

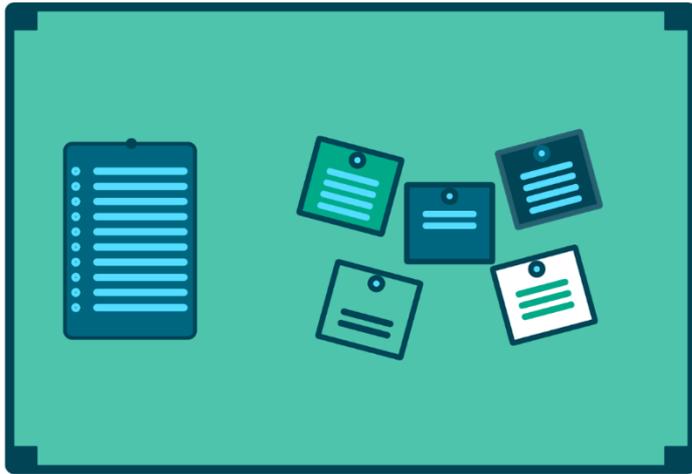


Introduction aux données de la recherche

Principes, outils, méthodes et bonnes pratiques

Au programme

1. Que sont les données de la recherche ?
2. Pourquoi s'intéresser aux données de la recherche ?
Quelques bonnes pratiques de gestion de données
3. Rendre ses données FAIR
4. Le cadre juridique des données de la recherche (introduction)
5. Diffuser, valoriser et réutiliser des données



Consultez en direct les sites internet mentionnés : https://frama.link/DR_URFIST

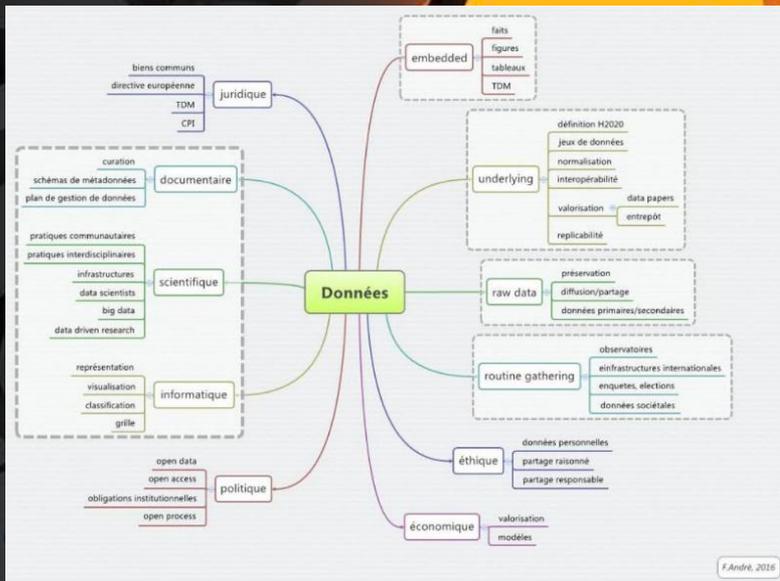
1- Que sont les données de la recherche ?



“Données” de la recherche, les mal-nommées

S. Fayet, « « Données » de la recherche, les mal-nommées », *UrfistInfo*, 2013 ([en ligne](#)).

De quoi parle-t-on ?



F. André, « Les données de la recherche dans le contexte de l’open science », 2016 ([en ligne](#)).

« Pour commencer, pourriez-vous définir ‘données de la recherche’ ? »
Une tentative de réponse

Joachim Schöpfel, Eric Kergosien, Hélène Prost

J. Schöpfel et al., « « Pour commencer, pourriez-vous définir ‘données de la recherche’ ? » Une tentative de réponse », 2017 ([en ligne](#)).

Matériaux scientifiques au statut peu clair



Projets de recherche

Cahier de laboratoire



Notes et relevés effectués par
le chercheur



Échantillons,
prélèvements



Tentative de définition

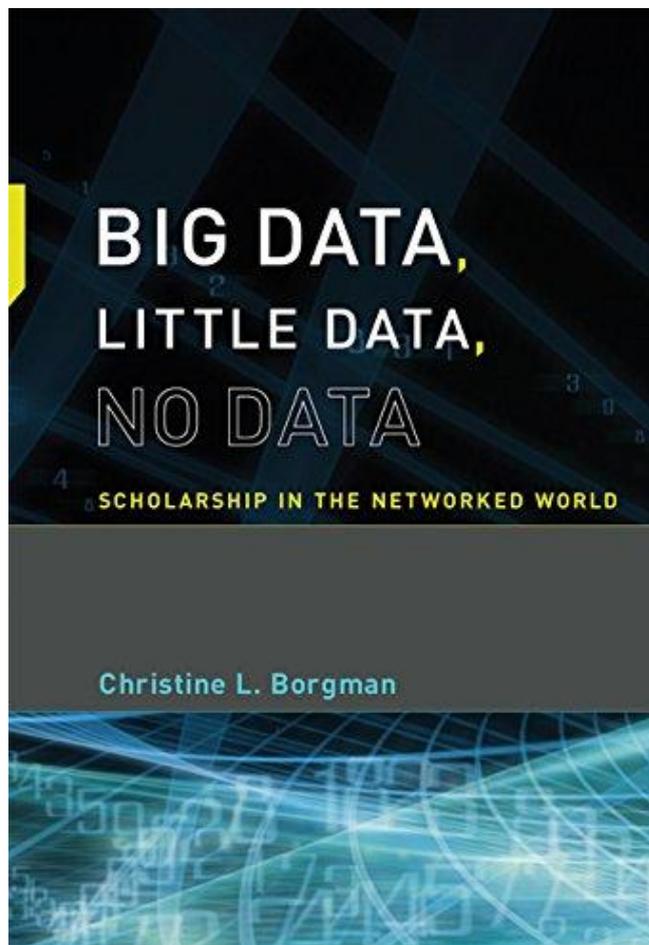
« Les données de la recherche sont définies comme des **enregistrements factuels** (chiffres, textes, images et sons), qui sont utilisés comme **sources principales** pour la recherche scientifique et sont généralement **reconnus** par la communauté scientifique comme nécessaires pour **valider** des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

OCDE, *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, 2007 ([en ligne](#)).



- Données = enregistrements ?
- Données = factuelles ?
- Données = ce qui est nécessaire à la validation de résultats ?
- Données = construites ?

Autres définitions possibles



« Une observation, un objet, un document ou toute autre entité devient une donnée de recherche, dès lors qu'elle est utilisée comme preuve d'un phénomène, c'est-à-dire qu'elle est collectée, analysée et interprétée »

« 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. »

SECTION AURORE



**association
des archivistes
français**

[Source](#)

Un consensus ?

- Les « données » sont en fait des « obtenues »
- Le contexte disciplinaire joue un rôle très important dans l'approche que l'on a de ses données
- Les données englobent finalement « tous les matériaux analysés, interprétés ou utilisés lors d'une recherche, quelle que soit leur origine et la méthode utilisée » ([M. Saby](#))
- Dans certains cas et pour certaines tâches de gestion, cette approche pragmatique des données devra être conciliée avec une approche plus juridique



(Les deux seules) questions à se poser pour identifier ses données

- Quels sont les éléments, numériques ou non, auxquels je tiens vraiment et qui seraient irremplaçables ou très longs à remplacer en cas de perte, de vol ou de problème technique ?
- Si je devais relire et évaluer les travaux de collègues qui travaillent sur un sujet de recherche similaire au mien, de quoi aurais-je besoin pour vérifier leurs résultats ?



Proposition de typologie

D'après les travaux de A. Ciolek-Figiel & F. Cosserat citant F. André ([source](#)), et V. Rebouillat citant C. Borgman ([source](#))

Données d'observation

- capturées en temps réel
- habituellement uniques, impossible à reproduire

Ex.: mesures sismiques, images d'une étoile en fin de vie, enquêtes sociologiques, fouilles archéologiques...

Données computationnelles

- générées par des modèles informatiques
- souvent reproductibles si le modèle est correctement documenté

Ex.: modélisation du changement climatique, « reproduction » du Big Bang, modèles économiques...

+ Données compilées ou dérivées

- issues du traitement de données brutes
- souvent reproductibles mais coûteuses

Ex.: bases de données compilées, corpus textuel préparé pour le TDM...

Données d'expérimentation

- obtenues à partir d'équipements de laboratoire
- souvent reproductibles, parfois coûteuses

Ex.: séquences peptidiques, résultats de réactions chimiques, observations sur des individus en situation de test...

Records (C. Borgman)

- documents témoignant d'un phénomène ou d'une activité humaine
- uniques ou non

Ex.: fonds de photographies, documents d'archives, textes de loi, ouvrages littéraires...

+ Données « de référence »

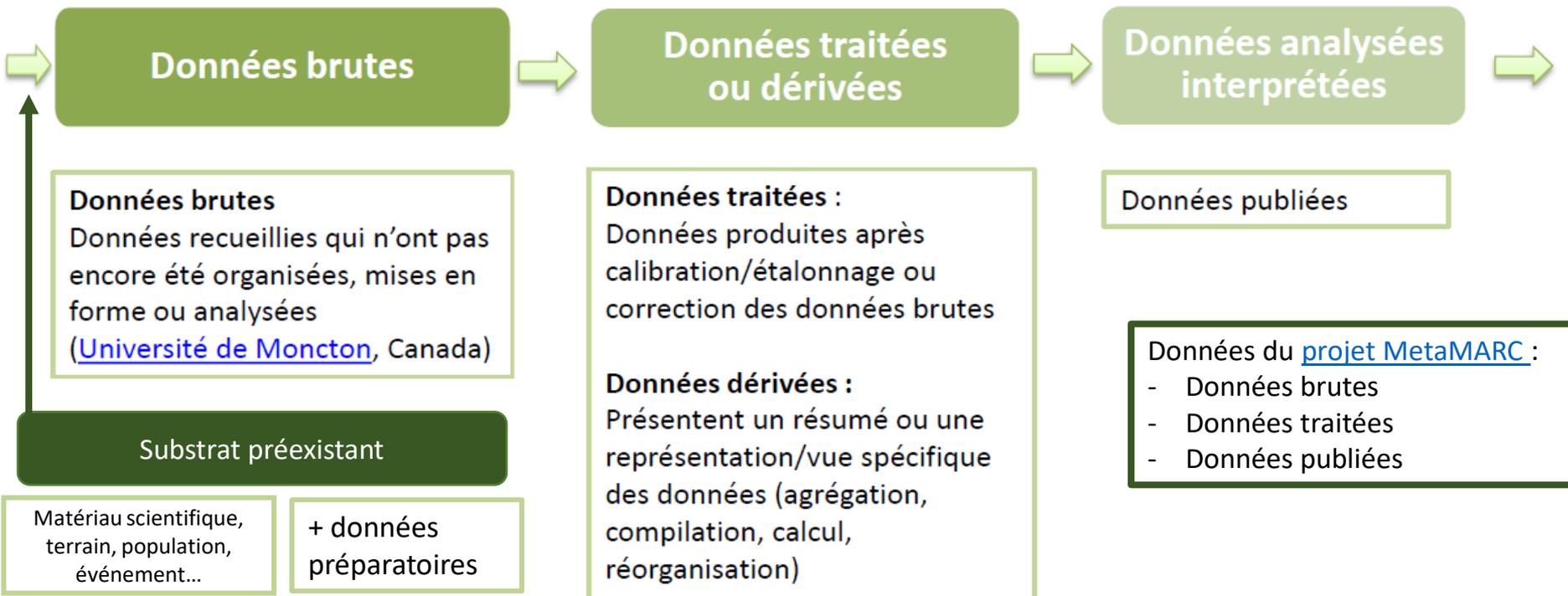
- validées par la communauté
- réutilisables

Ex.: décodage du génome humain, certaines données astronomiques...

➤ MetaMARC, un exemple de projet de recherche en microbiologie utilisant des données à la fois expérimentales, compilées et de référence (Lakin et al., 2019) :

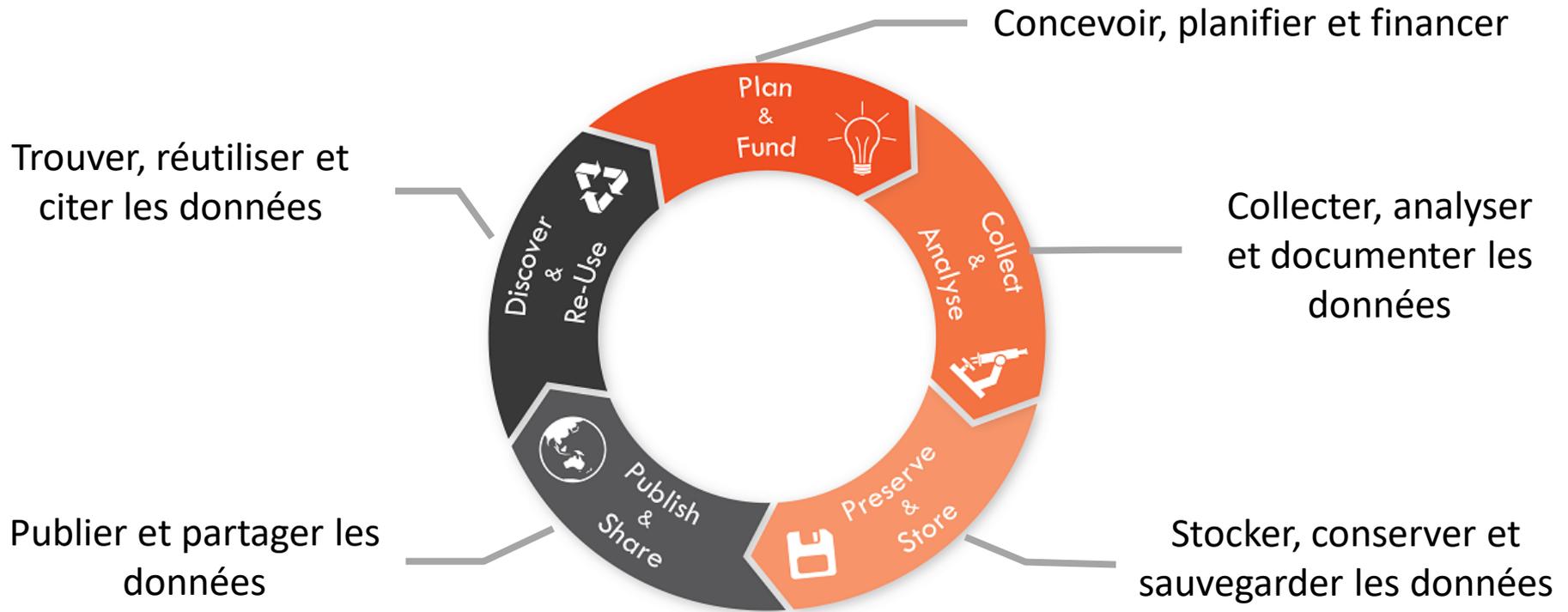
- [Descriptif](#) du projet
- [Données de référence](#)
- [Données produites](#)
- [Code source du projet](#)
- [Article de recherche](#)

Continuum données brutes ---> analysées



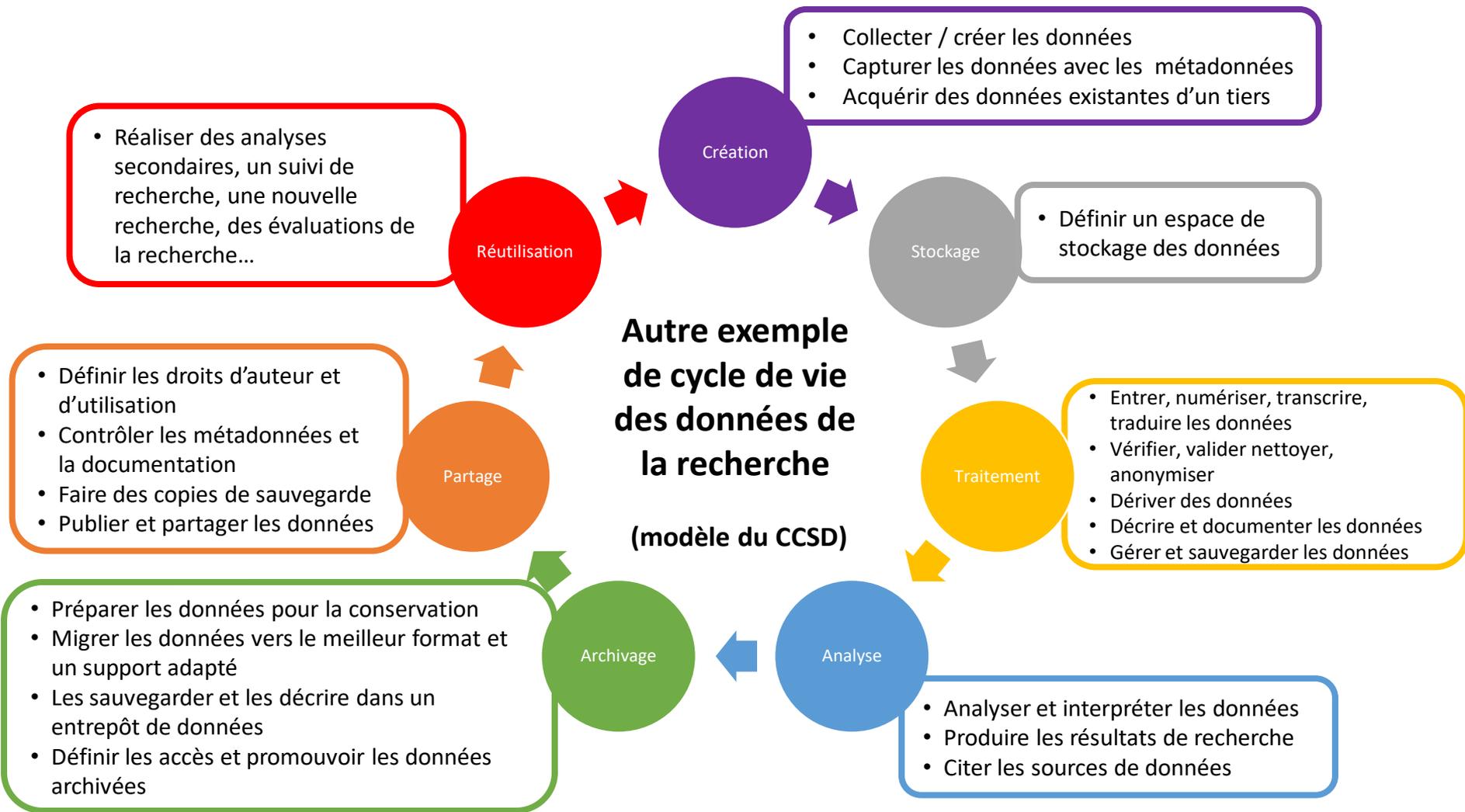
- **Données primaires** « données collectées par la personne qui effectue la recherche »,
Données secondaires « données collectées par une personne différente de l'utilisateur »
([Wikipédia](#), 2016)
- **Données sources** « données utilisées par les chercheurs pour leur recherche »
Données résultats « données produites comme résultats de recherche » (Prost, 2015)

Le cycle de vie des données de recherche



[Université de Sydney](#)

(voir aussi A. Ciolek-Figiel et F. Cosserat, « Gestion et diffusion des données de la recherche », 2016 ([en ligne](#)))

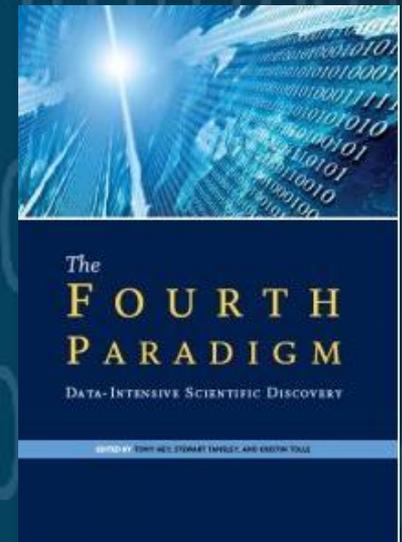




2- Pourquoi s'intéresser aux données de la recherche ?

Le 4^e âge de la science : la science des données ?

- Science classique : empirique (décrire les phénomènes naturels)
- XXe siècle : essor de la science théorique (modélisation)
- Ces dernières décennies : science computationnelle (simulations complexes)
- Science actuelle : *data-driven science* à partir de données collectées par des instruments toujours plus puissants ou générées par simulation.

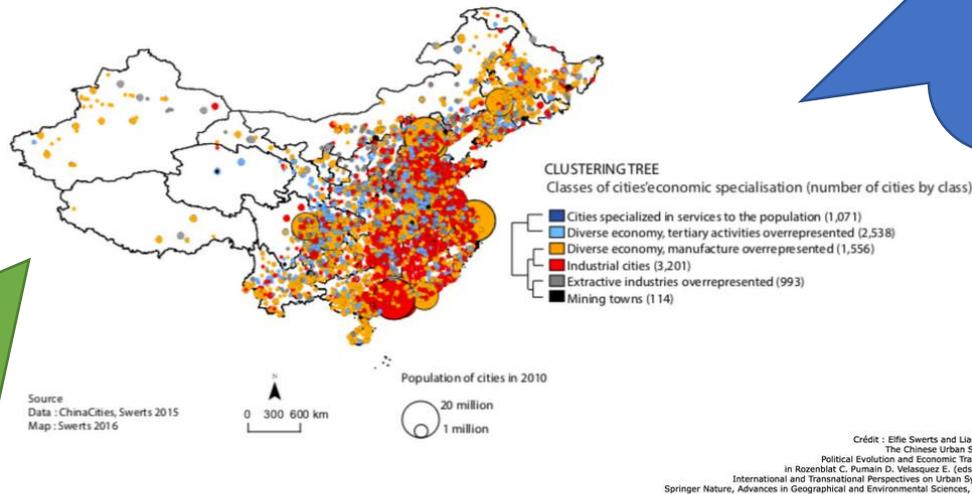


STM et big data : le Trou noir M87*

- Télescope virtuel géant Event Horizon Telescope (EHT) formé par 8 observatoires répartis sur le globe
- Calculateurs surpuissants au Haystack Observatory (USA) et au MPIfR (Allemagne)
- 4 équipes de recherche, 3 algorithmes de traitement des données
 - 8 Go de données collectés par seconde pendant 3 nuits (18 heures)
 - 5 pétaoctets de données (5 000 000 000 000 000 octets, soit 5 millions de milliards...)
 - 5 000 disques durs d'1 To chacun = 6 m³ de supports physiques (impossible à stocker dans le cloud : transfert par avion et non par internet)

SHS et big data : le projet Geodiversity

« Nous avons testé et validé les théories de dynamique d'évolution des villes et des systèmes de villes sur 20 000 agglomérations urbaines de 7 régions du monde (Europe, Etats-Unis et BRICS). Pour cela, nous avons réuni d'immenses bases de données que nous avons harmonisées pour qu'elles soient comparables. »



« On y étudie les spatialités et les territorialités géographiques à partir de concepts et de modélisations théoriques selon des perspectives historiques grâce à des enquêtes quantitatives et qualitatives, des réflexions méthodologiques et épistémologiques. »

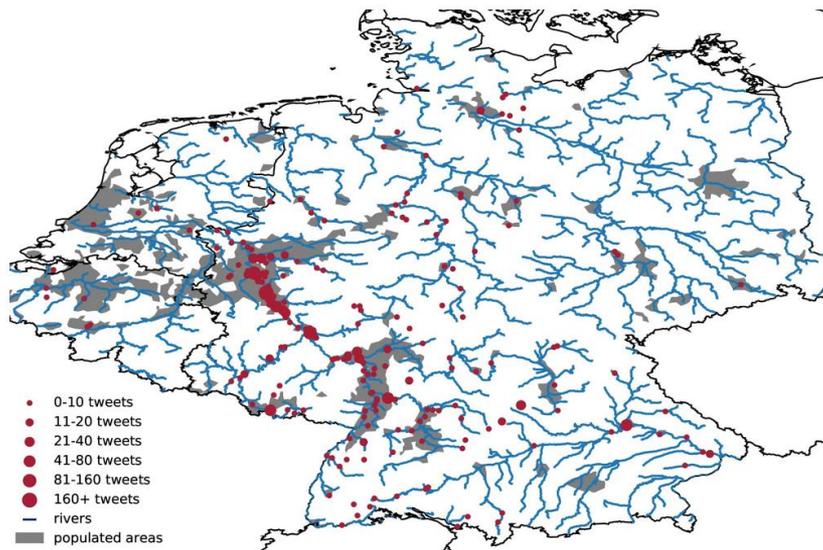
12 JUILLET 2018 / A LA UNE, DOSSIERS THÉMATIQUES, NEWS, SCIENCES HUMAINES ET SOCIALES ET DATA
UN PROJET INTERDISCIPLINAIRE EXEMPLAIRE EN GÉOGRAPHIE

« Pour Geodiversity, nous avons ainsi réalisé un demi-milliard de simulations. »

[DataAnalytics](#) : interview de Denise Pumain, responsable du projet (laboratoire Géographie-Cités)

« Tous les éléments de preuve sont vérifiables : toutes nos données sont accessibles dans la revue européenne de géographie [Cybergeog](#). »

Les données guident la recherche



Tweets during the January 2018 floods in Germany, the Netherlands and Belgium. The tweets filtered tweets mentioning a settlement are shown in red overlaying populated areas (in grey) and the river network (in blue). Note that tweets that mention the country (i.e., Germany) or admin1 regions (i.e., states) are not shown here.

J.A. de Bruijn et al., « A global database of historic and real-time flood events based on social media », *Scientific Data*, 2019 ([en ligne](#)).

R. Haunschild et al., « Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags », *Journal of Informetrics*, 2019 ([en ligne](#)).



F. Auclert, « Cette IA aide à compléter d'anciennes recherches scientifiques », *Futura-Sciences*, 2019 ([en ligne](#)).

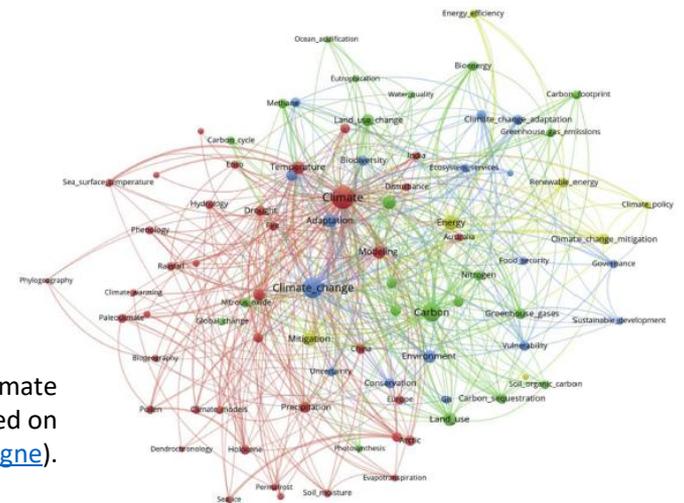
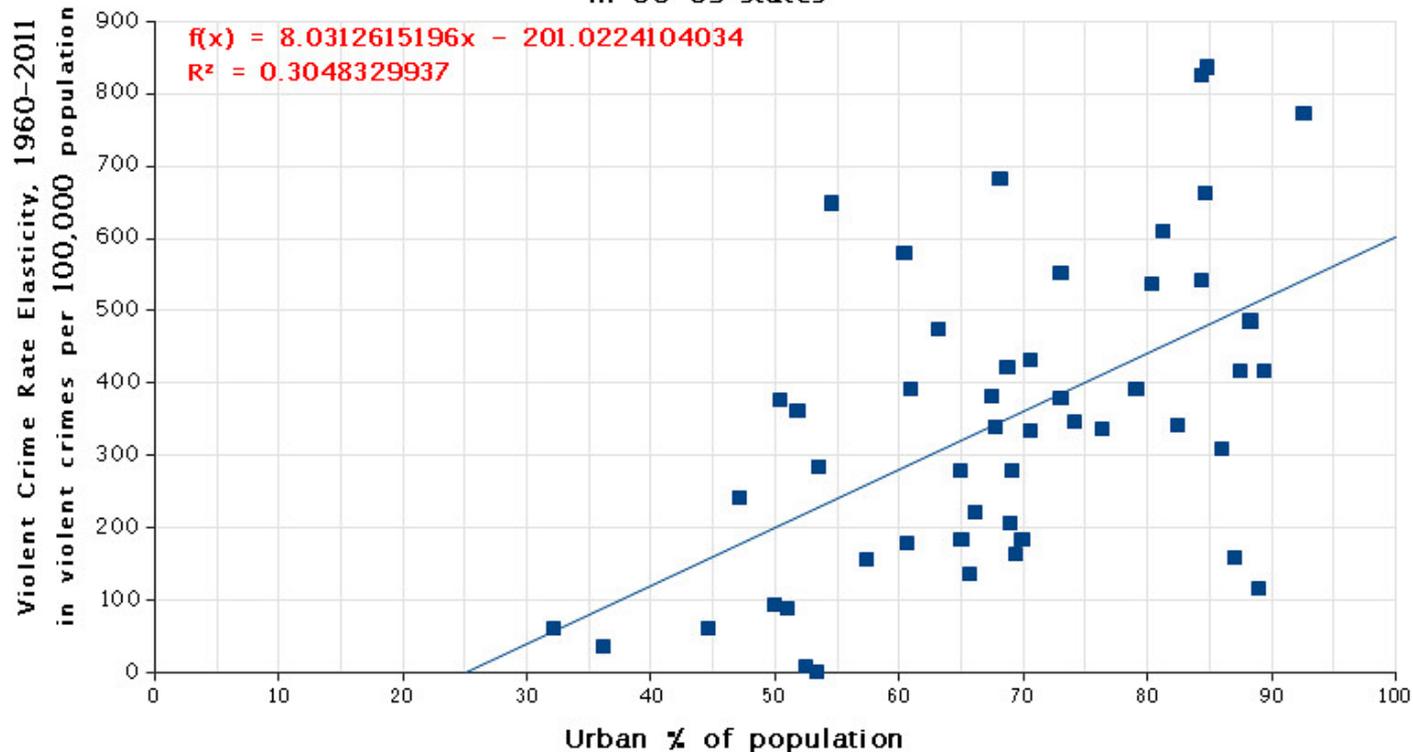


Fig. 2. Top-85 author keywords of climate change research papers published between 2011 and 2017. A resolution parameter of 0.54 with a minimum cluster size of 1 was used. An interactive version of this network can be viewed at <http://rayuri.com/2nu5va>. Note that the color scheme may be different in the interactive version.

- Recherches inductives : exemple de croisement de données et *data-driven science*

Urbanity vs. Violent Crime Elasticity in 50 US states



Source for Urbanity%: Census.gov via AllCountries.org - retrieved 23 Dec 2012

http://www.allcountries.org/usensus/37_urban_and_rural_population_and_by.html

Source for Elasticity: a particular analysis of FBI crime statistics taken from DisasterCenter.com/crime

Source de l'image : [Flickr](#)

- Recherches hypothético-déductives : les *Registered Reports* pour lutter contre les biais de confirmation

Registered Reports

Participating Journals

Details & Workflow

Resources for Editors

For Funders

FAQ

Allied Initiatives

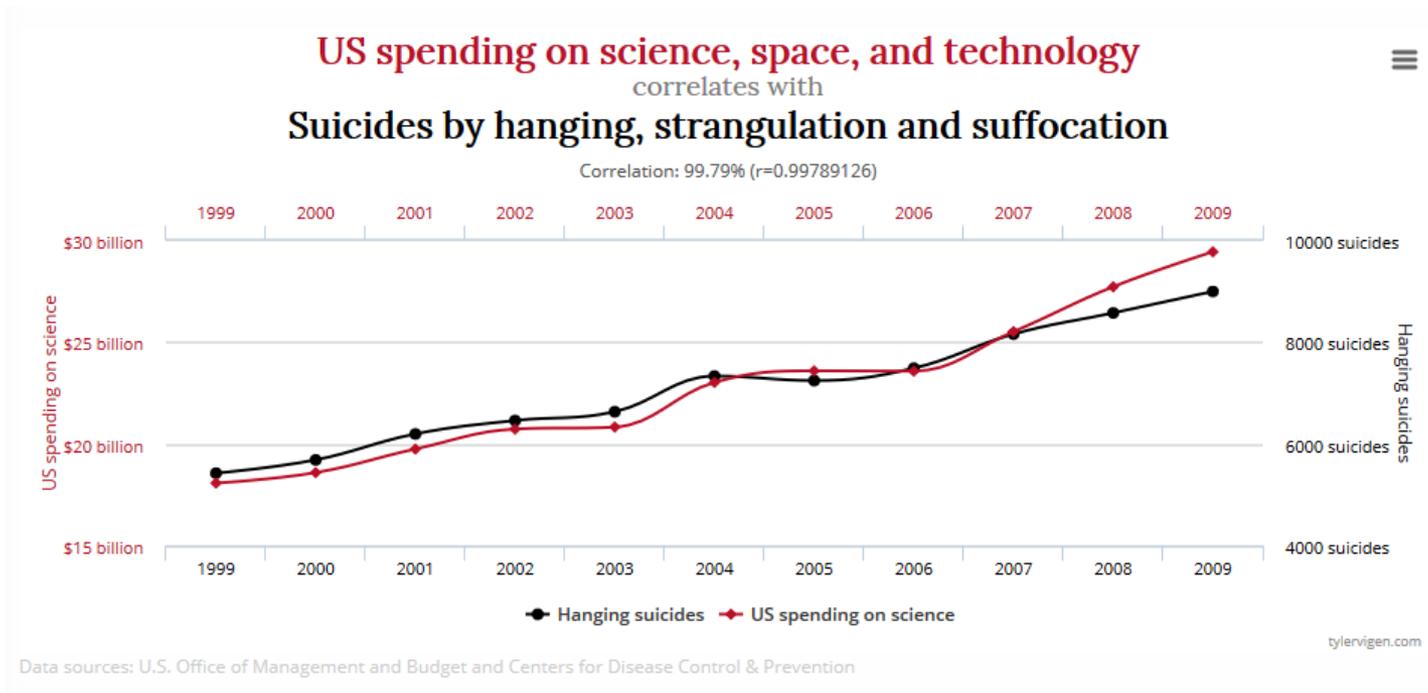
Registered Reports is a publishing format used by over 200 journals that emphasizes the importance of the research question and the quality of methodology by conducting peer review prior to data collection. High quality protocols are then provisionally accepted for publication if the authors follow through with the registered methodology.

This format is designed to reward best practices in adhering to the hypothetico-deductive model of the scientific method. It eliminates a variety of questionable research practices, including low statistical power, selective reporting of results, and publication bias, while allowing complete flexibility to report serendipitous findings.



Source : [Center for Open Science](#)

Un autre exemple de croisement de jeux de données (!)



Source: [Spurious correlations](#)

Les données sont à la base de l'intégrité scientifique

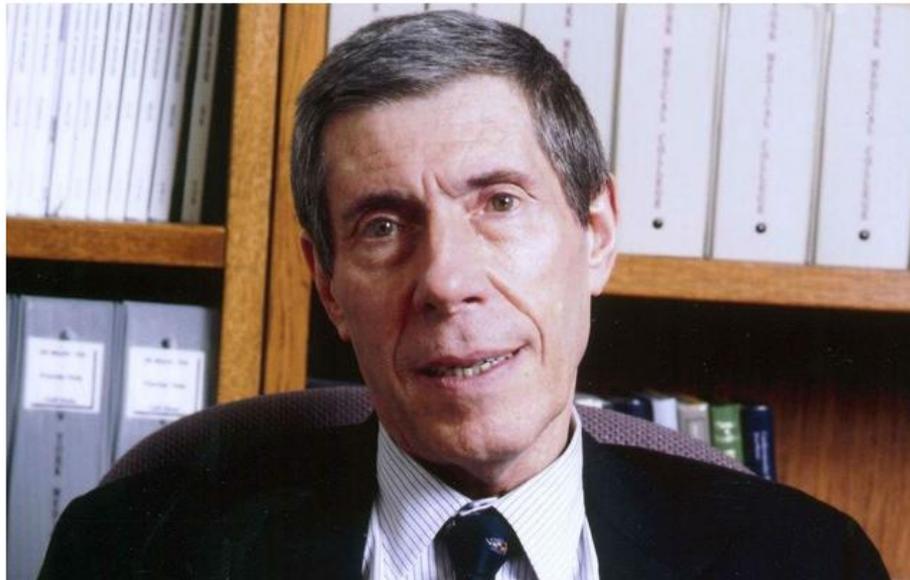
ACCUEIL > MONDE

Etats-Unis: Un chercheur aurait publié une trentaine d'études aux données falsifiées ou inventées

ESCROQUERIE Piero Anversa a reçu des dizaines de millions de dollars de subventions publiques pour ses « découvertes »...

20 Minutes avec agences | Publié le 19/10/18 à 14h25 — Mis à jour le 19/10/18 à 14h25

7 COMMENTAIRES 0 PARTAGE



Le chercheur Piero Anversa, dont 31 articles de recherche ont été retirés des revues médicales spécialisées. — AP/SIPA

À LIRE AUSSI



25/09/18 | **FAKE OFF**
82% des Français sont-ils opposés à la «PMA sans père»?



27/07/18 | **FAKE OFF**
Le paracétamol, «inefficace et dangereux»? Alerte à l'intox



20/07/18 | **RECHERCHE**
Une revue scientifique piégée par une fausse étude sur le cancer

 D'ACTU

Source : [20 Minutes](#)
Voir aussi une liste de fraudes scientifiques sur [Wikipédia](#)

Données et rétractations

Retraction Watch

Tracking retractions as a window into the scientific process

Unreliable Data	The accuracy or validity of the data is questionable
Unreliable Image	The accuracy or validity of the image is questionable

Concerns/Issues About Data	Any question, controversy or dispute over the validity of the data
Concerns/Issues About Image	Any question, controversy or dispute over the validity of the image

Duplication of Data	Also known as “self-plagiarism”. Used when the all or part of the data from an item written by one or all authors of the original article, are repeated in the original article without appropriate citation.
---------------------	---

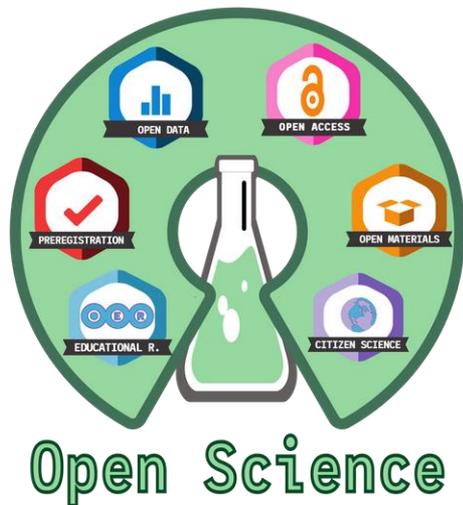
Falsification/Fabrication of Data	Intentional changes to data so that it is not representative of the actual finding
Falsification/Fabrication of Image	Intentional changes to an image so that it is not representative of the actual data

Plagiarism of Data	Used when the all or part of the data from an item not written by one or all authors of the original article, are repeated in the original article without appropriate citation.
--------------------	--

Manipulation of Images	The changing of the presentation of an image by reversal, rotation or similar action
------------------------	--

[Retraction Watch Database User Guide Appendix B : reasons](#)

Données de la recherche et science ouverte



« L'intégrité de la recherche, définie comme sa *reproductibilité* ou par le fait de *réduire la marge d'erreur possible*, est le principal argument en faveur de l'intégration des pratiques de science ouverte dans les processus de recherche scientifique. Car si les résultats de la recherche ne peuvent pas être vérifiés, à quoi sert d'y donner accès ? »

= Les données de la recherche sont indispensables pour vérifier les résultats scientifiques

Source: [Foster Open Science Learning Objectives](#) (2015)

The Mercury News

News

Tech accidentally wipes out info on Alaska's \$38 billion fund



By ASSOCIATED PRESS |
March 20, 2007 at 1:37 am

Get Morning Report and other email newsletters

SIGN UP

JUNEAU, Alaska (AP) – Perhaps you know that sinking feeling when a single keystroke accidentally destroys hours of work. Now imagine wiping out a disk drive containing information for an account worth \$38 billion.

That's what happened to a computer technician reformatting a disk drive at the Alaska Department of Revenue. While doing routine maintenance work, the technician accidentally deleted applicant information for an oil-funded account – one of Alaska residents' biggest perks – and mistakenly reformatted the backup drive, as well.

There was still hope, until the department discovered its third line of defense, backup tapes, were unreadable.

Follow Us



RECOMMENDED FOR YOU >

1 PG&E reaches \$11 billion deal with California wildfire insurers

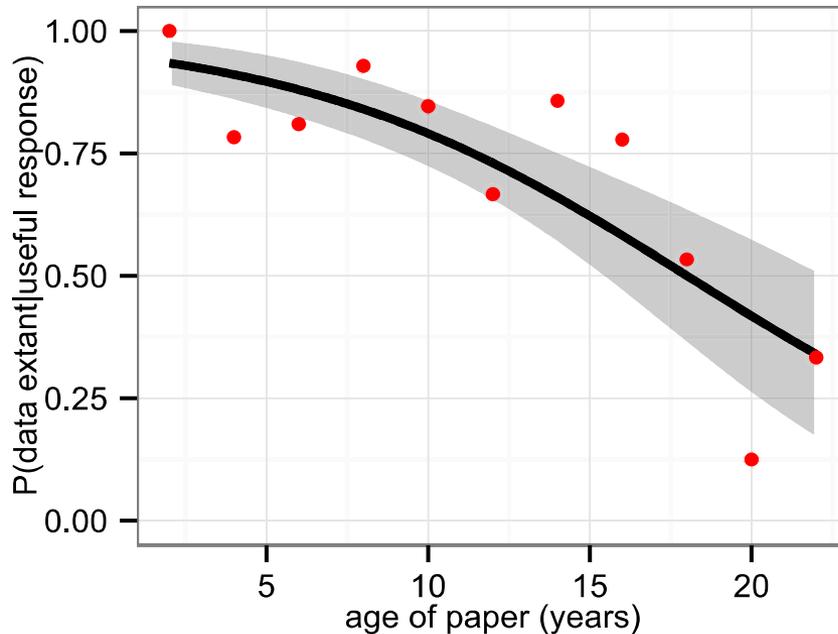
Un technicien écrase les données fiscales d'un fonds de financement représentant 38 milliards de dollars en reformatant un disque dur lors d'une opération de routine. Les deux sauvegardes sont inutilisables ([source](#)).

Les données sont fragiles

- Risque de perte de données, d'altération ou d'obsolescence
 - Risque de faille de sécurité
 - Besoin de constituer un patrimoine scientifique et technique
- Nécessite de prévoir très tôt dans le projet des modalités de tri, gestion, sauvegarde, conservation, archivage...



« La bibliothèque des données brûle »



(D) Predicted probability that the data were extant (either “shared” or “exist but unwilling to share”) given that we received a useful response. In all panels, the line indicates the predicted probability from the logistic regression, the gray area shows the 95% CI of this estimate, and the red dots indicate the actual proportions from the data.

T. Vines et al., « The availability of research data declines rapidly with article age », *Journal of Current Biology*, 2014.

Tous les ans :

- -17% de chances que les données sous-tendant une publication soient disponibles
- -7% de chances que l’adresse mail fournie pour contacter l’auteur soit valable



Où sont les données ? Idéal...

Intégration des données pour une meilleure compréhension des textes, publications interactives, résultats reproductibles

Données intégrées aux publications

Seulement pour certaines données impossibles à intégrer à une publication

Données annexées aux publications

Excellente conservation et valorisation des données, liens réciproques entre données et publications

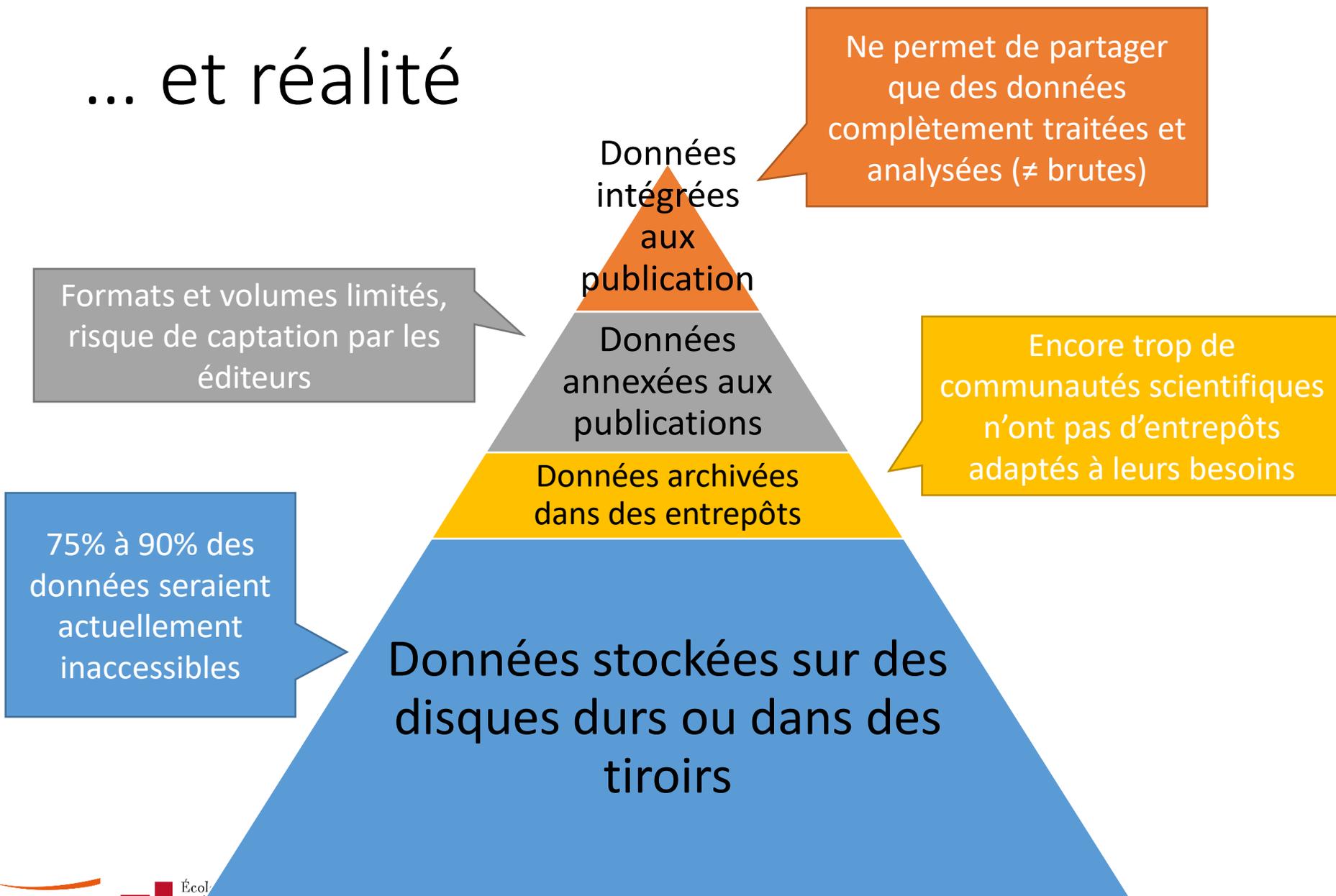
Données archivées dans des entrepôts

Multiplication des documents qui valorisent les données (PGD, *data papers*...)

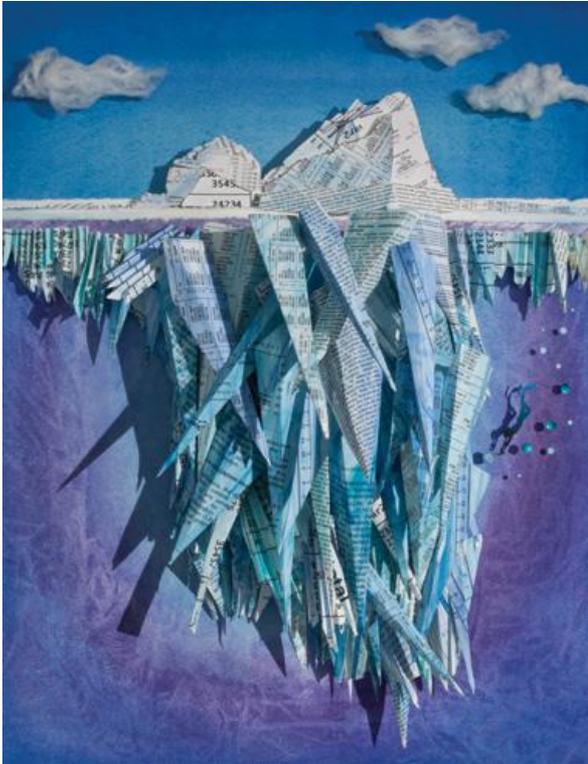
Données stockées sur des disques durs ou dans des tiroirs

À éviter au maximum

... et réalité



Le risque de sombrer dans l'oubli

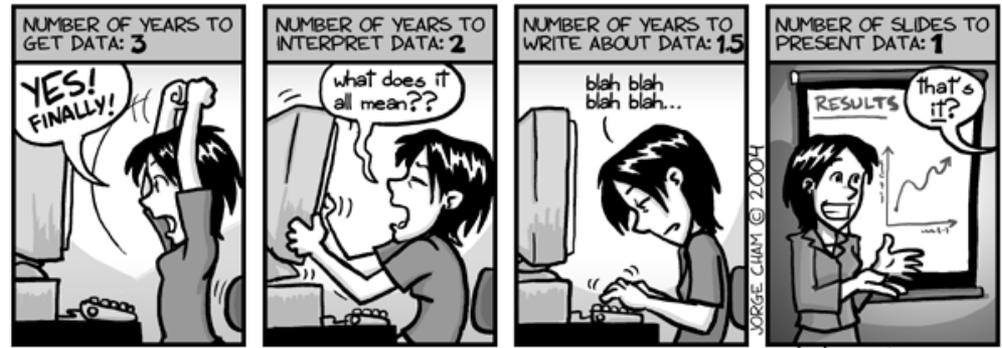


Tip of the iceberg, [The Scientist](#)

On estime que :

- 10% seulement des données produites sont mises à disposition
- 90% restent stockées sur les disques durs sans valorisation

DATA: BY THE NUMBERS



www.phdcomics.com

J. Cham, « [Data : by the numbers](#) », PhD Comics, 31/05/2004

Quelques bonnes pratiques de gestion de données

- Réfléchir à l'organisation de ses fichiers de données
- Décrire ses données de façon à les identifier sans erreur (nommage, versionnage, organisation...)
- Choisir des formats de fichier adaptés pour la conservation à long terme
- Opter pour une solution de stockage et d'archivage pérenne
- Planifier la sauvegarde de ses données

Qu'est-ce qu'un « jeu de données »?

- « Agrégation [...] de données brutes ou dérivées présentant une certaine "unité", rassemblées pour former un ensemble cohérent » (R. Gaillard)
- Jeu de données = dataset = multiples fichiers
 - de type, format, taille, contenu variés
 - évolutifs (accroissement, traitement...)
 - en rapport les uns avec les autres



Ex. d'arborescence décrite dans un DMP ([source](#))

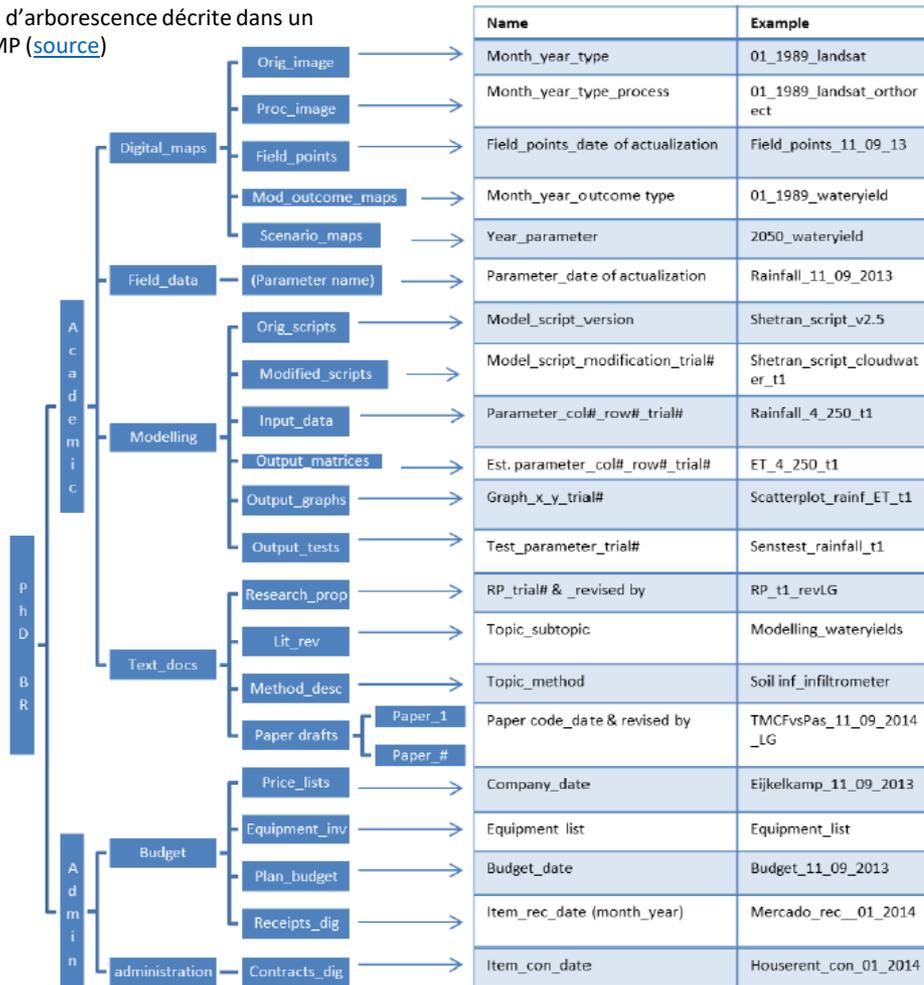


Figure 1. Proposed folder arrangement and file naming strategy.

L'arborescence

- 3 à 5 niveaux de dossiers
- Exemples de dossiers:
 - 1 dossier + sous-dossiers aspects supports (admin, finance...)
 - 1 dossier pour la doc scientifique et technique (méthodo, publications, rapports d'étape...)
 - 1 dossier par dataset

D'après C. Morel-Pair, « Formats et métadonnées », 2016 ([en ligne](#)).

Bonnes pratiques de nommage

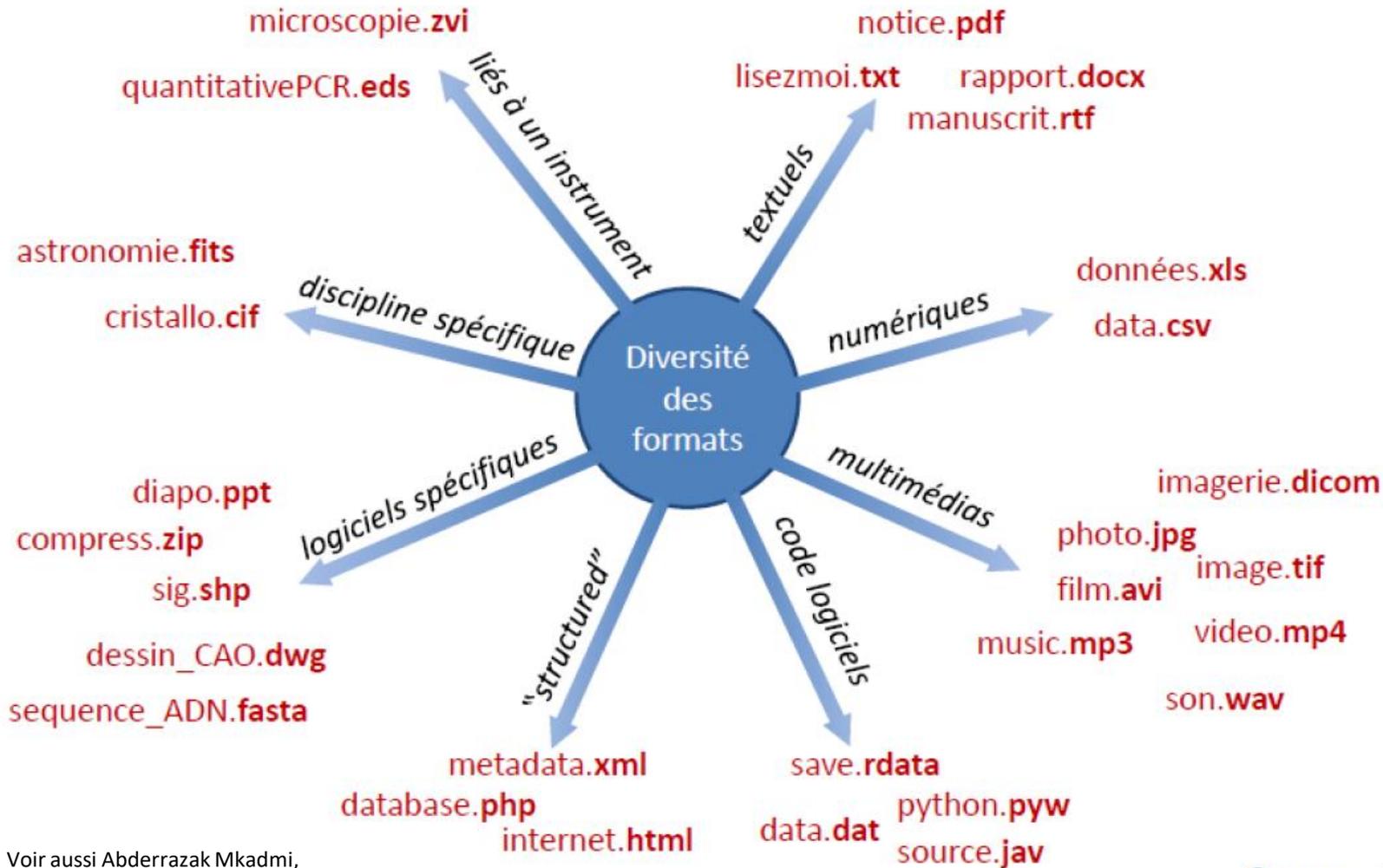


- Caractères signifiants + extension de format
- Environ 25 caractères max., sans accents ni symboles
- Éléments classiques : date de création, nom du créateur, description du contenu, versionnage, numéro/nom/code du projet...
- Définir clairement
 - ordre de ces éléments dans le nom du fichier
 - mode de saisie (ex.: format des dates)
- Pas de termes trop génériques (ex.: « brouillon », « manip », « test » etc.)
- Ne pas hésiter à utiliser un logiciel de renommage

Bonnes pratiques de nommage (2)

- Pas d'espaces (préférer _ ou majuscule initiale à chaque mot)
- Pas de caractères accentués (ée à ù ç) ni de symboles (+ ' @ ° [] : < / * » & ! \$)
- Utiliser des normes, comme ISO 8601 pour les dates : aaaammdd
- Nombres toujours sur au moins 2 unités (en fonction du nombre de fichiers concernés: 01, 002, 0003 etc.)
- Versionnage précis

Formats



bmp
cgm
emf
example
fits
g3fax
gif
ief
jp2
jpeg
jpm
jpx
ktx
naplps
png
prs.btif
prs.pti
pwg-raster
svg+xml
t38
tiff
tiff-fx
vnd.adobe.photoshop
vnd.airzip.accelerator.azv
vnd.cns.inf2
vnd.dece.graphic
vnd.djvu
vnd.dwg
vnd.dxf
vnd.dvb.subtitle
vnd.fastbidsheet
vnd.fpx
vnd.fst
vnd.fujixerox.edmics-mm
vnd.fujixerox.edmics-rlc
vnd.globalgraphics.pgb

Voir aussi Abderrazak Mkadmi,
[Les différents formats de fichiers](#)

Source : (10)

Source: C. Morel-Pair, « Participer à l'organisation du management des données de la recherche : gestion de contenu et documentation des données (formats et métadonnées) », 2016 ([en ligne](#)).

Différents types de formats

- Format ouvert: spécification publiquement accessible. Ex.: Portable Network Graphic
- Format normalisé: fait l'objet d'une normalisation par une institution publique ou internationale (AFNOR, ISO, W3C). Ex.: OpenDocument
- Format propriétaire: élaboré par une entreprise dans un but commercial. Ex.: Portable Document Format
- Format conteneur: contient et fait fonctionner ensemble différents types de données (ex.: flux audio + flux vidéo + sous-titres dans un conteneur vidéo comme .mkv, .mpeg4 etc.)
- Format fermé: spécification secrète; souvent associé à un seul logiciel. Ex.: Photoshop Document
- Format standard de fait: format populaire et largement utilisé, mais non normalisé. Ex.: Doc

Quel format de fichier choisir ?

- Le format idéal:

- Ouvert
- Normalisé (si possible)
- Largement diffusé
- Compatible avec de nombreux logiciels



[Quiz sur les formats de fichiers les plus courants](#)



- Facile à transformer (migration)
- Indépendant (vis-à-vis d'autres formats, d'une entreprise particulière, d'un matériel particulier...)

Les trois niveaux de sauvegarde des données

1.

Sauvegarde et stockage sécurisé durant le projet

?



Serveur sécurisé de l'institution

Espace collaboratif de travail



Le chercheur



L'équipe de recherche

2.

Sauvegarde pour le partage

?



Entrepôt généraliste ou disciplinaire



L'équipe de recherche



Autre équipe de recherche (même domaine)



Autre équipe de recherche (autre domaine)

3.

Archivage pérenne

?



Plateforme d'archivage (ex : CINES)



L'équipe de recherche



Autre équipe de recherche (même domaine)



Autre équipe de recherche (autre domaine)

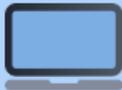
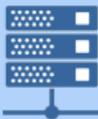


Stocker ses données

- Préférer un espace de stockage institutionnel
- Périphériques externes de type disques durs ou clés USB sont fragiles et faciles à perdre
- Attention à entretenir son matériel (mises à jour etc.)
- Stockage possible sur le *cloud*, mais attention aux propriétaires de ce genre de solutions (regarder CGU notamment)

COMPARATIF DES DIFFÉRENTS SUPPORTS DE STOCKAGE

Sans aller dans les extrêmes, il vous faut quand même être en mesure d'**anticiper ces risques** et de prendre des mesures adaptées à la grandeur des dommages potentiels. **Chaque support a ses avantages et inconvénients**, notamment en terme de coût, de sécurité, de capacité de stockage, et d'accès.

Support de stockage	Sécurité	Accès	Coût	Remarque d'utilisation
 Ordinateur professionnel	★★☆☆ Sujet au piratage informatique, aux détériorations et pannes	★☆☆☆ Pas adapté au partage, nécessite l'utilisation d'un support externe ou d'Internet (mail, cloud...)	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter les données confidentielles et sensibles
 Support externe	★☆☆☆ - Sujet au vol, à la perte du support - Durée de vie limitée (dégradation du matériel)	★★★★★ Facilement transportable, il permet de transférer les données vers un autre ordinateur	★★★★★ Pas de coût supplémentaire ou coût peu important	- Pour un stockage temporaire - Nécessité de crypter ou de sécuriser physiquement les données confidentielles et sensibles
 Serveur institutionnel	★★★★★ Stockage fiable, durable et sécurisé (contre le vol, le piratage, les incendies...)	★★☆☆ La connexion au serveur institutionnel ne facilite pas le travail avec des personnes extérieures	★★☆☆ Coût assez important mais pas forcément répercuté sur l'usager	- Pour un stockage plus pérenne - Adapté pour le stockage de données sensibles et des versions « stables » de vos données - Toutes les institutions ne proposent pas ce service
 Serveur Cloud	★★☆☆ On ne sait pas vraiment où sont stockées les données, ni ce qu'elles deviennent	★★★★★ Permet un travail synchronisé avec toutes les personnes ayant été autorisées au partage	★★☆☆ Payant à partir d'une certaine limite de stockage	- Pour un partage avec des personnes externes à l'institution - Ne pas y mettre de données sensibles ou confidentielles - Pas de contrôle sur la procédure de sauvegarde des données

Le choix parmi ces solutions dépend des possibilités offertes par votre institution et de vos données. Toutefois, lorsqu'il s'agit de données sensibles ou confidentielles, la **sécurité du stockage** devient un aspect important. On peut distinguer une **sécurité physique** et une **sécurité informatique**.

Source : [DoRANum](#)

Les 3 règles d'or de la sauvegarde

3

Sauvegardes

- Avoir au moins 3 copies de ses données (exactement la même version)

2

Supports

- Garder ces copies sur au moins 2 supports différents (serveur, disque dur d'ordinateur, clé USB...)

1

Ailleurs

- Stocker au moins un de ces supports à un endroit différent des autres (bureau, domicile, cloud...)

Outils de sauvegarde automatique



SincBack



XXClone



Cobian
Backup



Macrium Reflect

Outils de synchronisation automatique



Dropbox



Google Drive



OneDrive



iCloud

Quelle stratégie de sauvegarde ?

- Comment allez-vous sauvegarder vos données ?
- À quelle régularité ferez-vous des sauvegardes ?
- Sauvegarderez-vous **toutes** vos données (sauvegarde complète) ou uniquement celles qui ont été ajoutées/modifiées depuis la dernière sauvegarde (sauvegarde incrémentielle) ?
- Combien de temps garderez-vous vos sauvegardes ?
- Combien d'espace-disque sera nécessaire pour ces sauvegardes ?
- Comment garder trace des différentes sauvegardes effectuées, notamment si elles sont faites sur des supports différents ?
- Comment identifier les sauvegardes les plus récentes / complètes ?

Diffuser des données ne s'improvise pas

- Si on veut confier la diffusion ou la conservation de ses données à un tiers, celui-ci a le plus souvent des exigences précises quant à la préparation de ces données
- Choisir son entrepôt / son éditeur en fonction des pratiques adoptées pour la collecte de ses données ?
- Ou plutôt l'inverse...



➤ Pour aller plus loin : A.G. LeBlanc et al., « Scientific sinkhole: The pernicious price of formatting », *PLOS ONE*, 2019 ([en ligne](#)).

Les données sont précieuses

Les data d'entreprise, une mine d'or non exploitée

JULIE LE BOLZER | Le 29/11/2016 à 07:26 |     



FT Rethink

Les données : nouvel or noir

12 novembre 2018



Hannah Kuchler
correspondante du Financial Times à San Francisco

Les Données, la Richesse du 21ème siècle

OPEN DATA



Le 28.01.2013

Les données, plus importantes que le pétrole

Les données personnelles constituent les briques de base de la société de l'information.

Le 28.01.2013

Retour sur le lancement de l'exploration Data, le nouvel or noir ?

Découvrez le Storify de la soirée de lancement d'Inriality, par celyagd.



#échanges



Accueil > Economie > «La donnée, le nouvel or noir des banques, et surtout des fintechs?», par Jonathan Herscovici (WeSave)

Tribune libre

«La donnée, le nouvel or noir des banques, et surtout des fintechs?», par Jonathan Herscovici (WeSave)

Jonathan Herscovici • 30 mars 2018 à 12h45

Likely cost of not having FAIR research data

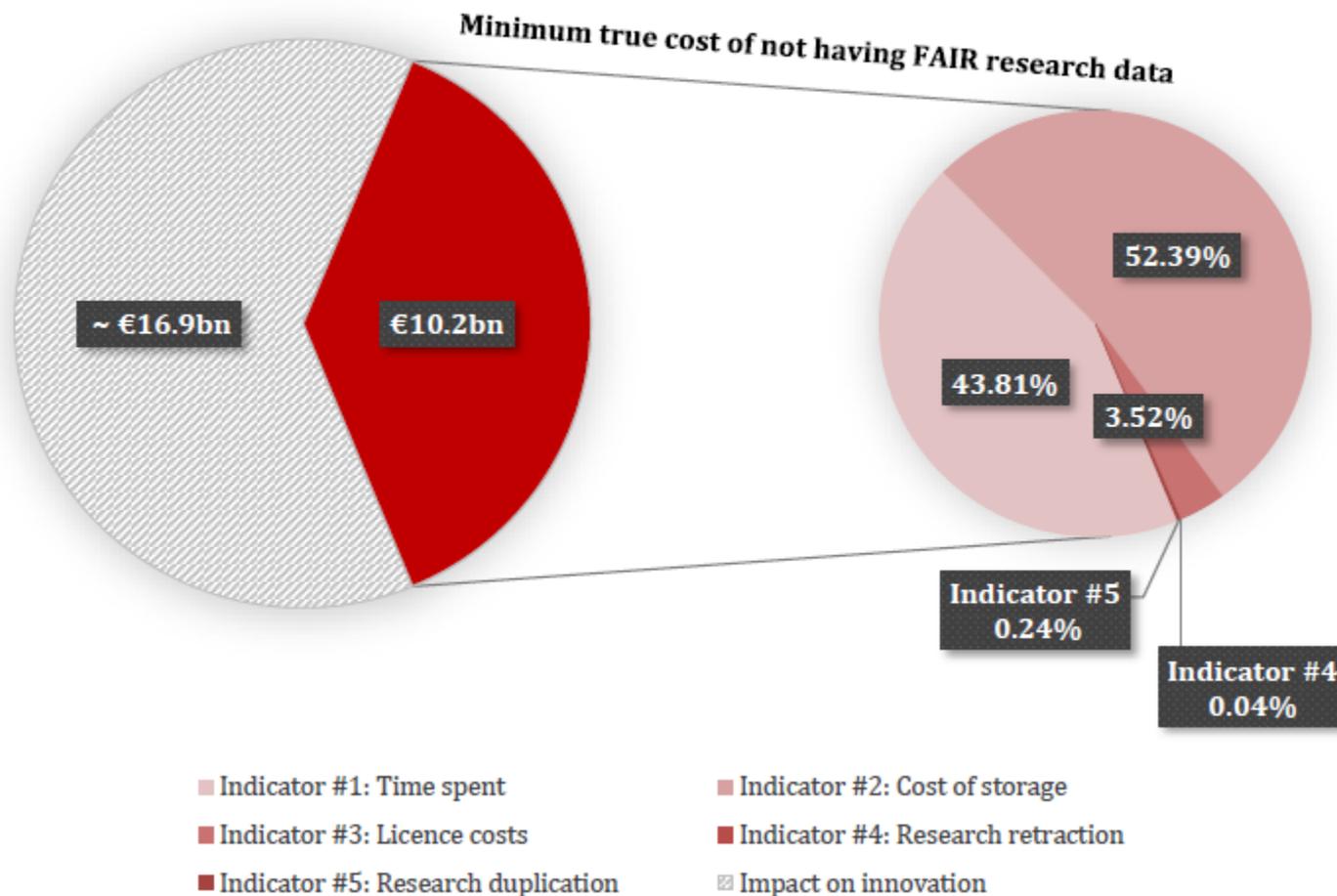


Figure 5: Cost breakdown

Commission européenne et PwC EU Service, *Cost-benefit analysis for FAIR research data : Cost of not having FAIR research data*, Commission européenne ([en ligne](#)).

Data Mining : quand Elsevier écrit sa propre loi...

© février 8, 2014 Data-mining, Industries académiques, Licences, Ouverture des données scientifiques Pierre-Carl Langlais

Présentation

Sciences Communes est un objet éditorial un peu à part, tantôt journal d'investigation (avec plusieurs "fuites" à son actif), tantôt carnet de recherche, tantôt lieu de réflexion et d'expérimentation sur des politiques publiques...

Il plaide la cause d'une science structurée par les communs. La mise à disposition des



WWW.PHDCOMICS.COM

J. Cham, « [Like Gold](#) », PhD Comics, 06/07/2015

P.-C. Langlais, « Data Mining : quand Elsevier écrit sa propre loi... », 2014 ([en ligne](#)).

Comment partager ?
Éviter la mainmise des éditeurs

Pays-Bas : des négociations avivent la crainte d'une "privatisation" de la science

Antoine Oury - 07.11.2019 | Edition - International - elsevier pays bas - negociations pays bas - elsevier edition

Tweet



Aux Pays-Bas, les universités, réunies en consortium, négocient régulièrement avec les éditeurs de contenus scientifiques et académiques. C'est dans le cadre d'une de ces sessions de négociations qu'a été mise sur la table la possibilité d'un accès ouvert total aux publications du groupe Elsevier, en échange d'une exploitation élargie des données liées à l'édition scientifique.

A. Oury, « Pays-Bas : des négociations avivent la crainte d'une "privatisation" de la science », *ActuaLitté*, 2019 ([en ligne](#)).



De bonnes raisons de s'intéresser à la gestion de ses données

- Les données guident la recherche (*data-driven science*)
 - Comment les manipuler ?
- Les données servent à valider des résultats de recherche
 - Comment les (faire) vérifier ?
- Les données sont fragiles
 - Comment les préserver et les (faire) réutiliser ?
- Les données sont précieuses
 - Comment les partager en évitant que d'autres ne les accaparent ?
- Comment les gérer convenablement tout au long d'un projet, et en faire le socle de futures autres recherches ?



Source: <http://www.kiwidatascience.com/>

Un *vade-mecum* utile : le Plan de gestion des données (PGD) ou *Data Management Plan (DMP)*

Document qui décrit la façon dont les données seront :

- obtenues,
 - traitées,
 - organisées,
 - stockées,
 - sécurisées,
 - préservées,
 - partagées,
 - etc.
- au cours et à l'issue d'un projet.

- Comment la gestion et le partage des données sont-ils financés, en particulier à long terme ?

Ressources

- En quoi consiste le projet ?
- Qui sont les partenaires ?
- Quelle est la politique de gestion des données ?
- Qui est responsable de la gestion des données ?

Responsabilités dans le projet

- Quelles données seront produites/utilisées au cours du projet ? (type, format, volume et accroissement...).
- Comment seront-elles produites ou transformées ?

Collecte des données

- Comment les données seront-elles identifiées, décrites ?
- Quels standards de métadonnées utilisera-t-on ?
- Comment seront générées les métadonnées ?

Documentation des données

- Qui pourra accéder aux données ?
- Les données seront-elles publiées ?
- Comment ?
- Dans quel délai ?
- Sous quelle licence ?

Accès et partage des données

What's the plan?

D'après S. Cocaud et D. L'Hostis, « Pourquoi et comment rédiger un plan de gestion de données ? », 2019 ([en ligne](#)).

- Comment, où, par qui, seront stockées, sauvegardées et sécurisées les données ?

Sauvegarde des données

- Qui sera propriétaire des données produites ?
- Des données externes seront-elles utilisées ?

Propriété intellectuelle

- Des données sensibles seront-elles produites ou utilisées ?
- Comment sera assurée leur anonymisation ?

Éthique

- Quel est le plan d'archivage et de préservation à long terme ?

Archivage et préservation des données

De l'utilité des PGD : retours d'expérience

« Cette expérience m'a appris à **adopter les bonnes pratiques** de gestion de données et m'a appris à réfléchir à l'accessibilité des données, à leur stockage et leur réutilisation. En somme, cette expérience m'a permis de **me poser les bonnes questions** quant à la gestion des données et a permis au projet de **se rapprocher des recommandations** en la matière. »

L. Locatelli, « Rédaction d'un plan de gestion de données : retour d'expérience. », *Le blog d'Huma-Num et de ses consortiums*, 2020 ([en ligne](#)).

« Cet exercice nous a permis [...] de nous confronter à la **nécessité de bien planifier** et penser sa recherche (aussi) **dans ce domaine.** »

D. Thonney et D. Dällenbach, « Data Management Plan : retour d'expérience », *Recherche d'IdéeS*, 2019 ([en ligne](#)).

« La mise en place [d'un PGD] en amont de toute recherche est un atout **précieux dans la construction même de son objet.** [Il] constitue donc une trame précieuse **sans laquelle nous ne nous serions pas posés, avec autant de précisions, un ensemble de questions pourtant essentielles** pour tout projet de recherche.

G. Torterat, « Rédaction d'un plan de gestion de données : retour d'expérience. Partie 2. », *Le blog d'Huma-Num et de ses consortiums*, 2020 ([en ligne](#)).

Pourquoi rédiger un PGD ?

- Se poser les bonnes questions avant de démarrer son projet
- Formaliser les décisions prises
- Mettre en place de bonnes pratiques
- Éviter des drames (panne technique, faille de sécurité...)
- Faciliter le travail à plusieurs
- Faire en sorte que les données « finales » soient d'une qualité irréprochable, pour tous les usages qu'on veut en faire



Comment rédiger un PGD ?

- En faisant intervenir les spécialistes de chaque aspect abordé
- En suivant un modèle (imposé, suggéré ou librement choisi)
- En gardant à l'esprit les principes FAIR
- En respectant quelques principes de base
 - Commencer la rédaction le plus tôt possible
 - Bien lire les questions, et chercher de l'aide si on ne les comprend pas
 - Être à la fois précis et bref
 - Oser dire qu'on ne sait pas mais qu'on va y réfléchir (et y réfléchir)
 - Mettre à jour souvent



LIBER Data Management Plan Catalogue
Un catalogue de DMPs évalués dans différentes disciplines.
Donnez votre avis à l'aide de ce questionnaire

<https://dmp.opidor.fr>

Bienvenue !

DMP OPIDoR vous accompagne à travers l'élaboration et la mise en pratique de plans de gestion de données et de logiciels.



Accessible à la communauté scientifique de l'ESR et à ses partenaires français ou étrangers



Personnalisable par tout organisme de recherche pour la mise en place de sa politique de données



Enrichi par des exemples et des recommandations adaptés à l'environnement de recherche



Collaboratif : il facilite les échanges entre les partenaires d'un même projet et les services d'accompagnement

DMP OPIDoR évolue grâce à vos retours. Les développements s'inscrivent dans le cadre d'une collaboration internationale autour du logiciel open source DMPRoadmap

[Connexion](#)[S'inscrire](#)

Accès institutionnel

Connexion

Accès individuel

* Courriel

* Mot de passe

[Mot de passe oublié ?](#)

Remember me



3- Rendre ses données FAIR

De bonnes raisons de s'intéresser à la gestion de ses données (*rappel*)

- Les données guident la recherche (*data-driven science*)
 - Comment les manipuler ?
- Les données servent à valider des résultats de recherche
 - Comment les (faire) vérifier ?
- Les données sont fragiles
 - Comment les préserver et les (faire) réutiliser ?
- Les données sont précieuses
 - Comment les partager en évitant que d'autres ne les accaparent ?
- Comment les gérer convenablement tout au long d'un projet, et en faire le socle de futures autres recherches ?



Source: <http://www.kiwidatascience.com/>

Des données bien gérées sont des données FAIR



<https://www.go-fair.org/>

Faciles à trouver



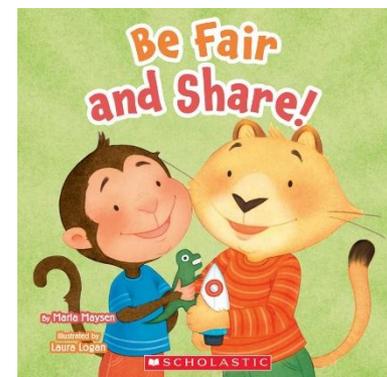
Accessibles



Interopérables



Réutilisables



[Crédits image](#)

Source : [Australian National Data Service](#)

Voir aussi [l'animation DoRANum](#) sur les principes FAIR

Let's assume you have research data almost ready for uploading to a repository: do you already know how you and the repository can work together to make the data as findable, accessible, interoperable and reusable (FAIR) as possible? By guiding you through the assessment process, the FAIR-Aware tool can help you to better understand the FAIR Principles [↗](#) and how making data FAIR can increase the potential value and impact of your data.

<https://fairaware.dans.knaw.nl/>

FAIR questions

FINDABLE

- 
1. Are you aware that a dataset should be assigned a globally unique and persistent identifier when deposited with a data repository?  Yes No
 2. Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others?  Yes No
 3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans?  Yes No

ACCESSIBLE

- 
4. Are you aware that access to your dataset may need to be controlled and that metadata should include licence information under which the data can be reused?  Yes No
 5. Are you aware that metadata should remain available over time, even if the data is no longer accessible?  Yes No

Que sont les métadonnées?



conditionnement: ● 2 portions

poids net: 400 g poids net égoutté: 265 g

Ingrédients

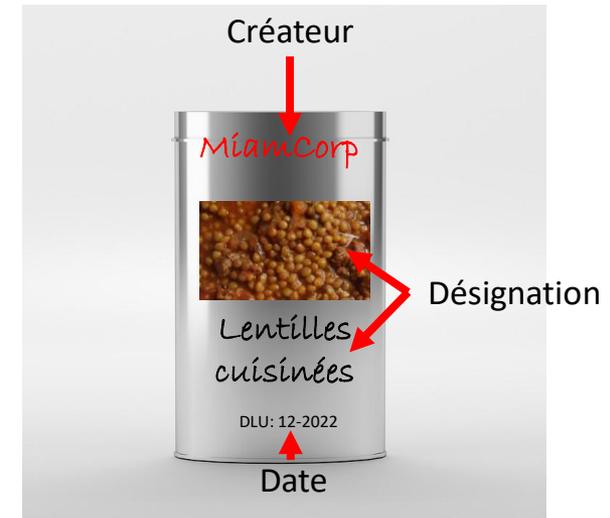
Lentilles sèches trempées, eau, double concentré de tomates, sel, saindoux, oignons, arôme naturel, arôme naturel d'oignons avec autres arômes naturels, ail, poivre. Peut contenir des traces de BLE.

Valeurs Nutritionnelles

pour 100g et par portion de 130g

Energie en kJ par portion

529 kJ



Source de l'image

D'après C. Morel-Pair, « Participer à l'organisation du management des données de la recherche : gestion de contenu et documentation des données (formats et métadonnées) », 2016 ([en ligne](#)).

Mes données sont-elles faciles à trouver ?

- **Saviez-vous qu'en plus de rendre vos données consultables par tout le monde, vous avez intérêt à fournir certains détails supplémentaires (appelés métadonnées) afin de rendre vos données faciles à trouver, compréhensibles et réutilisables par des tiers ?**
- Saviez-vous qu'un jeu de données gagne à se voir attribuer un identifiant unique et pérenne ?
- Saviez-vous que l'instance qui héberge vos données doit fournir les métadonnées associées dans un format lisible par des machines aussi bien que par des humains ?

L'importance des métadonnées

Contexte
Provenance

Évaluation

Reproduc-
tibilité

Confiance

Mots-clés

Découverte

Réutilisation

Identifiants

Accessibilité

Lien
données -
publications

Citabilité

Différents types de métadonnées

Description du contenu intellectuel

- Titre
- Résumé
- Mots-clés
- Etc.

Caractéristiques techniques

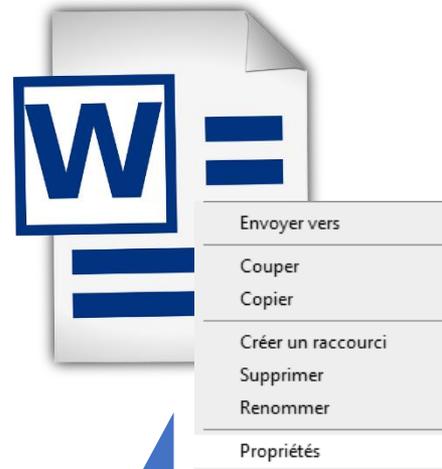
- Format
- Logiciel utilisé
- Version
- Taille
- Etc.

Droits liés

- Producteur
- Droits d'accès
- Droits de réutilisation
- Etc.

Relations entre données

- A pour partie...
- Est une partie de...
- Est une version de...



ReadMe.txt

Métadonnées
embarquées
*(embedded
metadata)*

Métadonnées
externes

Documenter ses données



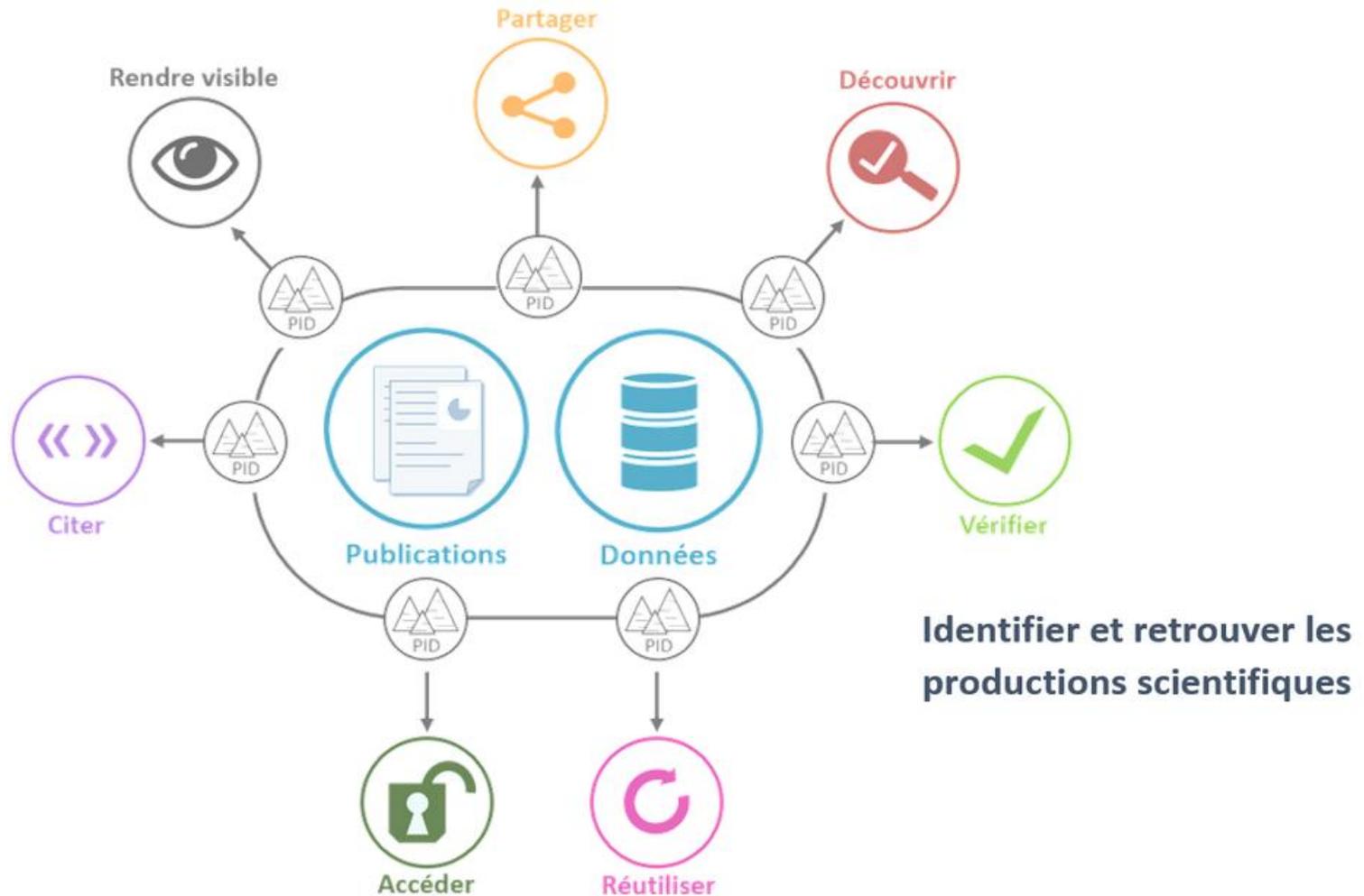
- Fichier texte (.txt, .pdf...) lisible par des humains
- Explications sur comment les données ont été obtenues, compilées, transformées etc.
- 3 niveaux :
 - Documenter le projet (contexte, hypothèse de départ, méthodologie, outils utilisés...). Ex.: cahiers de laboratoire (cf. [Open Lab Notebooks](#))
 - Documenter les jeux de données (contenu, relations entre les fichiers qui les constituent). Ex.: [fichiers ReadMe](#) dans Dryad
 - Documenter les données (glossaire, signification des sigles, échelles...) Ex.: [fichiers ReadMe](#) dans Dryad

Mes données sont-elles faciles à trouver ?

- Saviez-vous qu'en plus de rendre vos données consultables par tout le monde, vous avez intérêt à fournir certains détails supplémentaires (appelés métadonnées) afin de rendre vos données faciles à trouver, compréhensibles et réutilisables par des tiers ?
- **Saviez-vous qu'un jeu de données gagne à se voir attribuer un identifiant unique et pérenne ?**
- Saviez-vous que l'instance qui héberge vos données doit fournir les métadonnées associées dans un format lisible par des machines aussi bien que par des humains ?



Les identifiants pérennes (PID)



Source : [DoRANum](https://www.doanum.org/)



Un identificateur d'objet numérique (DOI) est une chaîne alphanumérique affectée à l'identification unique d'un objet. Il est lié à une description de métadonnées de l'objet ainsi qu'à un emplacement numérique, tel qu'une URL, où tous les détails concernant l'objet sont accessibles.



Pour créer de nouveaux DOI et les affecter à un contenu, il faut s'affilier à DataCite ou coopérer avec l'un de ses membres (l'INIST en France). Grâce à l'interface Web ou à l'API de l'entrepôt de métadonnées DataCite, on est en mesure de soumettre un nom, une description de métadonnées suivant le schéma de métadonnées DataCite et au moins une URL de l'objet pour créer un DOI. Une fois créées, les informations sur un DOI sont disponibles les différents services de Datacite, OAI-PMH et autres.

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

Attribuer un identifiant pérenne aux données de la recherche

[Contactez nous](#)

DataCite est un consortium international à but non lucratif dédié à l'attribution d'identifiants pérennes (**DOI**) pour les données de la recherche. Membre du consortium, **Inist-CNRS** est agence d'attribution de l'identifiant pérenne **DOI en France**. Pour pouvoir attribuer des DOI à vos données, contactez nous pour vous créer un compte.

[Accéder au service](#)[doiⁿ for DataCite](#)

doiⁿ for DataCite est une Application Web opensource destinée aux organismes ou institutions de recherche disposant d'un compte auprès d'un membre de DataCite et souhaitant attribuer et/ou mettre à jour **en grand nombre** des DOI. L'utilisation de **doiⁿ for DataCite** est une alternative à l'API de DataCite. Cette dernière nécessite des ressources informatiques dont ne disposent pas nécessairement certaines unités.

[Accéder au service](#)

We recently updated our privacy policy. You can read the [summary of the changes and full policy here](#).



Distinguish yourself in three easy steps

ORCID provides a persistent digital identifier (an ORCID iD) that you own and control, and that distinguishes you from every other researcher. You can connect your iD with your professional information — affiliations, grants, publications, peer review, and more. You can use your iD to share your information with other systems, ensuring you get recognition for all your contributions, saving you time and hassle, and reducing the risk of errors.

[FIND OUT MORE ABOUT OUR MISSION AND VALUES](#)



REGISTER

Get your unique ORCID identifier. It's free and only takes a minute, so [register now!](#)



USE YOUR ORCID ID

Use your iD, when prompted, in systems and platforms from grant application to manuscript submission and beyond, to ensure you get credit for your contributions.



SHARE YOUR ORCID iD

The more information connected to your ORCID record, the more you'll benefit from sharing your iD - so give the organizations you trust permission to update your record as well as adding your affiliations, emails, other names you're known by, and more.

Mes données sont-elles faciles à trouver ?

- Saviez-vous qu'un jeu de données gagne à se voir attribuer un identifiant unique et pérenne ?
- Saviez-vous qu'en plus de rendre vos données consultables par tout le monde, vous avez intérêt à fournir certains détails supplémentaires (appelés métadonnées) afin de rendre vos données faciles à trouver, compréhensibles et réutilisables par des tiers ?
- **Saviez-vous que l'instance qui héberge vos données doit fournir les métadonnées associées dans un format lisible par des machines aussi bien que par des humains ?**



Une syntaxe d'échange très utilisée : XML

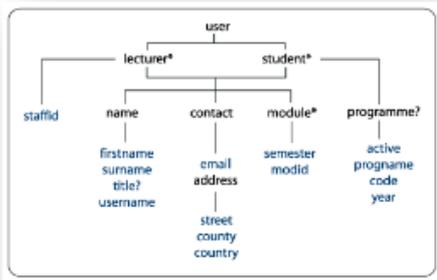
- XML : format d'échange très répandu, 1998
- Ici, éléments Dublin Core en XML

```
<dc:title>La très haute nappe alluviale d'âge pléistocène inférieur de la vallée de la Creuse  
cryoturbation, habitats préhistoriques et datations absolues</dc:title>  
<dc:title>The very high alluvial formation of early Pleistocene age in the Creuse River Valley  
cryoturbation patterns, préhistorique occupation sites and absolute dating</dc:title>  
<dc:creator>Despriée, J.</dc:creator>  
<dc:creator>Gageaonnet, R.</dc:creator>  
<dc:subject>Vallée de la Creuse</dc:subject>  
<dc:subject>Formations alluviales</dc:subject>  
<dc:subject>Phénomènes de cryoturbation</dc:subject>  
<dc:subject>Tectonique récente</dc:subject>  
<dc:subject>Datations RPE</dc:subject>  
<dc:subject>Très ancien Paléolithique</dc:subject>  
<dc:subject>Plateau d'Aigurande</dc:subject>  
<dc:subject>Massif central</dc:subject>  
<dc:description>Malgré des conditions de prospections difficiles, de nombreux témoins  
de la Creuse ont été reconnus dans la région d'Eguzon, sur la bordure nord, très bocagère, du Massif
```

- Syntaxe ...
 - Sans importance pour déposer ou rechercher des données via formulaire de saisie ...
 - A connaître pour réaliser des imports par lots, utiliser des exports, résultats de requête API

Deux approches pour fournir des métadonnées sur le Web

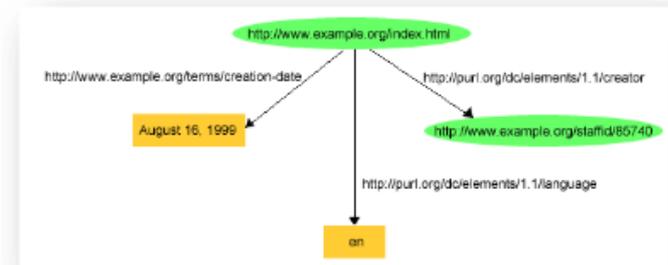
XML (Tree/container approach)



```
<?xml version="1.0"?>
<!DOCTYPE user SYSTEM "users.dtd">
<user>
  <student>
    <name>
      <firstname>Joe</firstname>
      <surname>Smith</surname>
      <title>Mr.</title>
      <username>smithj</username>
    </name>
    <contact>
      <address>
        <street>54 Maple Rise, Santry</street>
        <country>Dublin</country>
        <country>Ireland</country>
      </address>
      <email>smithj@dou.ie</email>
    </contact>
    <programme active="true">
      <programme>M.Eng in Electronic Systems</programme>
      <code>MEN</code>
      <year>1</year>
    </programme>
    <module semester="2">
      <modid>EE557</modid>
    </module>
    <module semester="1">
      <modid>EE553</modid>
    </module>
  </student>

```

RDF (approche basée sur le triplets)



```
ex:index.html dc:creator exstaff:85740 .
ex:index.html exterm:creation-date "August 16, 1999" .
ex:index.html dc:language "en" .
```

Source : European Data Portal, « Introduction à la gestion des métadonnées », 2013 ([en ligne](#)).

Schéma de métadonnées

« Un système de labellisation, marquage et codage utilisé pour l'enregistrement d'information de catalogue ou pour structurer des notices descriptives. Un schéma de métadonnées établit et définit les éléments de données et les règles d'application pour l'usage de ces éléments de données pour décrire une ressource. »

Schéma XML

```
<?xml version="1.0" encoding="UTF-8"?>
<DC:TYPE RDF>
- <rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dcmi="http://purl.org/dc/dcmi/" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">
- <rdf:Description rdf:about="http://purl.org/dc/terms/">
  <dcterms:title xml:lang="en">DCMI Metadata Terms - other</dcterms:title>
  <dcterms:publisher rdf:resource="http://purl.org/dc/aboutdcmi#DCMI"/>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-06-14</dcterms:modified>
- </rdf:Description>
- <rdf:Description rdf:about="http://purl.org/dc/terms/title">
  <rdfs:label xml:lang="en">Title</rdfs:label>
  <rdfs:comment xml:lang="en">A name given to the resource.</rdfs:comment>
  <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/"></rdfs:isDefinedBy>
  <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-11-11</dcterms:modified>
  <rdfs:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
  <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#title1-002"/>
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/title"/>
- </rdf:Description>
- <rdf:Description rdf:about="http://purl.org/dc/terms/creator">
  <rdfs:label xml:lang="en">Creator</rdfs:label>
  <rdfs:comment xml:lang="en">An entity primarily responsible for making the resource.</rdfs:comment>
  <dcterms:description xml:lang="en">Examples of a Creator include a person, an organization, or a service.</dcterms:description>
  <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/"></rdfs:isDefinedBy>
  <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-11-11</dcterms:modified>
  <rdfs:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
  <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#creator1-002"/>
  <rdfs:range rdf:resource="http://purl.org/dc/terms/Agent"/>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/contributor"/>
  <owl:equivalentProperty rdf:resource="http://xmlns.com/foaf/0.1/maker"/>
- </rdf:Description>
- <rdf:Description rdf:about="http://purl.org/dc/terms/subject">
  <rdfs:label xml:lang="en">Subject</rdfs:label>
```

Schéma RDF

```
<?xml version="1.0" encoding="UTF-8"?>
<DC:TYPE RDF>
- <rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:dcmi="http://purl.org/dc/dcmi/" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">
- <rdf:Description rdf:about="http://purl.org/dc/terms/">
  <dcterms:title xml:lang="en">DCMI Metadata Terms - other</dcterms:title>
  <dcterms:publisher rdf:resource="http://purl.org/dc/aboutdcmi#DCMI"/>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2012-06-14</dcterms:modified>
- </rdf:Description>
- <rdf:Description rdf:about="http://purl.org/dc/terms/title">
  <rdfs:label xml:lang="en">Title</rdfs:label>
  <rdfs:comment xml:lang="en">A name given to the resource.</rdfs:comment>
  <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/"></rdfs:isDefinedBy>
  <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-11-11</dcterms:modified>
  <rdfs:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
  <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#title1-002"/>
  <rdfs:range rdf:resource="http://www.w3.org/2000/01/rdf-schema#Literal"/>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/title"/>
- </rdf:Description>
- <rdf:Description rdf:about="http://purl.org/dc/terms/creator">
  <rdfs:label xml:lang="en">Creator</rdfs:label>
  <rdfs:comment xml:lang="en">An entity primarily responsible for making the resource.</rdfs:comment>
  <dcterms:description xml:lang="en">Examples of a Creator include a person, an organization, or a service.</dcterms:description>
  <rdfs:isDefinedBy rdf:resource="http://purl.org/dc/terms/"></rdfs:isDefinedBy>
  <dcterms:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2008-01-14</dcterms:issued>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2010-11-11</dcterms:modified>
  <rdfs:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property"/>
  <dcterms:hasVersion rdf:resource="http://dublincore.org/usage/terms/history/#creator1-002"/>
  <rdfs:range rdf:resource="http://purl.org/dc/terms/Agent"/>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
  <rdfs:subPropertyOf rdf:resource="http://purl.org/dc/terms/contributor"/>
  <owl:equivalentProperty rdf:resource="http://xmlns.com/foaf/0.1/maker"/>
- </rdf:Description>
- <rdf:Description rdf:about="http://purl.org/dc/terms/subject">
  <rdfs:label xml:lang="en">Subject</rdfs:label>
```

Du formulaire à la syntaxe

Title*
Enter a short, descriptive title for this data file.
2019 Dataset test

Description
Please describe your data file.
Résumé décrivant mon jeu de données

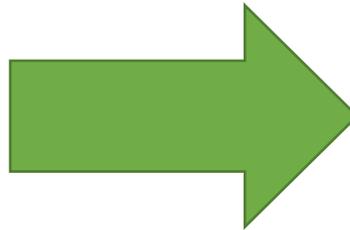
README file
Please upload a README file if needed.
Parcourir... Aucun fichier sélectionné.

Author of data file
Manually add an author:
Last name, e.g. *Smith* First name + initial, e.g. *Donald F.*
Fabre, Justine

Taxonomic names (e.g., *Thalassia hemprichii*, *Panthera pardus*)
Allium schubertii

Subject keywords

Un exemple de saisie :
formulaire de dépôt dans
[l'entrepôt Dryad](#)



dc.contributor.author	Angelier, Frédéric
dc.contributor.author	Brischoux, François
dc.coverage.spatial	Nouvelle-Aquitaine
dc.coverage.spatial	France
dc.date.accessioned	2019-02-11T20:28:01Z
dc.date.available	2019-02-11T20:28:01Z
dc.date.issued	2019-02-21
dc.identifier	doi:10.5061/dryad.gh2hh05
dc.identifier.citation	Angelier F, Brischoux F (2019) Are house sparrow populations limited by the lack of cavities in urbanized landscapes? An experimental test. Journal of Avian Biology.
dc.identifier.uri	http://hdl.handle.net/10255/dryad.206808
dc.relation.haspart	doi:10.5061/dryad.gh2hh05/1
dc.relation.isreferencedby	doi:10.1111/jav.02009
dc.subject	urbanization
dc.subject	house sparrow
dc.subject	cavities
dc.title	Data from: Are house sparrow populations limited by the lack of cavities in urbanized landscapes? an experimental test.
dc.type	Article
dwc.ScientificName	Passer domesticus
dc.contributor.correspondingAuthor	Angelier, Frédéric
prism.publicationName	Journal of Avian Biology
dryad.dansTransferDate	2019-03-01T10:40:52.498+0000
dryad.dansEditIRI	https://easy.dans.knaw.nl/sword2/container/d232131c-283e-486f-a407-3baf93e64d9d
dryad.dansArchiveDate	2019-03-01T12:02:14.612+0000

Mes données sont-elles accessibles ?

- Savez-vous que l'accès à votre jeu de données peut devoir être contrôlé et que vos métadonnées doivent inclure des informations sur les licences selon lesquelles les données peuvent être réutilisées ?
- Savez-vous que vos métadonnées doivent rester disponibles au fil du temps, même si les données ne sont plus accessibles ?

All versions

Found 63354 results.

< 1 2 3 4 5 6 7 8 9 >

Sort by:

Best match
asc.

Access Right

- Open (1577397)
- Closed (30373)
- Restricted (3679)
- Embargoed (1117)

File Type

- Pdf (856640)
- Jpg (357105)
- Png (200883)
- Zip (71423)
- Html (39852)
- Hdf5 (15102)
- Xml (11524)
- Docx (10839)
- Txt (8183)
- Xlsx (7426)

December 18, 2019 (v1) Dataset Closed Access

View

Triple Gem: Materials assembled for an exhibition in 2010 (part 7) Labels, label texts, wall panels

Willis, Michael;

Triple Gem: Materials assembled for an exhibition in 2010 (part 7) Labels, label texts, wall panels.

Uploaded on December 18, 2019

May 2, 2017 (v1) Dataset Closed Access

View

This is the title for testFolder1

NP;

no_description_available

Uploaded on May 2, 2017

May 1, 2015 (v1) Dataset Open Access

View

PowerTAC 2015-04 Finals Game 139

PowerTAC;

Log and boot files of game 139

Uploaded on December 6, 2017

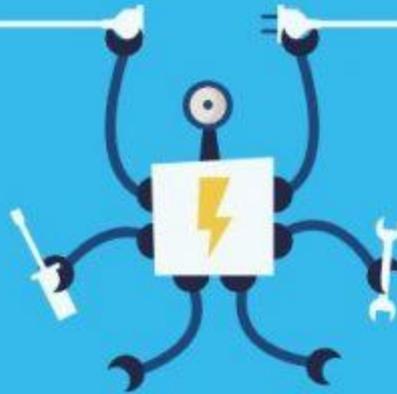
March 23, 2018 (v1) Dataset Open Access

View

Voir aussi [ces exemples](#) de formulations (en anglais) concernant les conditions d'accès et de réutilisation possibles (Université de Manchester)

Oops...

LOOKS LIKE SOMETHING WAS WRONG
WE'RE WORKING ON IT



404

Page not found

• BACK TO HOME •



[Source](#)

File(s) under permanent embargo

[Source](#)

Reason: Contact andy.mcdonald@northumbria.ac.uk for access

Introduction to RDM 080920.pptx

[Cite](#)[Share](#)[+ Collect](#)

186
views

0
downloads

0
citations

Presentation posted on 17.09.2020, 11:33 by Andy McDonald

Presentation outlining the various levels of service and support offered to researchers at Northumbria University in the field of research data management, and the importance and function of RDM in the greater research life cycle.

 **Read the peer-reviewed publication**

[Introduction to RDM Presentation](#)

HISTORY

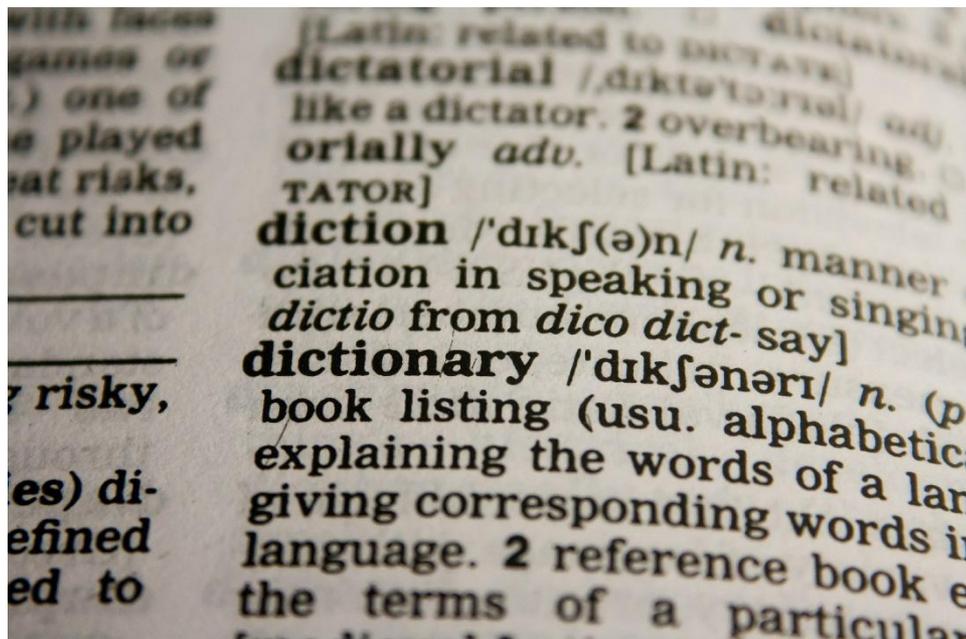
- **11.09.2020** - Submission date
- **17.09.2020** - First online date, Posted date



**Northumbria
University
NEWCASTLE**

Mes données sont-elles interopérables ?

- Savez-vous que les métadonnées qui décrivent vos jeux de données doivent utiliser des vocabulaires contrôlés ?



Standards sémantiques

Référentiels sujets

- Thésaurus (MeSH, GeMET...)
- Classifications (Dewey, CIB...)
- Ontologies (Gene Ontology...)
- Référentiels métiers (RAMEAU...)

Formes normalisées

- Pays (ISO 3166...)
- Mesurage de l'isolation acoustique des immeubles et des éléments de construction (NF EN ISO 140-4 et 140-5)

Identifiants pérennes

- URL (DOI, handle, ARK...)
- Ressources papier (ISBN, ISSN...)
 - Personnes (ORCID, IdRef...)
- Structures (hal_id...)

Référentiels techniques

- Matériel d'étude
 - Appareils
 - Méthodes
- Types de données

Référentiels de droits d'usage

- Licences (Creative Commons, BSL...)

BARTOC FAST

[Basic](#) [Advanced](#) [Expert](#) [API](#) [About](#) [BARTOC.org](#)

Basic FAST

FAST v1.0.3 is a Federated Asynchronous Search Tool for remote vocabulary resources ([learn more](#)).

Search word:



Valid.

Truncation (color*) and wildcards (colo*r) are supported, Boolean operators are not supported

Search

Interrogation fédérée de référentiels sujets

<https://bartoc-fast.ub.unibas.ch/bartocfast/>

<https://datu.ehess.fr/fr/>

datu

Partout ▾

français ▾

✕ 🔍 Chercher

Datu

[À propos](#)

[Vos commentaires](#)

[Aide](#)

[en - fr](#)

**L'ÉCOLE
DES HAUTES
ÉTUDES EN
SCIENCES
SOCIALES**

Le service de thésaurus et
vocabulaires contrôlés de l'Ehess, Datu !

Datu est le terme austronésien qui signifie "souverain".



VOCABULAIRES INSTITUTIONNELS

[Aires culturelles EHESS](#)

[Domaine HCERES](#)

[Domaines EHESS](#)



VOCABULAIRES SCIENTIFIQUES

[Thésaurus Savoirs](#)

[Thésaurus Timel](#)



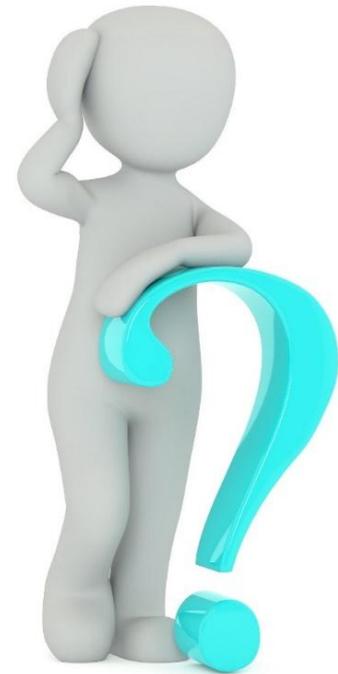
Mes données sont-elles réutilisables ?

- Savez-vous que vos métadonnées doivent inclure des informations relatives à la provenance des données collectées et/ou générées ?
- Savez-vous que les métadonnées décrivant vos données doivent suivre des standards approuvés par votre communauté de recherche ?
- Savez-vous que vos données doivent être déposées de préférence dans un format de fichier ouvert - pour permettre leur réutilisation - et pris en charge par votre entrepôt pour une conservation à long terme ?
- Savez-vous que la conservation de vos jeux de données à long terme exige une gestion et une préservation par des professionnels ?



Choisir un standard de métadonnées

- Quel(s) type(s) de données sont manipulé(s) ?
- Existe-t-il des préconisations :
 - dans mon établissement ?
 - dans ma discipline ?
 - dans l'entrepôt dans lequel je souhaite déposer mes données ?
- Quelles ressources ai-je à ma disposition pour générer des métadonnées ?



Pour mieux comprendre : parallèle avec un logiciel de gestion bibliographique

- Dans Zotero, certaines métadonnées sont indispensables pour décrire certaines ressources, mais tout à fait superflues pour d'autres
 - Idem pour les métadonnées qui décrivent des jeux de données (ex.: pas besoin de préciser les réactifs utilisés pour décrire un entretien oral)
- Dans Zotero, certains champs sont alignés même si le type de document diffère (ex.: le champ « auteur » d'un article est aligné sur le champ « metteur en scène » d'un film)
 - Idem pour les standards de métadonnées interopérables

Exemples de « standards » dans Zotero

Type de document Rapport

Titre

▼ Auteur (Nom complet)

Résumé

N° du rapport

Type de rapport

Titre de la coll.

Lieu

Institution

Date

Pages

Langue

Titre abrégé

URL

Consulté le

Archive

Loc. dans l'archive

Catalogue de bibl.

Cote

Autorisations

Extra

Date d'ajout 12/05/2021 à 11:41:03

Modifié le 12/05/2021 à 11:41:03

Type de document Carte

Titre

▼ Cartographe (Nom complet)

Résumé

Type

Échelle

Titre de la coll.

Édition

Lieu

Maison d'édition

Date

Langue

ISBN

Titre abrégé

URL

Consulté le

Archive

Loc. dans l'archive

Catalogue de bibl.

Cote

Autorisations

Extra

Date d'ajout 12/05/2021 à 11:44:23

Modifié le 12/05/2021 à 11:44:23

Type de document Émission de TV

Titre

▼ Metteur en scène (Nom complet)

Résumé

Titre du programme

N° de l'épisode

Format

Lieu

Réseau

Date

Durée

Langue

Titre abrégé

URL

Consulté le

Archive

Loc. dans l'archive

Catalogue de bibl.

Cote

Autorisations

Extra

Date d'ajout 12/05/2021 à 11:44:54

Modifié le 12/05/2021 à 11:45:07

Type de document Logiciel

Titre

▼ Programmeur (Nom complet)

Résumé

Titre de la coll.

Version

Date

Système

Lieu

Société

Langage de programmation

ISBN

Titre abrégé

URL

Autorisations

Archive

Loc. dans l'archive

Catalogue de bibl.

Cote

Consulté le

Extra

Date d'ajout 12/05/2021 à 11:44:54

Modifié le 12/05/2021 à 11:47:48

Quel standard pour mes métadonnées ?

Le [RDA Metadata Directory](#) est un annuaire de standards classés par domaine et fournissant des cas d'usages.

- *Dans quelle discipline utilise-t-on le standard FITS ?*
- *Quel standard utiliser pour décrire des données relatifs à des sites archéologiques ?*
- *Quel standard vous semblerait convenir pour vos propres données ?*



Metadata

RDA | Metadata Directory

Edit this page

View the standards

View the extensions

View the tools

View the use cases

Browse by subject areas

Contribute

Add standards

Add extensions

Add tools

Add use cases

Arts and Humanities [Edit](#)

- [Archaeology](#) [Edit](#)
- [Creative art and design](#) [Edit](#)
- [Heritage Studies](#) [Edit](#)
- [Historical and Philosophical Studies](#) [Edit](#)
- [History by Area](#) [Edit](#)
- [History](#) [Edit](#)
- [Information Services](#) [Edit](#)
- [Law](#) [Edit](#)
- [Music](#) [Edit](#)

Engineering [Edit](#)

- [Architecture](#) [Edit](#)
- [Building Conservation](#) [Edit](#)

Life Sciences [Edit](#)

- [Agricultural Economics](#) [Edit](#)
- [Agricultural Science](#) [Edit](#)
- [Animal pathology](#) [Edit](#)
- [Animal physiology](#) [Edit](#)
- [Biochemistry](#) [Edit](#)

(Faire) citer ses données

Recommandation DataCite :

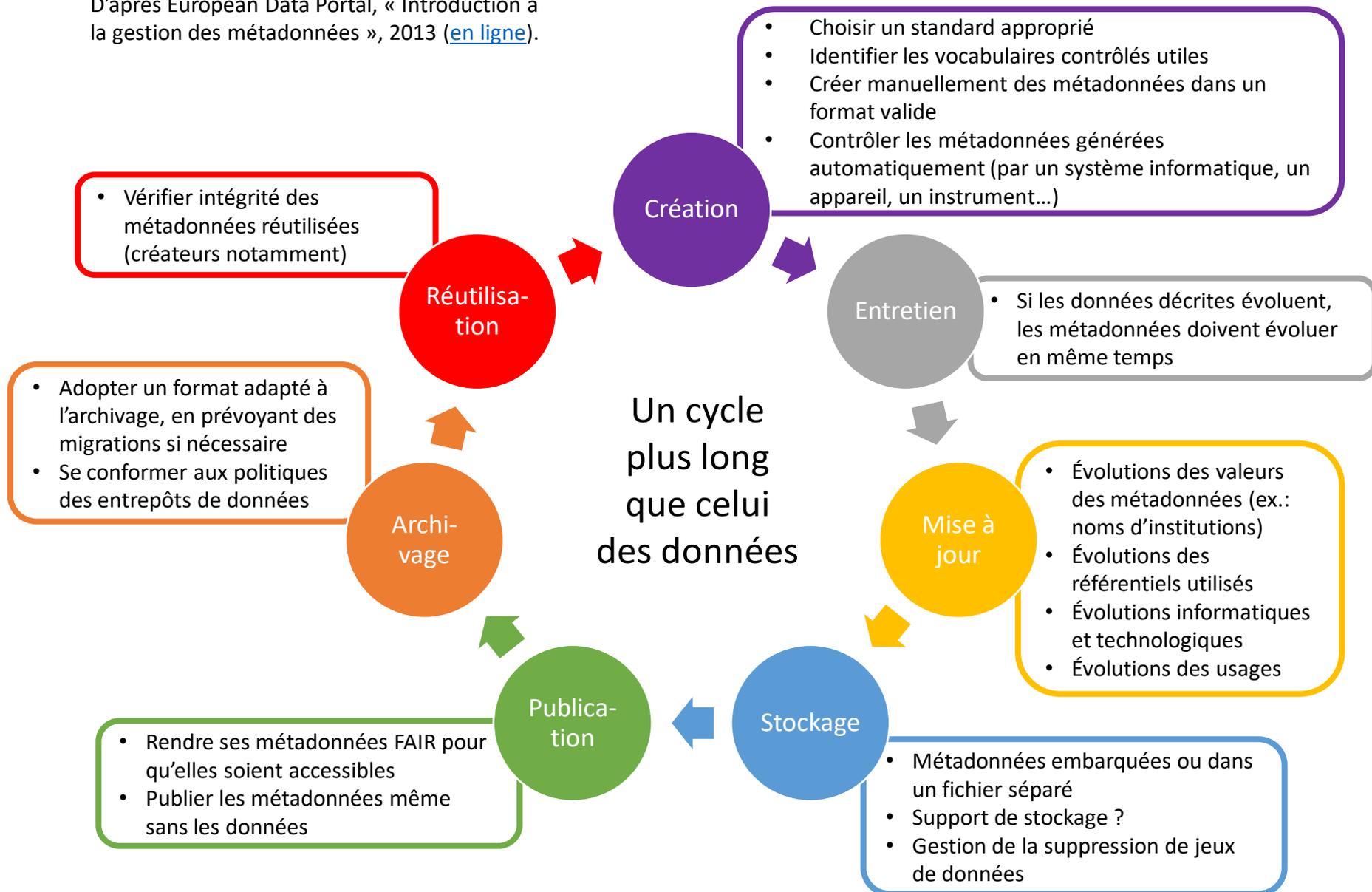
1. Créateur
 2. Année de publication
 3. Titre (Quoi, Quand, Où, Qui, Échelle)
 4. Source / Éditeur
 5. Identifiant pérenne
- + éventuellement version et type de ressource

➤ Outil utile : [DOI Citation Formater](#)

Ex.: Jaillard, B., Deleporte, P., Loreau, M., & Violle, C. (2018). Data from: A combinatorial analysis using observational data identifies species that govern ecosystem functioning [Data set].
<https://doi.org/10.5061/DRYAD.1V1K6RJ>

Le cycle de vie des métadonnées

D'après European Data Portal, « Introduction à la gestion des métadonnées », 2013 ([en ligne](#)).



Cet outil vous permet de générer des métadonnées pour les données de recherche que vous publiez. En complétant les champs situés à gauche, vous verrez apparaître sur la colonne de droite un document xml qui se génère automatiquement.

Plus les éléments de métadonnées sont fournis, plus vos données peuvent être facilement découvertes, réutilisées et citées.

DataCite Metadata Generator - Kernel 4.0

Mandatory Elements (éléments obligatoires)

DOI:
 Vos données ont-elles déjà un DOI?

Title(s): +
 Indiquez le nom ou le titre de votre ressource. Si besoin, ajoutez des variantes au titre en précisant le type de titre.

[titleType] ▾

AlternativeTitle ▾ -

Creator(s): +
 Précisez les producteurs des données ou auteurs d'une publication. Au format « Nom, Prénom ». Des sous-propriétés sont facultatives : noms et prénoms séparés, identifiant auteur (ORCID) et affiliation.
[CREATOR NAME]
 Ancelin-Fabre
 ORCID
 +
 +

Publisher:
 Indiquez l'éditeur, si vos données ont déjà été publiées (nom de l'entité détenant une archive, publiant un document, distribuant, diffusant ou produisant la ressource).

Metadata

```
<?xml version="1.0" encoding="UTF-8"?>
<resource xmlns="http://datacite.org/schema/kernel-4" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4/metadata.xsd">
  <identifier identifierType="DOI">10.00125/64893</identifier>
  <undefined>
    <title>Jeu de données très intéressant</title>
    <title titleType="AlternativeTitle">Titre alternatif de mon jeu de données</title>
  </undefined>
  <undefined>
    <creator>
      <givenName>Justine</givenName>
      <familyName>Ancelin-Fabre</familyName>
      <nameIdentifier Schéma identifiant auteur (ex: ORCID)="ORCID" Schéma URI (ex: https://orcid.org/)="https://orcid.org">orcid.org/0000-0000-0000-0000</nameIdentifier>
      <affiliation>URFIST de Paris</affiliation>
    </creator>
  </undefined>
  <publisher>Zenodo</publisher>
  <publicationYear>2019</publicationYear>
  <language>FR</language>
  <undefined>
    <size>86Go</size>
  </undefined>
  <undefined>
    <undefined>XML</undefined>
  </undefined>
  <version>2.3</version>
  <undefined>
    <rights URI des droits="http://creativecommons.org/licenses/by/3.0/fr/deed.en">Creative Commons Attribution 3.0 French License</rights>
  </undefined>
  <undefined>
    <undefined>
      <funderName>European Commission</funderName>
      <undefined funderIdentifierType="Crossref Funder ID">0123-456-76</undefined>
    </undefined>
  </undefined>
</resource>
```

Générer des métadonnées

Select All Reset Ready to Save

Autres outils pour créer des métadonnées



[Modèle OTELO-Inist CNRS](#) pour créer des métadonnées sous forme de tableur (format CSV)

À utiliser en suivant le [Guide des bonnes pratiques : gestion et valorisation des données de la recherche](#)



CatMDEdit

[CatMDEdit](#), pour générer des informations géographiques respectueuses des normes ISO



Morpho

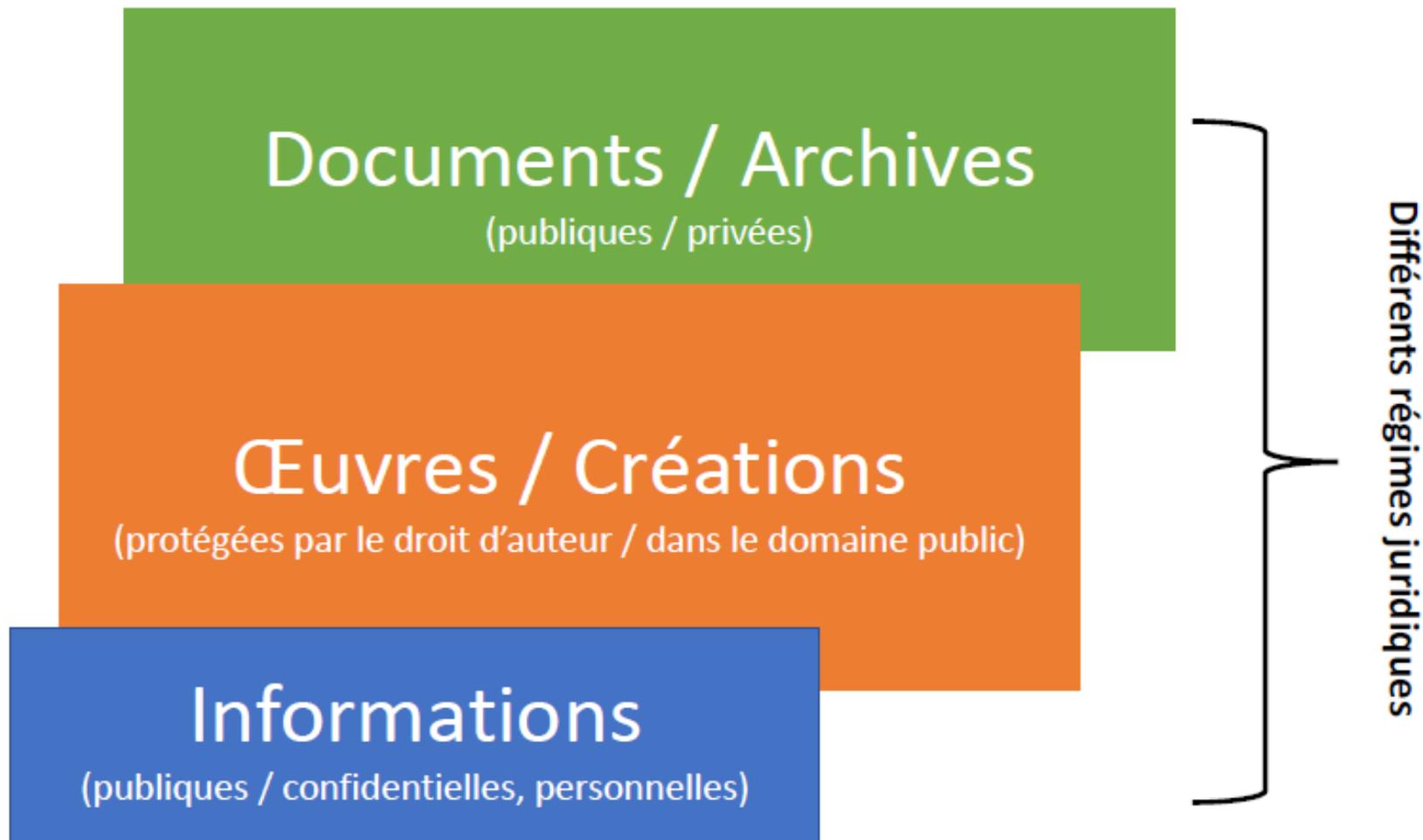
[Morpho](#), pour créer des métadonnées selon standard EML

➤ Voir [Outils de création de métadonnées](#), DoRANum



4- Le cadre juridique
des données de la
recherche

Une définition juridique des données de la recherche ?



Source : L. Maurel, « Quelles règles juridiques applicables aux données de la recherche ? », 2019 ([en ligne](#)).

Données et archives

« Les **archives** sont l'ensemble des documents, **y compris les données**, quels que soient leur date, leur lieu de conservation, leur forme et leur support, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité. »

[Art. L. 211-1](#) du *Code du Patrimoine*, modifié par la loi n°2016-925 du 7 juillet 2016 – art. 59, qui introduit les termes « y compris les données »

Données et archives publiques

« Les archives publiques sont les documents qui procèdent de **l'activité** de l'État, des collectivités territoriales, **des établissements publics** et des autres personnes morales de droit public. »

[Art. L. 211-4](#) du *Code du Patrimoine*

Données et documents administratifs

« Sont considérés comme **documents administratifs** [...], quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public [...]. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions, codes sources et décisions. »

[Art. L.300-2](#) du *Code des relations entre le public et l'administration* (CRPA)

Quel cadre réglementaire suivre ?

Est-ce que je produis des documents administratifs ?



Oui si mon employeur est public

- ❖ Tout ce que je fait dans le cadre de ma mission peut être considéré comme un document administratif.
- ❖ Attention ! Pour les **doctorants** : si la thèse est cofinancée ou réalisée en collaboration avec un partenaire privé => il faut se reporter au contrat.

- ❖ Cas particulier : je suis **chercheur ou enseignant-chercheur** : mes écrits, cartes, photographies, plans qui sont originaux et donc soumis au droit d'auteur m'appartiennent (exception – loi DADVSI 2006-961)

Mais le reste appartient bien à mon employeur.

Données et Open Data

« De façon générale, l'expression « **Open Data** » peut être définie comme une démarche de communication des **documents ou données publics**, afin qu'ils soient diffusés de manière structurée selon une méthode garantissant leur libre accès et leur réutilisation par tous, sans restriction technique, juridique ou financière injustifiée. »

N. Becard et al., *Ouverture des données de recherche. Guide d'analyse du cadre juridique en France*, INRA, 2017 ([en ligne](#)).



Les établissements d'ESR sont-ils concernés ?

- Toutes les données (y compris codes sources) produites par des établissements d'ESR dans le cadre de leur mission de service public (y compris EPIC) sont des documents administratifs, communicables à toute personne qui en fait la demande, sauf exceptions légales.
- Condition : documents « achevés » uniquement (ex.: pas les cahiers de laboratoires), sauf information environnementale



/!\ La notion juridique de données de la recherche a une acception plus restreinte que la notion scientifique.

LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique

[Masquer le panneau de navigation](#)

[Imprimer](#)

Navigation

LOI n° 2016-1321 du 7 octobre 2016

- ▶ **Version initiale**
- ▶ [Version en vigueur au 10 avril 2020](#)

Version consolidée à la date du ...

Jour Mois Année
Ex: 2020

Sommaire

Titre Ier : LA CIRCULATION DES DONNÉES ET DU SAVOIR

[En savoir plus sur ce texte...](#)

JORF n°0235 du 8 octobre 2016
texte n° 1

LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)

NOR: ECFI1524250L

ELI: <https://www.legifrance.gouv.fr/eli/loi/2016/10/7/ECFI1524250L/jo/texte>

Alias: <https://www.legifrance.gouv.fr/eli/loi/2016/10/7/2016-1321/jo/texte>

Article 38 [En savoir plus sur cet article...](#)

Le code de la propriété intellectuelle est ainsi modifié :

1° Après le second alinéa du 9° de l'article L. 122-5, il est inséré un 10° ainsi rédigé :

« 10° 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** ;

2° Après le 4° de l'article L. 342-3, il est inséré un 5° ainsi rédigé :

« 5° Les copies ou reproductions numériques de la base réalisées par une personne qui y a licitement accès, en vue de fouilles de textes et de données incluses ou associées aux écrits scientifiques dans un cadre de recherche, à l'exclusion de toute finalité commerciale. La conservation et la communication des copies techniques issues des traitements, au terme des activités de recherche pour lesquelles elles ont été produites, sont assurées par des organismes désignés par décret. Les autres copies ou reproductions sont détruites. »

Article 1 [En savoir plus sur cet article...](#)

I.-Sous réserve des articles [L. 311-5](#) et [L. 311-6](#) du code des relations entre le public et l'administration et sans préjudice de l'article L. 114-8 du même code, les administrations mentionnées au premier alinéa de l'article L. 300-2 dudit code sont tenues de communiquer, dans le respect de la [loi n° 78-17 du 6 janvier 1978](#) relative à l'informatique, aux fichiers et aux libertés, les documents administratifs qu'elles détiennent aux autres administrations

La libre réutilisation des données de la recherche

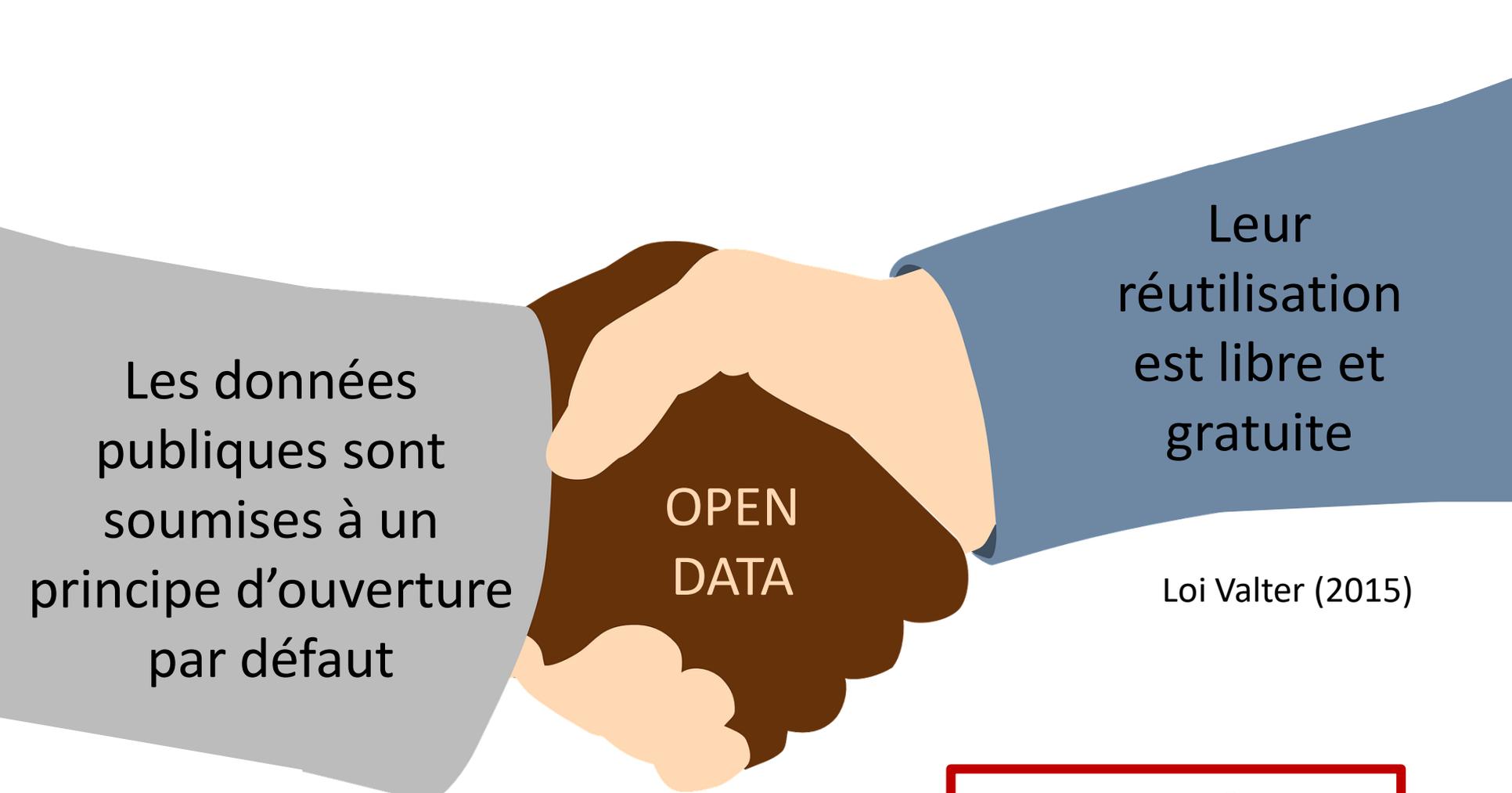
« Dès lors que les **données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'État**, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été **rendues publiques** par le chercheur, l'établissement ou l'organisme de recherche, leur **réutilisation est libre**. »

[Art. L533-4](#) du *Code de la Recherche*, créé par la loi pour une République numérique

La LRN et les DR

- Ébauche d'un statut spécifique pour les données de recherche
- Protection puissante contre la captation par les éditeurs
- Des données de recherche « libres » dès la publication
- Limites :
 - Données protégées par un autre droit spécifique (secret industriel et commercial, droit d'auteur...)
 - Données soumises à une réglementation particulière (données personnelles, données relatives à la Défense nationale...)
- Droit *sui generis* des bases de données :
 - N'appartient en réalité que rarement au chercheur
 - Est neutralisé par défaut si appartient à un établissement d'ESR
- Open Data par défaut applicable aux données de la recherche dès lors qu'elles appartiennent aux établissements d'ESR

Voir L. Maurel, « Quel statut pour les données de la recherche après la loi numérique ? », - *S.I.Lex* -, 2016 ([en ligne](#))



Les données
publiques sont
soumises à un
principe d'ouverture
par défaut

Loi pour une République numérique (2016)

Leur
réutilisation
est libre et
gratuite

Loi Valter (2015)

OPEN
DATA

Sauf
exceptions

Repérer les écueils des contrats d'édition

Pistes bibliographiques :

- S. Le Strujon, « Le chercheur face au contrat d'édition », 2017 ([en ligne](#)).
- A.-L. Stérin, « Le/la chercheur.e publie », *Ethique et droit*, 2017 ([en ligne](#)).
- Consortium Couperin, « Les contrats d'édition », *Science Ouverte France* (en ligne <https://scienceouverte.couperin.org/contrat/>).
- Consortium Couperin, « Les avenants au contrat d'édition », *Science Ouverte France* (en ligne <https://scienceouverte.couperin.org/les-avenants-au-contrat-dedition/>).

Les documents administratifs sont librement communicables à toute personne qui en fait la demande, sauf exceptions légales.

*Si possible, mise à disposition exhaustive et immédiate, pour tous
Sinon, mise à disposition ciblée, après examen et sur demande*

Communication obligatoire

- Données géographiques informatisées (directive INSPIRE): communication obligatoire d'office
- Données relatives à des émissions de substances dans l'environnement (convention d'Aarhus): communication obligatoire sur demande

Communication interdite

- Données présentant des risques pour la défense nationale
- Données à risque pour la sécurité de l'État, la sécurité publique, la sécurité de l'établissement
- **Données protégées par des secrets (industriel et commercial, médical, financier...)**

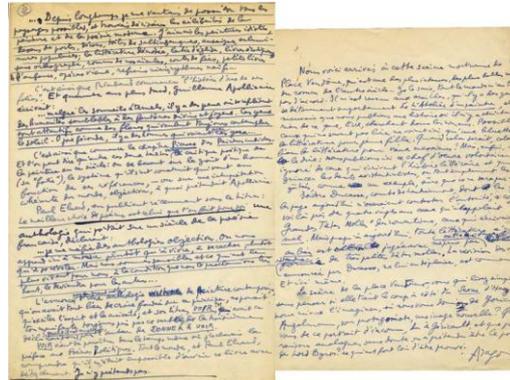
Communication sous conditions

- Données présentant des risques pour la protection du potentiel scientifique et technique de la nation
- **Données protégées par le droit d'auteur** ou autre droit de propriété intellectuelle
- **Données personnelles**
- Données statistiques
- Données liées à un contrat avec un tiers

Diffuser des données protégées par le droit d'auteur de tiers



Images (y compris photos)



Œuvres littéraires (y compris inédites)

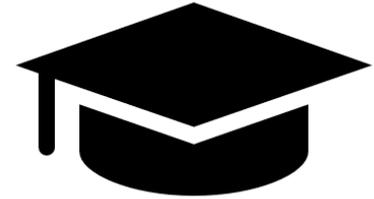


Logos, marques, designs

La diffusion publique de données protégées par le droit d'auteur de tiers est autorisée si :

- L'œuvre est dans le domaine public
- L'auteur a donné son autorisation expresse (attention : qui ne dit mot REFUSE)
- L'auteur a par ailleurs déjà autorisé ce type d'usage par le biais d'une licence

L'exception pédagogique et de recherche



- Maximum 20 images protégées
 - résolution max.: 72 ppp
 - définition max.: 400x400 pixels
 - mieux vaut intégrer des liens vers les ressources citées
- Aucune partition, même sous forme d'extrait
- Les œuvres sous droits ne doivent pas pouvoir être extraites de la publication
- Pas d'utilisation commerciale possible

➤ En savoir plus : A.-L. Stérin, « L'exception pédagogique et de recherche », *Ethique et droit*, 2017 ([en ligne](#)).

Diffuser des données personnelles

- Données directement identifiantes : nom, prénom, adresse, photo, voix...
- Données indirectement identifiantes : numéro de téléphone, croisement d'informations
- Ouverture et/ou communication sur demande uniquement si au moins une de ces 3 conditions :
 - Avoir recueilli le consentement des personnes
 - Avoir convenablement anonymisé les données
 - Obéir à un autre texte législatif ou réglementaire

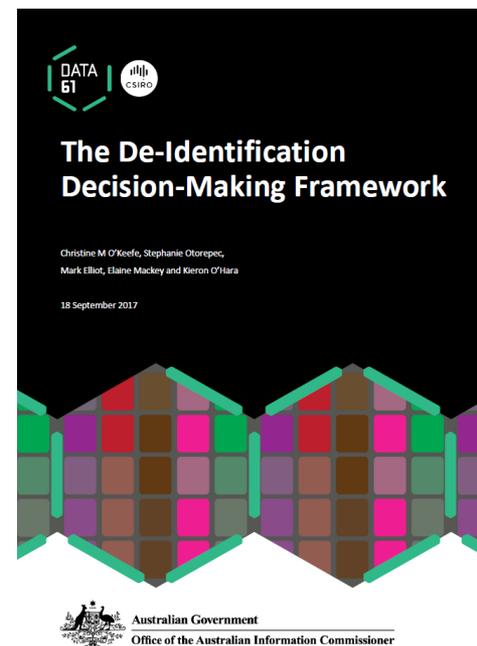


Anonymiser ses données



<http://theconversation.com/donnees-anonymes-bien-trop-faciles-a-identifier-123157>

Voir aussi B. Nguyen, « L'anonymisation : théorie et pratique », 2019 ([en ligne](#)).



[De-identification DMF](#)

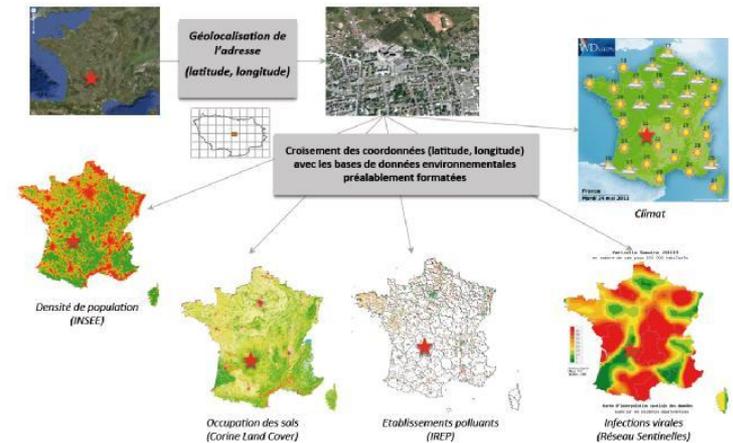
Données relatives au « secret industriel et commercial »

- Données peuvent concerner :
 - Secret des procédés
 - Secret des informations économiques et financières
 - Secret des stratégies commerciales ou industrielles
 - Tous les éléments sensibles ayant notamment un impact sur l'environnement concurrentiel de l'établissement et de ses partenaires
- Origine possible des données :
 - Données développées par l'établissement public
 - Données transmises à un établissement public mais appartenant à des partenaires industriels
- Communication interdite en dehors des intéressés
- Ou communicables après occultation des informations couvertes par ce secret



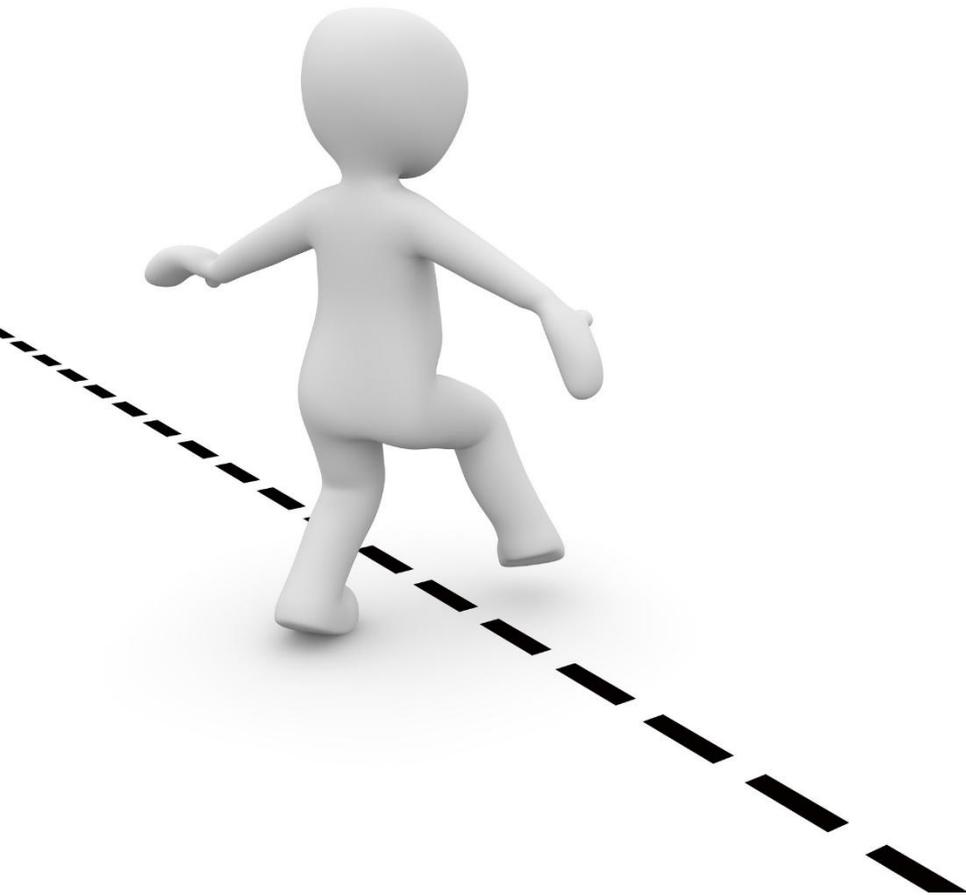
Cas particulier des données environnementales

- De plus en plus de données environnementales (relevés météo, images satellites, relevés topographiques, données sur les ressources naturelles...)
- Ces données sont géoréférencées, organisées au sein de [systèmes d'information géographiques](#)
- Convention d'Aarhus (1998) : autorisation pour toute personne d'accéder aux informations relatives à l'environnement détenues par les autorités publiques et de participer au processus décisionnel concernant l'environnement
- Directive européenne Inspire de mars 2007 : ouverture systématique de 34 types de données géographiques dès lors qu'elles sont disponibles sous forme numérique



Source : SIG 2011

+ Protocole d'Aarhus sur les [polluants organiques persistants](#) : ouverture immédiate obligatoire pour les données concernant l'émission de substances polluantes dans l'environnement



Au-delà des textes de loi

Les exigences des
financeurs

Exigences de financeurs : la Commission européenne



- Programme-cadre Horizon Europe (2021-2027) :
 - Les données doivent être rendues FAIR ;
 - Les données doivent être déposées dès que possible dans un entrepôt de confiance ;
 - Les données doivent être « aussi ouvertes que possible, aussi fermées que nécessaire », et autant que possible, réutilisables sans condition autre qu'une éventuelle reconnaissance de la paternité des auteurs (licence CC-BY ou CC-0) ;
 - Toutes les informations et outils permettant de réutiliser les données ou d'en vérifier la qualité doivent être fournis ou décrits ;
 - Les métadonnées décrivant les données doivent être mises à disposition sous licence CC-0, et inclure des identifiants pérennes ;
 - Les bénéficiaires des financements Horizon Europe doivent fournir des plans de gestion de données à jour.

Source : [Horizon Europe \(HORIZON\) General Model Grant Agreement](#), v1 (provisoire), 25/02/2021, p. 108-109

L'accès aux données

Le chercheur obtient un financement européen

Rédaction d'un plan de gestion des données (PGD ou DMP) dans les 6 mois :
Quelles données peuvent être partagées ? Dans quelles conditions ?

Restrictions : données sensibles (sécurité, industrie...), données sous droits, données à caractère personnel...

Rendre les données
les plus FAIR possible :

- Faciles à trouver
- Accessibles
- Interopérables
- Réutilisables

Déposer les données qui
accompagnent les
publications scientifiques
(+ éventuellement autres
données) dans un entrepôt
de confiance

Au plus tard au moment de la
publication de l'article associé

Fournir tous les matériaux
nécessaires à la compréhension
et à la réutilisation des données
+ métadonnées riches (PID,
licences...) sous licence CC-0

Données **aussi ouvertes que possible**, **aussi fermées que nécessaire**
Réutilisation libre (téléchargement, exploitation, TDM...)

PLAN NATIONAL POUR LA SCIENCE OUVERTE

[Le texte du Plan national pour la science ouverte](#) (axe n°2 : Structurer et ouvrir les données de la recherche)

[Le collège Données de la recherche du CoSO](#)

MERCREDI 4 JUILLET 2018

« Lors de l'annonce du plan « Intelligence artificielle » au Collège de France, le 29 mars 2018, le Président de la République a annoncé la mise en place d'un principe d'**ouverture par défaut pour toutes les données publiées dans le cadre d'appels à projet sur fonds publics.** »

05/09/2019

L'ANR met en place un plan de gestion des données pour les projets financés dès 2019



Dans le cadre de sa politique science ouverte, l'Agence nationale de la recherche (ANR) demande l'élaboration d'un Plan de Gestion des Données (PGD) pour les projets financés à partir de 2019. Ce document vise à décrire la façon dont les données de recherche seront produites, collectées, réutilisées, stockées, protégées, disséminées ou encore conservées à long terme. Objectif : une meilleure anticipation

[Source de l'article](#)



[Plan d'action 2019 de l'ANR](#), p. 9
Voir aussi les [recommandations du CoSO](#)

Plan d'action 2019

Date de publication :

26 juillet 2018

Ce document a été adopté le 03 juillet 2018 par le Conseil d'administration de l'ANR

L'ANR et les données



Plan d'action 2020

Appel à projets générique 2020

(version 1.1)

Date de publication 25 juillet 2019

ETAPE 1
CLOTURE DE LA SOUMISSION
DES PRE-PROPOSITIONS (PRC, PRCE, JCJC)
ET DE L'ENREGISTREMENT (PRCI)

Le jeudi 24 octobre 2019 à 17h00 (heure de Paris)

ETAPE 2 : CLOTURE DE LA SOUMISSION
DES PROPOSITIONS DÉTAILLÉES

Avant de déposer une (pré-)proposition de projet de recherche, il est nécessaire de lire attentivement l'ensemble du présent document ainsi que le Guide de l'AAPG 2020 (disponible fin Août 2019 sur le site de l'ANR) et le règlement relatif aux modalités d'attribution des aides de l'ANR (<http://www.anr.fr/9F>)

« Le coordinateur ou la coordinatrice s'engage [...] à fournir un plan de gestion des données (DMP) actualisé tout au long de la vie du projet selon les modalités communiqués lors du conventionnement. » ([Guide de l'AAPG 2020](#), p. 20)

C.4. Publications scientifiques et données de la recherche

Dans le cadre de la contribution de l'ANR à la promotion et à la mise en œuvre de la science ouverte, et en lien avec le plan national pour la science ouverte, les partenaires au projet déposé s'engagent en cas de financement :

- à déposer leurs publications scientifiques (texte intégral) issues du projet de recherche, de développement et d'innovation dans une archive ouverte, soit directement dans HAL soit par l'intermédiaire d'une archive institutionnelle locale, dans les conditions de l'article 30 de la Loi « Pour une République numérique »^{20,21}
- à fournir dans les 6 mois qui suivent le démarrage du projet un plan de gestion des données (PGD)²² selon des modalités communiquées dans les Conditions particulières et le Règlement financier de l'ANR.

Par ailleurs, l'ANR recommande de privilégier la publication dans des revues ou ouvrages nativement en accès ouvert.²³



5- Diffusion et réutilisation des données



Comment diffuser
ses données ?

Un ensemble de bonnes pratiques

S'intéresser à la gestion courante des données : classement, nommage, format, métadonnées, documentation, stockage, sécurité...

Rendre ses données les plus FAIR possible

Réfléchir au cadre juridique, aux obligations et/ou interdictions qui régissent la diffusion des données

Réfléchir à l'audience souhaitée : accès ouvert ou restreint ?
Quelle(s) réutilisation(s) possible(s) (dans le respect de la législation) ?

3 façons de diffuser ses données

Source

COMMUNICATIONS BIOLOGY

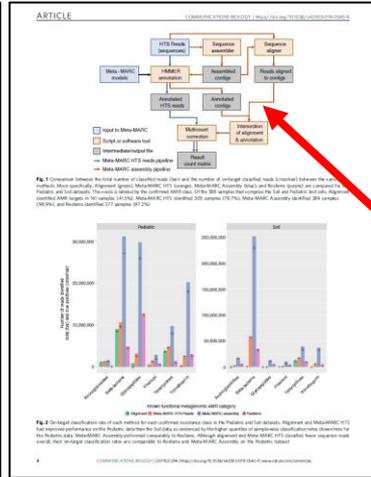
ARTICLE

<https://doi.org/10.1038/s43772-020-00445-8> OPEN

Hierarchical Hidden Markov models enable accurate and diverse detection of antimicrobial resistance sequences

Steven M. Labin¹, Alan Kuhn², Bahar Abanzadeh³, Nicole R. Reyes⁴, Chris Diaz⁵, Martin Muggli⁶, Bob Kaye⁷, Zaid Aboel⁸, Mattia Prosperini⁹, Keith E. Bekk¹, Paul S. Morley⁷ & Christina Boucher¹⁰

The characterization of antimicrobial resistance genes from high-throughput sequencing data has become fundamental to public health research and regulation. The massive amount of sequence reads to database of known antimicrobial resistance genes to determine the genes present in the sample. Mapping sequence reads to known genes is traditionally accomplished using alignment. Alignment methods have high specificity but are limited in their ability to detect sequences that are divergent from the reference database, which can result in a substantial loss of sequence data. We address this shortcoming through the creation of Meta-MARK, which enables detection of diverse resistance sequences using hierarchical DNA-based Hidden Markov Models. We first describe Meta-MARK and then demonstrate its efficacy on simulated and functional metagenomic datasets. Meta-MARK has higher sensitivity relative to competing methods. This sensitivity allows for detection of sequences that are divergent from known antimicrobial resistance genes. This functionality is important to accurately identify antimicrobial resistance genes. This functionality is important to accurately identify antimicrobial resistance genes.



1- Intégrer les données directement dans une publication

ScienceDirect Journals & Books

Source

Download PDF Share Export

Register Sign in

Brought to you by Université de recherche Paris sciences et lettres – PSL Research University

Outline

Highlights

- Tables (1)
- Table 1
- Extras (4)
- Download all
- Supplementary Table S1
- Supplementary Table S2
- Supplementary Table S3
- Supplementary material

Journal of Proteomics Volume 217, 15 April 2020, 103645

Intermittent fasting from dawn to sunset for 30 consecutive days is associated with anticancer proteomic signature and upregulates key regulatory proteins of glucose and lipid metabolism, circadian clock, DNA repair, cytoskeleton remodeling, immune system and cognitive function in healthy subjects

Recommended articles

- Nutritional Ketosis and photobiomodulation: Prodiagnosis and Photodynamic Therapy
- Proteomics reveals... cancer differ...
- The golden mussel proteome and its...

2- Joindre les données en annexe / matériel supplémentaire



3- Déposer ses données sur une plateforme dédiée

Données supplémentaires sur les plateformes de preprints et d'éditeurs

2. [arXiv:1904.11122](https://arxiv.org/abs/1904.11122) [pdf, other] [math.GT](#) [math.DG](#)

Thurston's sphere packing on 3-dimensional manifolds, I

Authors: [Xiaokai He](#), [Xu Xu](#)

Abstract: Thurston's sphere packing on 3-dimensional manifolds is a generalization of Thurston's circle packing on surfaces, allowing adjacent spheres to intersect with non-obtuse angles. In this paper, we prove that the discrete Laplacian for a large class of Thurston's Euclidean sphere packings is negative semi-definite. As an application, we prove a local rigidity theorem and an infinitesimal rigidity th... [▽ More](#)

Submitted 24 April, 2019; originally announced April 2019.

Comments: 31 pages, two [ancillary files](#) with Mathematica codes and outputs

Supplementary Data



2013 Supplementary Data

January 2013, Volume 28, Number 1

 Paleobiology and paleoecology of the early Turonian (Late Cretaceous) ammonite *Pseudaspidoceras flexuosum*
Christina Ilrim
[Supplementary Data](#) [Abstract](#) 9

February 2013, Volume 28, Number 2

 Preservational and morphological variability of assemblages of agglutinated eukaryotes in Cryogenian cap carbonates of northern Namibia
Lilly A. Dalton, Tanja Bosak, Francis A. Macdonald, Daniel J.G. Lahr, and Sara B. Pruss
[Supplementary Data](#) [Abstract](#) 67

Vinegaroon (Arachnida: Thelyphonida: Thelyphonidae) trackway production and morphology: Implications for media and moisture control on trackway morphology and a proposal for a novel system of interpreting arthropod trace fossils

Joshua D. Schmerge, David J. Riese, and Stephen T. Hasiotis

Supplementary Data Key

Windows Media Video (.wmv files) QuickTime Movie files (.mov)

Supplementary Data 1	Supplementary
Data 1	Supplementary
Supplementary Data 2	Supplementary
Data 2	Supplementary
Supplementary Data 3	Supplementary
Data 3	Supplementary
Supplementary Data 4	Supplementary
Data 4	Supplementary
Supplementary Data 5	Supplementary
Data 5	Supplementary
Supplementary Data 6	Supplementary
Data 6	Supplementary
Supplementary Data 7	Supplementary
Data 7	Supplementary
Supplementary Data 8	Supplementary

Images présentées dans M.-L. Malingre, « Gérer, diffuser ses données de recherche en SHS : introduction aux jeux, méthodes, pratiques », 2018 ([en ligne](#)).

Les données de la recherche dans les archives ouvertes

Données de recherche	41416	4	98	41518
Image	38718		4	38722
Vidéo	2388			2388
Carte	185			185
Logiciel	24	4	94	122
Son	101			101

[HAL au 17 mai 2019](#)

Sextant
Infrastructure de données géographiques marines et littorales

CATALOGUE CARTE MES TÉLÉCHARGEMENTS S'IDENTIFIER

Résultats 1 à 20 sur 7313 : 20 par page Tri par : Mise à jour

The relationship between genetic diversity and reef state in the honeycombworm ...

Profils de vulnérabilité des zones de production du département de la Charente-Mar...

[Sextant \(Ifremer\) au 17 mai 2019](#)

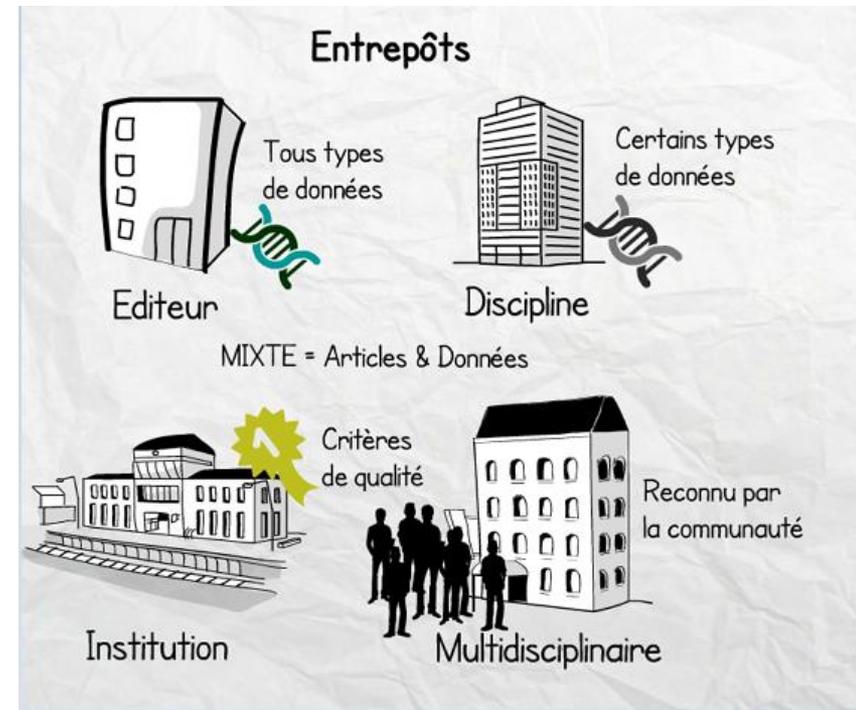
Entrepôt ?



- « Un entrepôt de données (*Data repository, digital repository*) est un réservoir constitué majoritairement de données de recherche, brutes ou élaborées, qui sont décrites par des métadonnées de façon à pouvoir être retrouvées. » ([Pascal Aventurier](#))
- « Service en ligne permettant la collecte, la description, la conservation, la recherche et la diffusion des jeux de données » (Pelletier)
- « Un *entrepôt de données* est une archive numérique qui collecte et diffuse des jeux de données et leurs *métadonnées*. Un grand nombre d'entrepôts de données acceptent également des publications et permet de lier les publications afférentes. » (définition OpenAIRE, citée par [Marilou Pain](#) en 2016)

Différentes sortes d'entrepôts

- Périmètre :
 - Tous jeux de données
 - Données liées à des publications
 - Publications scientifiques et/ou données
- Typologie :
 - Entrepôts d'éditeurs
 - Entrepôts disciplinaires ou thématiques
 - Entrepôts institutionnels
 - Entrepôts généralistes



Source : [La minute entrepôts](#) (Doranum)

Les critères de choix

- Dans mon périmètre disciplinaire ?
Recommandé par mes collègues ? Les financeurs ? Mon employeur ?
- Accepte mes fichiers (format, taille...) ?
- Accepte mes métadonnées ? Fournit un identifiant pérenne ?
- Permet de modifier mes données et accéder aux anciennes versions ?
- Propose licences de diffusion ? Permet de restreindre l'accès ?
- Procédure de validation du dépôt ?
Coûts éventuels associés ?
- Certifié / reconnu ? [Core Trust Seal](#) ?
[Principes TRUST](#) ?



- Autant que possible, privilégier un entrepôt (par ordre de préférence) :
1. Spécialisé dans la gestion de votre type spécifique de données
 2. Spécialisé dans votre discipline
 3. Généraliste ou institutionnel

[Re3data](#), un annuaire des entrepôts disponibles

Voir aussi :

- L'annuaire spécialisé [FAIRsharing](#), qui permet de cibler très finement votre thématique de recherche
- La [liste OAD](#) classe les entrepôts par disciplines
- Le [Trusted Digital Repository](#) qui recense les entrepôts européens certifiés

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES



- Dans quel entrepôt trouver des données relatives aux énergies nouvelles ?
- Quelle(s) institution(s) s'occupe(nt) de l'entrepôt OAGR ?
- L'entrepôt DAIS fournit-il des identifiants pérennes ? Si oui, lesquels ?

Astuce : à partir du menu Browse, choisissez l'affichage Texte au lieu de Graphique

Permet de savoir où déposer, mais aussi où trouver des données à réutiliser !

Databases



A catalogue of databases, described according to the [BioDBcore guidelines](#), along with the standards used within them; partly compiled with the support of Oxford University Press ([NAR Database Issue](#) and [DATABASE Journal](#)).

[Contribute by adding a database](#) [Any problems? Please tell us!](#)

Search Databases

Search

Search

Reset

Advanced

Showing records 1 - 50 of 1415.

« 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 »

28 29 »

View as Table View as Grid

Sort by
Name

Recommended Records
Recommended

Associated Publication?
No Publication Has Publication

Claimed?
No Maintainer Has Maintainer

Record Status
Uncertain Deprecated In developn Ready

Domains

Chemical Entity	387
Molecular Entity	280
Data Transformation	271
Sequence Feature	249

Show More

Registry	Name	Abbreviation	Type	Subject	Domain	Taxonomy	Related Database	Related Standard	Related Policy	In Co
	3D interacting domains	3DID	Database	Life Science	Protein Interaction Protein Structure Molecular Interaction Protein Structure	All	wwPDB	PDB GO HMMER Profile File Format FASTA	None	None
	4DNucleome Data Portal	4DN	Database	Life Science	Experiment Metadata Experimental Measurement Nucleic Acid Sequence Chromosome Data Transformation	Drosophila melanogaster Gallus gallus Homo sapiens Mus musculus	GEO SRA ENCODE	EFO FASTQ Sequence and Sequence Quality Format BigBed bigWig Track Format BAM Plus 4 more...	None	None
	4TU.Centre for Research Data	4TU.ResearchData	Database	Aerospace Engineering Architecture Building Design Civil Engineering Computer Science Plus 11 more...	Data Storage	All	None	DataCite Metadata Schema Francis.org DC OAI-ORE DOI Plus 2 more...	Elsevier's Data Policy Taylor and Francis Data Policy	Cher

All versions

Found 71387 results.

< 1 2 3 4 5 6 7 8 9 >

Sort by:

Best match
asc.

Access Right

- Open (1682547)
- Closed (32390)
- Restricted (4381)
- Embargoed (1175)

File Type

- Pdf (887061)
- Jpg (361103)
- Png (217531)
- Html (80454)
- Zip (77683)
- Xml (15132)
- Hdf5 (15115)
- Docx (12001)
- Txt (9425)
- Xlsx (9198)

July 10, 2018 (v1) Dataset Open Access

View

Supplementary material 5 from: Seppälä S, Henriques S, Draney M, Foord S, Gibbons A, Gomez L, Kariko S, Malumbres-Olarte J, Milne M, Vink C, Cardoso P (2018) Species conservation profiles of a random sample of world spiders II: Gnaphosidae to Nemesiidae. Biodiversity Data Journal 6: e26203. <https://doi.org/10.3897/BDJ.6.e26203>

Seppälä, Sini; Henriques, Sérgio; Draney, Michael; Foord, Stefan; Gibbons, Alastair; Gomez, Luz; Kariko, Sarah; Malumbres-Olarte, Jagoba; Milne, Marc; Vink, Cor; Cardoso, Pedro;

Distribution of *Leptodrassus croaticus* Dalmas, 1919

Uploaded on July 10, 2018

July 10, 2018 (v1) Dataset Open Access

View

Supplementary material 9 from: Seppälä S, Henriques S, Draney M, Foord S, Gibbons A, Gomez L, Kariko S, Malumbres-Olarte J, Milne M, Vink C, Cardoso P (2018) Species conservation profiles of a random sample of world spiders II: Gnaphosidae to Nemesiidae. Biodiversity Data Journal 6: e26203. <https://doi.org/10.3897/BDJ.6.e26203>

Seppälä, Sini; Henriques, Sérgio; Draney, Michael; Foord, Stefan; Gibbons, Alastair; Gomez, Luz; Kariko, Sarah; Malumbres-Olarte, Jagoba; Milne, Marc; Vink, Cor; Cardoso, Pedro;

Distribution of *Zelotes anthereus* Chamberlin, 1936

Uploaded on July 10, 2018

August 18, 2015 (v1) Dataset Open Access

View

TARL_41CE19_2006-4-84

<https://zenodo.org/>

<https://www.datadryad.org/>

DRYAD About ▾ For researchers ▾ For organizations ▾ Contact us Log in Sign up

Search for data

Search terms

Enter keyword, author, title, DOI, etc. Example: herbivory **Go**

[Advanced search](#)

Refinements

Subject: urban

Results/page Sort items by in order

Now showing items 1-20 of 180 [1](#) [2](#) [3](#) [4](#) ... [9](#) [Next Page >>](#)

[Dryad Data Packages \(180\)](#) [KNB \(174\)](#) [Dryad Data Files \(0\)](#) [TreeBASE \(0\)](#)

Angelier F, Brischoux F (2019) Data from: Are house sparrow populations limited by the lack of cavities in urbanized landscapes? an experimental test. *Journal of Avian Biology*
<https://doi.org/10.5061/dryad.gh2hh05>

DeCandia AL, Henger CS, Krause A, Gormezano LJ, Weckel M, Nagy C, Munshi-South J, vonHoldt BM (2019) Data from: Genetics of urban colonization: neutral and adaptive variation in coyotes (*Canis latrans*) inhabiting the New York metropolitan area. *Journal of Urban Ecology*
<https://doi.org/10.5061/dryad.c0282c8>

Refine Search

Author

- [Brans, Kristien I. \(4\)](#)
- [Frank, Steven D. \(4\)](#)
- [Munshi-South, Jason \(4\)](#)
- [Cacccone, Adalgisa \(3\)](#)
- [Costa, Federico \(3\)](#)
- [De Meester, Luc \(3\)](#)
- [Revell, Liam J. \(3\)](#)
- [Stoks, Robby \(3\)](#)
- [Altizer, Sonia \(2\)](#)
- [Angelier, Frédéric \(2\)](#)
- [... View More](#)

Subject

- [urbanization \(50\)](#)
- [urban ecology \(33\)](#)
- [urbanisation \(16\)](#)
- [urban \(8\)](#)
- [conservation \(7\)](#)
- [urban evolution \(7\)](#)
- [adaptation \(6\)](#)

store, **share**, discover **research**

get more citations for all of the outputs of your academic research
over 30,000 citations of **figshare** content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

Agricultural and Veterinary Sciences

Astronomy, Astrophysics, Space Science

Biological Sciences

Built Environment and Design

Chemistry

Commerce, Management, Tourism and Services

Earth and Environmental Sciences

Engineering

Health Sciences

Humanities

Information and Computing Sciences

Language, Communication and Culture

Mathematics

Meta Science

Physics

Psychology

Social Science

Studies in Creative Arts and Writing

Studies in Human Society

Technology

Uncategorised

<https://figshare.com/>

Accueil > Blog > Faisabilité d'un service générique d'accueil et de diffusion des données...

Faisabilité d'un service générique d'accueil et de diffusion des données simples : ambitions et scénarios

OUVRIR
LA SCIENCE !



Source : [actualités du CoSO](#)

ACTUALITÉS DU COMITÉ 26/03/2021

La troisième étude du groupe « Service générique d'accueil et de diffusion de données simples » du Comité pour la science ouverte est mise en ligne. Elle évalue différents scénarios et niveaux d'ambition fonctionnels relatifs à la création d'un service générique national de données. En partant des besoins identifiés lors de la première phase, le document propose différentes modalités d'implémentation qui sont présentées au travers de deux parcours utilisateurs : celui d'un producteur de données et celui d'un réutilisateur de données. Pour chacun d'eux, trois niveaux



Retrouver des données
dans des entrepôts

Être FAIR pour être indexé

- Dès lors que des données sont FAIR (et notamment, Faciles à trouver et Accessibles), elles apparaissent dans les résultats de moteurs de recherche généralistes ou spécialisés
 - Suppose que les métadonnées qui les décrivent soient de bonne qualité, et dans un standard adapté à leur découvrabilité par des robots ou par des API
- Tout entrepôt de qualité est censé faciliter la FAIRisation des données qu'il héberge, et notamment leur indexation / référencement
 - Pas besoin de savoir où chercher précisément

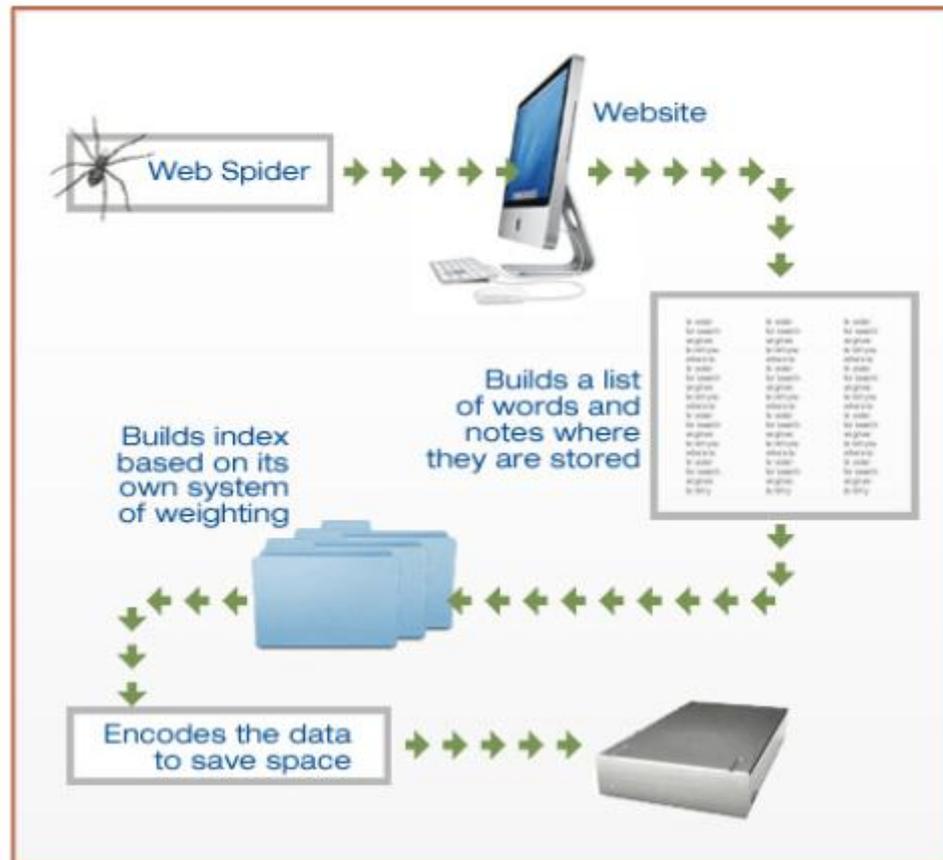
Trouver des données sur internet

- Il est donc (presque) plus important de disposer d'outils pour découvrir / retrouver / accéder aux données que d'outils pour les conserver / partager (même si l'un ne va pas sans l'autre)
- Outils de base :
 - Moteurs de recherche généralistes
 - Moteurs de recherche spécialisés (en données ou en ressources académiques)
 - Méta-moteurs de recherche : outils qui interrogent simultanément plusieurs moteurs de recherche
- Aucun moteur de recherche n'est forcément meilleur que tous les autres dans tous les domaines, nécessaire de croiser les sources

➤ Voir A. Bouchard, « Recherche d'informations sur Internet (perfectionnement) : méthodologie et outils disponibles », 2019 ([en ligne](#)).

Moteurs et métamoteurs de recherche

- Comment fonctionne un moteur de recherche ?



Source : A. Bouchard, « Recherche d'informations sur Internet (perfectionnement) : méthodologie et outils disponibles », 2019 ([en ligne](#)).

! chaque moteur a sa propre couverture du web, ses propres index, algorithmes et fonctionnalités

Choisir son outil pour trouver des données

- Combien et quels types d'entrepôts de données sont moissonnés pour être indexés par l'outil ? Combien de jeux de données sont référencés ?
- Le formulaire de recherche avancée et les facettes de tri des résultats conviennent-ils à mes besoins ? Ex.: interrogation par nom de producteur, tri par organisme d'affiliation, par mots-clés contrôlés...?
- Les fonctions d'affichage et d'export des résultats conviennent-ils à mes besoins ? Ex.: compatibilité avec Zotero, accès direct aux données, lien vers la publication associée quand elle existe...?
- L'outil propose-t-il d'autres fonctionnalités utiles pour moi ? Ex.: favoris, recherches enregistrées...

➤ D'après M.-C. Deboin, « Trouver des jeux de données via des bases pluridisciplinaires et des moteurs de recherche », 2020 ([en ligne](#)).

Dimensions

- Méta-moteur de recherche académique, lancé en 2018 : n'indexe pas que des jeux de données
- Interroge plus de 800 entrepôts de données par l'intermédiaire de 2 outils : DataCite et Figshare.
- Plus de 8,3 millions de jeux de données indexés en mars 2021
- N'indexe que des « jeux de données » (définis comme tels dans leurs entrepôts d'origine) et pas des « données de recherche » (définis comme images, sons, logiciels, etc. dans les entrepôts)

FILTERS FAVORITES

- > PUBLICATION YEAR
- > RESEARCHER
- > RESEARCH CATEGORIES
- > SOURCE TITLE
- REPOSITORY
 - Global Biodiversity Informa 929,793
 - The Cambridge Structural I 893,696
 - Public Library of Science 727,360
 - Plutof. Data Management e 707,410
 - Harvard Dataverse 636,109
 - University of Southern Cali 455,127
 - DSMZ 408,603
 - PANGAEA 381,438
 - WSL Landesforstinventar 250,841
 - UCD Digital Library 232,274
 - American Chemical Society 143,845

Search

PUBLICATIONS	DATASETS	GRANTS	PATENTS	CLINICAL TRIALS
116,971,519	8,326,747	5,799,216	135,472,831	627,020
POLICY DOCUMENTS				
583,548				

Show abstract Sort by: Relevance

Title, Author(s), Year, Repository

FirmFinancial

Kang, Ya, National University Of Singapore, National University Of Singapore

2021 - National University Of Singapore

Supplemental_materials: Spatial co-fragmentation pattern of cell-free DNA recapitulates in vivo chromatin organization and identifies tissues-of-origin

Liu, Yaping

2021 - Zenodo

The intermediate processed files for the manuscript: Spatial co-fragmentation pattern of cell-free DNA recapitulates in vivo chromatin organization and identifies tissues-of-origin

Supplemental_materials: Spatial co-fragmentation pattern of cell-free DNA recapitulates in vivo chromatin organization and identifies tissues-of-origin

Liu, Yaping

2021 - Zenodo

The intermediate processed files for the manuscript: Spatial co-fragmentation pattern of cell-free DNA recapitulates in vivo chromatin organization and identifies tissues-of-origin

Identifying factors determining the distribution of the invasive pest fall armyworm, Spodoptera frugiperda

, Fan

2021 - Zenodo

The raw data of Identifying factors determining the distribution of the invasive pest fall armyworm, Spodoptera frugiperda

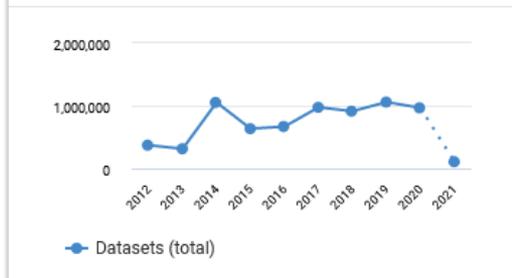
Identifying factors determining the distribution of the invasive pest fall armyworm, Spodoptera frugiperda

ANALYTICAL VIEWS

RESEARCH CATEGORIES

08 Information and Computing Sciences	1,105,459
0806 Information Systems	1,063,988
04 Earth Sciences	910,681
0403 Geology	837,806
06 Biological Sciences	661,470

OVERVIEW



RESEARCHERS

Mark A Scheel	11,435
California Institute of Technology, United States	
Harald Paul Pfeiffer	10,870
Max Planck Institute for Gravitational Physics, Germany	
Lawrence E Kidder	10,850
Cornell University, United States	
Bela Szilagyi	10,642
California Institute of Technology, United States	
Scott E Field	9,971
University of Massachusetts Dartmouth, United States	

DataCite Search

- Les institutions ou chercheurs qui souhaitent obtenir un DOI pour leurs productions scientifiques doivent en échange fournir à DataCite les métadonnées qui décrivent leurs productions
- Ces métadonnées sont compilées par DataCite dans un énorme index, interrogeable à travers une [interface dédiée](#)
- Fonctionnalité particulière : liens de citations qui concernent les jeux de données
- La valeur ajoutée de DataCite réside surtout dans l'exposition procurée aux métadonnées en question (moissonnage OAI-PMH par d'autres outils)

grammar

Search

7,163 Works

Developing Didactic Principles For Annotation Based Learning

Julia Hübner & Marc Schutzeichel

Poster published via Zenodo

Learners often conceive grammar as an irrelevant and even boring subject. As a result they significantly lack the key competence to describe and analyze grammatical structures (Nänny 2014: 183; Habermann 2013:35). When conveying grammatical knowledge, corpus based methods such as manual annotation by the learners still provides a new approach to access grammatical categories. The main purpose of annotation is the assignment of features and attributes to individual segments of an utterance. Acquiring knowledge about...

i No citations were reported. No usage information was reported.

<https://doi.org/10.5281/zenodo.59686> **“** Cite

НАУЧНО - ПРОСВЕТИТЕЛЬСКАЯ ДЕЯТЕЛЬНОСТЬ ОСНОВОПОЛОЖНИКА ИНГУШСКОЙ ПИСЬМЕННОСТИ З.К. МАЛЬСАГОВА

М.Б. Долгиева

Journal Article published via Международный научно-исследовательский журнал

В статье рассматривается научно - просветительская деятельность известного ингушского ученого начала XX века. Значительное внимание в работе уделяется научной, педагогической и творческой деятельности З.К. Мальсагова. Анализируются его прогрессивные взгляды по вопросам распространения образования, науки, ликвидации неграмотности среди ингушей. Освещаются успехи научных экспедиций, в горную Ингушетию в составе которых был и З.К. Мальсагов. В исследовании говорится о его вкладе в процесс становления и дальнейшего развития театрального искусства в Ингушетии.

“ 1 citation **i** No usage information was reported.

<https://doi.org/10.18454/irj.2016.53.175> **“** Cite

ПРИМЕНЕНИЕ МЕТОДА РОЛЕВОЙ ИГРЫ В ПРОЦЕССЕ ОБУЧЕНИЯ ГРАММАТИКЕ АНГЛИЙСКОГО ЯЗЫКА В ВУЗЕ

Registration Year

<input type="checkbox"/> 2021	333
<input type="checkbox"/> 2020	1,590
<input type="checkbox"/> 2019	1,335
<input type="checkbox"/> 2018	1,427
<input type="checkbox"/> 2017	849
<input type="checkbox"/> 2016	516
<input type="checkbox"/> 2015	667
<input type="checkbox"/> 2014	197
<input type="checkbox"/> 2013	104
<input type="checkbox"/> 2012	60

Resource Types

<input type="checkbox"/> Text	4,694
<input type="checkbox"/> Dataset	803
<input type="checkbox"/> Audiovisual	313
<input type="checkbox"/> Other	280
<input type="checkbox"/> Collection	202
<input type="checkbox"/> Software	147
<input type="checkbox"/> Image	116
<input type="checkbox"/> Workflow	49
<input type="checkbox"/> Interactive Resource	38
<input type="checkbox"/> Sound	17
<input type="checkbox"/> Data Paper	7
<input type="checkbox"/> Physical Object	2

Affiliations

<input type="checkbox"/> Ludwig-Maximilians-Universität München	4
---	---

BASE

- Moteur de recherche académique géré par la bibliothèque de l'Université de Bielefeld (Allemagne)
- Sélection des sources à interroger (supposément en libre accès, mais pas toujours, cf. notices bibliographiques sans texte intégral dans les archives ouvertes) se fait manuellement
- Indexe plus de 8,5 millions de jeux de données issues de plus de 8800 sources, [dont la liste est connue](#)
- Intérêt de la recherche multilingue
- Une partie des sites que Base indexe sont alimentés par les chercheurs eux-mêmes, et les dépôts peuvent être référencés de manière incomplète ou imparfaite

Entire Document 


 Verbatim search Additional word forms Multi-lingual search

 Boost open access documents

 Retain my current filters

29 hits in 267,785,217 documents

 1. MarsaGram


Author: MOREAU DE MONTCHEUIL, Grégoire [[claim](#)]

Description: Tool to explore treebanks. Extract a context-free grammar (CFG) and properties over the CFG rules. Generate HTML pages to explore the rules and their properties. Versatile : for constituents or dependencies treebanks. ; Herramienta de exploración...

Publisher, Year: Laboratoire parole et langage - UMR 7309 (LPL, Aix-en-Provence FR), 2015-04-09

Source: source data ; datos de origen ; données source ; 源数据

Document Type: info:eu-repo/semantics/dataset ; Software ; [Software ; Dataset]

Content Provider: Speech and Language Data Repository (SLDR/ORTOLANG)  

[Detail View](#) | [Email this](#) | [Add to Favorites](#) | [Check in Google Scholar](#) | [Export Record](#) 

 2. Éléments de la grammaire Mandchoue

Publisher, Year: Brill, 2014-02-20T15:03:43Z

Source: Western Books on China up to 1850 Online

Document Type: dataset ; [Dataset]

Sort Your Results



Remove Filters

 Document Type: Dataset

Refine Search Result

















Mendeley Data

- Lancé en 2015 par l'éditeur Elsevier
- Aussi bien moteur de recherche qu'entrepôt de données
- Liste complète des sources [consultable en ligne](#)
- Va plus loin que les métadonnées : [interroge le texte intégral](#) des données qu'il entrepose
- Indexe plus de 27,4 millions d'ensembles de données en mars 2021, y compris les données des articles publiés par Elsevier (tables, figures, etc.)
- À la fois producteur et hébergeur : pas neutre ?

grammar 

[Advanced search help](#)

Filter Results 11282 results

DATA TYPES 

SOURCE TYPES 

SOURCES 

- Zenodo (3005)
- ScienceDirect (1815)
- figshare Academic Research System (771)
- arXiv (683)
- Crossref Citations (462)
- KiltHub Repository (208)
- Mendeley Data (193)
- University of Southern California Digital Library (183)
- Rutgers University Community Repository (RUcore) (173)
- University of Vienna (173)
- California Digital Library (161)
- Open Collections (154)
- ANU Data Commons (116)
- SOAS University of London (111)
- Columbia University Libraries (109)

[zenodo](#) (Unlimited ebook) Italian Grammar (Barron's Grammar Series) Full-Access

Contributors: (Unlimited ebook) Italian Grammar (Barron's Grammar Series) Full-Access

Date: 2020-11-02

Source: Zenodo

Read Or Download Italian Grammar (Barron's Grammar Series) Full Books By by Marcel Danesi Ph.D. (Author) Read Online => Read Italian Grammar (Barron's Grammar Series) Download Book => Download Italian Grammar (Barron's Grammar Series) Italian Grammar (Barron's Grammar Series) pdf download Italian Grammar (Barron's Grammar Series) read online Italian Grammar (Barron's Grammar Series) epub Italian Grammar (Barron's Grammar Series) vk Italian Grammar (Barron's Grammar Series) pdf Italian Grammar (Barron's Grammar Series) amazon Italian Grammar (Barron's Grammar Series) free download pdf Italian Grammar (Barron's Grammar Series) pdf free Italian Grammar (Barron's Grammar Series) pdf Italian Grammar (Barro...

Data Types:

Document

Export: [APA](#) [BibTeX](#) [DataCite](#) [RIS](#)

[zenodo](#) Download !PDF Italian Grammar (Barron's Grammar Series) *Full Online

Contributors: Download !PDF Italian Grammar (Barron's Grammar Series) *Full Online

Date: 2020-11-02

Source: Zenodo

Read Or Download Italian Grammar (Barron's Grammar Series) Full Books By by Marcel Danesi Ph.D. (Author) Read Online => Read Italian Grammar (Barron's Grammar Series) Download Book => Download Italian Grammar (Barron's Grammar Series) Italian Grammar (Barron's Grammar Series) pdf download Italian Grammar (Barron's Grammar Series) read online Italian Grammar (Barron's Grammar Series) epub Italian Grammar (Barron's Grammar Series) vk Italian Grammar (Barron's Grammar Series) pdf Italian Grammar (Barron's Grammar Series) amazon Italian Grammar (Barron's Grammar Series) free download pdf Italian Grammar (Barron's Grammar Series) pdf free Italian Grammar (Barron's Grammar Series) pdf Italian Grammar (Barro...

Google Dataset Search

- Lancé en 2018 (version beta) puis 2020, gratuit
- Indexe les pages web dont les données sont décrites selon un standard de métadonnées bien particulier : schema.org
- N'intègre pas les jeux de données dont les données sont disponibles dans d'autres standards
- 31 millions de jeux de données indexés [en août 2020](#)
- Des erreurs dans l'indexation
- Critères d'interrogation et de tri des résultats assez basiques

Dataset Search

Rechercher des ensembles de données



<https://datasetsearch.research.google.com/>

Google

grammaire



Mis à jour

Format de téléchargement

Droits d'usage

Thème

Gratuits

Ensembles de données enregistrés

28 ensembles de données trouvés



Data from: CLE International
amp.fr.google-info.com

Dernière mise à jour : May 16, 2020



Polynesian Segmented Data

zenodo.org
figshare.com
+1plus

tsv

Dernière mise à jour : Nov 29, 2018



Data from: Exercices,
problèmes, situations et tâches
comme lieux de...

explore.openaire.eu

Dernière mise à jour : Jan 1, 2015

Data from: CLE International

Article associé

Découvrir sur amp.fr.google-info.com

Ensemble de données mis à jour May 16, 2020

Licence

[Attribution-ShareAlike 3.0 \(CC BY-SA 3.0\)](#)

Les informations sur la licence ont été dérivées automatiquement

Description

CLE International est une maison d'édition créée en 1973, spécialisée dans le français langue étrangère FLE, c'est-à-dire la langue française apprise par des non-francophones. CLE International est une des maisons d'édition spécialistes de la discipline. Plus de 100 titres nouveaux sont publiés chaque année et viennent enrichir un catalogue de près de 1 000 titres qui sont diffusés dans plus d'une centaine de pays. Ses concurrents principaux, Hachette FLE Hachette Éducation et Didier FLE Éditions Didier, appartiennent tous deux au groupe Hachette, et sont des départements dédiés de maisons d'édition

← Ce n'est pas un jeu de données... ([source](#))



Un résultat que vous attendiez ne s'affiche pas ?

[Découvrez](#) comment ajouter de nouveaux ensembles de données à notre index.

Voir A. Tay, « Three research dataset related updates - Dimensions, Google Dataset search and DataCite search », 2020 ([en ligne](#)).

<https://explore.openaire.eu/>

Research outcomes ▾ Search in OpenAIRE for scholarly works

Open Access

Include: Publications Research data Software Other research products

SEARCH

Advanced Search

1M datasets interlinked with publications



Engaging The Community Toward a Data Discovery Index (DataMed v3.0)

Search for data through bioCADDIE



Search for data set Search for repository

[Advanced Search](#) [help](#)

Looking for COVID-19 datasets?



covid19dataindex.org

To help researchers find COVID-19 datasets, we have developed the COVID-19 Data Index covid19dataindex.org, which collects and indexes all types of COVID-19 datasets from major data repositories, publications, and individual online sources.

Statistics



76 REPOSITORIES



16 DATA TYPES

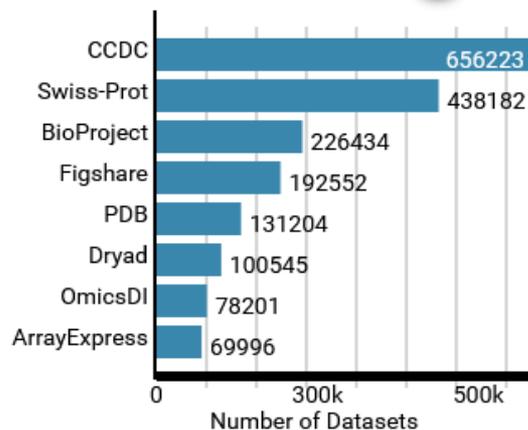


2,336,403 DATASETS



4 PILOT PROJECTS

Top 8 Repositories



Voir X. Chen et al., « DataMed – an open source discovery index for finding biomedical datasets », *Journal of the American Medical Informatics Association*, 2018.

1.2

Milestones and Millstones

Uses and Reuses of Scientific Data: The Data Creators' Advantage

by Irene V. Pasquetto, Christine L. Borgman, and Morgan F. Wofford

now on branch
#public



Contributors (3) 

Published Nov 15, 2019

DOI 10.1162/99608f92.fc14bf2d

[Show All Details](#)

- Les chercheurs réutilisent fréquemment des données (publiques ou d'autres chercheurs)
- Ils s'en servent surtout pour les comparer aux leurs, mais rarement pour les analyser à nouveau
- L'ajout de métadonnées et de documentation favorise grandement la compréhension des données
- La connaissance que les producteurs de données ont de leurs données peut servir de monnaie d'échange pour de futures collaborations

Valoriser et réutiliser des données : l'importance de la citation

- La citation correcte des données est une bonne pratique indispensable en recherche
 - Dans le corps du texte, en note et/ou en bibliographie
 - Dans le même style bibliographique que la littérature
 - En citant également toute publication associée aux données
- Ne pas hésiter à fournir soi-même un modèle de citation, et réutiliser ceux fournis par les autres
 - Permet d'identifier les données (titre, date, version...)
 - Permet de donner crédit aux producteurs (noms)
 - Permet d'accéder aux données (identifiant pérenne)

Versions

Version 1.0 May 20, 2020
10.5281/zenodo.3835956

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.3835955](https://doi.org/10.5281/zenodo.3835955). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Share

Cite as

Anastasia A. Minervina, Ekaterina A. Komech, Aleksei Titov, Meriem Bensouda Koraichi, Elisa Rosati, Ilgar Z. Mamedov, ... Mikhail V. Pogorelyy. (2020). Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T cell memory formation after mild COVID-19 infection (Version 1.0) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3835956>

Start typing a citation style...

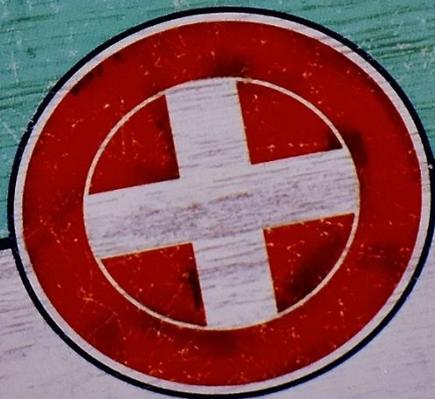
Export

[BibTeX](#) [CSL](#) [DataCite](#) [Dublin Core](#) [DCAT](#)
[JSON](#) [JSON-LD](#) [GeoJSON](#) [MARCXML](#)
[Mendeley](#)

[Exemple](#) d'outils de citation de données dans Zenodo

Medicines & Bandages

FIRST AID CASE

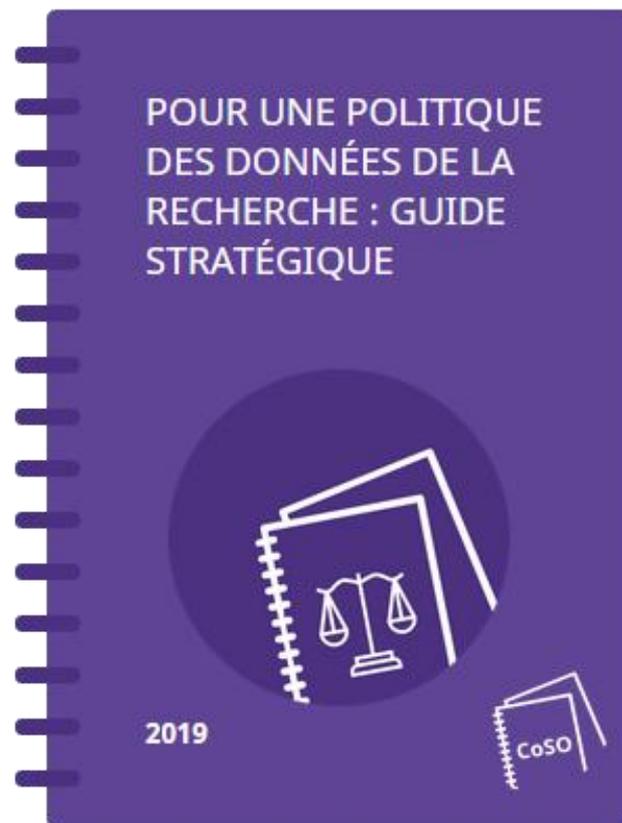


Ressources

Mettre en place une politique des données de la recherche à l'échelle d'un établissement

Recommandations du collège Données de la recherche du Comité pour la Science Ouverte :

1. Faire adopter un texte-cadre sur la science ouverte au niveau de l'établissement
2. Nommer un chargé de mission science ouverte auprès de la gouvernance
3. **Encourager l'écriture d'article de données** (*Data papers*)
4. Développer une culture de la science ouverte par la sensibilisation et la formation des différents acteurs de la recherche
5. Mise en œuvre d'un guichet unique
6. Création d'un **poste d'administrateur des données** de l'établissement
7. Inciter l'écriture de **plan de gestion de données** dans tout projet de recherche



Et en pratique :

Autres recommandations du collège Données du CoSO, pour faire avancer la science ouverte dans les laboratoires et permettre très concrètement l'ouverture pérenne de plus de données :

- Définir ce qui doit être conservé
- Décliner la science ouverte par discipline
- Donner aux chercheurs.es la possibilité de se former en matière juridique
- Donner aux chercheurs.es la possibilité de se former en matière de curation des données de la recherche
- Organiser les soutiens humains nécessaire à l'ouverture des données
- Structurer les infrastructures pour répondre aux besoins de la science ouverte



Collège Données de la recherche du Comité pour la science ouverte,
« Usage et gouvernance des données », *Ouvrir la Science*, 2019 ([en ligne](#)).

Les personnes ressources

- Aspects juridiques : DPO (ex-CIL)
- Aspects éthiques : Comités de protection des personnes (recherche biomédicale notamment)
- Protection du potentiel scientifique et technique et de l'innovation : fonctionnaire de sécurité de défense et SATT
- Aspects informatiques : service SSI
- Mais aussi :
 - Archivistes (réseau [AURORE](#) à l'origine du [Référentiel de gestion des archives](#) de la recherche)
 - Professionnels IST

S'auto-former

S'initier en ligne aux données de la recherche et à leur gestion

- 1 - Familiarisez-vous avec le concept de « données de la recherche »
- 2 - Pourquoi vous intéresser à la gestion des données de la recherche ?
- 3 - Le concept de données de la recherche : sites de référence en français
- 4 - Découvrez des tutoriels en ligne gratuits sur les données de la recherche

M.-C. Deboin, « S'initier en ligne aux données de la recherche et à leur gestion, en 4 points », 2020 ([en ligne](#)).

The screenshot shows the Coursera interface for the course 'Research Data Management and Sharing'. The course is offered by the University of North Carolina at Chapel Hill and the University of Edinburgh. It has a rating of 4.6 stars based on 221 notes and 53 reviews. The page includes buttons for 'Aller au cours' and 'Déjà inscrite(e) Aide financière disponible', and indicates that 12,065 people are already enrolled. Navigation links for 'À propos', 'Programme de cours', 'Avis', 'Enseignants', 'Options d'inscription', and 'FAQ' are visible at the bottom.

À propos de ce cours

10 065 consultations récentes

This course will provide learners with an introduction to research data management and sharing. After completing this course, learners will understand the diversity of data and their management needs across the research data lifecycle, be able to identify the components of good data management plans, and be familiar with best practices for working with data including the

100% en ligne
Commencez dès maintenant et apprenez aux horaires qui vous conviennent.

[MOOC Research Data Management and Sharing](#)

The screenshot shows the MANTRA website, which is a free online course for those who manage digital data as part of their research project. The website features a navigation menu with links for Home, About, Acknowledgements, DIY Training Kit for Librarians, Feedback, and Contact Us. Below the menu, there are four icons representing different user roles: Research Student, Career Development, Digital Academic, and Information Professional. The main content area displays 'Learning Units: Select one to start' with a grid of nine units, each with an icon and a title: 'Research data explained', 'File formats & transformation', 'Data protection, rights & access', 'Data management plans', 'Documentation, metadata, citation', 'Sharing, preservation & licensing', 'Organising data', 'Storage & security', and 'Data handling tutorials'.

The screenshot shows the Research Data Management Librarian Academy (RDMLA) website. The main heading is 'RESEARCH DATA MANAGEMENT LIBRARIAN ACADEMY [RDMLA]' with the start date 'Started Oct 7, 2019'. Below the heading, there is a large 'ENROLL' button. The website features a grid of six information boxes: 'COURSE DATE: Open', 'DURATION: Ongoing', 'COMMITMENT: 1.5-2 hrs/Unit', 'REQUIREMENT: None', 'COURSE TYPE: Self-paced', and 'CREDENTIAL: Certificate; Continuing Education Credits (CEUs)'. The background image shows a person looking at a large digital display with various data visualizations.

[Research Data Management Librarian Academy \(RDMLA\)](#)

DONNÉES DE LA RECHERCHE APPRENTISSAGE NUMÉRIQUE

THÉMATIQUES ABORDÉES

PARCOUREZ LES RESSOURCES CLASSÉES PAR THÉMATIQUES

www.doranum.fr

ENJEUX & BÉNÉFICES

Pourquoi partager les données ?
Qu'est-ce que l'Open Science ?



ASPECTS JURIDIQUES, ÉTHIQUES, INTÉGRITÉ SCIENTIFIQUE

Que puis-je partager, réutiliser ?
Quelles pratiques devrais-je
respecter ?



PLAN DE GESTION DE DONNÉES

Pourquoi et comment rédiger un
plan de gestion des données ?



MÉTADONNÉES

Comment décrire les données ?



IDENTIFIANTS PÉRENNES

Comment associer durablement
des données à son auteur ?



DÉPÔT & ENTREPÔTS

Comment et où déposer mes
données ?



STOCKAGE & ARCHIVAGE

Quelles données conserver à long
terme et comment ?



DATA PAPERS & DATA JOURNALS

Comment publier mes données
comme un article scientifique ?



ACCÈS & VISUALISATION

Où et comment extraire et
visualiser les données qui
m'intéressent ?



**S'AUTOFORMER
AUX DONNÉES DE
LA RECHERCHE :
GUIDE À
DESTINATION DES
PROFESSIONNELS
DE
L'INFORMATION
ET DE LA
DOCUMENTATION**

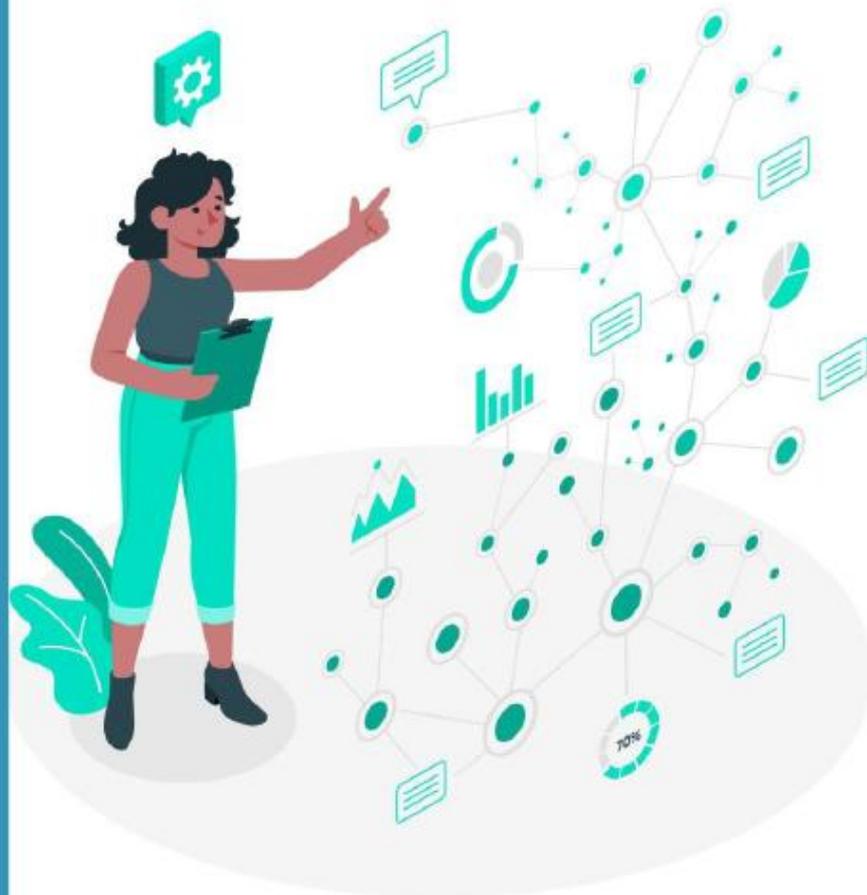
2020



[Une production](#) du gTIGRE (groupe de travail inter-établissements Grand Est)

**FORMER LES
PROFESSIONNELS DE
L'INFORMATION ET
DE LA
DOCUMENTATION
AUX DONNÉES DE LA
RECHERCHE EN
45 MINUTES**

2020



L. Bracco et al., « Former les professionnels de l'information et de la documentation aux données de la recherche en 45 minutes », 2020 ([en ligne](#)).

The Data Management Skillbuilding Hub contains resources for better data management and is open to community input and update. These resources are adaptable across a range of contexts and intended for use by researchers, teachers, librarians, or anyone who wants to learn better data management practices. Each tile below links to community contributed education materials, such as best practices and lesson plans

The resources presented on the Data Management Skillbuilding Hub can be updated by users to promote a current, well-maintained, and sustainable educational tool. Learn more about how you can [contribute](#).

Using This Resource

Click individual tiles to learn more and use each resource. You can limit resources by content type and [Data Life Cycle](#) stage. Comprehensive information is available in the [FAQ](#).

» Filter by content type:

ALL

TEACHING MODULE

BEST PRACTICE

VIDEO

» Filter by stage of the Data Life Cycle

All



<https://dataoneorg.github.io/Education/>

Voir aussi: <https://www.dataone.org/for-librarians>

Mener une « interview-données »

1. Quelle est l'histoire de ces données ? Leur contexte de production / collecte ? Le projet de recherche auquel elles sont liées ?
2. De quel(s) type(s) et dans quel(s) format(s) sont-elles ?
3. Quelle est la durée de vie supposée du jeu de données ?
4. Comment pourraient-elles être (ré)utilisées, et pourquoi ?
5. Quelle est la taille du jeu de données ? Quel accroissement ?
6. Quelle audience potentielle pour ce jeu de données ?
7. À qui appartiennent les données ?
8. Y a-t-il des données sensibles ?
9. Quelles publications ou découvertes découlent de ces données ?
10. Comment rendre ces données accessibles ?

Source : M. Witt et J. Carlson, « Conducting a Data Interview », 2007 ([en ligne](#)).

GOpenDoRe, pour jouer avec les données de la recherche



Merci pour votre attention

T₁

H₄

A₁

N₁

K₅

S₁

Justine Ancelin-Fabre

justine.ancelin-fabre@chartes.psl.eu

<https://urfist.chartes.psl.eu>