

MERIC AKDOGAN
Chargé de projets Données de la recherche

27 Février 2024

L'OUVERTURE DES DONNÉES DE LA RECHERCHE EN STM



L'ouverture des données de la recherche



1

- Introduction

2

- Et vous ? Bonnes pratiques de gestion

3

- Comment diffuser ses données ?

1

**DONNÉES DE LA RECHERCHE
UNE INTRODUCTION ↴**

C'est quoi la science ouverte ?

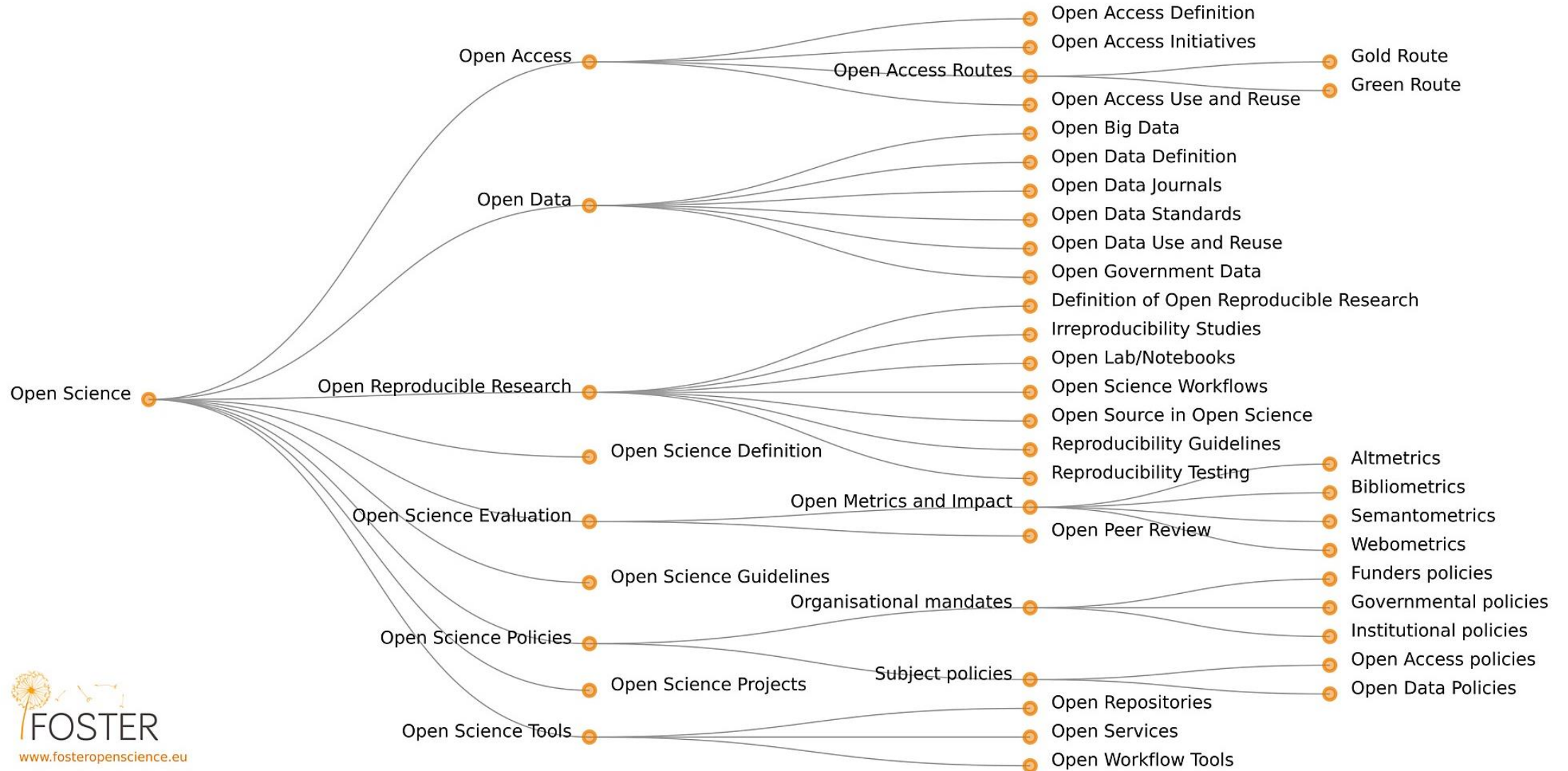
...construire un **écosystème** dans lequel la **science** est **plus cumulative**, plus fortement **étayée par des données**, plus **transparente**, plus rapide et **d'accès plus universel**

Réf : <https://www.ouvrirlascience.fr/plan-national-pour-la-science-ouverte/>

... **partager les connaissances et productions scientifiques**, non seulement avec l'ensemble de la communauté scientifique, mais aussi **avec tous les acteurs de la société**.

Réf : <https://www6.inrae.fr/dipso/Zoom-sur/Pour-une-definition-de-la-science-ouverte>

Taxonomie de la Science Ouverte



La science ouverte : entre initiatives des pionniers, dark open access et politiques institutionnelles



1991 : lancement d'ArXiv, développé par Paul Ginsparg

1999 : lancement de E-biomed, aujourd'hui PubMed Central, sous l'impulsion de Harold Varmus

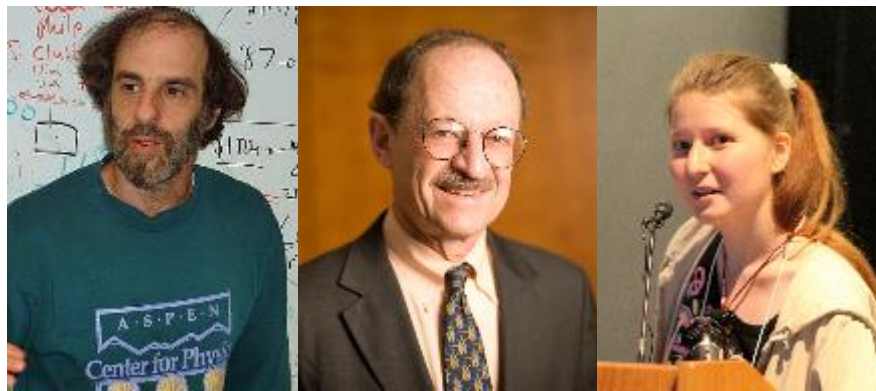
2001 : ouverture de HAL

2002 : Budapest Open Access Initiative

2011 : lancement de Sci-Hub...

2012 : San Francisco Declaration on Research Assessment (**DORA**)

2016 : Amsterdam Call for Action on Open Science



Un mouvement en faveur de l'ouverture des données

1972 : Création du centre de Données astronomiques de Strasbourg

1992 : Genbank

2006 : échelle 5 étoiles de Tim Berners-Lee

2007 : travaux sur l'Open government data à l'initiative de Lawrence Lessig, Tim O'Reilly, Ethan Zuckermann, Joseph Hall, Aaron Schwartz et Carl Mamamud

Principes : accessibilité, données non-propriétaires, licences ouvertes , etc.



Silvio Tanaka, « Tim Berners-Lee, CC BY 2.0, https://commons.wikimedia.org/wiki/File:Tim_Berners-Lee_CP.jpg
Lessig 2016, CC BY 2.0, https://commons.wikimedia.org/wiki/File:Lawrence_Lessig_Headshot.jpg
Christopher Michel, "Tim O'Reilly", CC BY 2.0, [https://commons.wikimedia.org/wiki/File:Tim_O%27Reilly_-_2017_\(38700700672\)_cropped.jpg](https://commons.wikimedia.org/wiki/File:Tim_O%27Reilly_-_2017_(38700700672)_cropped.jpg)
Sage Ross, CC BY-SA 2.0, https://commons.wikimedia.org/wiki/File:Aaron_Swartz_2_at_Boston_Wikipedia_Meetup,_2009-08-18.jpg
David Orban, CC:BY, Open Government Working Group, <https://www.flickr.com/photos/davidorban/2095565103/in/album-72157603410393877/>

Un mouvement en faveur de l'ouverture des données

2007 : travaux sur l'[Open government data](#) à l'initiative de Lawrence Lessig, Tim O'Reilly, Ethan Zuckermann, Joseph Hall, Aaron Schwartz et Carl Mamamud

Principes :

1. [Complete](#)
All public data is made available. Public data is data that is not subject to valid privacy, security or privilege limitations.
2. [Primary](#)
Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms.
3. [Timely](#)
Data is made available as quickly as necessary to preserve the value of the data.
4. [Accessible](#)
Data is available to the widest range of users for the widest range of purposes.
5. [Machine processable](#)
Data is reasonably structured to allow automated processing.
6. [Non-discriminatory](#)
Data is available to anyone, with no requirement of registration.
7. [Non-proprietary](#)
Data is available in a format over which no entity has exclusive control.
8. [License-free](#)
Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restrictions may be allowed.

Le mouvement de l'Open data

Répondre à une attente citoyenne

- Un mouvement concernant la société dans son ensemble
- [Open government partnership](#), 2011
 - 78 pays »
 - « OGP's vision is that more governments become sustainably **more transparent, more accountable, and more responsive to their own citizens**, with the ultimate goal of improving the quality of governance, as well as the quality of services that citizens receive. »

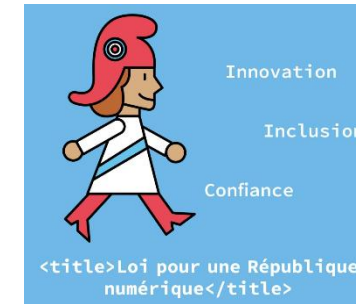


Le mouvement de l'Open data : En France



- 1978 - Loi n°78-753 du 17 juillet 1978, dite « loi CADA »
- Création de la mission interministérielle Etalab avec l'ouverture du portail de data.gouv.fr en 2011
- Loi pour une République numérique en 2016
- Pour une action publique transparente et collaborative : plan d'action national pour la France, 2015-2017, 2018-2020, puis 2021-2023...

Loi pour une République numérique (2016)



Open research data : article 30

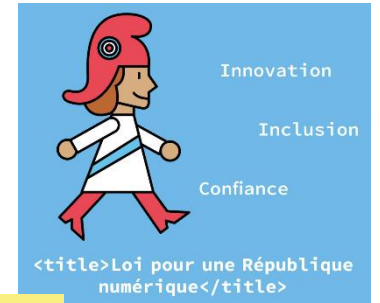
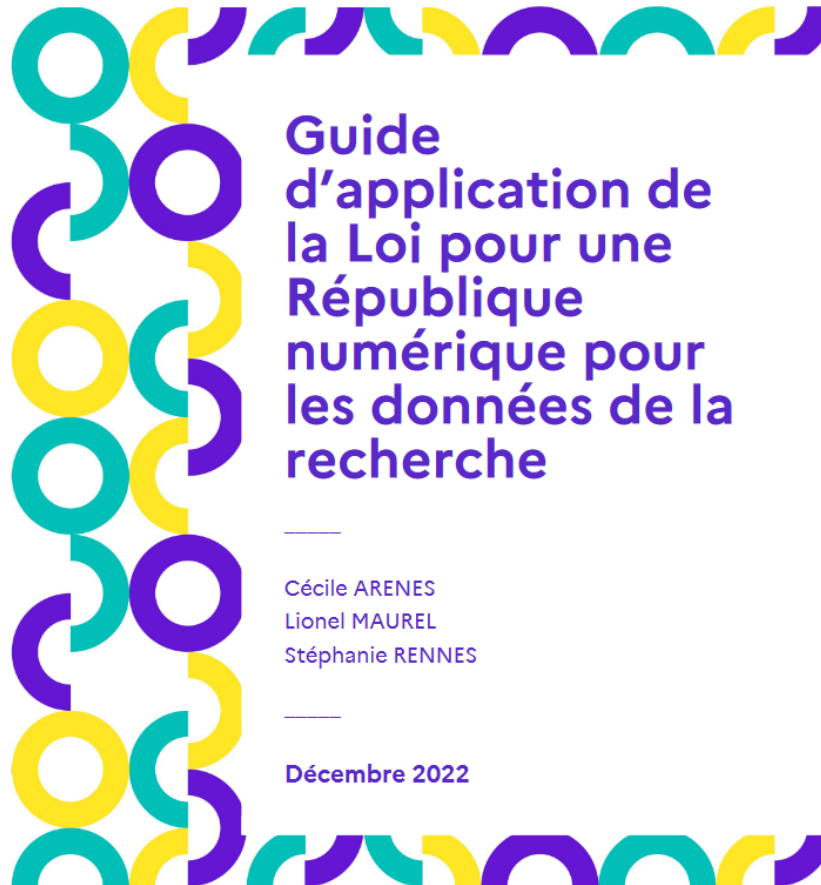
« II.- Dès lors que les **données** issues d'une activité de recherche **financée** au moins pour **moitié** par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne **ne sont pas 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**.

« III.- L'**éditeur** d'un écrit scientifique mentionné au I **ne peut limiter la réutilisation** des données de la recherche rendues publiques dans le cadre de sa publication.

Image : https://www.economie.gouv.fr/files/files/PDF/DP_LoiNumerique.pdf

Loi pour une République numérique (2016)

Ou
Ouvrir
la science !



Sommaire

Synthèse

Cadre général

Quel contexte de production des données ?

Exceptions à l'ouverture et à la libre réutilisation des données

Quand ouvrir les données ?

Rôle du chercheur et des établissements

Notion de données achevées

Qui peut réutiliser les données et pour quels usages ?

Licences

Pratiques des éditeurs et Loi pour une République numérique

La science ouverte en France

[Plan national pour la science ouverte](#), 2018.

« La France s’engage pour que les résultats de la recherche scientifique soient ouverts à tous, chercheurs, entreprises et citoyens, sans entrave, sans délai, sans paiement. »

Axe 2, structurer et ouvrir les données de la recherche :

- **Rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics.**
- Créer la fonction **d’administrateur des données** et le réseau associé au sein des établissements.
- Créer les conditions et promouvoir l’adoption **d’une politique de données ouvertes associées aux articles publiés** par les chercheurs.



Le plan national pour la science ouverte – 2

[Plan national pour la science ouverte 2](#), 2021-2024.

Axe 2 : structurer, partager et ouvrir les données de la recherche :

- Mettre en œuvre l'**obligation de diffusion des données de recherche financées sur fonds publics**
- Créer **Recherche Data Gouv**, la plateforme nationale fédérée des données de recherche
- Promouvoir l'adoption d'une **politique de données** sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, intéropérables et réutilisables (**FAIR**)

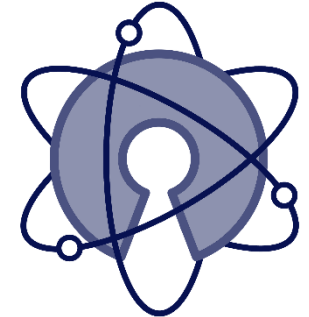
Axe 3 : ouvrir et promouvoir les codes sources produits par la recherche

- **Valoriser et soutenir la diffusion sous licence libre des codes sources** issus de recherches financées sur fonds publics
- **Mettre en valeur la production des codes sources** de l'enseignement supérieur, de la recherche et de l'innovation
- Définir et promouvoir une **politique en matière de logiciels libres**



La science ouverte

A Sorbonne Université



- Page web dédié : [L'engagement pour la science ouverte](#)

Concernant les données :

- Une [Charte de Sorbonne Université pour le libre accès aux publications](#)
- Une Déclaration sur le [droit des données de la recherche](#), signée en janvier 2020
- Une [Politique d'ouverture des données de la recherche](#) au sein de l'Alliance Sorbonne Université
 - inciter leurs équipes et personnels de recherche à ouvrir les données issues de la recherche autant que possible ;
 - promouvoir l'élaboration de plans de gestion des données dès la conception des protocoles de recherche ;
 - valoriser l'engagement des personnels de recherche qui partagent leurs données et œuvrent pour la science ouverte ;
 - mettre en place des programmes de formation et mobiliser les moyens nécessaires afin de créer un environnement favorisant l'ouverture des données.

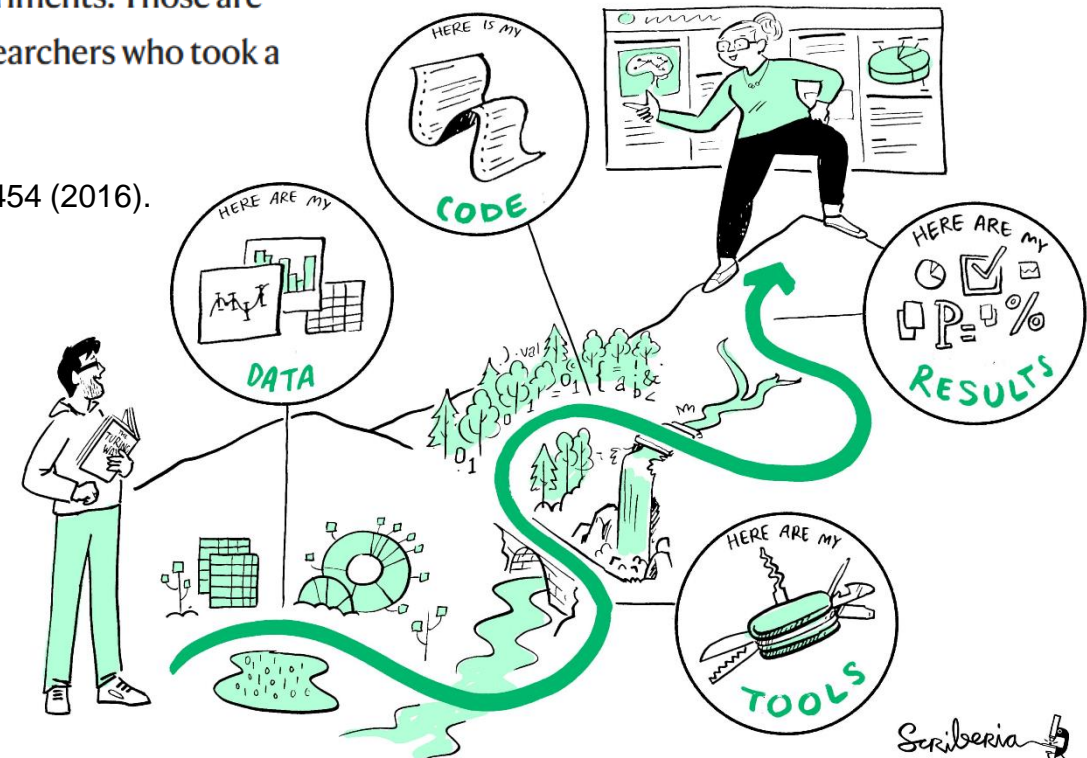
La recherche reproductible ?

More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature's* survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016).
<https://doi.org/10.1038/533452a>

La recherche reproductible

- garantit que le processus scientifique reste ouvert et vérifiable.
- implique de documenter et d'organiser la recherche de manière à ce que d'autres puissent reproduire indépendamment les résultats, favorisant ainsi la transparence et la fiabilité.



The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807).

Pourquoi gérer et ouvrir les données ?



- **50 % des expériences** sont considérées comme **non-reproductibles**.
- **80 % des données** produites ces 20 dernières années seraient **perdues**.
- **Plus de 80 % des données produites** sont **stockées ailleurs** que dans des entrepôts.

DATAACC. « Gestion des données : une nouvelle exigence, de nouvelles compétences », 2020. <https://www.dataacc.org/bonnes-pratiques/adopter-un-plan-de-gestion-des-donnees/gestion-des-donnees-une-nouvelle-exigence-de-nouvelles-competences/>.

D'où vient la nécessité d'une recherche reproductible ?

1. Difficulté de Reproduction Complète
2. Exigence d'une Norme Minimale
3. Vérification des Résultats Publiés
4. Complexité Croissante des Analyses
5. Avancées Technologiques
6. Préoccupations Concernant la Recherche Frauduleuse



<https://the-turing-way.netlify.app/reproducible-research/reproducible-research>

Partager que des données ?

- “Research object” is a term that is gaining in popularity, not least in the humanities where the relevance of the term ‘data’ is not always recognised
- Research objects can comprise any supporting material which underpins or otherwise enriches the (written) outputs of research
 - Data (numeric, written, audiovisual....)
 - Software code
 - Workflows and methodologies
 - Slides, logs, lab books, sketchbooks, notebooks, you name it!
- See <http://www.researchobject.org/> for more info

Ouvrir les données : des risques ?



Plagiat ?

Mettre sa thèse en ligne garantit un horodatage

Risques liés à la vie privée ?

Respecter le RGPD

Sécurité des données ?

Données concernées par la notion de secret ?

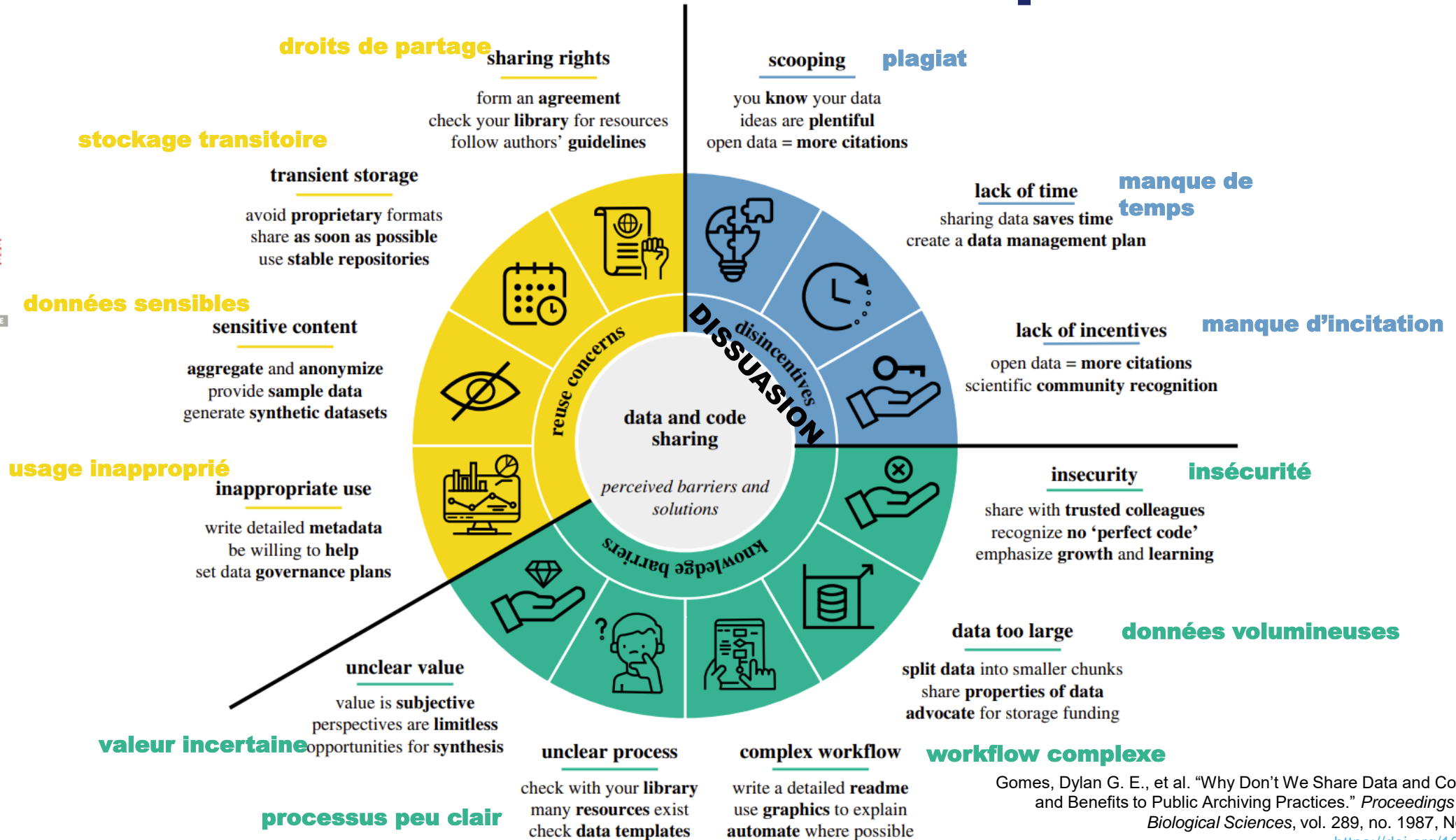
Utilisation malveillante des données ?

Le plus compliqué à anticiper

Un principe :

« **Ouvrir** autant que possible, **fermer** autant que nécessaire »

Ouvrir les données : des risques ?



Gomes, Dylan G. E., et al. "Why Don't We Share Data and Code? Perceived Barriers and Benefits to Public Archiving Practices." *Proceedings of the Royal Society B: Biological Sciences*, vol. 289, no. 1987, Nov. 2022, p. 20221113, <https://doi.org/10.1098/rspb.2022.1113>.

Un exemple de réutilisation des données ouvertes

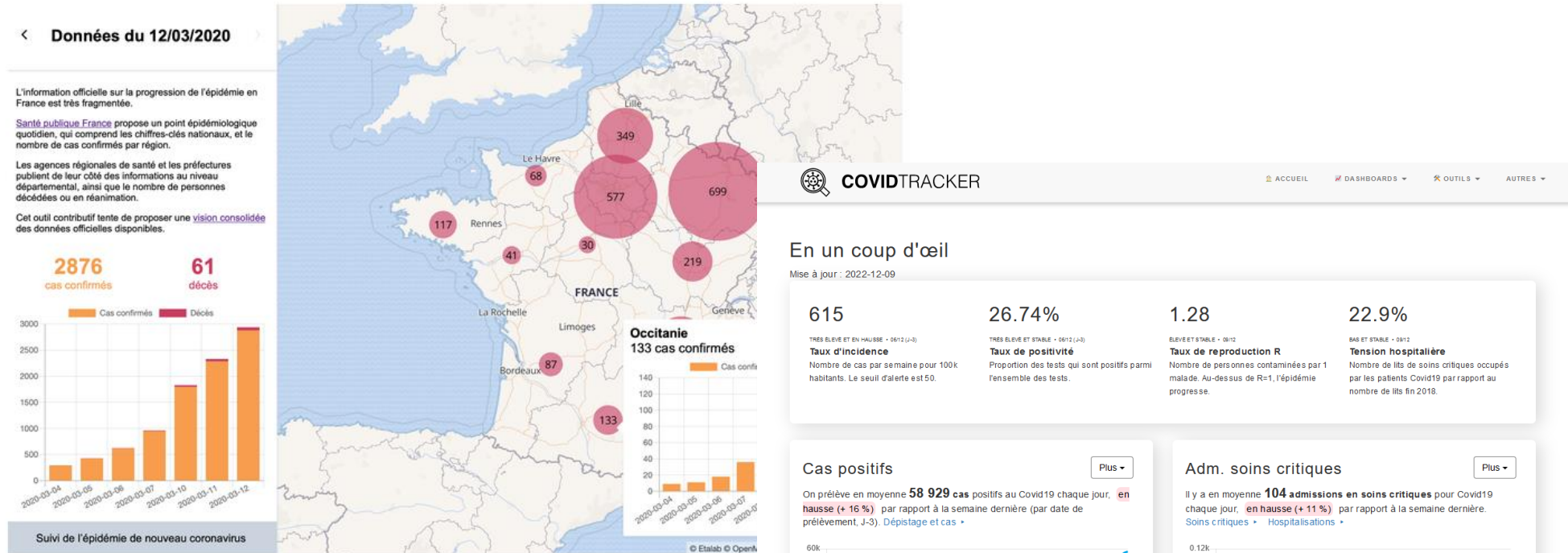
Tableau de bord de suivi de l'épidémie de COVID19

Publié le 12 mars 2020



BIBLIOTHÈQUE
UNIVERSITAIRE

De l'ouverture
des données
publiques à des
initiatives
citoyennes



Source : <https://www.data.gouv.fr/fr/reuses/tableau-de-bord-de-suivi-de-lepidemie-de-covid19/>

Source : <https://covidtracker.fr/>

Plus d'exemples

Jeux de données

Rechercher parmi les 47 058 jeux de données sur data.gouv.fr

[Rechercher dans les réutilisations](#)

Ex : élections présidentielles 2022

Recherche

Filtres

Organisations

Toutes les organ...

Mots clés

Tous les mots-clés

Formats

Tous les formats

Licences ⓘ

Toutes les licenc...

Schéma ⓘ

Tous les schémas

Couverture Spatiale ⓘ

Toutes les couve...

Granularité spatiale ⓘ

Toutes les granul...

10000 résultats

Trier par : Pertinence



Base Sirene des entreprises et de leurs établissements (SIREN, SIRET)

Par [Institut National de la Statistique et des Etudes Economiques \(Insee\)](#)

Pour vous abonner à notre lettre d'information Sirene open data actualités, cliquez ici Pour consulter nos lettres d'information Sirene open data actualités, ...

48 réutilisations

113 favoris

Qualité des métadonnées: — Mis à jour il y a 26 jours



Demandes de valeurs foncières DVF

Par [Ministère de l'Économie, des Finances et de la Souveraineté industrielle et numérique](#)

Propos liminaires Conformément au décret n° 2018-1350 du 28 décembre 2018 relatif à la publication sous forme électronique des informations portant sur les val...

81 réutilisations

140 favoris

Qualité des métadonnées: — Mis à jour le 10 octobre 2023



Répertoire National des Associations RNA

Par [Ministère de l'Intérieur et des Outre-Mer](#)

⚠ Des emails frauduleux usurpant l'identité de data.gouv.fr peuvent circuler à destination des associations en demandant des données personnelles. N'y répondez...

10 réutilisations

70 favoris

Qualité des métadonnées: — Mis à jour le 1 janvier 2024

Plus d'exemples

Réutilisations

Rechercher parmi les 3 725 réutilisations sur data.gouv.fr

[Rechercher dans les jeux de données](#)

Rechercher...

- Tout
- Santé
- Transports et mobilité
- Aménagement du territoire et logement
- Alimentation et agriculture
- Culture et loisirs
- Économie et entreprise
- Environnement et énergie
- Emploi et formation
- Politique et vie publique
- Sécurité
- Éducation et recherche
- Société et démographie
- Droit et justice
- Outils open data
- Autres

3725 résultats

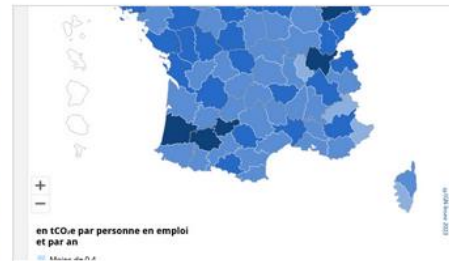
Trier par : **Les plus récents**



Article de blog

Découvrez la liste de tous les lieux où vont se...

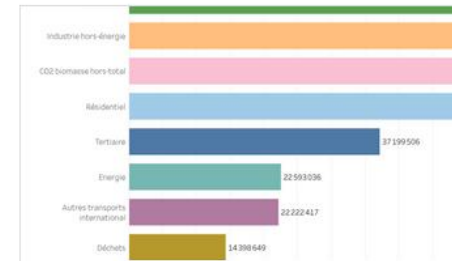
Publié le 25 février 2024 par [Actu-Rando](#)



Papier

Déplacements domicile-travail : des émissions de gaz...

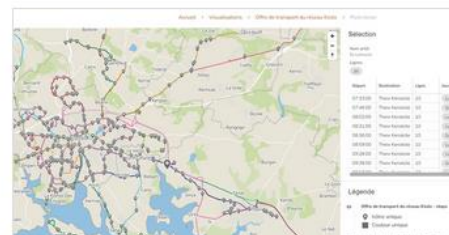
Publié le 24 février 2024 par [Institut National de la Statistique et des Etudes Economiques \(Insee\)](#)



Article de blog

Le secteur tertiaire pollue 1.7x plus que les...

Publié le 22 février 2024 par [Jady Nekena R](#)



Entreprises vendues à découvrir 20/02/2024

France | Classement quotidien top 10, pourcentage de positions courtes

Entreprise	Pourcentage	Évolution
→ ATOS SE	15,06 %	+0,00
→ VALLOUREC	11,65 %	+0,00
→ CLARIANE SE	8,31 %	+0,00
→ ALSTOM	5,63 %	+0,00
→ CGG	5,59 %	+0,00
→ EUTELSAT COMMUNICATIONS	5,56 %	+0,00
→ VALEO	5,53 %	+0,00
→ EUROAPI	4,90 %	+0,00
→ EUROFINS SCIENTIFIC SE	4,27 %	+0,00
→ CASINO GUILLET-EPICERIE	4,25 %	0,00

2

ET VOUS ?

BONNES PRATIQUES DE GESTION ↴

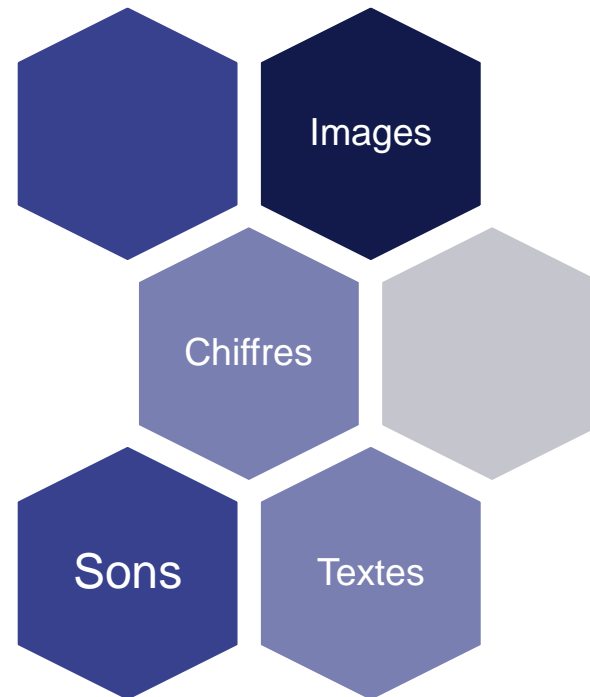
Données de la recherche

Sources principales

**Nécessaires pour
valider les résultats**

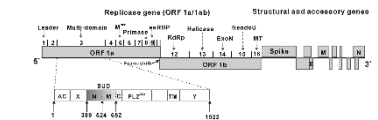
Source: [OCDE, 2007](#)

Enregistrements factuels



Quelques exemples

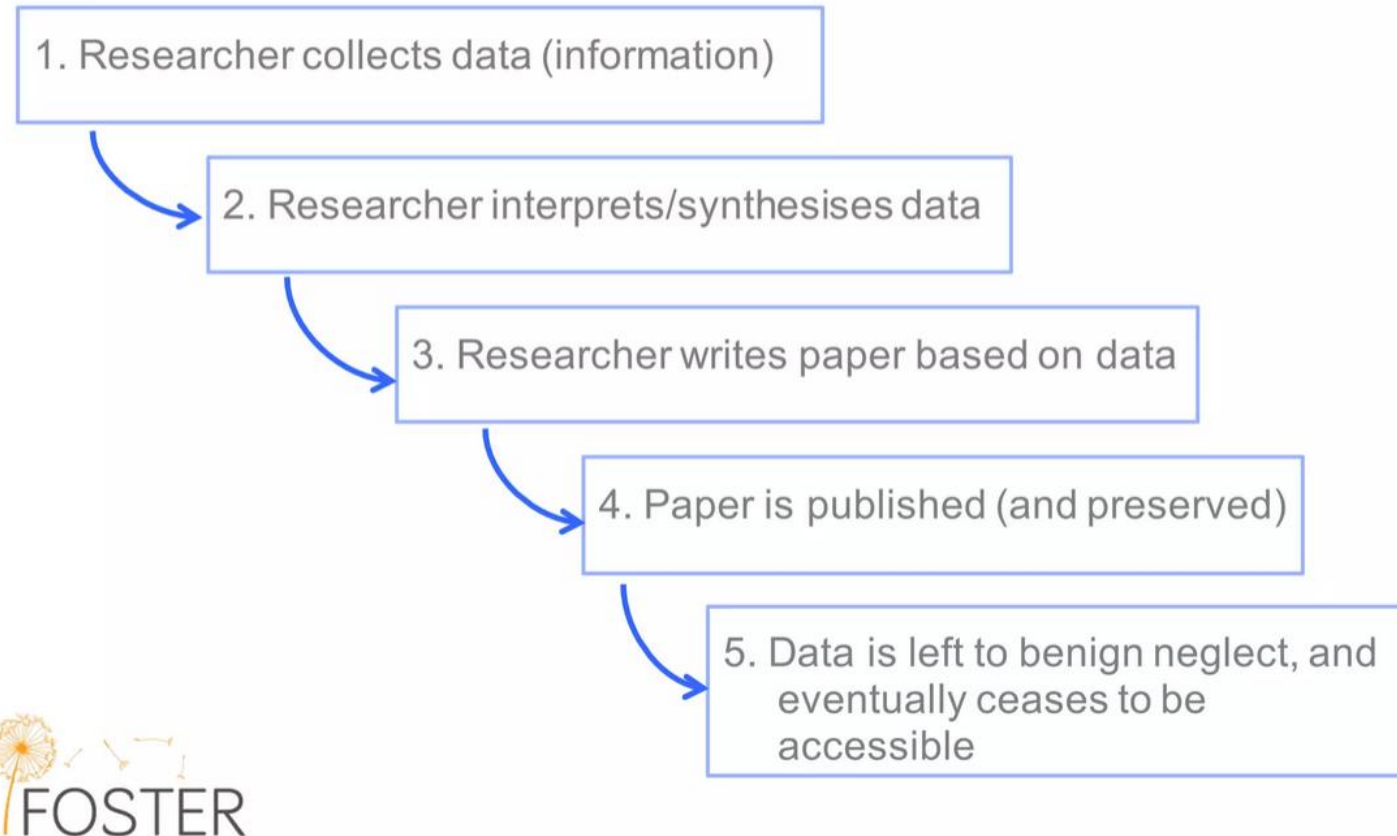
Types de données	Caractéristiques	Exemples
Données d' observation	Collectées en temps réel, impossibles à reproduire. Conservation pérenne	Cliché astronomique, neuro-imagerie, relevés de précipitations
Données expérimentales	Obtenues en laboratoires, reproductibles à des coûts élevés	Séquence de génome, chromatographie, spectres RMN
Données computationnelles , de modèles ou de simulation	Données générées à partir de modèles test. Reproductibles, nécessité de décrire le modèle, le logiciel associé et d'en archiver toutes les informations qui permettront sa ré-exploitation	Modèles météorologiques, modèle de simulation sismique
Données dérivées ou compilées	Données brutes qui ont fait l'objet de traitements et d'analyses successifs, conservation recommandée	Fouille de textes ou de données d'une base, modèles 3D
Données de référence ou données canoniques	Collections statiques ou organiques de jeux de données validées	Banques de données sur le génome, structure chimique, portail de données spatiales



Year	Author	Title	Year	Author	Title
1998	2000
1999	2001
2000	2002
2001	2003
2002	2004
2003	2005
2004	2006
2005	2007
2006	2008
2007	2009
2008	2010
2009	2011
2010	2012
2011	2013
2012	2014
2013	2015
2014	2016
2015	2017
2016	2018
2017	2019
2018	2020
2019	2021
2020	2022



Cycle de vie des données



Cycle de vie des données

Décrire / Documenter

Collecter / Créer

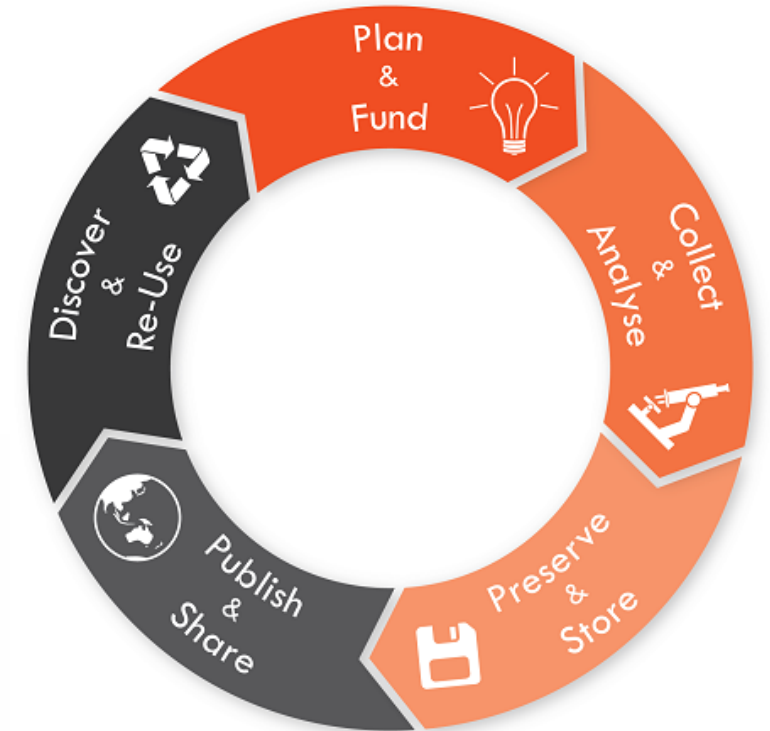
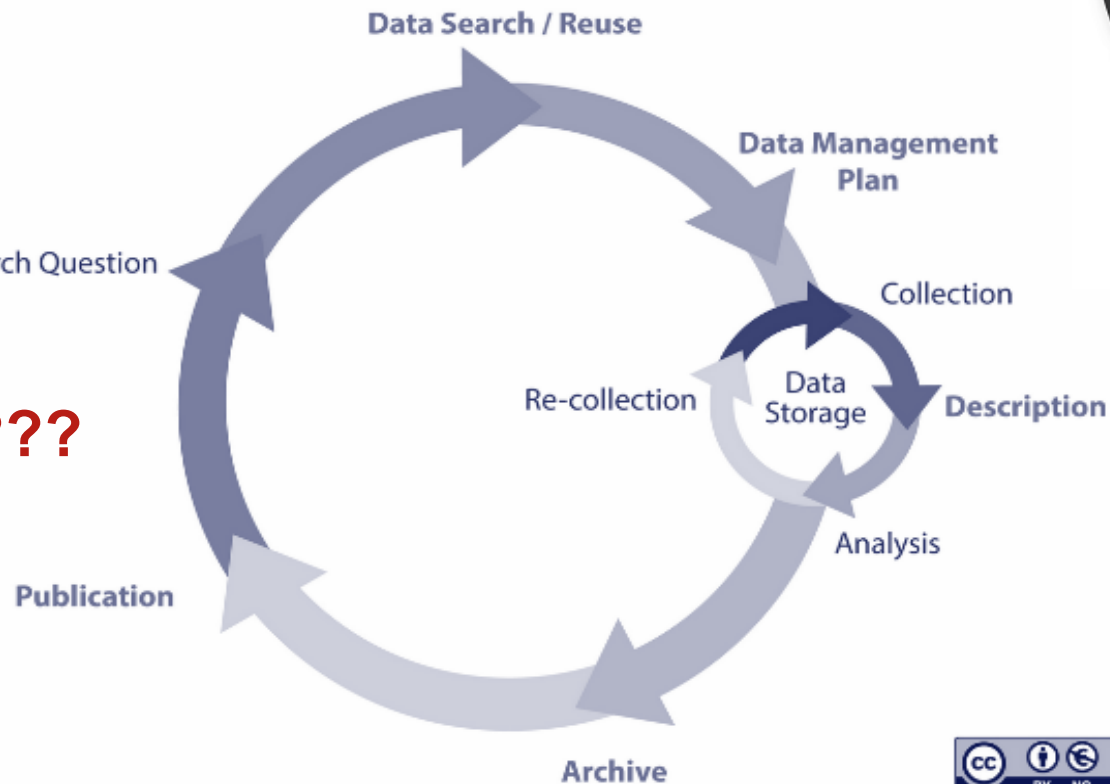
Publier

Archiver

Analyser

Stocker Research Question

Réutilisation ???



Plan de gestion des données

- 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 ?

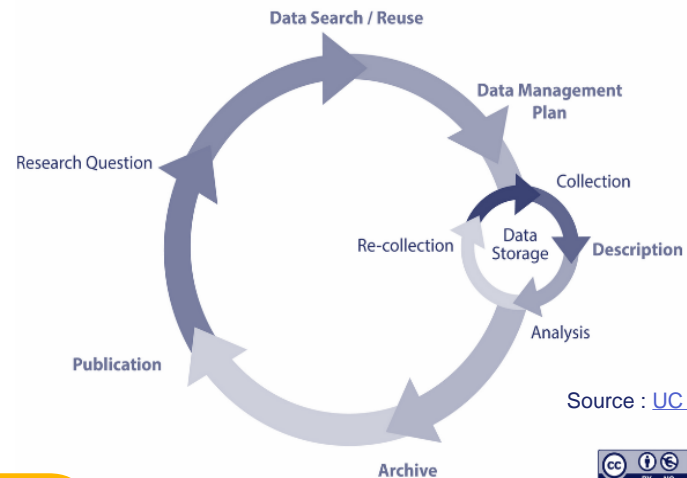
Collecte des données

- 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



- 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 ?

Ethique

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

Archivage et préservation des données

Plan de gestion des données



The screenshot shows a video player interface. The video content features a dark blue background with stylized light blue leaves. Several icons are overlaid on the leaves: a play button, a document with a magnifying glass, a microphone, and a document with a list. Two hands are shown holding up some of these icons. The video player includes a 'Version audio' button and social media sharing icons. Below the video player, the title 'LA GESTION DES DONNÉES ET LE CYCLE DE VIE DES DONNÉES DE LA RECHERCHE' is displayed. To the right of the video player, there is a 'Notice' section with details about the video, including language, credits, and citation information. At the bottom right of the notice section, there is a 'Contacter' button.

Notice

Sous-titrage : Français (Affichés par défaut) Anglais
Langue : Français
Crédits : 4 minutes 34 (Réalisation), Comité pour la science ouverte - Ministère de l'Enseignement supérieur et de la Recherche (Production), Université de Lille (Production), Egon Heuson (Intervenant)
Conditions d'utilisation : [CC BY-SA \(Attribution - Partage dans les Mêmes Conditions\)](#)

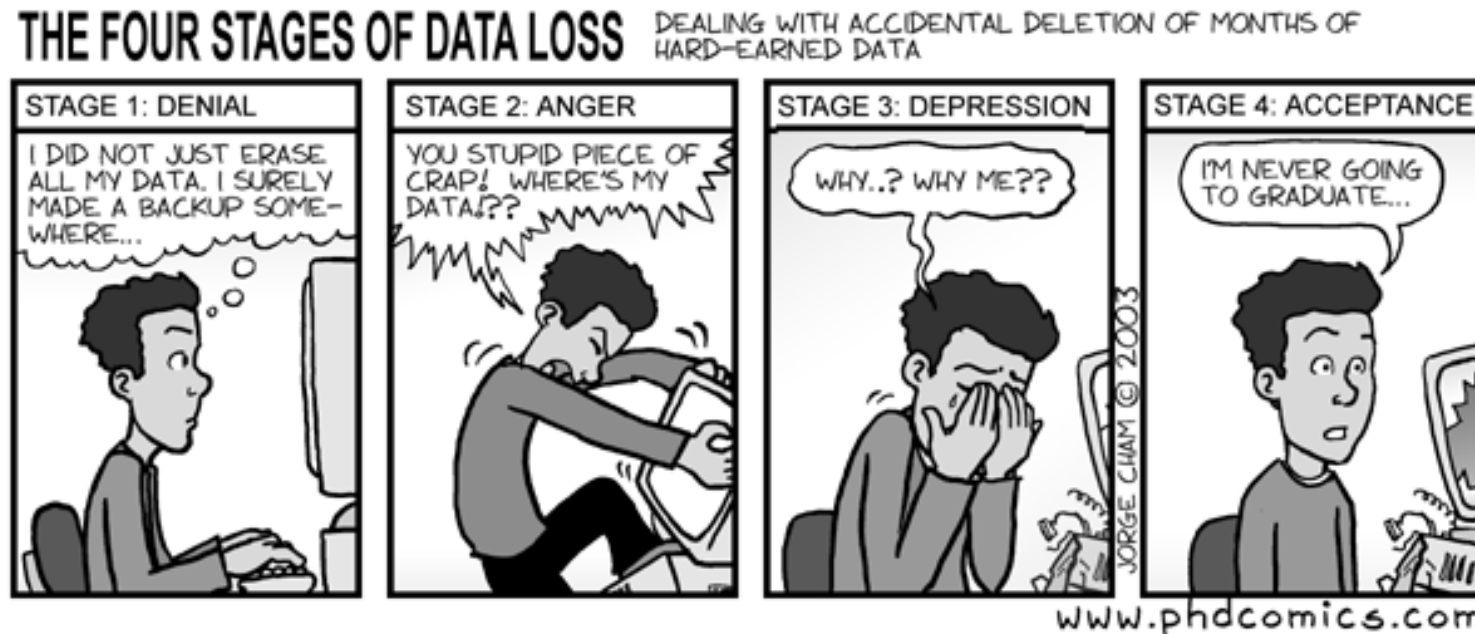
Citer cette ressource :
Egon Heuson, ouvrirlascience. *La gestion des données et le cycle de vie des données de la recherche*, in *Passeport pour la science ouverte*. [Vidéo]. Canal-U. <https://www.canal-u.tv/138419>. (Consultée le 28 novembre 2023)

 Ouvrir la science! Ouvrir la science 

LA GESTION DES DONNÉES ET LE CYCLE DE VIE DES DONNÉES DE LA RECHERCHE

Source : <https://www.canal-u.tv/chaines/ouvrirlascience/passeport-pour-la-science-ouverte/la-gestion-des-donnees-et-le-cycle-de-vie>

Bonnes pratiques de gestion



Source : <http://phdcomics.com/comics/archive.php/images/archive.php?comid=382>

Ca n'arrive qu'aux autres ?

Elle perd sa thèse dans un train Bordeaux-Nîmes, son père lance un appel

Le Figaro

A la Une - Météo



LE FIGARO · fr
étudiant

Nos événements

Études sup | Bac | Collège - Lycée

Après avoir perdu sa thèse dans un TGV, elle lance un appel pour la retrouver

Par Figaro Etudiant • Publié le 04/04/2019 à 16:46



CASH REWARD

for returning my lost backpack



383 Adventure.com

- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is VERY important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: 07804430054 (ar456@cam.ac.uk)

Thank you!!

Stocker ses fichiers – 1



- **Conserver plusieurs copies de ses données sur différents lieux de stockage. Règle de la sauvegarde 3-2-1 : 3 copies sur 2 supports différents, avec au moins 1 une copie à distance**
- **Gérer les accès aux données : accès sécurisé pour les partenaires, vérification des accès en fin de projet**
- **Chiffrer les données si elles ont un caractère sensible**
- **Utiliser un cloud pour les accès distants**
 - **Dropbox**
 - **Google drive**
 - **Etc.**

Stocker ses fichiers – 2



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'utilisateur	- 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

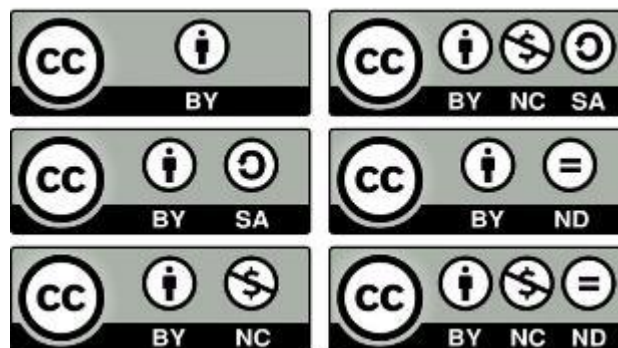
Source : <https://doranum.fr/stockage-archivage/stockage-donnees/>

Deux points de vigilance



Réutilisation ou partage de données en ligne, attention à leur cadre juridique :

- Vérifier si la licence permet la réutilisation d'une image, etc.



Utiliser par exemple des données sous Creative commons

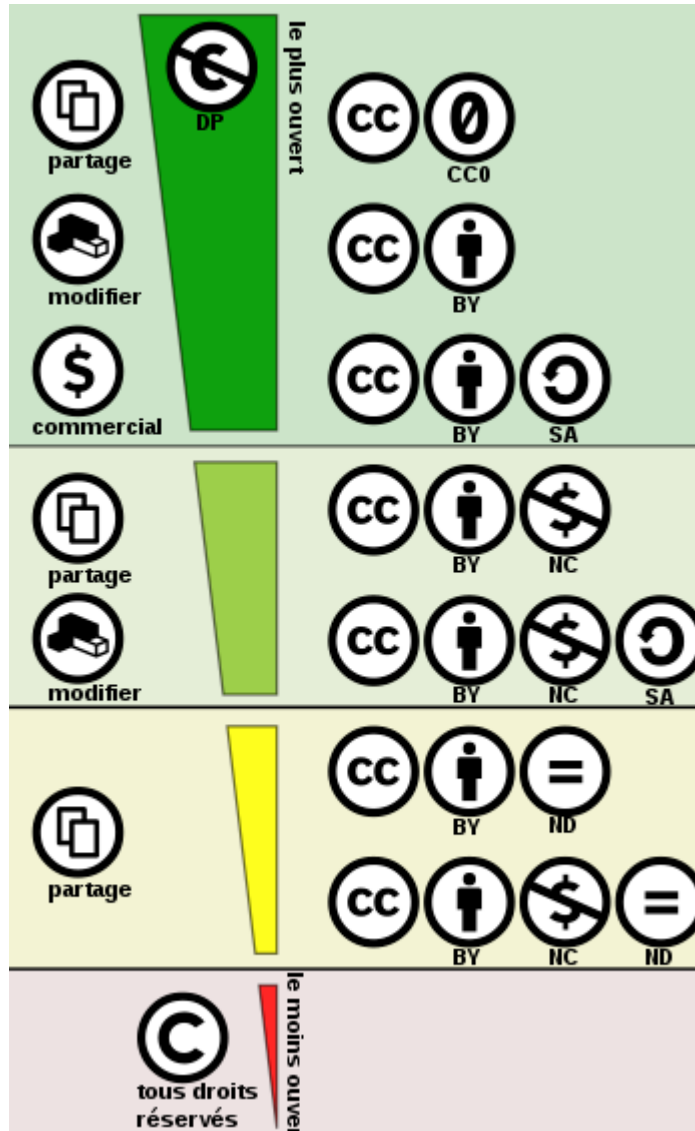
- Ne pas diffuser de données personnelles



Se conformer au RGPD, par exemple en anonymisant les données

Licences

- Creative commons (international)



- Licence ouverte (France)



LICENCE OUVERTE
OPEN LICENCE

Source : FOSTER. Introduction to RDM concepts and tools / S. Venkataraman. 2018.

<https://www.fosteropenscience.eu/node/2514>

Obligations liées au RGPD



Université Paris Lumières. 2019. « Fiches pratiques sur le Règlement Général pour la Protection des Données ».

<https://www.u-plum.fr/wp-content/uploads/2019/09/Guide-RGPD-2019-web.pdf>

Santé

CNIL. 2018. « Recherche médicale : quel est le cadre légal ? »
2018. <https://www.cnil.fr/en/node/24981>

Travailler avec son DPD

A Sorbonne Université, documentation sur l'intranet :

- Un [guide](#),
- Des exemples de fiches de registre d'activité par faculté.

Contact de la DPD :

dpd@sorbonne-universite.fr

UMR :

- Le DPD a été choisi par la/le DU, soit au CNRS/INSERM, soit à l'université.

Fiche de registre d'activité

MISE EN ŒUVRE D'UN TRAITEMENT PORTANT SUR DES DONNÉES PERSONNELLES	
Nom du traitement	_____
Date de création Date de révision de la fiche	_____
Service chargé de la mise en œuvre du traitement et nom du responsable	_____
Coordonnées du sous-traitant ou du service informatique chargé de la sécurité des données	_____
Population concernée (étudiants, personnels...)	_____
Finalité principale: décrire les objectifs poursuivis et l'usage pour le service	_____
Énumérer chacune des tâches automatisées par le traitement: collecte, transfert, anonymisation, traitement comparatif, statistiques...	<input type="checkbox"/> _____ <input type="checkbox"/> _____ <input type="checkbox"/> _____ <input type="checkbox"/> _____ <input type="checkbox"/> _____
Information des personnes concernées ou recueil du consentement éclairé (conserver les justificatifs)	<input type="checkbox"/> par mention sur le questionnaire de collecte <input type="checkbox"/> par affichage <input type="checkbox"/> par la remise d'un document <input type="checkbox"/> par une mention sur le site internet <input type="checkbox"/> par envoi de courrier <input type="checkbox"/> par intranet <input type="checkbox"/> Autres: _____
Échanges de données	• Existe-t-il un échange de données avec d'autres fichiers? OUI <input type="checkbox"/> NON <input type="checkbox"/> Si oui, le(x)quel(s): _____ _____ _____ • Existe-t-il des transferts d'informations hors de l'Union européenne? OUI <input type="checkbox"/> NON <input type="checkbox"/>

Envoyer à dpd@sorbonne-universite.fr

Anonymisation des données



Deux exemples d'outil :

- **ARX**
<https://arx.deidentifier.org/>
- **Amnesia**
<https://amnesia.openaire.eu/>



PERSONAL DATA		ANONYMOUS DATA
Fully identifiable data	Pseudonymous data	Data which cannot be traced back to an individual person. Data collected from human subjects, either directly or when re-used for secondary research, is rarely ever anonymous.
Any data from which a person could immediately be identified with little to no effort (e.g. names, faces, unique characteristics).	Data from which a person cannot be immediately identified, however it remains possible to identify a person from the data.	
EXAMPLE Patient number 90210 City Leeuwarden Date of birth 27-4-1967 Income 7.861 Job Judge Car DeLorean License Plate SN-09-HN	Study subject 47110009 Region Friesland Year of birth 1967 Income 7.500-10.000 Job Legal Car DeLorean	Country Netherlands Age 51-60 Income 5.000 – 15.000 Job Legal Car sportwagen

3

**POURQUOI ET COMMENT DIFFUSER SES
DONNÉES ? ↓**

Principe d'ouverture des données



« As **open** as possible, as **closed** as necessary. »

- Un principe d'**ouverture par défaut** ;
- Des **limitations** en fonction du type de données.

Mise en œuvre : production de données FAIR

Principes



« As **open** as possible, as **closed** as necessary. »



Détail des principes FAIR :
<https://www.force11.org/group/fairgroup/fairprinciples>

Logo FAIR : SangyaPundir. CC: BY-SA 4.0.

Image : [ANDS](#). CC: BY 4,0,

Findable



ARCs



Persistent identifiers



Metadata

Cassandra. 1947. "This World of Women."

DOI : 10.34847/nkl.5f554adl version 2 Hdl : 11280/a1fa2a8f Public

Author : Anonymous

Cassandra. 1947. "This World of Women." Southern Nigerian

Files

- SND 1948 (56) Cassa.... 3..png
- Southern-Nigerian-...A-07.odt

Viewer



Southern Nigerian Defender (2019)
«Cassandra. 1947. "This World of Women."» [Image] NAKALA.
<https://doi.org/10.34847/nkl.5f554adl>

License

Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0)

> View all metadata

Cassandra. 1947. "This World of Women."

Title	English	Cassandra. 1947. "This World of Women."
Author		Anonymous
Created		1948-02-06
License		Creative Commons Attribution Non Commercial Share Alike 4.0 International (CC-BY-NC-SA-4.0)
Type		Image
Http://purl.org/dc/terms/creator		Southern Nigerian Defender
Description	English	Cassandra. 1947. "This World of Women." Southern Nigerian Defender, February 6, 1948, sec. "Women's Realm" (p. 3).
Http://purl.org/dc/terms/date		1948-02-06+01:00
Http://purl.org/dc/terms/type		Photos et images
Language		English
Http://purl.org/dc/terms/spatial	English	Nigeria
Http://purl.org/dc/terms/bibliographiccitation	English	Cassandra. 1947. "This World of Women." Southern Nigerian Defender, February 6, 1948, sec. "Women's Realm" (p. 3).
Http://purl.org/dc/terms/license		https://creativecommons.org/licenses/by-nc-sa/4.0



Accessible



How to access data
(Repositories)

Access conditions

Beardsley, Marc, Martinez-Moreno, Judit, Santos, Patricia, & Hernández-Leo, Davinia. (2019). Supporting students in making informed data sharing decisions: from comprehension to consenting [Data set]. Zenodo.

<https://doi.org/10.5281/zenodo.3604060>

Supporting students in making informed data sharing decisions: from comprehension to consenting

Beardsley, Marc; Martinez-Moreno, Judit; Santos, Patricia; Hernández-Leo, Davinia

Whether deciding to participate in research or gain access to online applications, studies on consent suggest that individuals often do not read or fail to understand the text outlining the conditions they agree to when giving their consent. Limited comprehension related to the risks of big data and habituation may contribute to consent being given in a manner that counters the best interests of individuals (i.e. a privacy paradox). Are those asking students to use new technologies and share their data ethically responsible for helping students learn to make informed decisions regarding the sharing of their data? As educational technology researchers, we conducted a quasi-experimental study involving 127 undergraduate students in which we explored the effects on study enrollment of supporting students in making informed data sharing decisions. Prior to reading the consent form for a study involving the evaluation of a collaborative learning application, one group of participants was primed to consider data sharing risks whereas the other group was not. The primed group scored lower on a consent form comprehension quiz yet had higher enrollment rates and were more likely to cite trust as reasoning for enrolling in the study.

The data set is composed by:

- 'Classes' spreadsheet: including the relation of flows, group of students, number of students, instructor id, day and time
- 'Participants' spreadsheet: including the relation of ID (of each participant), if s/he consented, group class, flow and language
- 'Prior knowledge Survey' spreadsheet: participant's ID and ID, relevant prior knowledge questionnaire data (questions included).
- 'Quiz data' spreadsheet: participant's ID, quiz data and survey data (questions included)
- PDF including:
 - Table S1. Summary of participating classes
 - Table S2. Consent form subsections and summaries
 - Table S3. Consent form quiz questions, question types, and assigned points
 - Table S4. Guiding questions in the prior knowledge questionnaire
 - Table S5. Comparison of Flow A and Flow B quiz scores per question

Files

Restricted Access

You may request access to the files in this upload, provided that you fulfil the conditions below. The decision whether to grant/deny access is solely under the responsibility of the record owner.

185 views, 3 downloads

Indexed in **OpenAIRE**

Publication date: May 22, 2019

DOI: [10.5281/zenodo.3604060](https://doi.org/10.5281/zenodo.3604060)

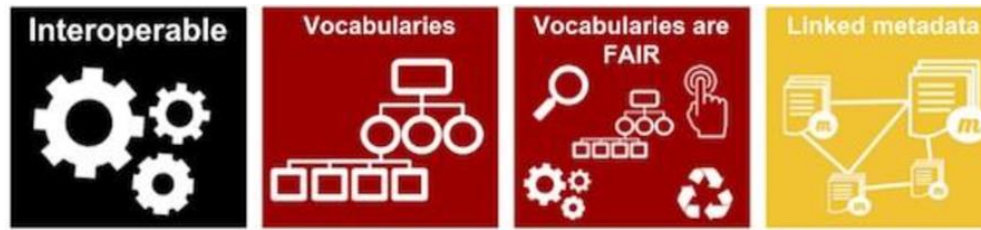
Keyword(s): Educational Technology, Ethics, Responsible Research, Informed Consent, Data Sharing Risks

Communities: Educational Data Analytics, Department of Information and Communication Technologies, UPF, Barcelona

Versions: Version 2 (May 22, 2019), Version 1 (May 22, 2019)

Cite all versions? You can cite all versions by using the DOI

604060	May 22, 2019
132921	May 22, 2019
You can cite all versions by using the DOI 10.5281/zenodo.3604060. This DOI represents all versions up to the latest one. Read more.	
Martinez-Moreno, Judit, Santos, Patricia, & Hernández-Leo, Davinia. (2019). Supporting students in making informed data sharing decisions: from comprehension to consenting [Data set]. Zenodo. https://doi.org/10.5281/zenodo.3604060	



Interoperability



BIBLIOTHÈQUE
UNIVERSITAIRE

Open formats

Ontologies

Vocabulaires

Beja DoReCo dataset
DOI : 10.34847/nkl.edd011t1 | version 8 | Publique

Auteur : Martine Vanhove

dataset for the language Beja in DoReCo v1.2b

Request URL: `https://api.nakala...`

Server response Code: 200

Fichiers

- doreco_README.txt
- doreco_CONVENTIONS.txt
- doreco_beja1238...info.txt
- doreco_beja123...data.csv**
- doreco_beja1238...ions.csv
- doreco_beja1238...ions.csv

Visualisation

Afficher 250 entrées

id	name	spk_code	spk_age	spk_age_c	spk...
T001	BEJ_MV_CONV_01_RICH	MA/AH	49/35	certain/certain	m
T002	BEJ_MV_NARR_01_SHELTER	AA	40	certain	m

doreco_beja1238_metadata.csv - Bloc-notes

```

id,name,spk_code,spk_age,spk_age_c,spk_sex,rec_date,rec_date_c,genre,genre_stim,gloss,transl,sound_quality,background_noise,word_tokens,extended
T001,BEJ_MV_CONV_01_RICH,MA/AH,49/35,certain/certain,m/m,2011-02-05,certain,conversation,na,all,English,medium,punctual,932,no
T002,BEJ_MV_NARR_01_SHELTER,AA,40,certain,m,2007-12-25,certain,traditional narrative,na,all,English,good,none,249,no
T003,BEJ_MV_NARR_02_farmer,AA,40,certain,m,2007-02-04,certain,traditional narrative,na,all,English,good,constant,538,no
T004,BEJ_MV_NARR_03_camel,AA,40,certain,m,2007-12-30,certain,personal narrative,na,all,English,good,none,394,no
T005,BEJ_MV_NARR_04_djinn,AA,40,certain,m,2006-02-08,certain,personal narrative,na,all,English,medium,constant,249,no
T006,BEJ_MV_NARR_05_eritrea,AA,40,certain,m,2003-02-10,certain,traditional narrative,na,all,English,good,constant,710,no
T007,BEJ_MV_NARR_06_foreigner,AA,40,certain,m,2007-12-25,certain,traditional narrative,na,all,English,good,punctual,108,no
T008,BEJ_MV_NARR_07_cold,AA,40,certain,m,2007-12-25,certain,traditional narrative,na,all,English,good,none,135,no
T009,BEJ_MV_NARR_08_drunkard,AA,40,certain,m,2007-12-25,certain,traditional narrative,na,all,English,good,none,342,no
T010,BEJ_MV_NARR_09_jewel,AA,40,certain,m,2007-12-25,certain,traditional narrative,na,all,English,good,none,113,no
  
```

Source : Vanhove, Martine (2023) «Beja DoReCo dataset» [Dataset] NAKALA.

<https://doi.org/10.34847/nkl.edd011t1>



Reusable



Documentation

Licenses

Beja DoReCo dataset

DOI: 10.34847/nkl.edd011t1 version 8 Publique

Fichiers

- doreco_README.txt
- doreco_CONVENTIONS.txt
- doreco_beja1238...info.txt
- doreco_beja123...data.csv
- doreco_beja1238...ions.csv
- doreco_beja1238...ions.csv
- doreco_beja1238...iers.csv
- doreco_beja1238_wd.csv
- doreco_beja1238_ph.csv
- doreco_beja123...ment.pdf
- doreco_beja123...TextGrid
- doreco_beja1238...elle.eaf
- doreco_beja123...TextGrid

Visualisation

```

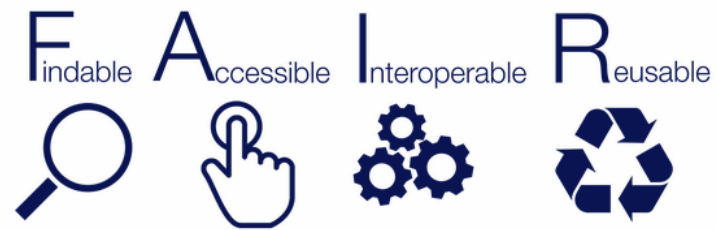
The Beja DoReCo dataset was compiled by Martine Vanhove between 2003 and 2011, originally annot
Beja DoReCo annotation files are distributed under an Attribution-NonComercial 2.0 Generic (CC BY-NC 2.0) L
Beja DoReCo audio files are distributed under an Attribution-NonComercial 2.0 Generic (CC BY-NC 2.0) Licens
The Beja DoReCo dataset should be cited as follows:
Vanhove, Martine. 2022. Beja DoReCo dataset. In Seifart, Frank, Ludger Paschen and Matthew Stave (eds.). La
Please note that when citing this dataset, or any number of DoReCo datasets, it is NOT sufficient to refer
A web interface for easy access to this dataset as well as other DoReCo datasets and further information ca
          
```

Copier l'ID Copier l'url d'intégration Copier l'url de téléchargement

Licence
Creative Commons Attribution Non Commercial 4.0 International (CC-BY-NC-4.0)

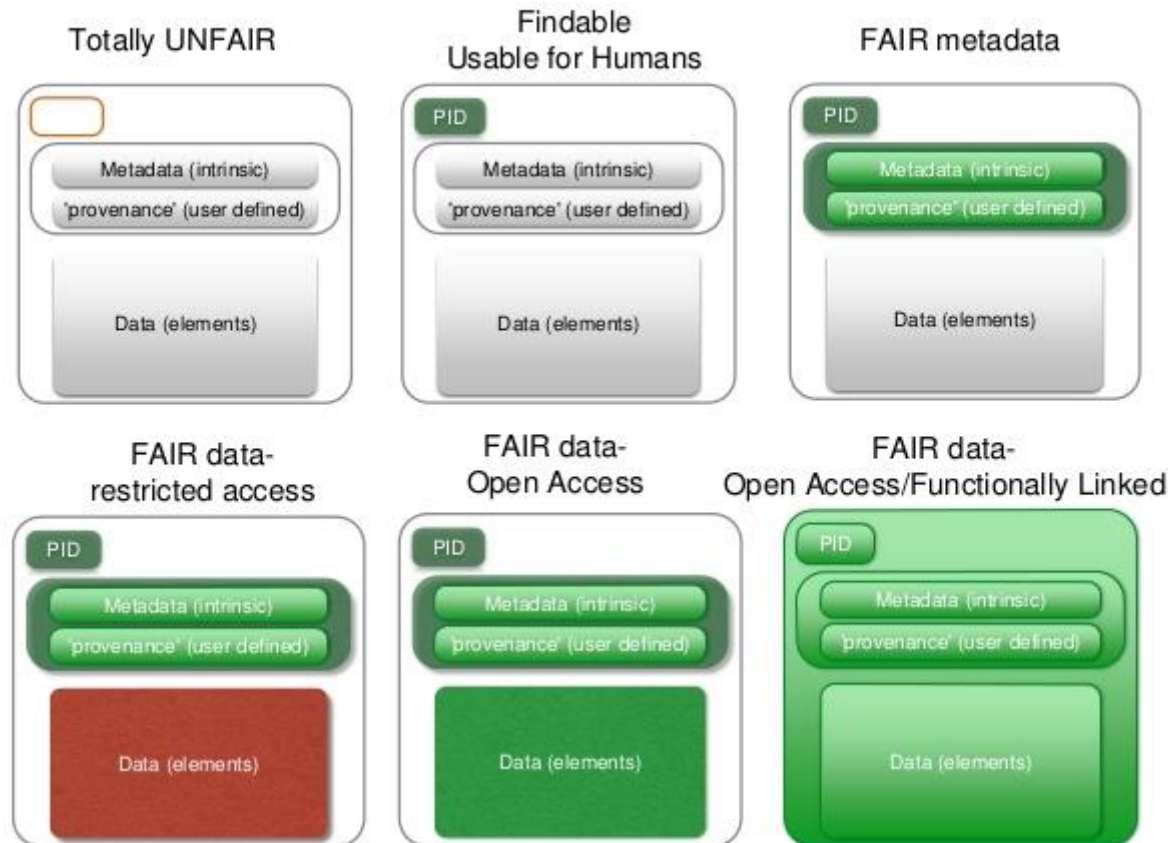
Source : Vanhove, Martine (2023) «Beja DoReCo dataset» [Dataset] NAKALA. <https://doi.org/10.34847/nkl.edd011t1>

Citer



≠ **Open Data**

Data as increasingly FAIR Digital Objects



Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LOB, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information Services & Use. 2017;37(1):49-56. <https://doi.org/10.3233/ISU-170824>

Management differences

digitized object and its metadata

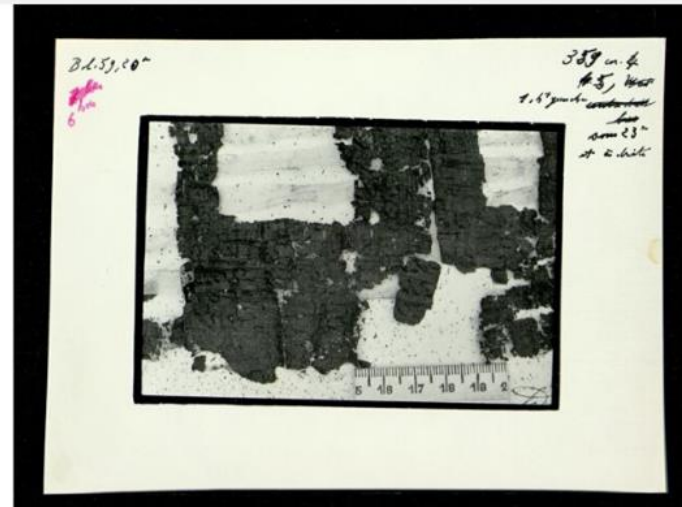
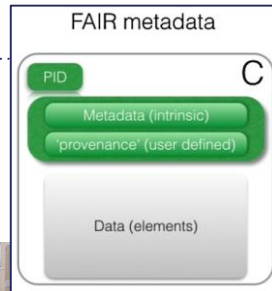
Physical objects

BIBLIOTHÈQUE UNIVERSITAIRE

- A. Paléographie antique et médiévale
- Évolution des écritures
 - Écritures grecques et latines du 1er au 5e siècle
 - Alphabet latin
 - Nomenclature des écritures livresques
 - B "baroque"
 - "Onciale" épigraphique
 - III ème - Xème siècles
 - XI ème - XII ème siècles
 - Pecia XII ème - XIV ème siècles
 - XV ème siècle, Humanistique
 - XVI ème siècle
 - Études et transcriptions, par lieux:
 - Écritures romaines
 - 2.1.1. Ostie
 - 2.1.2. Rome
 - 2.1.3. Palatin
 - 2.1.4. Catacombes
 - 2.1.5. Herculanium
 - 2.1.6. Pompéi
 - 2.2. Afrique du Nord
 - Études et transcription, par support d'écriture
 - 3.1. Graffiti et dipinti
 - 3.2. Tablettes
 - 3.3. Amphores
 - Autres inscriptions
 - Vocabulaire militaire
 - Généralités
 - Vocabulaire par ordre alphabétique.
 - Écritures: Suisse
 - Kaiseraugst
 - Vindonissa



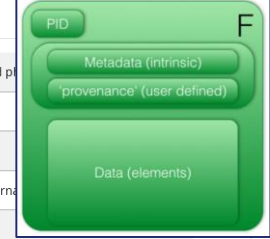
Source: Scappaticcio, *Per i testi latini*.



FAIR data-
Open Access/Functionally Linked

Rustic Capital Writing: typed drafts and photos

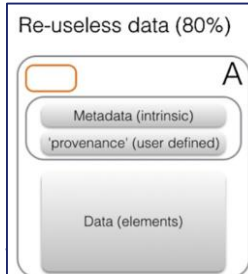
Titre	Rustic Capital Writing: typed drafts and photos
Auteur	Robert Marichal
Date de création	Inconnue
Licence	Creative Commons Attribution 4.0 International
Type	Autre
Descriptions	vue 17
Http://purl.org/dc/terms/publisher	VP
Http://purl.org/dc/terms/type	text ## image
Http://purl.org/dc/terms/identifier	PLATINUM_4MAR_10_2_5-19.6(532)



source: Marichal, Robert (2015) «Rustic Capital Writing: typed drafts and photos» [text ## image] NAKALA. <https://doi.org/10.1280/1a32894f>

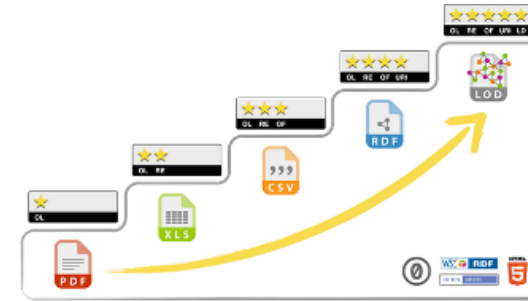
research data produced during project

X Copie de Marichal_20...
X Maricha_20160302.xl...
W Maricha_num_classm...
W plan-classement_Mari...
PDF PeritestLatiniMaricha...
W Marichal_num_classm...
W Marichal_num_classm...
X Maricha_20160220.xl...



Echelle 5 étoiles de Tim Berners-Lee

2006 : échelle 5 étoiles de Tim Berners-Lee



Prévision des températures pour Galway, Irlande

Jour	Température la plus basse (°C)
Samedi, 13 novembre 2010	2
Dimanche, 14 novembre 2010	4
Lundi, 15 novembre 2010	7



	A	B	C
1	Prévision des températures pour Galway, Irlande		
2			
3			
4	Jour	Température la plus basse (°C)	
5	Samedi, 13 novembre 2010	2	
6	Dimanche, 14 novembre	4	
7	Lundi, 15 novembre 2010	7	



Jour, Température la plus basse (C)
 "Samedi, 13 novembre 2010 ",2
 "Dimanche, 14 novembre 2010 ",4
 "Lundi, 15 novembre 2010 ",7

PRÉVISION DES TEMPÉRATURES POUR GALWAY, IRLANDE

Jour	Température la plus basse (°C)
Samedi, 13 novembre 2010	2
Dimanche, 14 novembre 2010	4
Lundi, 15 novembre 2010	7



PREVISION DES TEMPÉRATURES POUR GALWAY, IRLANDE

Jour	Température la plus basse (°C)
Samedi, 13 novembre 2010	2
Dimanche, 14 novembre 2010	4
Lundi, 15 novembre 2010	7



```
<div id="data" about="#Galway" typeof="meteo:Place">
  <table border="1px">
    <tr>
      <th>Jour</th>
      <th>Température la plus basse (&deg;C)</th>
    </tr>
    <tr rel="meteo:forecast" resource="#forecast20101113">
      <td>
        <div about="#forecast20101113">
          <span property="meteo:predicted" content="2010-11-13T00:00:00Z" datatype="xsd:dateTime">Samedi, 13 novembre 2010</span>
        </div>
      </td>
    </tr>
  </table>
</div>
```

```
<div id="data" about="#Galway" typeof="meteo:Place">
  <span rel="owl:sameAs" resource="http://dbpedia.org/resource/Galway"></span>
  <table border="1px">
    <tr>
      <th>Jour</th>
      <th>Température la plus basse (°C)</th>
    </tr>
    <tr>
      <td>
        <div about="#temp">
          <a rel="rdfs:seeAlso" href="http://en.wikipedia.org/wiki/Température" resource="http://dbpedia.org/resource/Température">Température</a> la plus basse (<span rel="owl:sameAs" resource="http://dbpedia.org/resource/Celsius"&deg;C</span>)
        </div>
      </td>
    </tr>
    <tr rel="meteo:forecast" resource="#forecast20101113">
      <td>
        <div about="#forecast20101113">
          <span property="meteo:predicted" content="2010-11-13T00:00:00Z" datatype="xsd:dateTime">Samedi, 13 novembre 2010</span>
        </div>
      </td>
    </tr>
    <tr rel="meteo:temperature">
      <td>
        <div about="#temp20101113">
          <span property="meteo:celsius" datatype="xsd:decimal">2</span>
        </div>
      </td>
    </tr>
  </table>
</div>
```

Open linked-data



Qu'est ce qu'un entrepôt de données ?

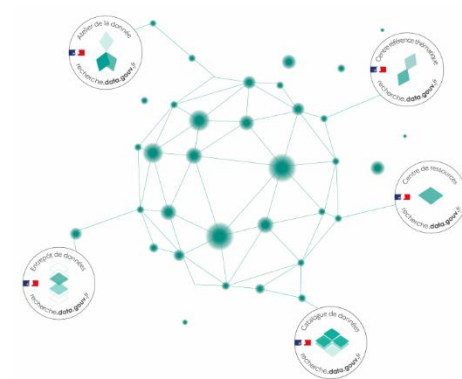
« Un service en ligne destiné à gérer la **description** d'ensembles de données en vue de leur **préservation** et de leur **réutilisation**. Un entrepôt expose, pour chaque ensemble de données, des **métadonnées** détaillées et facilite l'**accès aux données**. Parmi les informations [...], on trouve notamment les métadonnées essentielles, de provenance et les conditions d'utilisation. »

Partager sur un entrepôt de données



Un entrepôt de données ?

- Archives en ligne permettant le dépôt, la description, la conservation, la recherche et la diffusion de jeux de données

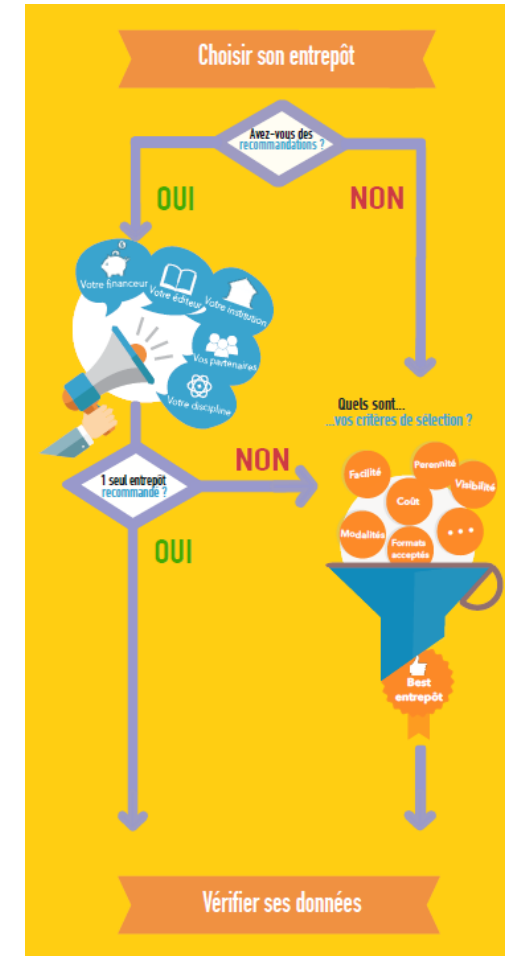


Choisir un entrepôt

- Éditeurs ou financeurs font souvent des recommandations en matière d'entrepôt de données.
- Pas de recommandations ? Avant de choisir un entrepôt, vérifier :
 - La facilité de dépôt
 - Le coût
 - Les modalités d'accès
 - La pérennité d'accès
 - Les formats acceptés
 - Les volumes acceptés
 - L'attribution d'un identifiant pérenne (DOI ou autre)
 - La certification : [CoreTrustSeal](#) ([critères](#))

Pour comparer, voir aussi le tableau de synthèse de l'université de Sherbrooke :

https://libguides.biblio.usherbrooke.ca/savoirs/choix_depot



Source : Doranum <https://doranum.fr/depot-entrepots/choix-entrepot-depot-donnees/>

Rechercher des entrepôts de données

- Trouver des entrepôts par disciplines



- Entrepôts français :
CAT OPIDoR



- Rechercher des jeux de données

[Search DataCite](#)

[OpenAIRE](#)



Partager logiciels et codes source



Software Heritage



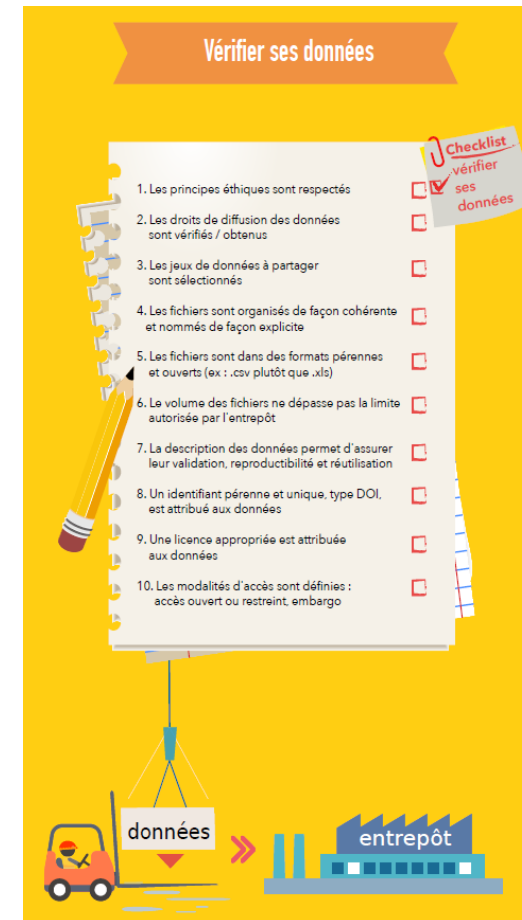
- Le projet Software Heritage, archive des logiciels et codes source
- A préparer avant le dépôt :
 - fichier README
 - liste normalisée d'auteurs
 - choisir une licence
- Chaque version du logiciel archivé dispose d'un identifiant unique et pérenne.

Partenariat avec HAL : dépôt des codes sources de logiciels via HAL, et archivage pérenne automatique dans Software Heritage.

Voir : <https://doc.archives-ouvertes.fr/deposer/deposer-le-code-source/>

Vérifier un jeu de données avant son partage

- Les **principes éthiques** sont respectés
- Les **droits de diffusion** sont vérifiés/obtenus
- Les modalités d'**accès** sont définies
- Une **licence** appropriée est attribuée aux données
- Les jeux de données à partager sont sélectionnés
- Les **fichiers** sont organisés et **nommés** de façon explicite
- Les fichiers sont dans des **formats pérennes et ouverts**
- Le **volume** des fichiers ne dépasse pas la limite autorisée
- Un **identifiant pérenne** est attribué aux données
- Les données sont **décrites et documentées**



Source : Doranum <https://doranum.fr/depot-entrepots/choix-entrepot-depot-donnees/>

Démonstration de dépôt



BIBLIOTHÈQUE
UNIVERSITAIRE



Zenodo : <https://zenodo.org/>



DRYAD

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

Pour tester, le bac à sable : <https://sandbox.zenodo.org/>

Un exemple de dépôt Zenodo – 1

Delete Save Publish

New upload

Instructions: (i) Upload minimum one file or fill-in required fields (marked with a red star). (ii) Press "Save" to save your upload for editing later. (iii) When ready, press "Publish" to finalize and make your upload public.

Files Choose files Start upload

Filename (1 files)	Size	Progress	Delete
DMP_OAW_20211027.pdf md5:896f6ce29e177dbff441ada39fe08e2	4.9 MB	✓	

Note: File addition, removal or modification are not allowed after you have published your upload. This is because a Digital Object Identifier (DOI) is registered with [DataCite](#) for each upload.
(minimum 1 file required, max 50 GB per dataset - [contact us](#) for larger datasets)
If you're experiencing issues with uploading larger files, read our [FAQ section](#) on file upload issues.

Communities recommended

Specify communities which you wish your upload to appear in. The owner of the community will be notified, and can either accept or reject your request. Please make sure your record complies with the content policy of the communities you add; reported abuse will be followed by account inactivation.

Upload type required

Zenodo :
<https://zenodo.org/>

**Pour tester, le bac
à sable :**
<https://sandbox.zenodo.org/>

Un exemple de dépôt Zenodo – 2

■ Digital Object Identifier

e.g. 10.1234/foo.bar

Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.

■ Reserve DOI

📅 Publication date *

2021-10-27

Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.

📄 Title *

Rédiger son plan de gestion des données

Required.

👤 Authors *

Arènes, Cécile

Sorbonne Université

0000-0002-1839-3530

Optional.

+ Add another author

📄 Description *



Présentation sur la rédaction des plans de gestion de données

Durée : 1h

Modalité : [visio](#), exposé et discussion

Required.


📄 Version


Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended. See semver.org for more information on semantic versioning.


Un exemple de dépôt Zenodo – 3


License required ▼

Access right *

 Open Access

 Embargoed Access

 Restricted Access

 Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

License *

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the *Other* licenses available (*Other (Open)*, *Other (Attribution)*, etc.). The supported licenses in the list are harvested from opendefinition.org and spdx.org. If you think that a license is missing from the list, please [contact us](#).

Funding recommended ▼ ×

Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](#). Specify grants which have funded your research, and we will let your funding agency know!

Grants

Optional. OpenAIRE-supported projects only. For other funding acknowledgements, please use the *Additional Notes* field.
Note: a human Zenodo curator will need to validate your upload - you may experience a delay before it is available in OpenAIRE.

[+ Add another grant](#)

Related/alternate identifiers recommended ▶

Contributors optional ▶

References optional ▶

Journal optional ▶

Conference optional ▼

Conference title


Optional.

Un exemple de dépôt Zenodo – 4

zenodo [Upload](#) [Communities](#) cecile.arenas@gmail.com

October 27, 2021 [Presentation](#) [Open Access](#) [Edit](#) [New version](#)

Rédiger son plan de gestion des données

 Arènes, Cécile


Présentation sur la rédaction des plans de gestion de données

Durée : 1h

Modalité : visio, exposé et discussion

Support : version .pptx originale et version .pdf


Preview



Files (10.5 MB)

Name	Size	
DMP_OAW_20211027.pdf	4.9 MB	Preview Download

Indexed in



Publication date:
October 27, 2021

DOI:
DOI [10.5281/zenodo.5603564](https://doi.org/10.5281/zenodo.5603564)

Keyword(s):
[data management plan](#) [research data management](#)

Meeting:
Open Access Week, October 25-31, 2021

License (for files):
[Creative Commons Attribution 4.0 International](#)

Versions

Version 1	Oct 27, 2021
10.5281/zenodo.5603564	

Pour des données faciles à trouver

Les identifiants pérennes



Le plus utilisé : le Digital Object Identifier (DOI)

- Identifiant pérenne, qui permet de **retrouver l'emplacement d'un document en ligne si son URL a changé**. Automatiquement généré lors du dépôt sur un entrepôt de données

Identifiants pour des logiciels : le Software Heritage Identifiers



Autres identifiants pérennes :

- **Handle system** : permet d'attribuer des identifiants pérennes à des objets numériques
- **ARK** : identifiant pérenne pour des objets numériques et physiques, très utilisé par les bibliothèques.

Pour des données accessibles et interopérables

Bien choisir les métadonnées



Dublin core : 15 champs à remplir

1. Titre (métadonnée)
2. Créateur (métadonnée)
3. Sujet (métadonnée) ou mots clé
4. Description (métadonnée)
5. Éditeur
6. Contributeur
7. Date (métadonnée)
8. Type
9. Format
10. Identifiant de la ressource
11. Source
12. Langue (métadonnée)
13. Relation (métadonnée)
14. Couverture (métadonnée)
15. Gestion de droits (métadonnée)

- Un standard générique, le **dublin core** : 15 champs
 - associé à des **standards disciplinaires** : **DDI** (SHS), **EML** (écologie), **DwC** (Darwin Core, biodiversité)
 - souvent transparent lors du dépôt, champs à remplir sur les **formulaire**s des entrepôts.
- Voir : DCC. **Disciplinary metadata**.
<http://www.dcc.ac.uk/resources/metadata-standards>
- Un outil pour **générer des métadonnées** :
https://doranum.fr/wp-content/uploads/datacite_metadata_generator_4.0.html



Pour des données réutilisables

Nommer les fichiers

Classement thématique

- par sujet :

sujet_type_date_version.extension

reunion_CR_20200227_V01.docx

- par typologie de document :

type_sujet_date_version.extension

CR_reunion_20200227_V01.docx

Classement chronologique

- par sujet :

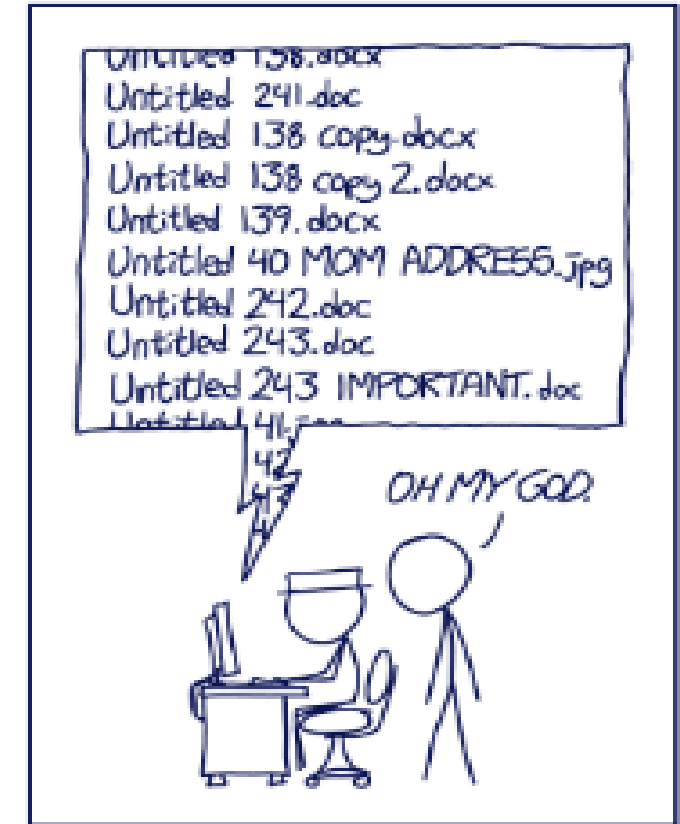
date_sujet_type_version.extension

20200227_reunion_CR_V0.1.docx

- par type de document :

date_type_sujet_version.extension

20200227_CR_reunion_V0.1.docx



PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

<https://xkcd.com/1459/>

Nommer les fichiers

Exemple





L'organisation du jeux de données



BIBLIOTHÈQUE
UNIVERSITAIRE

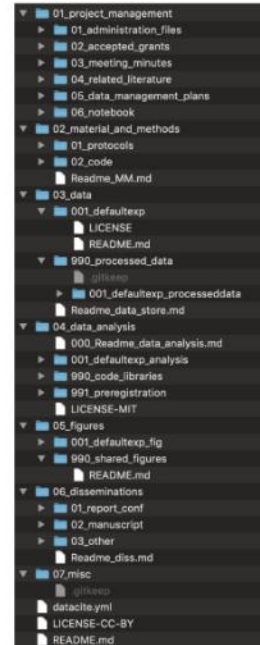
Colomb, Julien, Thorsten Arendt, Deepthi Mittal, et Keisuke Sehara. 2020. « Folder Structure Template for Research Repositories ». <https://doi.org/10.5281/zenodo.4410128>

Towards a Standardized Research Folder Structure – GenR. 2021. <https://web.archive.org/web/20230425045409/https://genr.eu/wp/towards-a-standardized-research-folder-structure/>

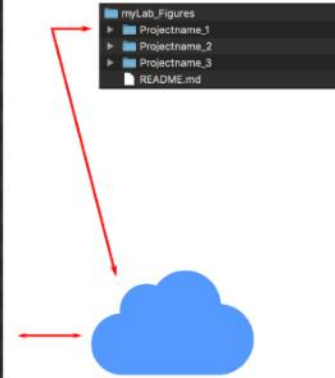
Experiment level:
add sub-folders for each experiment



Project level



Laboratory level:
mirror sub-folders in other structure



- template_par
 - ._LICENSE-CC-BY
 - 01_project_management
 - ._05_data_management_plans
 - ._06_notebook
 - 05_data_management_plans
 - ._DMP_main.txt
 - 06_notebook
 - ._.gitkeep
 - 04_data_analysis
 - ._LICENSE-MIT
- template_par
 - 01_project_management
 - 01_administration_files
 - .gitkeep
 - 02_accepted_grants
 - .gitkeep
 - 03_meeting_minutes
 - .gitkeep
 - 04_related_literature
 - .gitkeep
 - 05_data_management_plans
 - DMP_main.txt
 - 06_notebook
 - .gitkeep
 - 02_material_and_methods
 - 01_protocols
 - .gitkeep
 - 02_code
 - .gitkeep
 - Readme_MM.md
 - 03_data
 - 001_defaultexp
 - LICENSE
 - README.md
 - 990_processed_data
 - .gitkeep
 - 001_defaultexp_processeddata
 - .gitkeep
 - Readme_data_store.md
 - 04_data_analysis
 - 000_Readme_data_analysis.md
 - 001_defaultexp_analysis
 - .gitkeep
 - 990_code_libraries
 - .gitkeep
 - 991_preregistration
 - .gitkeep
 - LICENSE-MIT
 - 05_figures
 - 001_defaultexp_fig
 - .gitkeep
 - 990_shared_figures
 - Readme_sharedfig.md
 - 06_disseminations
 - 01_report_conf
 - Readme_conf.md

Pour des données réutilisables

Le fichier Readme

- Documente le dépôt
- Décrit le contenu,
- Précise les **logiciels** ou codes informatiques nécessaires pour l'utilisation

Trames de fichiers Readme :

- 4TU. **Guidelines for creating a README file.** https://researchdata.4tu.nl/fileadmin/user_upload/Dolines_for_creating_a_README_file.pdf
- Doranum. **Gabarit « Readme ».** https://doranum.fr/content/uploads/gabarit_readme.txt

Example of a good README which was submitted to the data repository:

*** Mode I fatigue delamination growth in composite laminates with fibre bridging ***
Authors: L. Yao, R.C. Alderliesten
Structural Integrity & Composites Group, Faculty of Aerospace Engineering, Delft University of Technology

Corresponding author: R.C. Alderliesten

Contact Information:

r.c.alderliesten@tudelft.nl

Delft University of Technology - Faculty of Aerospace Engineering
P.O. Box 5058
2600 GB Delft
The Netherlands

General Introduction

This dataset contains data collected during crack growth experiments at Delft University of Technology, as part of Liaojun Yao's PhD Thesis project (December 2015):
doi:10.4233/uuid:66e210e1-c884-45d6-b9d4-711907680452
It is being made public both to act as supplementary data for publications and the PhD thesis of Liaojun Yao and in order for other researchers to use this data in their own work.

The data in this data set was collected in the Structures and Materials Laboratory of the Delft University of Technology - Faculty of Aerospace Engineering, between October 2012 and December 2015.

This research project was made possible by a grant from the China Scholarship Council (CSC).

Purpose of the test campaign

The purpose of these experiments was to investigate delamination growth in CFRP composites. The Double Cantilever Beam (DCB) specimens were subjected to a variety of load cases, as detailed in "Text matrix.doc"

Test equipment

All tests were performed on a 10 kN MTS fatigue test machine. The crack length was measured by means of a camera system.
The applied force and displacement were measured by the fatigue test machine, and also sent as inputs to the camera, in order to facilitate synchronisation of the data.

Description of the data in this data set

The data included in this data set has been organised per specimen. The files follow the nomenclature system: Sp_X_Data_analysis_Y with
X = the specimen number 1 to 56
Y = indicating the number of runs with the same specimen.



Formats de fichiers

Formats ouverts, pour quoi faire ?

Pour que d'autres utilisateurs puissent lire mes documents sans entrave

Des utilisateurs s'échangent un rapport.



FORMAT FERMÉ, LOGICIELS IDENTIQUES

Alice utilise le logiciel « Carcera⁽¹⁾ ». Elle enregistre son rapport dans un format fermé (ne permettant pas l'interopérabilité), puis l'envoie à Bob qui possède le même logiciel. Celui-ci peut lire le document, le modifier et le renvoyer à Alice.



FORMATS FERMÉS, LOGICIELS DIFFÉRENTS

Le lendemain, Alice envoie son rapport à Albert. Il ne possède pas le même logiciel et celui-ci refuse de lire le document. Albert n'a pas d'autre solution que d'acquiescer le logiciel Carcera utilisé par Alice, en espérant qu'il soit adapté à son ordinateur.

Pour que vos documents puissent être lus facilement par d'autres personnes, sans avoir besoin de vous soucier du logiciel qu'elles utilisent, choisissez des formats ouverts.

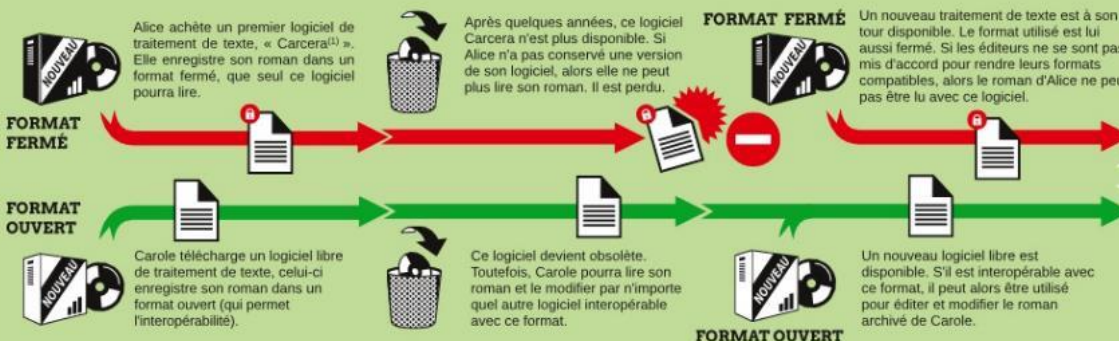


FORMATS OUVERTS, LOGICIELS DIFFÉRENTS

Carole, une autre utilisatrice, choisit d'enregistrer son rapport dans un format ouvert (permettant l'interopérabilité) et l'envoie à David. David peut lire le document, le modifier et l'enregistrer à son tour, soit en utilisant le même logiciel, soit en utilisant un autre logiciel interopérable.

Pour assurer la pérennité de mes documents dans le temps

Alice et Carole utilisent des logiciels de traitement de texte pour rédiger, enregistrer et conserver un roman.



La disponibilité et la pérennité de vos documents, enregistrés dans un format fermé, dépendent des décisions des éditeurs de logiciels.

Dans votre intérêt, choisissez des logiciels qui enregistrent vos documents dans des formats ouverts. Ils ne sont pas dépendants d'un logiciel en particulier.

Dans le cadre du partage et de la conservation à long terme il est indispensable d'utiliser des formats ouverts :


Format de fichier à privilégier
Comma Separated Values (.csv)
Texte ascii (.txt) ou PDF/A si formatage
PDF/A (.pdf)
TIFF (.tif, .tiff)
MPEG-4 (.mp4)

L'outil du CINES pour vérifier si vos formats sont conformes avec les exigences de l'archivage pérennes : <http://facile.cines.fr>

Rivet, Alain, Marie-Laure Bachèlerie, Auriane Denis-Meyere, et Delphine Tisserand. 2018. « Traçabilité des activités de recherche et gestion des connaissances : Guide pratique de mise en place ». https://qualite-en-recherche.cnrs.fr/wp-content/uploads/2021/08/guide_tracabilite_activites_recherche_gestion_connaissances.pdf


Publier un data paper

- Article décrivant un jeu de données, notamment les méthodes de recueil
- Détaille les potentiels de réutilisation de données
- Le DMP peut servir de trame
- Peut être publié dans une revue académique classique, plus souvent publié dans un data journal
- Les articles font l'objet d'un peer-reviewing

SCIENTIFIC DATA 

Data Descriptor | OPEN | Published: 10 April 2018

Long-term dataset on aquatic responses to concurrent climate change and recovery from acidification

Taylor H. Leach , Luke A. Winslow, Frank W. Acker, Jay A. Bloomfield, Charles W. Boylen, Paul A. Bukaveckas, Donald F. Charles, Robert A. Daniels, Charles T. Driscoll, Lawrence W. Eichler, Jeremy L. Farrell, Clara S. Funk, Christine A. Goodrich, Toby M. Michelena, Sandra A. Nierzwicki-Bauer, Karen M. Roy, William H. Shaw, James W. Sutherland, Mark W. Swinton, David A. Winkler & Kevin C. Rose

Scientific Data 5, Article number: 180059 (2018) | [Download Citation](#) ↓

Abstract

Concurrent regional and global environmental changes are affecting freshwater ecosystems. Decadal-scale data on lake ecosystems that can describe processes affected by these changes are important as multiple stressors often interact to alter the trajectory of key ecological phenomena in complex ways. Due to the practical challenges associated with long-term data collections, the majority of existing long-term data sets focus on only a small number of lakes or few response variables. Here we present physical, chemical, and biological data from 28 lakes in the Adirondack Mountains of northern New York State. These data span the period from 1994–2012 and harmonize multiple open and as-yet unpublished data sources. The dataset creation is reproducible and transparent; R code and all original files used to create the dataset are provided in an appendix. This dataset will be useful for examining ecological change in lakes undergoing multiple stressors.

Metadata summary

Design Type(s)	observation design • time series design • data integration objective
Measurement Type(s)	geographic location • Physical Phenomenon or Property • Inorganic Chemistry • planktonic material • weather
Technology Type(s)	digital curation • data acquisition system • Analytical Chemistry • light microscopy
Factor Type(s)	
Sample Characteristic(s)	Adirondack Park • freshwater lake biome

[Download metadata file](#) ↓

Leach TH, Winslow LA, Acker FW, Bloomfield JA, Boylen CW, Bukaveckas PA, et al. Long-term dataset on aquatic responses to concurrent climate change and recovery from acidification. Scientific Data [Internet]. 2018 [cité 28 mars 2019];5. Disponible sur: <https://www.nature.com/articles/sdata201859>

Publier un data paper

Trame

La particularité du texte d'un data paper porte sur la description fine à la fois de la méthode de production des données et des données elles-mêmes, ainsi que sur l'absence de résultat et discussion :

- **contexte** de la recherche et travaux antérieurs dans lesquels celle-ci s'inscrit, apport des données dans ce contexte et potentiel de réutilisation ;
- **protocole** de production des données : qualification du producteur des données, méthode de constitution de l'échantillon, matériel utilisé, procédures de traitement, mise en oeuvre du contrôle qualité sur les données, questions éthiques soulevées par la collecte de ces données (consentement de patients), etc.
- **description** du jeu de données : nature ou type de données, format ainsi que version du format le cas échéant, volume de données, date de publication des données dans l'entrepôt choisi par l'auteur ou préconisé par l'éditeur, identifiant des données (attribué par l'entrepôt), lien pérenne vers l'entrepôt choisi, licence d'utilisation attribuée aux données. L'article regroupe donc un texte spécifique à la description des données, ainsi qu'un lien vers le jeu de données décrit, dans l'entrepôt où celui-ci a été déposé.

Publier un data paper : quelques data journals

Quelques exemples de data journals :

- [Biomedical Data Journal](#) (Biomédical)
- [Journal of Open Health Data](#) (Données de santé)
- [Gigascience](#) (Sciences de la vie)
- [Scientific Data](#) (Multidisciplinaire)
- [Data in Brief](#) (Multidisciplinaire)

Susini, Vanessa, Laura Caponi, Veronica Lucia Rossi, Antonio Sanesi, Nadia Romiti, Aldo Paolicchi, et Maria Franzini. « Data about Performances of Whole and Monovalent Half-Fragments Antibodies in Immunosorbent Assays ». *Data in Brief* 35 (avril 2021): 106778. <https://doi.org/10.1016/j.dib.2021.106778>.

Data Article

Data about performances of whole and monovalent half-fragments antibodies in immunosorbent assays

Vanessa Susini ^a, Laura Caponi ^a, Veronica Lucia Rossi ^b, Antonio Sanesi ^b, Nadia Romiti ^a, Aldo Paolicchi ^a, Maria Franzini ^a

[Show more](#) ▾

[+](#) Add to Mendeley [↻](#) Share [↻](#) Cite

<https://doi.org/10.1016/j.dib.2021.106778>

[Get rights and content](#)

Under a Creative Commons license

[open access](#)

Refers to Vanessa Susini, Laura Caponi, Veronica Lucia Rossi, Antonio Sanesi, Nadia Romiti, Aldo Paolicchi, Maria Franzini

[Sensitivity and reproducibility enhancement in enzyme immunosorbent assays based on half fr...](#)

Analytical Biochemistry, Volume 616, 1 March 2021, Pages 114090

 [Download PDF](#)

Abstract

The data here presented are related to the research article entitled “Sensitivity and reproducibility enhancement in enzyme immunosorbent assays based on half fragment antibodies” [1] aimed to compare the performance in ELISA of whole antibodies and their corresponding monovalent half-fragments obtained by reduction. Half-fragment antibodies represent an interesting method to orient antibodies in high-sensitive immunoassays taking advantage of the free sulfhydryl groups of the hinge region [2], [3], [4] that allow their oriented binding on maleimide functionalized microplates. Data here presented describe the contribution of both chemical reduction and orientation on the antigen binding capacity of whole and half-fragments antibodies. For this purpose, monoclonal anti-

Un guide Science ouverte

- **guide** conçu pour accompagner les doctorants à chaque étape de leur parcours de recherche, quel que soit leur champ disciplinaire.
- Il propose une série de bonnes pratiques et d'outils directement activables.

PASSEPORT POUR LA **SCIENCE OUVERTE**



**GUIDE PRATIQUE
À L'USAGE DES
DOCTORANTES ET
DES DOCTORANTS**

Deux références sur les données de la recherche

Enjeux épistémologiques et enjeux techniques

- Borgman, C. L. (2020). *Qu'est-ce que le travail scientifique des données ? : Big data, little data, no data* (C. Matoussowsky, Trans.). OpenEdition Press. <http://books.openedition.org/oep/14692>

- Leonelli, S. (2019). *La recherche scientifique à l'ère des big data: cinq façons dont les big data nuisent à la science et comment la sauver*. Mimésis.



**LA BIBLIOTHÈQUE PEUT
VOUS AIDER !**

Services autour de la science ouverte à la BSU



Mission Données de la recherche & Humanités numériques :
data-bsu@sorbonne-universite.fr

Département Publications et Open Access :
hal@sorbonne-universite.fr



Service des Archives et du Recueil des Actes :
sara-archives@sorbonne-universite.fr

Vos questions

Par chat ou de vive voix !



BIBLIOTHÈQUE
UNIVERSITAIRE



Sources

- Cangadoba. *Vue d'artiste d'une exoplanète gazeuse type Jupiter froid. D'après le fichier commons.* 3 mai 2018. Own work.
https://commons.wikimedia.org/wiki/File:Exo_planete_gazeuse.jpg.
- « Criminocorpus ». Consulté le 23 novembre 2020. <https://criminocorpus.org/fr/>.
- Hilgenfeld, inzhi Tan, Clemens Vonrhein, Oliver S. Smart, Gerard Bricogne, Michela Bollati, Yuri Kusov, Guido Hansen, Jeroen R. Mesters, Christian L. Schmidt, Rolf. *English: Genome organisation of SARS-CoV.* 15 mai 2009.
<https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1000428.g001>.
https://commons.wikimedia.org/wiki/File:SARS_Coronavirus_Genome_Organization.png.
- INHESJ. « Open data, risques et crises ». INHESJ, 2019. <https://inhesj.fr/articles/open-data-risques-et-crisis>.
- Jebulon. *English: A seismometer, Mount Vesuvius, Italy.* 26 juillet 2015. Own work.
https://commons.wikimedia.org/wiki/File:Sismographe_V%C3%A9suve.jpg.
- Lemaire, Gilles G., et André Pflimlin. « Les sécheresses passées et à venir : quels impacts et quelles adaptations pour les systèmes fourragers ? » *Fourrages* 190 (2007): 163-80. <https://hal.inrae.fr/hal-02664009>.
- National spatial data infrastructure implementation plan.* Consulté le 23 novembre 2020.
[https://commons.wikimedia.org/wiki/File:National_spatial_data_infrastructure_implementation_plan_\(IA_nationalspatiald00unit\).pdf](https://commons.wikimedia.org/wiki/File:National_spatial_data_infrastructure_implementation_plan_(IA_nationalspatiald00unit).pdf).
- « Navigocorpus ». Consulté le 23 novembre 2020. <http://navigocorpus.org/>.
- Tanaka, Silvio. *English: Tim Berners-Lee at Campus Party Brasil, 2009.* 20 janvier 2009. originally posted to Flickr as Tim Berners-Lee.
https://commons.wikimedia.org/wiki/File:Tim_Berners-Lee_CP.jpg.

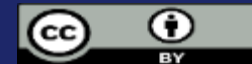
BIBLIOTHÈQUE DE SORBONNE UNIVERSITÉ

MERIC AKDOGAN

meric.akdogan@sorbonne-universite.fr



BIBLIOTHÈQUE
UNIVERSITAIRE



Sauf logos et mention contraire, cette présentation est mise à disposition selon les termes de la Licence Creative Commons Attribution 2.0 France.