

#DATAiASEMINARS



Alexandra Bensamoun
Pierre Zweigenbaum



«Text and data mining : pratiques scientifiques et droit applicable»



Wednesday, 3rd March 2021
17.00



<https://bluejeans.com/9352872428/9913>

Alexandra Bensamoun

Professeure de droit, Université Paris-Saclay

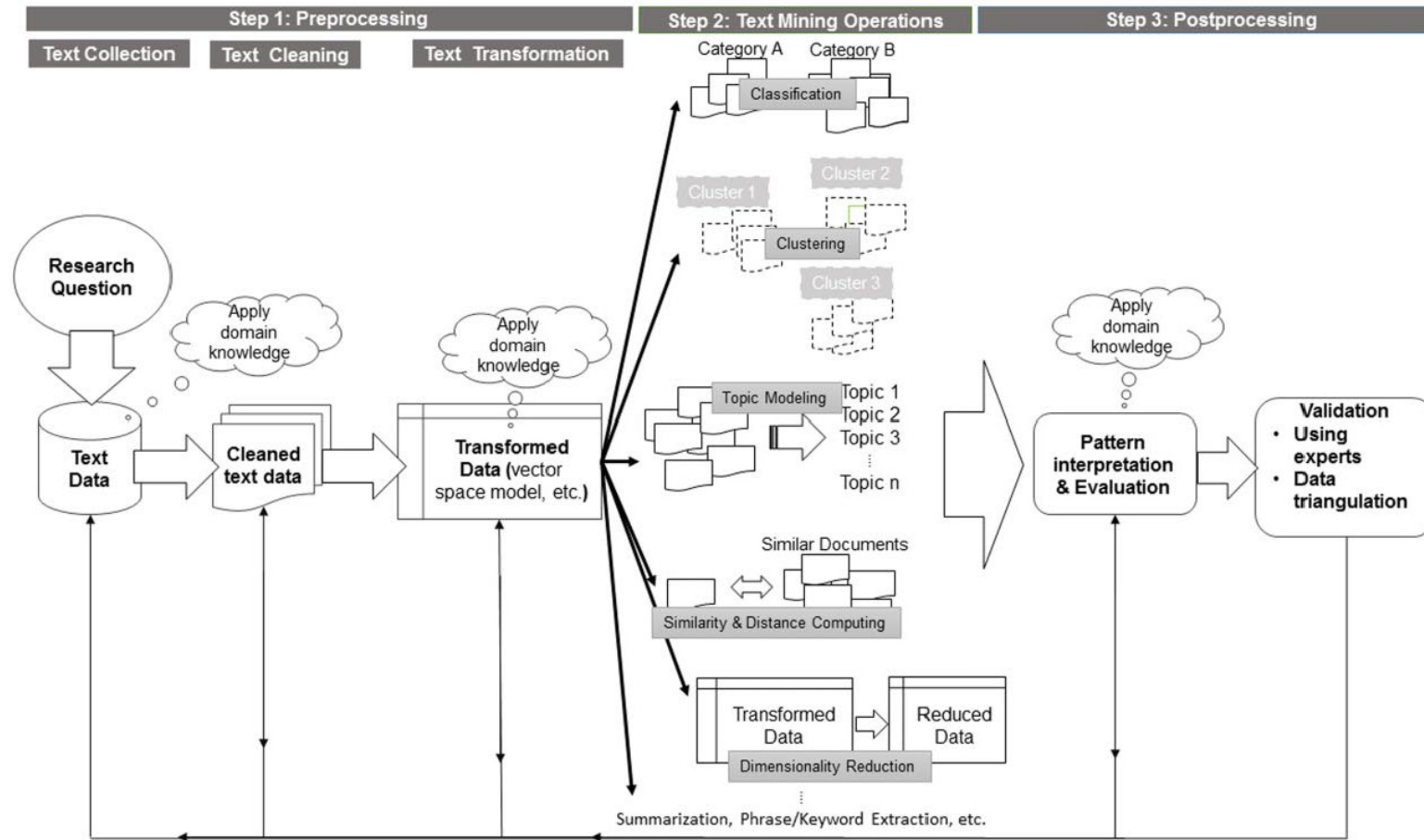
Membre du bureau de DATAIA

Dialogue avec Pierre Zweigenbaum

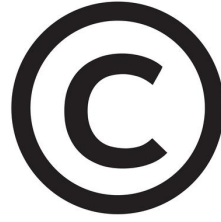
Directeur de recherche au LISN, CNRS, Université Paris-Saclay



Les 3 phases du TDM



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



Possible atteinte à des droits de PI du fait des utilisations réalisées sans autorisation des titulaires de droits.

- Articles scientifiques
- Articles de presse grand public
- Romans libres de droits
- Articles encyclopédiques, type Wikipédia
- Forums..

➔ Importance de la qualification juridique
Insécurité juridique de la recherche scientifique

Contexte

- **Loi pour une République numérique du 7 octobre 2016**

- Art. L. 122-5, 10° CPI : « 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 ».
- Système comparable pour le droit sui generis des bases de données, art. L. 342-3 CPI.
- Exception **très étroite** : réservée à la recherche publique, sans finalité commerciale, et aux « écrits scientifiques ».
- Décret d'application jamais intervenu : **incompatibilité** avec le droit européen (liste exhaustive des exceptions, dir. 2001/29).

Contexte

- **Lien avec le développement de l'IA, notamment l'apprentissage**
 - Rapport Villani, « *Donner un sens à l'intelligence artificielle* » de mars 2018: 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 elle est à ce titre « *porteuse de potentiels énormes pour la découverte scientifique et le développement de nouvelles connaissances* ».

Contexte

- **Directive 2019/790 dite « Digital Single Market », du 17 avril 2019 :**
 - Art. 1^{er} : « 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 ».
 - Art. 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.
 - Art. 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).

Contexte



- **Mission au CSPLA pour proposer une transposition de ces exceptions**

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

Rapport remis en déc. 2020

Disponible en ligne, sur le site du CSPLA

- **Loi DADDUE, 3 décembre 2020** : habilitation à légiférer par ordonnance pour transposer les exceptions, dans un délai d'un an

Art. 3 : le TDM « académique »

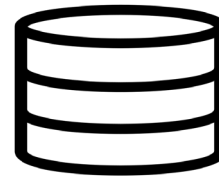
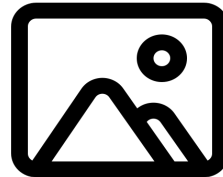
Article 3

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.

Champ d'application

- Une exception aux :
 - Droit d'auteur
 - Droits voisins (y compris le nouveau DV des éditeurs de presse)
 - Droit des bases de données
- Tous types de « données » : textes, sons, images, vidéo, etc.



Conditions

- Les **bénéficiaires** :
 - Organismes de recherche : universités (y compris BU), instituts de recherche... (non lucratif ou mission d'intérêt public reconnue par l'EM)
 - Institutions du patrimoine culturel : bibliothèques, musées, archives ou institutions du dépôt légal
- La finalité de **recherche scientifique**
- L'exclusion des **PPP** en cas de conclusion d'une convention avec une entreprise lui accordant un accès privilégié aux résultats de la fouille
- L'accès **licite** aux contenus protégés
- Le stockage et la conservation avec un niveau de **sécurité** approprié

Régime/articulation

- Une exception **obligatoire** : Les EM doivent obligatoirement transposer
- Une exception qui ne peut **pas être remise en cause** par le contrat ou par l'*opt-out* exercé par le titulaire de droits (art. 4), mais qui reste une exception, non un « droit »...
- Une exception soumise au **test en trois étapes**

Régime/articulation

- Une exception dont le bénéfice doit être **garanti** par les EM (Hadopi)
- La recommandation de l'élaboration d'une **charte** des bonnes pratiques entre les parties prenantes
- Une exception qui permet la conservation des corpus à des fins exclusives de recherche scientifique, y compris leur transmission aux fins de **vérification** des résultats

université
PARIS-SACLAY

INSTITUT
DATAIA
Science des données, Intelligence & Société

Merci !