

La Déclaration de La Haye sur l'EXTRACTION DES CONNAISSANCES
à l'ère du numérique

Vision

Les nouvelles technologies révolutionnent la manière dont les êtres humains peuvent acquérir des connaissances sur le monde et sur eux-mêmes. Ces technologies ne sont pas seulement un moyen de gérer les BigData¹, elles sont indispensables pour extraire des connaissances à l'ère du numérique ; leur pouvoir repose sur la disponibilité croissante des données. Des facteurs tels que l'augmentation de la puissance de calcul, la croissance de l'internet, et l'engagement des gouvernements en matière de libre accès² à la recherche financée sur fonds publics permettent d'accroître la disponibilité des faits, des données et des idées.

Cependant, les cadres législatifs actuels dans les différentes juridictions peuvent être formulés d'une manière qui ne favorise pas l'introduction de nouvelles approches pour entreprendre des travaux de recherche, en particulier la fouille de contenus. La fouille de contenus consiste à obtenir des informations à partir de matériel lisible par machine. Elle fonctionne en copiant de grandes quantités de matériel, en extrayant les données et en les recombinaut pour identifier des profils et des tendances.

Dans le même temps, les législations sur la propriété intellectuelle qui datent d'une époque bien antérieure au web limitent la portée des techniques d'analyse de contenus numériques telles que la fouille de textes et de données ou de fouille de contenus (pour ce qui est de l'analyse informatique de contenus dans toutes ses formes)³. Ces facteurs créent également des inégalités d'accès à l'extraction des connaissances à l'ère du numérique. La législation en question peut être celle sur les droits d'auteur, celle régissant les brevets ou celle sur les bases de données, chacune d'entre elles pouvant limiter la capacité de l'utilisateur à effectuer une analyse de contenu détaillée.

Les chercheurs devraient avoir la liberté de mener une analyse et de satisfaire leur curiosité intellectuelle sans crainte de la surveillance ou des répercussions. Cette liberté ne doit pas être restreinte dans l'environnement numérique. De même, l'éthique autour de l'utilisation de la fouille de données et de contenus continue à évoluer face aux changements technologiques.

L'analyse informatique de contenus sous toutes ses formes, ou fouille de contenus, permet l'accès à des connaissances inexplorées du domaine public et fournit d'importantes indications sur tous les aspects de notre vie économique, sociale et culturelle. L'analyse de contenus aura aussi une incidence profonde sur la compréhension de la société et des mouvements sociaux (par exemple, en prévoyant les soulèvements politiques, en analysant les changements démographiques). L'utilisation de telles techniques peut révolutionner les méthodes de la recherche aussi bien universitaire que commerciale.

¹ Concernant la Charte du G8 pour l'ouverture des données publiques, voir <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>. Voir également le rapport de la RDA sur la moisson de données http://europe.rd-alliance.org/sites/default/files/report/TheDataHarvestReport_%20Final.pdf

² Concernant la Déclaration de Berlin, voir <http://openaccess.inist.fr/?Declaration-de-Berlin-sur-le-Libre>

Voir également l'Initiative de Budapest pour l'Accès Ouvert : 10 ans après, <http://www.budapestopenaccessinitiative.org/boai-10-recommendations>

³ Quand l'expression « fouille de textes et de données » / TDM est utilisée dans la Déclaration, elle signifie la fouille de toutes les formes de données, qu'il s'agisse de textes, d'images, d'enregistrements sonores ou de films.

Avantages de la fouille de contenus

Les avantages potentiels de la fouille de contenus sont considérables, notamment :

- **Relever les grands défis comme le changement climatique et les pandémies**
- **Améliorer la santé des populations, accroître leur richesse et leur développement**
- **Créer de nouveaux emplois**
- **Augmenter de manière exponentielle le rythme et les progrès de la science grâce à des connaissances nouvelles et une efficacité accrue de la recherche**
- **Renforcer la transparence des gouvernements et de leurs actions**
- **Favoriser l'innovation et la collaboration et amplifier l'impact de la science ouverte**
- **Créer des outils à des fins d'éducation et de recherche**
- **Proposer des perspectives nouvelles et plus riches sur la culture**
- **Accélérer le développement économique et social dans le monde entier**

Les chercheurs, les PME (Petites et Moyennes Entreprises) et les grandes entreprises technologiques fouillent du contenu depuis au moins 10 ans, mais la capacité à tirer des avantages significatifs de ce travail est limitée en raison d'incertitudes juridiques et de restrictions persistantes. Cependant, à notre époque de tous les possibles, il est essentiel que tous les membres de la société bénéficient de manière égale des progrès en matière de disponibilité des technologies et contenus numériques. Cela nécessite la définition de nouveaux principes sur l'accès aux faits, aux données et aux idées.

Principes

Dans ces conditions, nous, soussignés, reconnaissant les avantages potentiels énormes pour l'économie et la société de l'extraction des connaissances à l'ère du numérique, approuvons les principes suivants :

1. LA PROPRIÉTÉ INTELLECTUELLE N'A PAS ÉTÉ CONÇUE POUR RÉGLEMENTER LA LIBRE CIRCULATION DES FAITS, DES DONNÉES ET DES IDÉES, MAIS ELLE A COMME OBJECTIF-CLÉ DE PROMOUVOIR LES ACTIVITÉS DE RECHERCHE

La libre circulation des informations et des idées est un droit individuel essentiel⁴. C'est un catalyseur pour la production de connaissances qui soutient le bien-être et la prospérité. Les entreprises du monde entier ont choisi de protéger certains droits restreints de propriété intellectuelle comme des mesures incitatives à la fois à l'innovation et à la diffusion des connaissances. Le droit de la propriété intellectuelle n'a jamais visé à couvrir les faits, les idées et les données pures et simples. Le droit de la propriété intellectuelle tel qu'il est appliqué actuellement entrave de plus en plus la création et le partage des connaissances, car il peut poser des contraintes et des limitations au libre partage même lorsque de simples composantes comme des faits et des données sont utilisées.

Dans certains pays, la législation sur le droit d'auteur⁵ a été interprétée en particulier pour restreindre la capacité d'appliquer la lecture par machine et l'analyse informatique à un contenu légalement disponible autrement. D'autres cadres législatifs comme le droit des brevets ou le droit des bases de données peuvent avoir un impact similaire. Quand le droit de la propriété intellectuelle autorise la lecture et l'analyse manuelles par des êtres humains et les refuse à leurs machines, il manque à ses objectifs initiaux.

⁴ <http://www.un.org/fr/documents/udhr/#a19>

⁵ Et dans l'Union européenne les législations sur les bases de données, également sous la forme de la Directive sur les bases de données.

2. LES CITOYENS DEVRAIENT AVOIR LA LIBERTÉ DE MENER UNE ANALYSE ET DE SATISFAIRE LEUR CURIOSITÉ INTELLECTUELLE SANS CRAINTE DE LA SURVEILLANCE OU DES RÉPERCUSSIONS

Les fournisseurs de contenus devraient respecter la vie privée intellectuelle des lecteurs individuels et devraient prendre des mesures pour protéger cette vie privée de l'ingérence de la part d'un organisme externe. Toute exception, qui par exemple entraînerait une intrusion dans le domaine privé d'une personne devra être justifiée comme étant nécessaire, proportionnée et prévue par la loi. L'utilisation de faits, de données et d'idées ne doit pas porter préjudice aux droits légitimes des personnes à la vie privée et à la confidentialité.

3. LES CLAUSES DES LICENCES ET CONTRATS NE DEVRAIENT PAS EMPÊCHER LES PERSONNES D'UTILISER DES FAITS, DES DONNÉES ET DES IDÉES

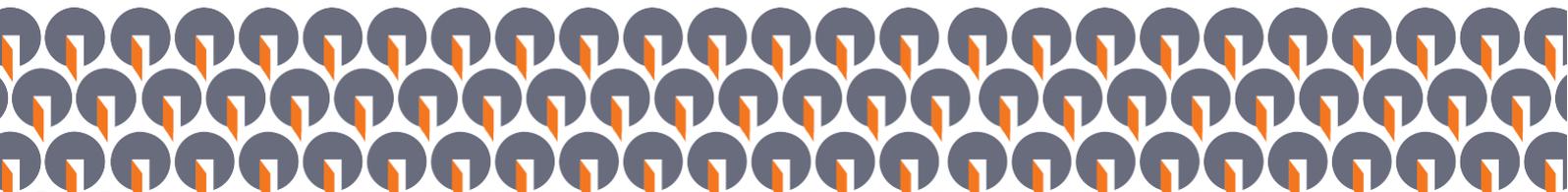
En général, les clauses des licences et contrats qui régissent et restreignent la manière dont les personnes peuvent analyser et utiliser les faits, les données et les idées sont inacceptables et freinent l'innovation et la création de nouvelles connaissances et, par conséquent, ne devraient pas être adoptées. De la même façon, il est inadmissible que des mesures techniques dans les systèmes de gestion des droits numériques empêchent le droit légitime de réaliser une fouille de contenus.

4. L'ÉTHIQUE AUTOUR DE L'UTILISATION DES TECHNIQUES DE FOUILLE DE CONTENUS DEVRA CONTINUER À ÉVOLUER FACE AUX CHANGEMENTS TECHNOLOGIQUES

Le respect de normes éthiques bien établies dans le monde de la recherche et de l'entreprise, ainsi que l'élaboration continue de normes et de lois, devront être soutenus et encouragés afin de garantir que les technologies de fouille de contenus soient déployées dans l'intérêt de la société.

5. L'INNOVATION ET LA RECHERCHE COMMERCIALE REPOSANT SUR L'UTILISATION DE FAITS, DE DONNÉES ET D'IDÉES NE DEVRAIENT PAS ÊTRE LIMITÉES PAR LE DROIT DE LA PROPRIÉTÉ INTELLECTUELLE

Comme les faits, les données et les idées ne sont pas soumis au droit d'auteur, il est totalement illogique de restreindre l'utilisation commerciale éthique de ces faits, données et idées extraits du contenu qui a été obtenu légalement. Il est établi que même si le droit des brevets est conçu pour protéger les innovations et les inventions, il n'entend pas inclure les faits et les données. Les restrictions concernant l'utilisation de faits, de données et d'idées peuvent avoir de sérieuses répercussions sur l'innovation et le développement économique à l'échelle mondiale. Cela peut aussi réduire la capacité d'utiliser des outils et des processus qui peuvent bénéficier aux citoyens dans les secteurs de la santé, de la science, de l'emploi, de la recherche, de l'environnement et de la culture.



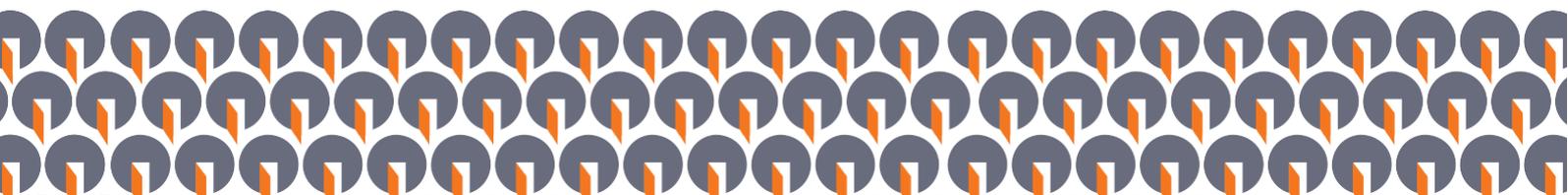
Feuille de route pour agir

ACCEPTATION DE LA VISION ET DES PRINCIPES

1. La vision inscrite dans cette Déclaration est que la propriété intellectuelle n'a pas été conçue pour réglementer la libre circulation des faits, des données et des idées, mais qu'elle a comme objectif-clé de promouvoir les activités de recherche.
2. Lorsque le cadre du droit d'auteur ne soutient pas une telle vision, les législateurs devraient immédiatement s'efforcer de favoriser l'introduction de changements qui permettraient aux utilisateurs d'entreprendre une fouille de contenus sur les éléments pour lesquels ils disposent d'un accès légal.
3. Lorsque des exceptions ou des limitations sont introduites dans la législation sur le droit d'auteur pour autoriser la fouille de contenus, elles doivent être obligatoires et ne peuvent pas être supplantées par les contrats.
4. Il est inadmissible que des mesures techniques dans les systèmes de gestion des droits numériques empêchent le droit légitime de réaliser une fouille de contenus.
5. La signature de licences distinctes devrait être inutile pour mener une activité de fouille de contenus, puisque le droit de lire est le droit de fouiller dans les cas où les personnes menant des activités de fouille disposent déjà d'un accès légal au contenu pertinent.

POLITIQUE ET MESURES INCITATIVES

6. Les dirigeants politiques devraient chercher à apporter de la clarté juridique en garantissant que la fouille de contenus n'est pas une violation du droit d'auteur ou des droits voisins. Nous sommes convaincus que le droit de lire inclut le droit de fouiller, mais uniquement lorsque les personnes ont un accès légal audit contenu.
7. Lorsque les financeurs de la recherche ou d'autres institutions exigent, et lorsque les auteurs le souhaitent, que les résultats de la recherche soient rendus accessibles sous des licences spécifiques, celles-ci devraient être du type CC-BY pour les publications et du type CC0 pour les données de la recherche.
8. Chaque université, organisme de recherche, financeur de la recherche ou entreprise commerciale devrait veiller à ce que sa politique recommande la fouille de contenus comme une méthodologie de recherche qui a le potentiel de transformer la manière dont la recherche est effectuée. La croissance du libre accès et des données ouvertes a été et restera un moteur-clé de la fouille de contenus.
9. Les chercheurs devraient reconnaître en tant que tel le droit des auteurs de publications et que ces auteurs soient respectés et reconnus comme producteurs de données, dans la mesure du possible.
10. Les politiques concernant la fouille de contenus devraient respecter les droits légitimes des auteurs et des éditeurs et être dictées par les besoins des chercheurs et des entreprises à l'ère du numérique.
11. Pour encourager l'adoption de l'activité de fouille de contenus, les universités, les organismes de recherche, les financeurs de la recherche et les entreprises devraient envisager de mettre en place des mesures incitatives pour récompenser ceux qui utilisent ces nouvelles techniques - par exemple, les activités de fouille de contenus devraient être signalées et valorisées au cours de procédures d'évaluation.



INFRASTRUCTURE ET OUTILS

12. Les organismes de recherche, les universités et les entreprises devraient veiller à maintenir et développer des infrastructures de dépôt pour permettre le stockage et l'accès aux publications qui peuvent être mises à disposition légalement pour la fouille de contenus. Les chercheurs indépendants devraient s'efforcer d'utiliser ces ressources lorsqu'elles sont disponibles.
13. Les organismes de recherche, les universités et les entreprises devraient donner accès à des infrastructures adéquates pour permettre la mise à disposition des données de la recherche à des fins de fouille de contenus, quand il est licite et possible sur le plan éthique de le faire. Les chercheurs indépendants devraient s'efforcer d'utiliser ces ressources lorsqu'elles sont disponibles.
14. Les développements impliquant des infrastructures techniques, des standards, des normes éthiques et des prescriptions de financement devraient rendre accessibles les résultats de la recherche comme des résultats ouverts.
15. Les standards ouverts tels que XML et JSON pour le transport des données, ORCID pour les identifiants d'auteurs et les licences CC pour des concessions de licence ouvertes devraient être utilisés partout où c'est possible.
16. Si le matériel est rendu accessible sous une licence CC BY, les créateurs de contenus devraient mettre à disposition les éléments suivants en téléchargement : l'XML ou un autre format de fichier de haute qualité, des images en haute résolution, des données à la base des images ; l'article accepté non formaté de l'auteur.

PROMOTION

17. Les institutions telles que les universités, les organismes de recherche, les associations de bibliothèques, la communauté médicale, les entreprises et les membres de la communauté pratiquant la fouille de contenus devraient défendre les avantages de cette fouille.
18. Les bibliothèques de recherche sont bien placées pour remplir ce rôle de promotion dans le cadre de leurs activités de soutien à la recherche.
19. Les bibliothèques devraient former les chercheurs à la culture de la fouille de contenus, y compris le conseil juridique.

ÉLABORATION ET SUIVI DE LA DÉCLARATION DE LA HAYE & FEUILLE DE ROUTE

20. LIBER (Ligue des Bibliothèques Européennes de Recherche) a travaillé de concert avec les parties prenantes pour élaborer la Déclaration de La Haye et la Feuille de route. LIBER continuera de suivre et superviser la progression du processus de signature de la Déclaration, et prônera la mise en œuvre de la Feuille de route.



Version française : Marc Rubio – Équipe Traduction - INIST-CNRS

Pour en savoir plus & signer la Déclaration :
www.thehaguedeclaration.com

