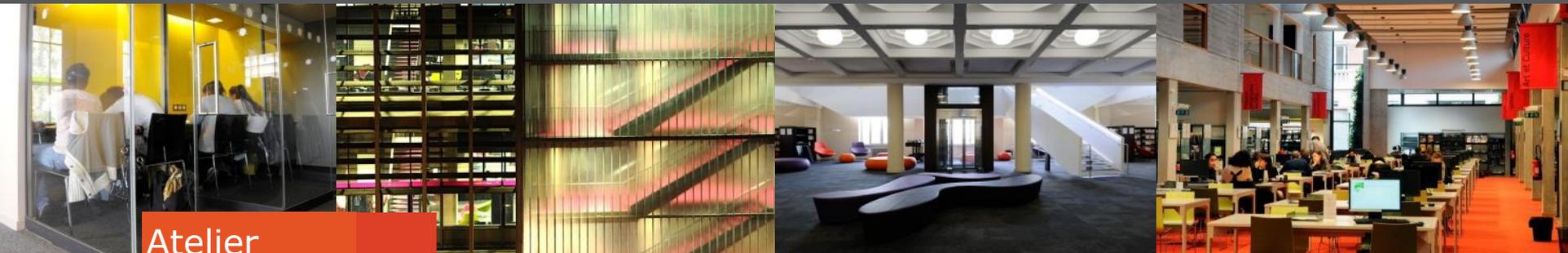


# Université Claude Bernard Lyon 1

## Service Commun de la Documentation



Atelier  
Sensibilisation  
aux Données de  
la Recherche

+

+

+

+

+

+

Données de la Recherche – version janvier 2020



## Sommaire

- Vers une définition des données de la recherche
- Contexte
- Où trouver des données ?
- Où déposer des données ?
- Valoriser ses données
- Qu'est-ce qu'un plan de gestion des données ?
- En conclusion

## Définition de l'OCDE :

« Enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour **valider des résultats de recherche.** »

## Définition de l'Association des archivistes français (AAF) :

« Les données de la recherche sont l'ensemble des informations et matériaux produits et reçus par des équipes de recherche et des chercheurs. Elles sont collectées et documentées à des fins de recherche scientifique. A ce titre, **elles constituent une partie des archives de la recherche.** »

## Et vous, quels types de données de recherche collectez-vous ?

- Données d'observation :
  - Descriptives, collectées à un instant T
  - Uniques, impossibles à reproduire
- Données expérimentales
  - Obtenues à partir d'équipements de laboratoire, suivant une méthodologie définie
  - Potentiellement reproductibles mais coût parfois prohibitif
- Données computationnelles ou de simulation
  - Générées par des modèles informatiques ou de simulation
  - Souvent reproductibles
- Données dérivées ou compilées
  - Issues du traitement ou de la combinaison de données brutes
  - Souvent reproductibles mais coûteuses
- Données de référence
  - Collection ou accumulation de petits jeux de données

# Contexte réglementaire (1/3)

## Une impulsion européenne :

« La stratégie de la Commission en matière de **libre accès aux données** et de circulation des connaissances se fonde sur une approche qui exclut tout nouveau paiement lors de l'accès à des informations déjà payées par des dépenses publiques. » (2012)

**H2020** : <http://www.horizon2020.gouv.fr/>



- Programme de financement de la recherche et de l'innovation de l'UE (2014)

## Avant une prise de relais par la France :

- Loi pour une République numérique (2016)
- Plan national pour la science ouverte (2018)
- Politique du principal financeur national de la recherche : l'ANR (2019)



# Contexte réglementaire (2/3)

## [Loi pour une République Numérique n° 2016-1321](#) :

Favorise l'accès ouvert aux publications et aux données de la recherche (art. 30), ainsi que la fouille de texte et de données (art. 38 – FTD ou en anglais Text Data Mining / TDM)

### **- Article 30 - consacré à l'Open Access**

Nouveau « **droit d'exploitation secondaire** » au profit des chercheurs leur garantissant la possibilité d'auto-archiver leurs écrits, en dépit des clauses des contrats d'exclusivité qu'ils auraient pu signer avec des éditeurs, au terme d'un délai de 6 à 12 mois selon les disciplines. Le texte contient aussi un passage sur les **données de la recherche**, qui instaure un mécanisme de protection contre leur captation par les éditeurs scientifiques : « L'éditeur d'un écrit scientifique (...) ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication. »

### **- Article 38 - nouvelle exception au droit d'auteur en faveur du Text et Data Mining**

Les « données de la recherche » vont constituer une nouvelle catégorie juridique.

### **Pour aller plus loin :**

Billet de blog : <https://scinfolex.com/2016/11/03/quel-statut-pour-les-donnees-de-la-recherche-apres-la-loi-numerique/>

## Contexte réglementaire (3/3)



[Plan pour une science ouverte](#) – Juillet 2018

Il intègre une partie sur l'ouverture des données :

« Les données de la recherche sont le pilier des connaissances futures... [il faut] mettre en place une **obligation de diffusion en accès ouvert des données issues de recherches financées par appels d'offres sur fonds publics** » ;

### Premières retombées :

[Appel Flash science ouverte](#) : pratiques de recherche et données ouvertes

Juillet 2019 : [25 projets lauréats](#) d'un appel à projets (2,3 millions €),

Exemple : Neuroweblab, une sorte de « wikipedia » de la neuroscience.

# Enjeux

- Enjeux pour la communauté scientifique :
  - Eviter la perte de données
  - Assurer la reproductibilité de la Science ([voir vidéo en lien](#))
  - Faciliter de nouvelles recherches
  
- Enjeu pour le chercheur :
  - Gagner en visibilité
  
- Enjeux économiques :
  - Rentabiliser les coûts (éviter de financer 2 fois le même projet)
  - Accélérer l'innovation
  
- Enjeux sociétaux et patrimoniaux

# Une ouverture encadrée

## Le principe des données « FAIR » comme standard de définition des données ouvertes

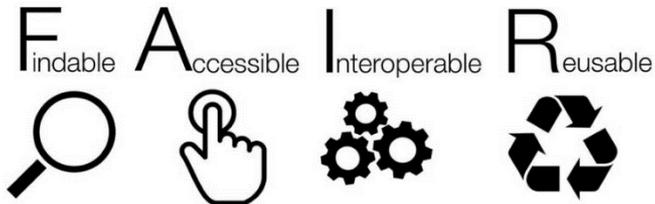
D'après : Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. [doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Findability (identifiant pérenne)

Accessibility (entrepôts, embargo)

Interoperability (formats)

Reusability (licences)



**Pour aller plus loin :** [Document INRAE sur FAIR](#)

# Où consulter des données ?

- **Sur Google... Datasetsearch :**

Google Dataset Search Bêta

- **Sur Data Cite...**

<https://search.datacite.org/>



- **Sur Elsevier...**

<https://datasearch.elsevier.com>

DataSearch Beta

<https://data.mendeley.com/>

- **Sur le Web of Science... Data citation index**

# Où déposer des données ?

- Choisir un entrepôt : annuaire [re3data](https://re3data.org)



- Les entrepôts multidisciplinaires :



Source : PAIN M., Critères d'évaluation d'un entrepôt de données, <https://www.ccsd.cnrs.fr/2016/11/comment-diffuser-mes-donnees-de-recherche-2/> [en ligne] . Consulté le 28/11/2019

# Où déposer des données ?

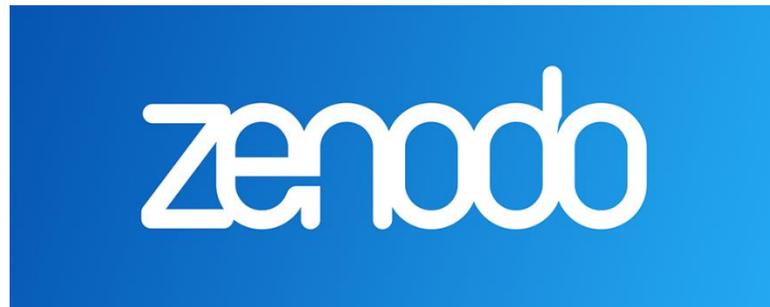
- **Un entrepôt libre pour les supplementary materials : Dryad**  
(<https://datadryad.org>)
- Entrepôt de dépôt pour les données en lien avec un article publié ou en cours de publication.
- Payant : 120\$ par dépôt.
- Obligatoirement sous licence Creative Commons CC0 (tous droits ouverts).
- Dépôt pour des données en biologie au départ, puis s'est ouvert aux autres domaines



*Données de la Recherche – mis à jour janvier 2020*

# Où déposer des données ?

- **Un entrepôt libre pour un projet à l'échelle européenne : Zenodo (<https://zenodo.org/>)**
- Approche multidisciplinaire, avec un système de « communautés ».
- Ajustement des droits d'accès : embargo ou accès à la demande.
- Jusqu'à 50 Go par dataset (dépôts multiples possibles, la capacité de dépôt est quasiment infinie)
- Lié aux capacités du CERN – pas d'engagement absolu en matière de pérennité (garantie déclarée sur 20 ans)



# Où déposer des données ?

- **Des entrepôts avec services mais non-neutres**

**Figshare** - <https://figshare.com/>

- Lié à l'éditeur Macmillian Publishers
- Services pour le dépôt, la gestion, la valorisation

**Mendeley Data** - <https://data.mendeley.com/>

- Géré par l'éditeur Elsevier (ScienceDirect)
- Services pour le dépôt, la gestion, lien avec les publications



# Valoriser ses données : data papers

- Data journals :

Scientific Data (Nature), Geoscience Data journals (Wiley), Data in brief (Elsevier),



- Stratégie des éditeurs commerciaux :

Les données de la recherche sont un terrain sur lequel ils veulent se positionner après la voie dorée (système auteur/payeur)

- Riposte des établissements qui créent leurs propres entrepôts de données (Inra, CIRAD, IRD etc.)

<https://data.inra.fr/>

Voir aussi outil [Data Paper de l'Inra](#)

## ***Pour en savoir plus :***

Chartron G. Stratégie, politique et reformulation de l'Open Access. Revue française des sciences de l'information et de la communication [En ligne]. 1 janv 2016 [cité le 17 nov. 2017];(8). Disponible: <https://rfsic.revues.org/1836>.

*Données de la Recherche – mis à jour janvier 2020*

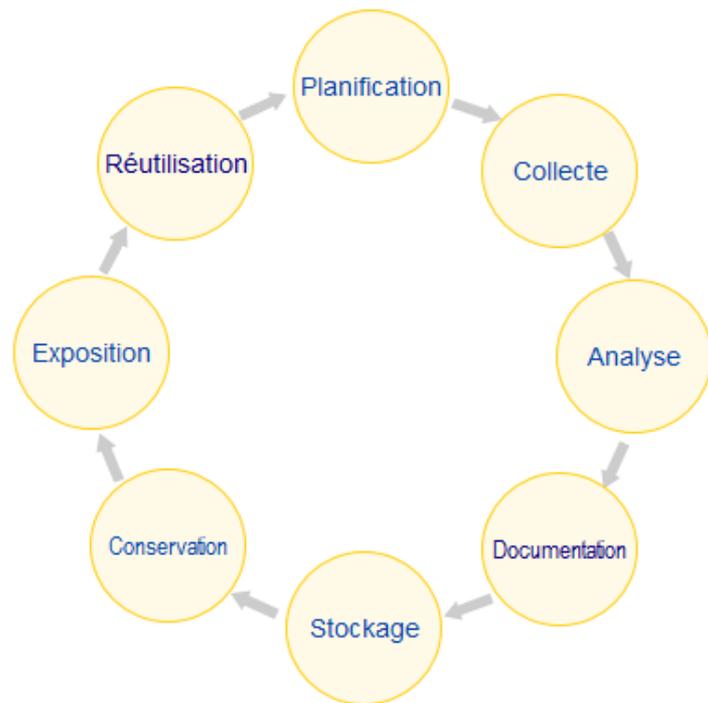
## Obstacles à surmonter

- Questions juridiques et éthiques
  - Protection des données à caractère personnel
  - Droit de propriété industrielle
- Objections pratiques
  - Manque de compétences
  - Manque de temps
  - Coûts
  - Manque d'infrastructures adaptées
- Réticences liées à des aspects sociaux-culturels
  - Concurrence ou coopération ?
  - Absence de reconnaissance dans le processus d'évaluation des chercheurs

# Les plans de gestion des données : une nouvelle donne ?



## Modèle de Plan de gestion des données (PGD)



# Les plans de gestion de données : comment faire ?

## Outils et méthodes

Un plan de gestion des données de la recherche - Méthodes

**DMP OPIDoR : Outil pour faciliter la rédaction de plans de gestion de données :**

<https://www.ouvrirlascience.fr/dmp-opidor-outil-pour-faciliter-la-redaction-de-plans-de-gestion-de-donnees/>



## Pour en savoir plus



**Bienvenue sur DoRANum**

Apprentissage Numérique sur les Données de la Recherche

LE PROJET DORANUM

# En conclusion

- Les données sont un sujet sensible car cela touche au savoir-faire du chercheur (cœur de métier)
- La gestion des données fait partie intégrante du processus de recherche
- La manière de gérer les données varie en fonction :
  - du type de projet
  - de la nature des données
- La France a pris un peu de retard dans l'accompagnement des chercheurs, mais le mouvement est en marche (INRA, IRSTEA, CNRS, certaines universités ...)
- Les doctorants et jeunes chercheurs : des [acteurs à part entière du processus](#).



*Données de la Recherche – mis à jour janvier 2020*