

Exploration des données : un environnement législatif en évolution

[Par Lionel Maurel](#)

Journée ADBU-AEF

« Quelle(s) stratégie(s) de recherche face
à la massification des données »

02/12/2014

A l'origine : « Information Wants to Be Free »

L'INFORMATION
VEUT ÊTRE
LIBRE *

*slogan non copyrighté

Le domaine public de l'information de l'UNESCO ([2011](#))

L'UNESCO soutient fortement l'accès au domaine public informationnel ou "indivis mondial de l'information". Ce domaine public informationnel est constitué par l'information publiquement accessible, dont l'utilisation ne porte atteinte à aucun droit légal, ne viole aucun autre droit communautaire (par exemple les droits des populations autochtones) ou n'enfreint aucune obligation de confidentialité.

Le domaine public informationnel englobe donc l'ensemble des oeuvres ou objets de droits apparentés qui peuvent être exploités par quiconque sans autorisation, par exemple parce que la protection n'est pas assurée en vertu du droit national ou international, en raison de l'expiration du délai de protection, ou à cause de l'absence d'un instrument international assurant la protection dans le cas des oeuvres ou objets de droits apparentés étrangers.

Les idées, les faits, les données restent « de libre parcours »

[Image par Geralt CCO](#)

KNOWLEDGE



INFORMATION INFORMATION INFORMATION



DATA DATA DATA DATA



Une condition fondamentale pour la liberté de la recherche

L'extraction des informations d'un livre...

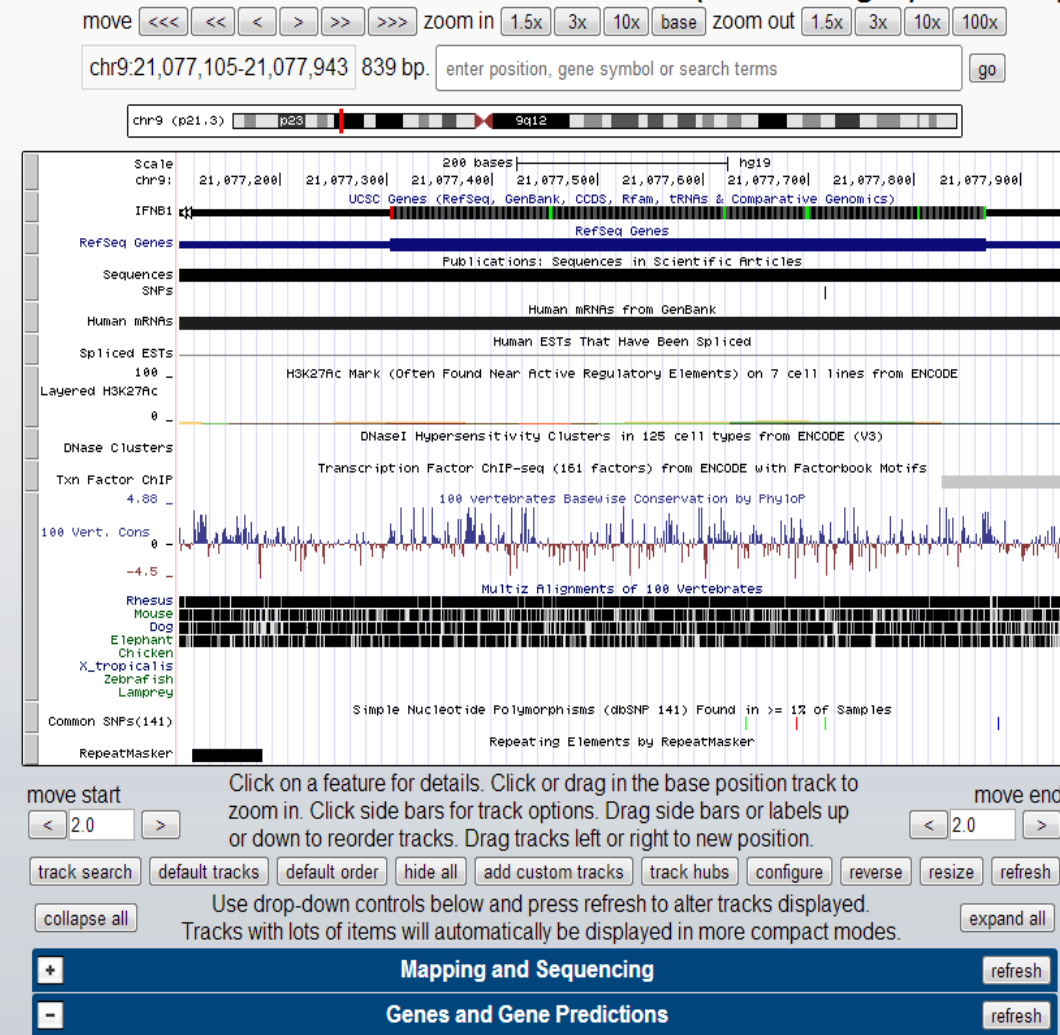


[Image par Valentin de Bourgoynes. Domaine Public.](#)

... est un acte fondamentalement libre, lié à la
« liberté de lire » elle-même.

Mais avec le numérique, tout change...

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly



Projet Text2Genome : 3 ans de négociations complexes pour pouvoir constituer un corpus de 3 millions d'articles.

Depuis les années 90, les données sont saisies par le droit...



Image par Postmasters. CC-BY-SA.

[Directive 96/9/CE sur la protection juridique des bases de données](#) : protège les investissements réalisés pour constituer une base de données par un droit *sui generis*, portant sur la structure de la base et empêchant les « extractions substantielles »

La fouille de données ou de textes nécessite généralement que des copies de contenus soient effectuées.



Or la réalisation de copies déclenche en principe l'application du droit d'auteur (copyright) et les exceptions traditionnelles conviennent mal (citation, illustration de la recherche, copies techniques provisoires, etc).

Dans ce contexte, volonté des éditeurs scientifiques d'imposer des licences pour le Text et Data Mining

Rights and permissions

- » [Foreign rights news](#)
- » [Obtaining permissions](#)
- » [Anti-piracy strategies for Springer eBooks](#)
- » [Springer's text- and data-mining policy](#)
- » [Rights and permissions at book fairs](#)
- » [Contacts](#)

Springer's text- and data-mining policy

Springer grants text- and data-mining rights to subscribed content, provided the purpose is non-commercial research.

Demand for TDM has been low up to now, but is steadily increasing. As the volume of scientific publications increases and TDM software tools improve, Springer acknowledges the need for a more formalized process to enable TDM, and strives to make this as simple as possible for researchers.

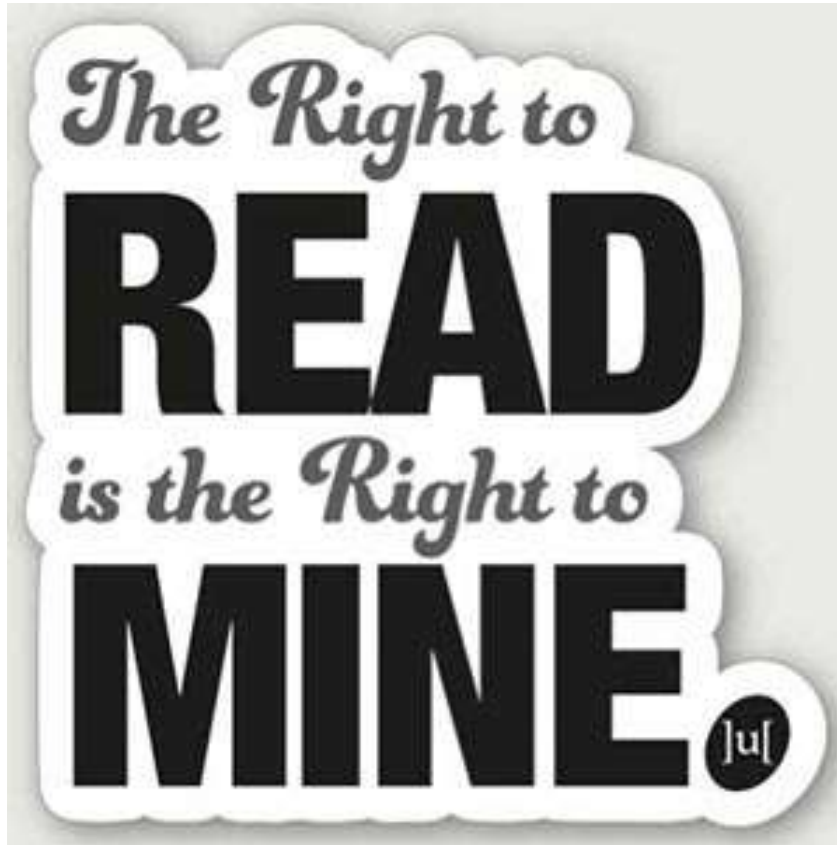
Springer grants text- and data-mining rights to subscribed content to researchers via their institutions, provided the purpose is non-commercial research. The selection and refinement of desired articles can be conducted by using existing search methods and tools, such as PubMed, Web of Science and Springer's Metadata API, among others.

Individual researchers are encouraged to download subscription and open access content for TDM purposes directly from the SpringerLink platform. No registration or API key is required. Full-text content can be accessed easily and programmatically at friendly URLs based on the content's Digital Object Identifier (DOI).

Researchers who do not have subscription access to SpringerLink can send requests for TDM access to the contact listed below. These inquiries will be considered on a case by case basis.

Cette solution contractuelle a été fortement poussée au niveau européen ([processus Licences For Europe](#) => Echec en 2013).

[Position de LIBER](#), la ligue des bibliothèques de recherche européennes en réaction à la politique de TDM d'Elsevier.



Nombreuses objections au système des licences de TDM. Volonté d'inscrire ces usages a minima dans une exception au droit d'auteur.

En Angleterre, [une nouvelle exception](#) introduite en 2014 pour sécuriser les « analyses computationnelles » effectuées dans un cadre non-commercial.



Aux Etats-Unis, de larges ouvertures en matière d'exploration de données, considéré comme un usage transformatif couvert par le fair use (usage équitable), y compris parfois dans un cadre commercial ([jurisprudence Google Books - 2013](#)).

Pendant ce temps, en France...

On ne peut ignorer que l'activité d'exploration de données se développe sur la création et l'investissement d'autrui, des bases de données et des oeuvres. Sans cet apport initial, l'exploration de données ne pourrait ni exister, ni développer de valeur. En ce sens, dans l'ordre des organismes vivants, l'exploration de données est de nature « parasitaire ».

[Rapport CSPLA sur le Text et Data Mining. 2014](#)



Et la conclusion controversée d'une licence nationale avec Elsevier comportant [une clause sur le Text et Data Mining.](#)

Prochain enjeu : la réforme de la directive européenne sur le droit d'auteur.



Révision annoncée par la Commission européenne à l'horizon 2016. Plusieurs pistes possibles pour sécuriser les pratiques d'exploration de données, mais contexte encore très incertain.

Sans attendre, une piste du côté de l'Open Science ?



Open Access
+ Open Data
+ Open
Source =
Open
Science ?

Lien direct avec la politique de diffusion des données de la recherche et le développement d'infrastructures de dépôt.

Les données de la recherche en Open Data ? Une solution pour le Text et Data Mining ?



[2015 : annonce d'une grande loi sur le numérique en France.](#) Un statut spécial serait créé pour des jeux de données publiques considérées comme « d'intérêt général », pour accélérer leur passage en Open Data. Mais les données de la recherche sont encore très largement exclues du processus d'ouverture et risquent de rester sous un régime d'exception.

Conclusion...



WE WANT YOU!