

Quel cadre juridique pour **mettre** **à disposition** des données publiques?

CC 4.0 BY-NC-ND



Danièle Bourcier

Directrice de recherche CNRS

Member of the Comité d’Ethique du CNRS
(COMETS)



LOI

Crédits : Jwyg - Flickr (CC BY-SA 2.0)

Data & Opendata & data sharing

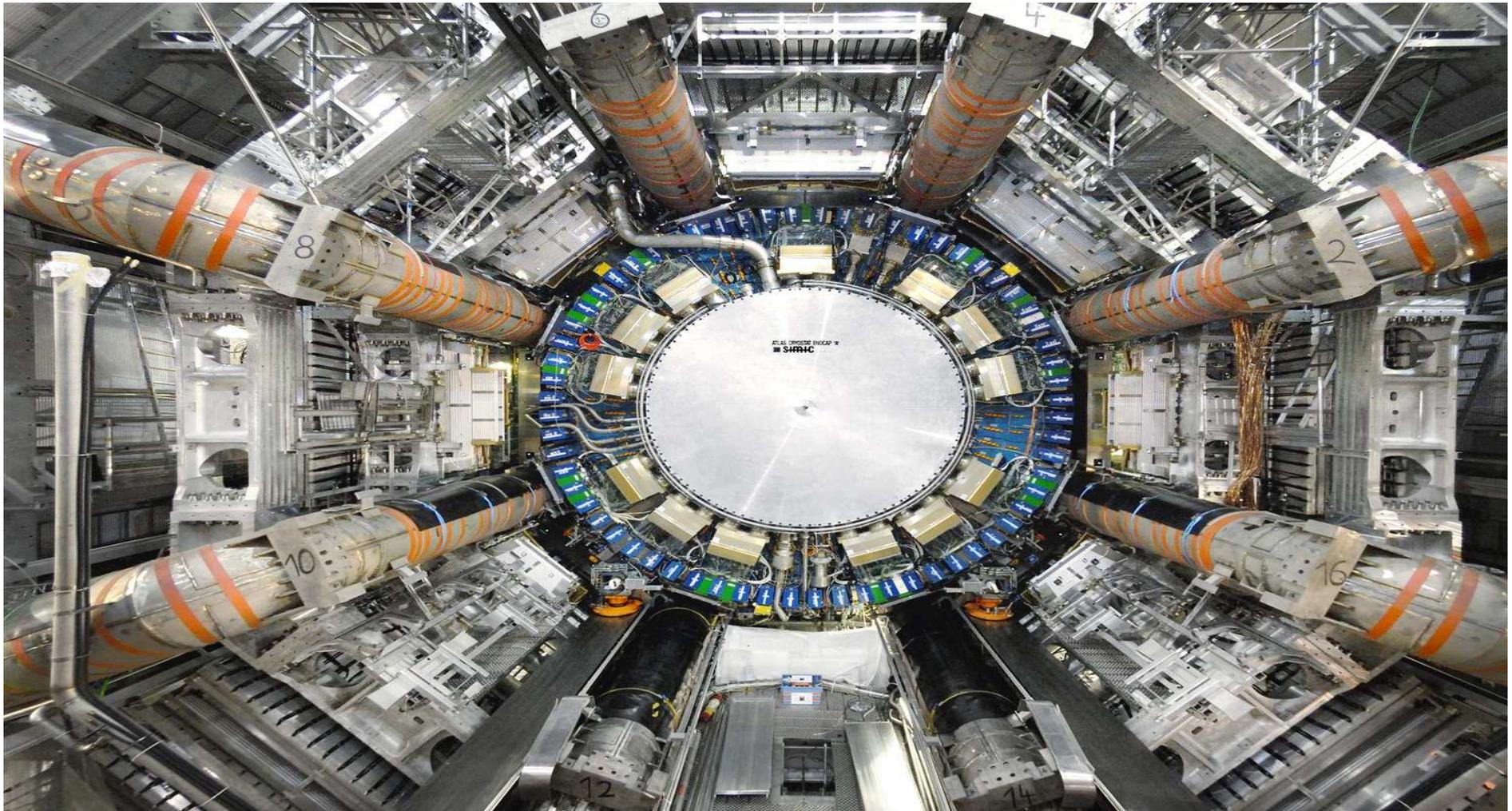
- **Open data** policies by governments, institutions, organisations

ex loi sur le numérique

- **Datasharing** policies by communities

ex Global alliance

very Big data?



“The *Global Alliance* *for genomics and health*”, 2013

- A **successful example** of open networking initiated by researchers is an initiative in the field of human biology
- 285 member institutions from 30 countries
- that have decided to contribute to facilitate the sharing of scientific data by establishing standards, sharing experience and best practices and establishing a framework for responsible data sharing.

today?

- a **paradigm shift** - from hypothesis - driven-research to data-driven research
- data as a (separate) **part of research**
- high value **of** data
- **from data to open data**

De la loi à la directive

Loi 17 juillet 1978 dite loi CADA
Liberté d'Accès aux documents et à la
réutilisation des informations du secteur public

Directive 2003/98/CR 17 nov 1978
Libre Réutilisation des données (informations)
publiques

De la directive à la loi

- 1- Ord. 6 juin 2005 a transposé la première directive PSI datant de 2003.
- 2- l'UE a publié une nouvelle directive en 2013 qui devait être transposée (« au rabais » d'après Regards citoyens) au plus tard le 18 juillet 2015...
- 3- Loi 2015 relative à la gratuité et aux modalités de réutilisation...

Un document devrait être considéré comme présenté sous un format lisible par machine s'il se présente dans un format de fichier structuré de telle manière que des applications logicielles puissent facilement identifier et reconnaître des données spécifiques qu'il contient et les en extraire.

Télescopage de deux textes

- Projet de Clotilde Valter

voté par ordonnance

« Réforme de l'Etat »

- Projet d'Axelle Lemaire

« l'open data ne vaut pas une loi »

voie normale

« République numérique »

Un cas de rétention de deux ans

La cagnotte d'environ **150 millions d'euros** inscrite au budget du ministère de l'Intérieur et répartie arbitrairement et en toute opacité entre les sénateurs et les députés échappait jusqu'en 2013 à toute publication.

Au terme de deux années de procédure judiciaire, [Hervé Lebreton](#), un professeur de mathématiques a finalement réussi à obtenir un **fichier PDF**.

Des documents aux données

- Loi CADA : documents
- Loi de 2003: informations
- Ord. de 2015 : données (geo, statistique,

Les textes sont ils des données ?

Qu'est ce que une donnée?

géographique, statistique
lisible par la machine

Un document est il une donnée?

Un document devrait être considéré comme présenté sous un **format lisible par machine** s'il se présente dans un format de fichier structuré de telle manière que des applications logicielles puissent facilement identifier et reconnaître des données spécifiques qu'il contient et les en extraire.

- **LOI n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public**

(11 articles)

3 Principes

format ouvert

gratuité

liberté de réutilisation (musées, archives,
bibliothèques OUI)

Format ouvert

« ces informations le sont, si possible, dans un standard ouvert et aisément réutilisable, c'est-à-dire lisible par une machine. »

Gratuité

- « La réutilisation d'informations publiques est gratuite. Toutefois, des administrations peuvent établir une redevance de réutilisation lorsqu'elles sont tenues de couvrir par des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service public. »

Réutilisation par défaut

- À des fins commerciales **ou** non commerciales
- Coûts **marginiaux** de reproduction, de mise à disposition, et de diffusion
- Exception: météo, données juridiques, statistiques

Rapport Trojette

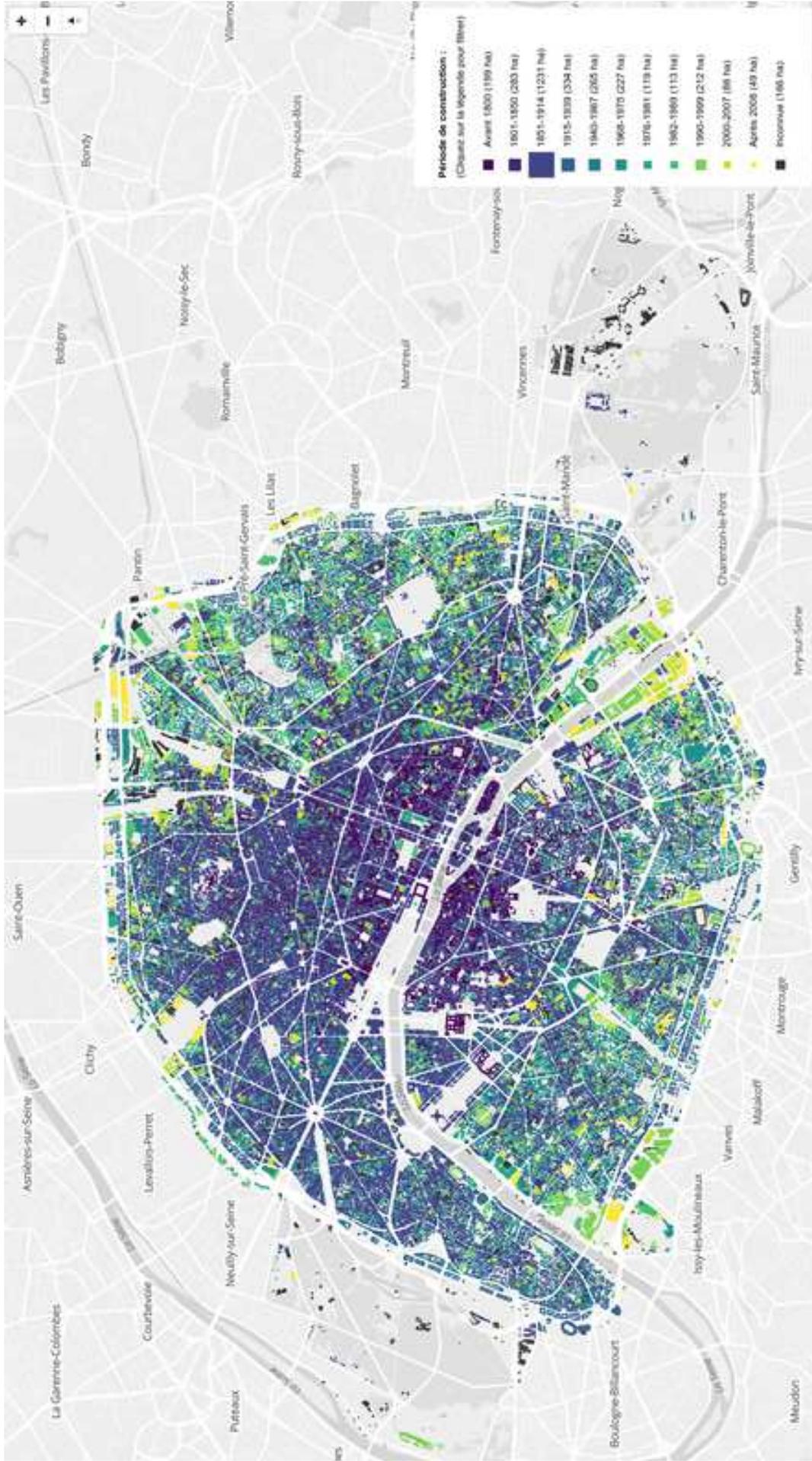
« Ouverture des données publiques
Les exceptions au principe de
gratuité sont-
elles toutes *légitimes*? »

- Les entreprises privées peuvent capturer des **biens publics** en associant aux redevances des **accords d'exclusivité** pendant 10 ans (art. 8).
- Régime d'exception pour la numérisation des **œuvres culturelles**, sans aucune limite temporelle d'exclusivité (art. 2.1).
- Non gratuité : reculade visant à protéger les quelques dizaines de **redevances survivantes**, que ce soit à l'IGN, Météo-France ou l'INSEE (art. 3).

Data.gouv

Appli format INSPIRE

- Tout producteur de données au « format Inspire » peut donc, en quelques clics, partager ses données sur data.gouv.fr et y référencer son portail.
- Il suffit de [se connecter sur la passerelle](#)



EX : Data.BNF.fr

- **technique** : respect des standards du [web sémantique](#) et ouverture sur le web de données définis par le W3C : [linked data](#).
- **juridique** : les données sont placées sous [Licence ouverte](#) de l'État autorisant la réutilisation libre, y compris commerciale, avec mention de la source.
- Le projet data.bnf.fr utilise des données produites dans des formats divers, notamment **Intermarc** pour les catalogues de livres, **XML-EAD** pour les inventaires d'archives, et **Dublin Core** pour la bibliothèque numérique.

- Ces codes de fonctions sont répertoriés :
<http://data.bnf.fr/vocabulary/roles>.

Exemple : Baudelaire est traducteur de « Dix contes d'Edgar Poe »

(<http://catalogue.bnf.fr/ark:/12148/cb311263053>).

La notice bibliographique du document est liée à la notice d'autorité de l'auteur « Charles

Baudelaire » avec un code de fonction 0680 qui correspond au rôle de traducteur. Exemple en

Intermarc : 700 \$311890582 \$w 0

b.....\$aBaudelaire\$mCharles\$d1821-1867\$40680

Questions juridiques concernant les données textuelles

Text & data mining

- ▶ La notion de TDM regroupe des techniques de collecte de données, traitement automatique de la connaissance issue de ces données, présentation automatique des résultats générés par une demande d'un utilisateur.
- ▶ Le TDM permet d'explorer un volume croissant de données en traitant, par des techniques d'intelligence artificielle et de TAL, les données et les métadonnées associées.
- ▶ Le TDM a pour objet l'extraction ou la mise en perspective d'un savoir ou d'une connaissance nouvelle.

Pas définition légale
Pas de statut légal

x



x



Droits d'auteur et droits d'éditeur sont-ils opposables au TDM ?

Disposition législative votée en février 2016

Une exception au droit d'auteur en faveur du
Text and Data Mining (TDM) – Ce nouvel article
18 bis ajoute à la liste de l'article L.122-5 du
Code de la propriété intellectuelle une dixième
exception au droit d'auteur

- ▶ Pour toute ressource il convient de **s'assurer préalablement des détenteurs de droits** sur la ressource initiale.
- ▶ Droits d'auteur et droits d'éditeur
 - ▶ un certain nombre de corpus textuels ne peuvent être mutualisés car cette condition n'est pas respectée ;
 - ▶ problème : les droits d'auteur ou d'éditeur ne sont pas toujours clairs en cas de butinage sur le Web ;
- ▶ Droits d'auteur et droits d'éditeur sont-ils opposables au TDM ?

- ▶ **Propriété intellectuelle**

- ▶ Licence d'utilisation des sources
- ▶ Droits du fournisseur sur les données
- ▶ Droits des producteurs de bases de données
- ▶ Droit d'auteur

- ▶ **Réglementations spécifiques applicables**

- ▶ Loi informatique et libertés
- ▶ Droits des tiers
- ▶ Secret statistique
- ▶ Secret commercial

Qui est titulaire des droits?

Le chercheur ?

Le laboratoire ?

Le consortium gérant la plateforme
?

L'institution de rattachement du
laboratoire ?

Le financeur ?

Différence entre droit moral et droit patrimonial?

Perpétuel et inaliénable

Limité dans le temps (?)

Une politique publique d'établissement à étudier :

transmettre les droits patrimoniaux au
laboratoire ou à l'organisme de rattachement.

- ▶ Qui est propriétaire des droits sur le numérique en cas de publication papier préalable il y a 10 ou 15 ans ?

Une recommandation à faire aux chercheurs : limiter autant que ce peut la cession de droits à des éditeurs sur des versions numériques.

- les droits d'auteur ou d'éditeurs sont-ils opposables à l'open data dans le domaine scientifique ?

Cas particulier FRANTEXT

Les enseignants chercheurs ont un statut d'exception par rapport au droit d'auteur

Le conseil d'état a appliqué la loi

Quels droits sur les résultats de TDM ?

- Les fournisseurs de données?

Les intervenants à la création

Language Resources	Content creator	LR creator	Lang. theorist	Legal basis
Original works	x			content
Corpora	x	x		arrangement/ Investment (sui generis)
Semantic resources		x	x	Content/ arrangement
Annotation		x		content

Consentement de l'auteur?

- Tout **usage** d'une ressource linguistique doit avoir le **consentement de l'auteur**
- **Usage** : Télécharger un corpus, annoter, réutiliser une ontologie, la fusionner, l'utiliser pour annoter un corpus, ...

Protection de l'œuvre dérivée

- Une création est dans le Domaine public (+ 70 ans) mais il existe de récentes éditions (traductions, etc)
- Elle est protégée

Collection d'annotations et d'ontologies

- Les items sont non protégeables
- La collection entière est protégeable

Suivant le droit européen
sur les bases de données

Questions pièges

- Une théorie linguistique n'est pas protégée
- Mais la forme de l'expression l'est
- Des schémas d'annotations formatées seraient protégeables

Recommendation of the COMETS, in May 2015

« Ethical challenges of the sharing
of scientific data »

Recommendation 1

- to join the movement of open access in accordance with the **practices specific to each discipline.**
- In certain disciplines, embargoes limited to the use of data by the researcher may be requested temporarily or minimum guarantees concerning the identification of the researcher or the team using the data and the purpose of their research.

Recommendation 2

- Researchers and research staff should be **trained** in the ethical dimensions of data management, especially respect for privacy, intellectual property and the **quality and integrity** of data.
- They should be **informed** of the current state of and changes in legal rules concerning the responsible sharing of data used.

Recommendation 3

- To Identify barriers to ethical data sharing (intellectual property of data and sui generis status of databases) to promote science commons and establish scientific data as general interest data.

Recommendation 4

- The work of making usable data from raw data must be recognised in the assessment and promotion of personnel involved in this work.
- A section be added on these activities in the activity report and the annual record of researchers' activity.

Recommendation 5

- the authors of publications should **include researchers** who have contributed data
- and correctly report the **data sources** used.

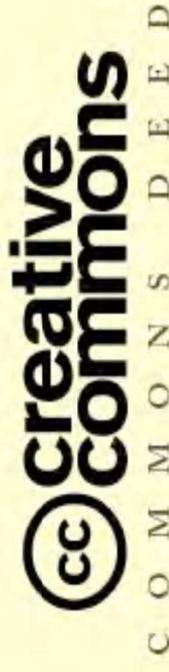
Recommendation 5

- Researchers should be careful when they **transfer proprietary rights** over **their data or databases** to third parties.
- **Access to data** relating to a publication must **not be limited by copyright** imposed by the journal. It must be **provided** with the article in an associated file but must **remain available** to researchers for repeat analysis and publication in any journal.

Recommendation 6

- The **HAL open archives** be preferred for submission of data on which to support the **publication of research results**.
- Researchers must be **able to choose the conditions for reuse**, through open licenses such as Creative Commons.

Commons deed



Attribution 2.5

You are free:

- to copy, distribute, display, and perform the work
- to make derivative works
- to make commercial use of the work

Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Recommendation 7

- CNRS encourage its researchers to participate in **international standards bodies** to address metadata.
- It strongly encourages the use of **unique and persistent identifiers** to enable tracking and data interoperability.

Licence ouverte

Pour les producteurs et aux réutilisateurs des données publiques

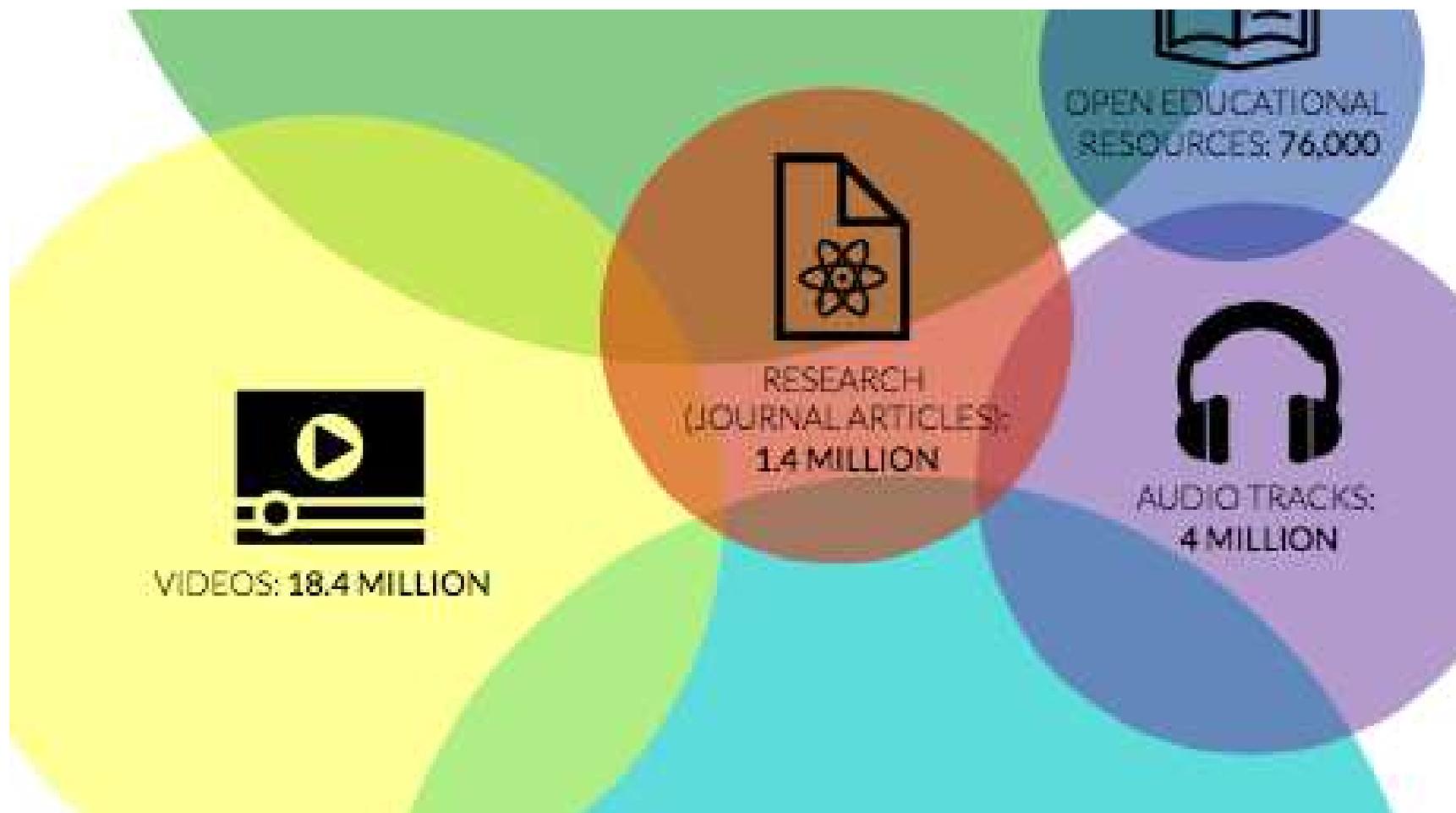
Une licence qui s'inscrit dans un contexte international en étant compatible avec les standards des licences Open Data développées à l'étranger et notamment celles du gouvernement britannique (Open Government Licence) ainsi que les autres standards internationaux (ODC-BY, CC-BY 2.0).

Licences Creative Commons

- un complément au droit d'auteur
 - pour encourager la création et le partage
 - pour accorder plus de libertés que le régime minimum du droit d'auteur
- une information au public que certaines utilisations sont autorisées à l'avance
- un service web qui associe des contrats flexibles à des métadonnées pour la recherche automatique



Œuvres sous CC par catégories



UE Directive 2014/02/26

DIRECTIVE 2014/26/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 26 February 2014

on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market

3. Rightholders shall have **the right to grant licences for non-commercial uses of any rights, categories of rights or types of works and other subject-matter that they may choose.**



ASCRIBE

Registration of works and automatic downstream uses tracking with blockchain technology. Advantage

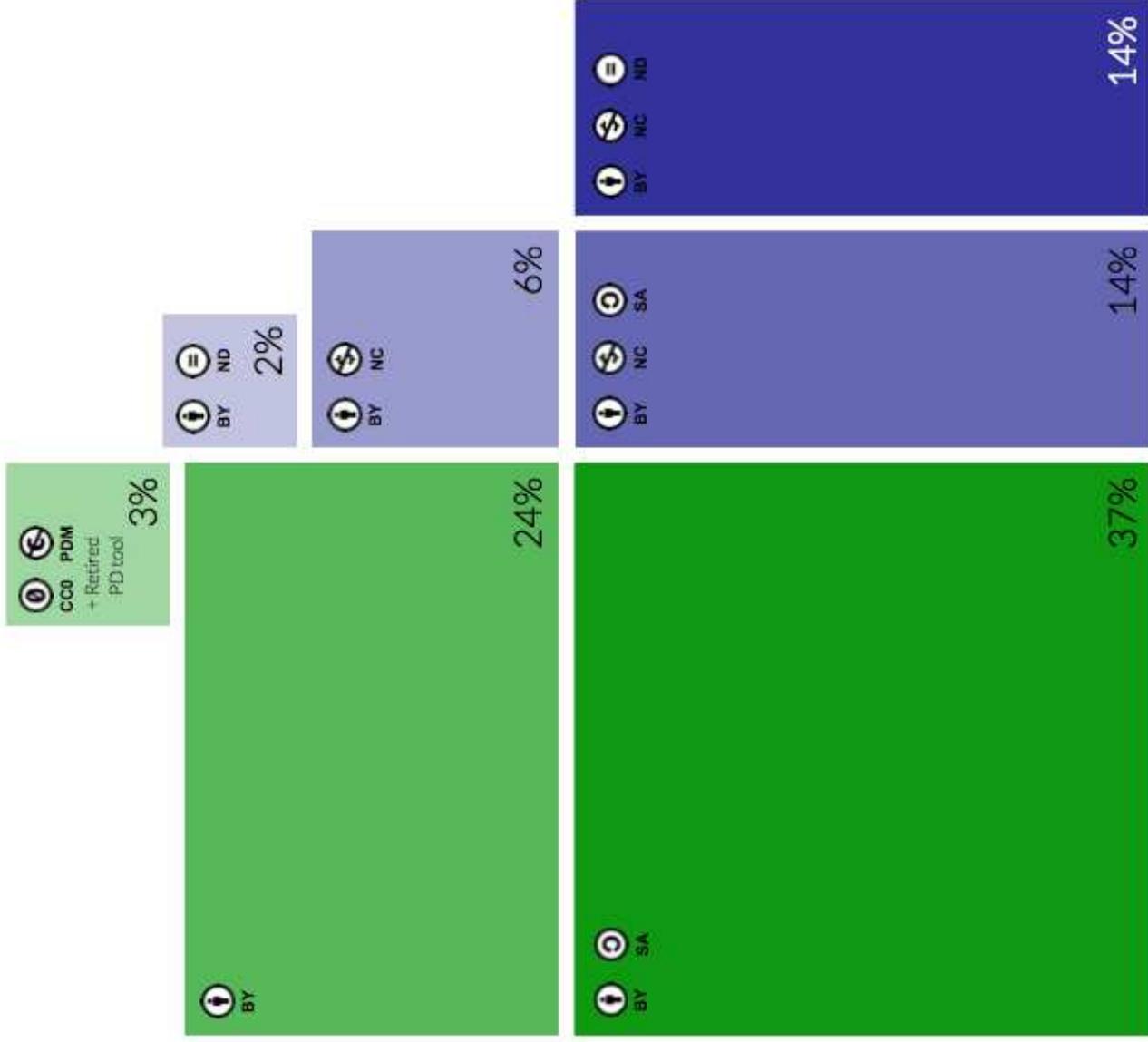
- Allocation procedure and verification easier.
- Best accessibility.
- Use monitoring.
- Share easily work.

Ethic of science committee (CNRS)

- **7.** COMETS advocates that, where there is no **Recommendations** open international database dedicated to this data, the HAL open archives be preferred for submission of data on which to support the publication of research results. Researchers must be able to choose the conditions for reuse, through open licenses such as **Creative Commons**.
- **8.** ...

Des utilisations collectives de la licence Creative Commons

- PLOS (<http://www.plos.org/>), CERN (<http://creativecommons.org/tag/cern>), Biomedcentral (<http://www.biomedcentral.com>) et certaines branches de la revue Nature. Le CNRS en France a commencé à signaler depuis 2014 à ses chercheurs qu'il est possible de partager leur travaux sous une licence CC (<https://hal.archives-ouvertes.fr/search/index>).
- - De plus en plus de gouvernements ont adopté et recommandé l'utilisation des licences CC pour la diffusion et l'accès à leurs données ou à leurs plateformes (<https://wiki.creativecommons.org/Government>). Depuis 2014, certaines administrations françaises ont commencé à diffuser données et plateformes publiques sous CC : <http://www.gouvernement.fr/> / <http://www.culturecommunication.gouv.fr/Mentions-legales> / <http://www.culture.fr/Mentions-legales>
- - Concernant le secteur culturel, public et/ou privé, ériger des murs autour des savoirs et des œuvres ne paraît pas être la meilleure solution pour le rayonnement de la culture nationale
- D'abord, les plateformes dédiées à la diffusion de contenus tel que <http://vimeo.com/creativecommons>, <https://www.flickr.com/commons>, <http://www.europeana.eu/>, ect... ont permis, *techniquement*, la publication de contenus sous licence CC. Aujourd'hui, en France, plusieurs expériences ont été lancées comme <http://www.sacem.fr/cms/home/createurs-editeurs/creative-commons/experience-pilote-sacem-creative-commons> ou encore <http://creativecommons.fr/conclusion-dun-partenariat-historique-entre-creative-commons-france-et-le-ministere-francais-de-la-culture/>



MORE OPEN MORE RESTRICTIVE

