



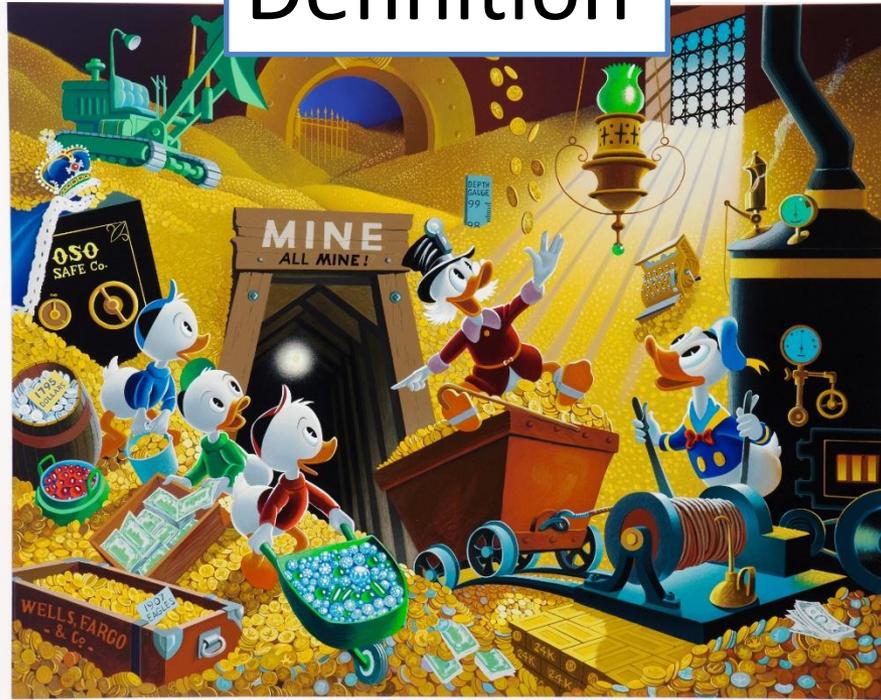
Text and Data Mining : éléments généraux et juridiques

Julien Baudry

Séminaire « Text and Data Mining :
explorons les cités enfouies »

21 juin 2023

Définition



Fouille de texte et de données (glossaire Doranum) :
« Analyse informatique de bases de données et de corpus de texte pour extraire des éléments d'information qui serviront à la construction de connaissances. »

Le processus TDM sollicite des « œuvres » dont certaines relèvent du droit d'auteur.

Presse
généraliste



Textes intégraux
d'articles scientifiques



Données
bibliographiques



Magazines
professionnels

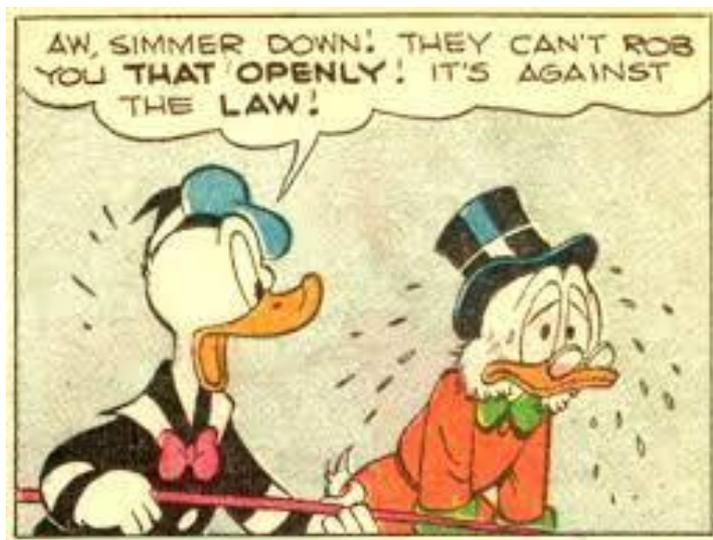
Le processus TDM sollicite des « œuvres » dont certaines relèvent du droit d'auteur.

Nécessite une exception au droit d'auteur !

Presse
généraliste



Textes intégraux
d'articles scientifiques



Données
bibliographiques



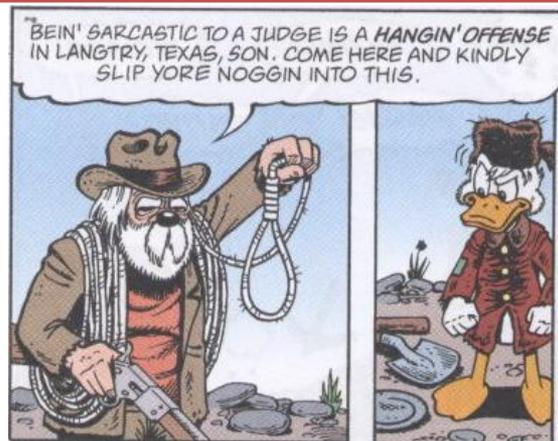
Magazines
professionnels

Cheminement législatif du TDM en France

2016 : Loi République Numérique (article 38)

« [sont autorisées] 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. »



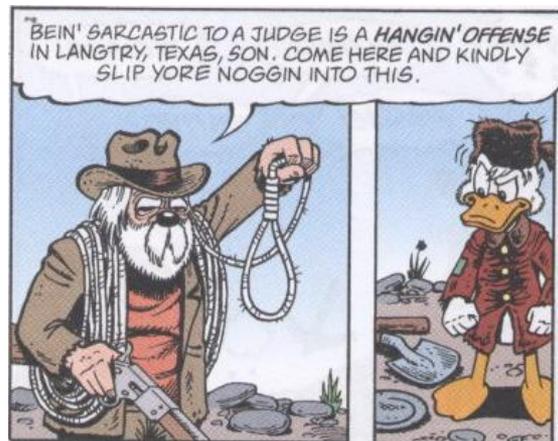
Cadre juridique (1)

Cheminement législatif du TDM en France

2016 : Loi République Numérique (article 38)

2019 : directive européenne sur le droit d'auteur

« L'insécurité juridique concernant la fouille de textes et de données devrait être traitée en prévoyant, au bénéfice des universités et autres organismes de recherche, ainsi que des institutions du patrimoine culturel, une exception obligatoire au droit de reproduction exclusif et au droit d'empêcher l'extraction à partir d'une base de données. »



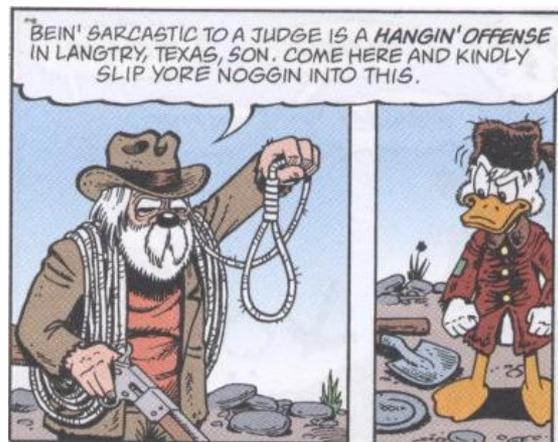
Cadre juridique (1)

Cheminement législatif du TDM en France

2016 : Loi République Numérique (article 38)

2019 : directive européenne sur le droit d'auteur

2021 : transposition de la directive UE par
l'ordonnance du 24/11/21 (article L.122-5-3)



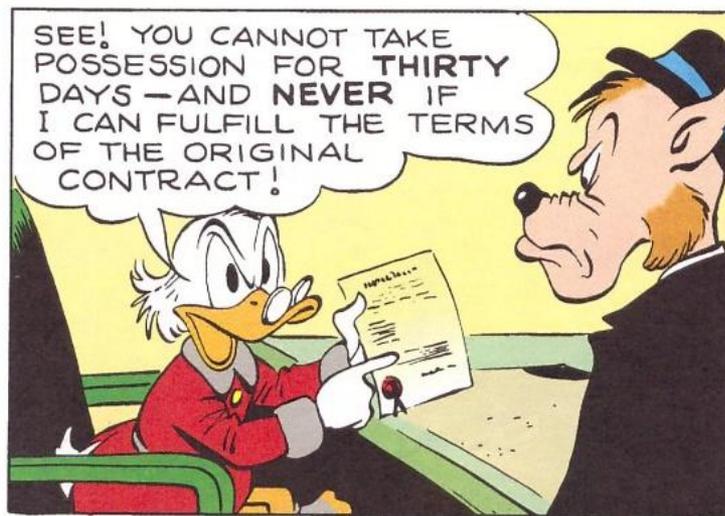
Cheminement législatif du TDM en France

Transposition de la directive UE par l'ordonnance du 24/11/21 (article L.122-5-3) :

- Confirme et rend opérationnel le TDM en France
- L'accès aux données doit être licite et le but non-commercial
- Limite les usages et les types d'établissements pouvant pratiquer ses usages (inclus partenariat public-privé)
- La conservation de copies est autorisée
- Donne peu d'éléments sur les dispositifs techniques et les formats

Cadre juridique (2)

Cas pratique : que disent les contrats avec les éditeurs scientifiques ?



Cadre juridique (2)

Cas pratique : que disent les contrats avec les éditeurs scientifiques ?

L'éditeur ne peut pas s'opposer au TDM.

L'éditeur peut proposer des modalités de récupération (API...). Mais peut-il les rendre obligatoires ?

L'éditeur peut étendre l'exception à d'autres usages.



Cadre juridique (3)

Des enjeux juridiques et éthiques :

- Trouver un accord autour de bonnes pratiques avec les éditeurs
- Un outil éditeur ne répond pas forcément aux enjeux d'interopérabilité
- Bien extraire le TDM des clauses de « failles de sécurité »
- Interdire explicitement les mesures de blocage





**Merci de votre
attention**

Crédits des images : Carl Barks sauf diapositive
« Cadre juridique (1) » : Keno Don Rosa