné à cet alinéa a conclu une convention avec une entreprise lui accordant un accès privilégié aux résultats de la fouille.

Les copies et reproductions numériques effectuées lors d'une fouille de textes et de données sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins exclusives de recherche scientifique, y compris pour la vérification des résultats de la recherche.

L'auteur peut mettre en œuvre des mesures proportionnées et nécessaires afin d'assurer la sécurité et l'intégrité des réseaux et des bases de données dans lesquels les œuvres sont hébergées.

Une charte définit les bonnes pratiques relatives à l'application de la présente exception.

Sans préjudice de leur droit de saisir le juge, les organismes de recherche et les institutions du patrimoine culturel peuvent saisir la Haute Autorité pour la diffusion des œuvres et la protection des droits sur internet de tout différend lié à l'application de ces dispositions et aux modalités d'exercice de l'exception. La Haute Autorité procède selon les dispositions de l'article L. 331-35.

Toute stipulation contraire aux dispositions du présent a) est nulle et non avenue.

b) par toute personne, quelle que soit la finalité de la fouille, sauf si l'auteur s'y est opposé au moyen de procédés appropriés.

Cette opposition est sans incidence sur l'exception mentionnée au a) du présent 10°.

Les copies et reproductions sont stockées avec un niveau de sécurité approprié puis détruites à l'issue de la fouille de textes et de données. »

II – A l'article L. 122-6-1, il est inséré un IV bis ainsi rédigé :

« IV bis. Les actes prévus aux 1° et 2° de l'article L. 122-6 ne sont pas soumis à l'autorisation de l'auteur lorsqu'ils sont nécessaires à la fouille de textes et de données, quelle que soit sa finalité, sauf si l'auteur s'y est opposé au moyen de procédés appropriés. Les copies et reproductions sont stockées avec un niveau de sécurité approprié puis détruites à l'issue de la fouille de textes et de données. »

III – A l'article L. 211-3, il est inséré un 8° ainsi rédigé :

« 8° Les copies ou reproductions numériques d'une interprétation, d'un phonogramme, d'un vidéogramme, d'un programme ou d'une publication de presse, accessible de manière licite en vue de la fouille de textes et de données réalisées :

a) par des organismes de recherche ou des institutions du patrimoine culturel, aux seules fins de la recherche scientifique.

L'exception mentionnée à l'alinéa précédent n'est pas applicable lorsque l'organisme ou l'institution mentionné à cet alinéa a conclu une convention avec une entreprise lui accordant un accès privilégié aux résultats de la fouille.

Les copies et reproductions numériques effectuées lors d'une fouille de textes et de données sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins exclusives de recherche scientifique, y compris pour la vérification des résultats de la recherche.

L'artiste-interprète, le producteur, l'entreprise de communication audiovisuelle, l'éditeur de presse ou l'agence de presse peuvent mettre en œuvre des mesures proportionnées et nécessaires afin d'assurer la sécurité et l'intégrité des réseaux et des bases de données dans lesquels l'interprétation, le phonogramme, le vidéogramme, le programme ou la publication de presse sont hébergés.

Une charte définit les bonnes pratiques relatives à l'application de la présente exception.

Sans préjudice de leur droit de saisir le juge, les organismes de recherche et les institutions du patrimoine culturel peuvent saisir la Haute Autorité pour la diffusion des œuvres et la protection des droits sur internet de tout différend lié à l'application de ces dispositions et aux modalités d'exercice de l'exception. La Haute Autorité procède selon les dispositions de l'article L. 331-35.

Toute stipulation contraire aux dispositions du présent a) est nulle et non avenue.

b) par toute personne, quelle que soit la finalité de la fouille, sauf si l'artiste-interprète, le producteur, l'entreprise de communication audiovisuelle, l'éditeur de presse ou l'agence de presse s'y est opposé au moyen de procédés appropriés.

Cette opposition est sans incidence sur l'exception mentionnée au a) du présent 8°.

Les copies et reproductions sont stockées avec un niveau de sécurité approprié puis détruites à l'issue de la fouille de textes et de données. »

IV – L'article L. 331-14 est complété par un alinéa ainsi rédigé :

« La Haute Autorité prend toute mesure, notamment par l'adoption de recommandations, de modèles et clauses types ainsi que de codes de bonne conduite, visant à favoriser la définition de bonnes pratiques par les titulaires de droits, les organismes de recherche et les institutions du patrimoine culturel, pour l'application de l'obligation et des mesures prévues aux a) du 10° de l'article L. 122-5, du 8° de l'article L. 211-3 et du 5° de l'article L. 342-3 ».

V – Le 2° de l'article L. 331-31 est ainsi modifié :

1° Au deuxième alinéa, les mots « et 8° » sont remplacés par les mots « , 8° et 10° » ;

2° Au troisième alinéa, les mots « et 7° » sont remplacés par les mots « 7° et 8° » ;

3° Au quatrième alinéa, les mots « et, à compter du 1^{er} janvier 2009, 4° » sont remplacés par les mots « , 4° à compter du 1^{er} janvier 2009 et 6° ».

VI – L'article L. 342-3 est ainsi modifié :

1° Le 5° est ainsi rédigé :

« 5° Les extractions, copies ou reproductions numériques d'une base de données accessible de manière licite en vue de la fouille de textes et de données réalisées :

a) par des organismes de recherche ou des institutions du patrimoine culturel, aux seules fins de la recherche scientifique.

L'exception mentionnée à l'alinéa précédent n'est pas applicable lorsque l'organisme ou l'institution mentionné à cet alinéa a conclu une convention avec une entreprise lui accordant un accès privilégié aux résultats de la fouille.

Les extractions, copies et reproductions numériques effectuées lors d'une fouille de textes et de données sont stockées avec un niveau de sécurité approprié et peuvent être conservées à des fins exclusives de recherche scientifique, y compris pour la vérification des résultats de la recherche.

Le producteur de la base de données peut mettre en œuvre des mesures proportionnées et nécessaires afin d'assurer la sécurité et l'intégrité des réseaux dans lesquels ces bases sont hébergées.

Une charte définit les bonnes pratiques relatives à l'application de la présente exception.

Sans préjudice de leur droit de saisir le juge, les organismes de recherche et les institutions du patrimoine culturel peuvent saisir la Haute Autorité pour la diffusion des œuvres et la protection des droits sur internet de tout différend lié à l'application de ces dispositions et aux modalités d'exercice de l'exception. La Haute Autorité procède selon les dispositions de l'article L. 331-35.

Toute stipulation contraire aux dispositions du présent a) est nulle et non avenue.

b) par toute personne, quelle que soit la finalité de la fouille, sauf si le producteur de la base de données s'y est opposé au moyen de procédés appropriés.

Cette opposition est sans incidence sur l'exception mentionnée au a) du présent 5°.

Les extractions, copies et reproductions sont stockées avec un niveau de sécurité approprié puis détruites à l'issue de la fouille de textes et de données. »

2° A l'avant dernière phrase, les mots « au 1° ci-dessus » sont remplacés par les mots : « aux présents 1° ou a) du 5° ».

Lettre de mission



MINISTÈRE DE LA CULTURE

Liberté
Égalité
Fraternité

Conseil supérieur de la propriété littéraire et artistique

Le Président

182 rue Saint-Honoré
75033 Paris Cedex 01

Téléphone : 01.40.15. 38.73

cspla@culture.gouv.fr

<https://www.culture.gouv.fr/Sites-thematiques/Propriete-litteraire-et-artistique/Conseil-superieur-de-la-propriete-litteraire-et-artistique>

Paris, le 10 juin 2020

Madame Alexandra Bensamoun,
Professeure des universités

Madame,

En 2014, le Conseil supérieur de la propriété littéraire et artistique (CSPLA) a conduit une mission sur l'exploration de données, également qualifiée de "*text and data mining*" (TDM). Le rapport réalisé par Maître Jean Martin, de nature exploratoire, insistait sur la nécessité de favoriser un développement de l'exploration de données, qui ne soit pas parasitaire de la création et des investissements dans les gisements de contenus, dans l'intérêt même de l'exploration de données.

L'article 38 de la loi du 7 octobre 2016 pour une République numérique a introduit dans le code de la propriété intellectuelle une nouvelle exception au droit d'auteur et au droit *sui generis* des producteurs de bases de données pour les activités d'exploration menées à des fins de recherche scientifique.

Ces initiatives conduites au niveau national doivent désormais prendre en compte la directive 2019/790 du 17 avril 2019 dite "*Droit d'auteur dans le marché unique numérique*" qui met en évidence l'importance des technologies de fouilles de données pour la communauté des chercheurs, au soutien de l'innovation, ainsi que pour les entités privées et publiques, notamment dans la prise de décisions commerciales. La directive insiste néanmoins sur la nécessité de concilier ces intérêts avec ceux des titulaires de droits et instaure à ce titre deux régimes juridiques de fouille de données.

Le premier, prévu à l'article 3 de la directive, consiste en une exception large au profit des activités d'exploration de données opérées à des fins de recherche scientifique exclusivement.

Le second, ouvert même aux entités animées par un objectif purement commercial, contient un mécanisme de limitation des droits patrimoniaux, inédit, consistant alors en une exception avec option de retrait (dite « *opt-out* ») pour le titulaire de droits. L'exception est ouverte, ne vise ni bénéficiaire spécifique, ni finalité particulière hormis la réalisation d'opérations de fouille de données dans ce qui apparaît

relever d'activités d'intelligence artificielle telles que l'apprentissage automatique (*machine learning*). Les conditions d'une telle exception sont posées à l'article 4 de la directive.

Sa transposition en droit français soulève des questions d'ordre technique, juridique et économique.

Ainsi, d'un point de vue technique, il importe de comprendre comment pourra s'exercer l'option de retrait. L'article 4 précise sur ce point que la réservation doit se faire de manière appropriée, "*notamment par des procédés lisibles par machine pour les contenus mis à la disposition du public en ligne*", certains procédés, variés, étant précisés par le considérant 18 de la directive.

Sur le plan juridique, l'exception soulève des questions de qualification et d'articulation.

D'abord, ce nouveau mécanisme de consentement présumé interroge sur sa qualification et sur sa compréhension à l'aune de la jurisprudence européenne.

L'exception pose en outre des questions d'articulation, non seulement avec le monopole lui-même, mais aussi avec l'exception obligatoire prévue à l'article 3 de la même directive, ainsi qu'avec l'exception pour les reproductions techniques provisoires de l'article 5 §1 de la directive 2001/29 du 22 mai 2001. Se pose enfin la question de l'articulation avec le nouveau droit voisin des agences et éditeurs de presse, instauré par l'article 15 de la directive 2019/790, afin que son efficacité ne soit pas mise à mal par l'usage du TDM.

Enfin, sur le plan économique, sans garantie de transparence du point de vue de la sécurité du site primaire hébergeant les contenus et des contenus eux-mêmes, il n'est pas exclu qu'un grand nombre de titulaires de droits souhaitent exercer leur option de retrait par défaut afin d'éviter une exploration (« *crawling* ») massive et systématique de leurs données, l'exception risquant alors d'instaurer un climat de méfiance généralisée. Il importe également de s'interroger sur la réaction des grands acteurs du numérique face à des titulaires de droits ayant exercé leur option de retrait de l'exception, notamment quant au référencement des contenus. Enfin, compte tenu de l'importance de l'intelligence artificielle pour le numérique et le secteur de la culture en général, il vous appartiendra de relever les autres éventuels aspects économiques que les travaux que vous conduirez vous amèneraient à juger utiles.

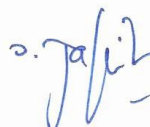
Vous pourrez, dans votre rapport, formuler des propositions dans le domaine juridique et technologique, qui pourront donner lieu, par la suite, à l'élaboration de textes législatifs et réglementaires.

Afin de mener à bien cette mission, vous procéderez à des auditions des membres du CSPLA qui le souhaitent ainsi que de toute personne dont vous jugerez utiles les contributions. Vous pourrez, si vous le souhaitez, bénéficier de l'appui d'un rapporteur.

Il serait souhaitable que vous puissiez présenter votre rapport à la réunion plénière du Conseil qui devrait se tenir au mois de décembre prochain.

Je vous remercie d'avoir accepté cette mission et vous prie de croire, Madame, à l'expression de mes sentiments distingués.

Amitiés,



Olivier Japiot

Questionnaire élaboré par la mission

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'un à des **fins de recherche** exclusivement (art. 3), à laquelle ne peuvent pas s'opposer les titulaires de droit concernés, et l'autre, dit « 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 ¹²⁷ (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 ¹²⁸ (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 :
Secteur concerné :

¹²⁷ <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>

¹²⁸ <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* »¹²⁹, 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 ?
- 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) 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 ?
- 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 données¹³⁰ ? Ces dispositions vous semblent-elle adaptées¹³¹ ?
- 5) Avez-vous été associés, directement ou indirectement, à l'élaboration des articles 3 et 4 de la directive 2019/790 ?
- 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 :

¹²⁹ Directive 2019/790, article 2, 2).

¹³⁰ 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°.

¹³¹ 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.

- 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. 7.1¹³²).

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 minime¹³³) vous semblent-ils satisfaisants ?

¹³² « Toute disposition contractuelle contraire aux exceptions prévues aux articles 3, 5 et 6 est non exécutoire. »

¹³³ V. cons. 17.

- 2) Les partenariats public-privé peuvent bénéficier de cette exception « académique »¹³⁴, 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 ?
- 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) ?
- 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 ?
- 5) Quelles modalités souhaiteriez-vous voir mises en place pour la conservation des copies¹³⁵ ?
- 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 ?
- 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 ?

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

¹³⁴ 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. »

¹³⁵ 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. »

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 ?
- 3) Comment comprenez-vous le droit d'option ouvert au titulaire et qui permet de se retirer du dispositif ?
- 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* ?
- 5) Pensez-vous exercer cette option de retrait ?
- 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) ?
- 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 ?
- 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 ?
- 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 ?
- 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 directive¹³⁶ ?
- 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 ?

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 ?

¹³⁶ 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. »

Liste des contributeurs

Autorités

Conseil national du numérique

Secteur numérique – commission consultative indépendante

Hadopi

Pauline Blassel, Secrétaire générale

Anna Butlen, Directrice des affaires juridiques, européennes et internationales

Carla Menaldi, Directrice adjointe des affaires juridiques, européennes et internationales

Denis Rapone, Président

Laura Salomon-Raimondo, Juriste propriété intellectuelle et droit du numérique, Direction des affaires juridiques, européennes et internationales

Ministère de la Culture

Sarah Jacquier, Chargée de mission auprès du Chef du service des affaires juridiques et internationales

Marie Véronique Leroi, Chargée de mission Web Sémantique, département de l'innovation numérique

Alban de Nervaux, Chef du service des affaires juridiques et internationales

Nicolas Orsini, Chef du département de l'innovation numérique

Ministère de l'enseignement supérieur, de la recherche et de l'innovation

Odile Contat, Cheffe du département de l'information scientifique et technique et réseau documentaire (DGRI-DGESIP)

Marin Dacos, Conseiller pour la science ouverte du Directeur général de la recherche et de l'innovation (DGRI)

Claire Leymonerie, Chargée de mission pour la science ouverte (DGRI)

Personnalité qualifiée au CSPLA

Jean Martin, Avocat, Membre d'honneur, Président du premier rapport CSPLA sur la fouille de données

Ayants droit et représentants

Alliance de la presse d'information générale

Samir Ouachtati, Directeur adjoint, Responsable du pôle juridique et social

CFC

Benoît Kerjean, Directeur juridique

Conseil permanent des écrivains

Bessora, Présidente, SNAC

FNPS

Laurent Bérard-Quélin, Président, Directeur général, Rédacteur en chef de la Société Générale de Presse

Boris Bizic, Directeur juridique

PROCIREP

Idzar Van der Puy, Secrétaire général

SACEM

David El Sayegh, Secrétaire général

SACD

Hubert Tilliet, Directeur juridique

SEPM

Julie Lorimy, Directrice générale

SGDL

Patrice Locmant, Directeur général

Maïa Bensimon, Directrice juridique

SNE

Catherine Blache, Chargée de Mission Relations Internationales, SNE

Liliane de Carvalho, Responsable juridique Madrigall (et rapporteur du premier rapport CSPLA sur la fouille de données)

Julien Chouraqui, Directeur juridique SNE

Frédéric Géraud de Lescazes, Directeur des relations institutionnelles du Groupe RELX (France et Europe du Sud)

François Gèze, Editeur, La Découverte, Président du groupe des éditeurs universitaires du Syndicat national de l'édition

Pauline Menez, Juriste Madrigall

Denis Mignan, Directeur juridique, Hachette Livre, branche illustré

Organismes dépositaires**INA**

Eléonore Alquier, Responsable du département collecte et conservation de la direction déléguée aux collections

Jean-François Debarnot, Directeur juridique

Aude Formagne, Juriste au département des affaires juridiques et réglementaires de la direction juridique

Nicolas Hervé, Chef de projet recherche et développement au sein du département recherche et innovation numérique de la direction déléguée à la diffusion et à l'innovation

Barbara Mutz, Responsable du département des affaires juridiques et réglementaires de la direction juridique

BNF

Harold Codant, Chef du service juridique

Emmanuelle Bermès, Adjointe scientifique et technique au Directeur des services et des réseaux

Chercheurs/universitaires

Droit

Nicolas Binctin, Professeur agrégé des facultés de droit, Université de Poitiers, CECOJI

Annie Blandin-Obernesser, Professeure de droit, IMT Atlantique

Christophe Geiger, Professeur de droit, Directeur du laboratoire de recherche du CEIPI, Université de Strasbourg

Jane C. Ginsburg, Professeur Morton L. Janklow de droit de la propriété littéraire et artistique, Columbia Law School

Agnès Robin, Maître de conférences, Université de Montpellier

Eleonora Rosati, Associate Professor in Intellectual Property Law, Stockholm University

Informatique

Gaël DIAS, Professeur, Université de Caen, co-directeur de GREYC (Groupe de recherche en informatique, image, automatique et instrumentation de Caen), membre du CA de l'Association française pour l'intelligence artificielle (AFIA)

Alexis Lechervy, Maître de conférences, Université de Caen, Responsable du master IDM, Laboratoire Greyc - Equipe Image - UMR CNRS 6072

Gaël Richard, Professeur, Head of the Image, Data, Signal Department (IDS), LTCI (Laboratoire Traitement et Communication de l'Information), Télécom Paris, Institut polytechnique de Paris

Mathieu Roche, Research Scientist (PhD, HDR), CIRAD - Environments and Societies Department, UMR TETIS

Pierre Zweigenbaum, Directeur de recherche, LIMSI, CNRS, Université Paris-Saclay

Secteur privé/Autres

ADBU

Christophe Pérales, Président, Directeur des bibliothèques, Université de Paris

BSA – Software Alliance

Conseils représentant les éditeurs STM (scientifiques, techniques et médicaux)

Carlo Scollo Lavizzari, Advocate (Switzerland), Solicitor (England & Wales), Attorney (non-practising, South Africa)

Mathilde Renou, Founder and Legal Consultant, Rendel Consulting

DASSAULT SYSTEMES SE

Rodolphe Pantanacce, General Counsel, Intellectual Property

EARE (European Alliance for Research Excellence)

ELDA

Khalid Choukri, CEO

France Digitale (association startups et investisseurs du numérique)

Julia Fenart, Responsable des affaires européennes

GOOGLE France

SATT Paris-Saclay

Xavier Apolinarski, Président

Cristina Bayona-Philippine, Directrice des Affaires Juridiques