

Zenodo, Figshare, etc.: rechercher et partager des données de recherche grâce aux entrepôts de données

Atelier des BU - Sciences et techniques





Programme

La mise à disposition des données associées aux articles ou données sous-jacentes (underlying data) est une demande croissante de la part des éditeurs et des agences de financement. Les entrepôts de données permettent de répondre à ces exigences, en assurant l'accessibilité, l'archivage et/ou la diffusion des données déposées, en conformité avec les principes FAIR. Cette session présente les différents types d'entrepôts de données et des recommandations pour identifier un entrepôt adapté.

- →Introduction
- →Trouver un entrepôt de données généraliste
- →Trouver un entrepôt de données spécialisé



Introduction

Terminologie : données de recherche Modes de partage des données de recherche Enjeux Entrepôts de données et principes FAIR

Terminologie : données de recherche

→Définition large OCDE

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

→ Définition NIH : NOT-OD-21-013: Final NIH Policy for Data Management and Sharing

"Scientific Data: The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens."

- → Un sous-ensemble : les données sous-jacentes à une publication ou *underlying data*
 - > Y compris la documentation et le code informatique associés à ces données

→ Répondre manuellement à des demandes



Exemple: https://doi.org/10.1186/s12888-020-03010-3

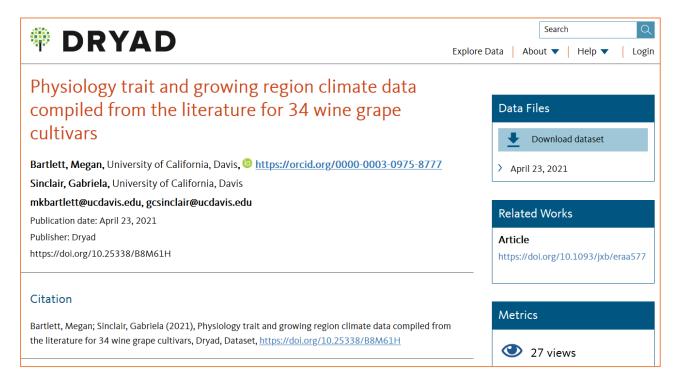
- → Annexer à un article des fichiers de Supplementary data, publiés sur le site de l'éditeur en même temps que l'article
- → Déposer les données dans un entrepôt de données



Exemple: https://doi.org/10.25338/B8M61H

Availability of data and materials

Readers and all interested researchers may contact Runsen Chen (Email address: runsen.chen@psych.ox.ac.uk) for details.

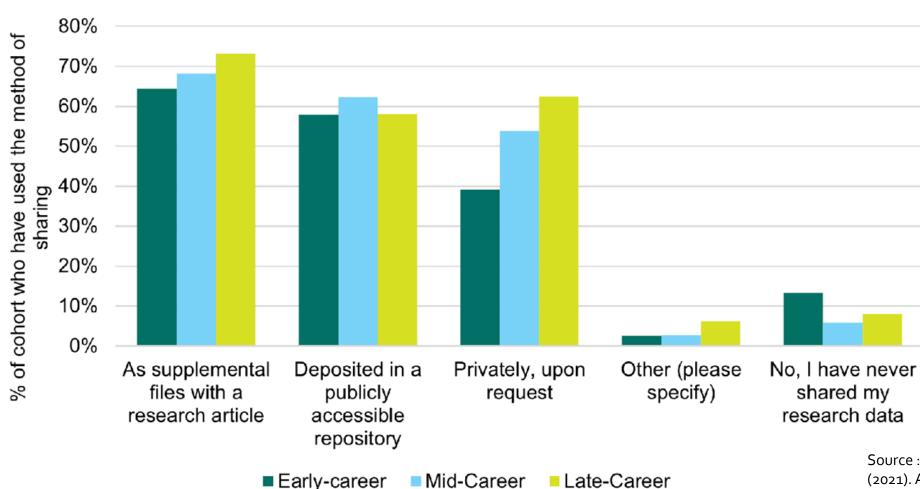


Quel mode de partage des données vous semble-t-il le plus fréquent?

- → Fichiers supplémentaires à un article scientifique
- → Dépôt dans un entrepôt de données accessible publiquement
- → Privé, sur demande
- \rightarrow Autre
- →Aucun partage

Quels avantages et quelles limites associez-vous à chacun d'entre eux?

Methods of data sharing previously used



Source: Hrynaszkiewicz, I., Harney, J., & Cadwallader, L. (2021). A Survey of Researchers' Needs and Priorities for Data Sharing. *Data Science Journal*, 20(1), 31.

- →Enquête menée par PLOS
- →1477 réponses
 - 20,9% Médecine et sciences de la santé
 - 30,4% Biologie et sciences de la vie
 - 9,7% Sciences sociales

"Sharing data as supplemental files alongside a research paper was the most common method for all career levels (67%), followed by deposition in a public repository (59%) and sharing privately on request (49%). Only 10% of respondents reported that they had never shared their research data – the largest proportion of whom (42%) work in Medicine and Health Science disciplines."

Source: Hrynaszkiewicz, I., Harney, J., & Cadwallader, L. (2021). A Survey of Researchers' Needs and Priorities for Data Sharing. *Data Science Journal*, 20(1), 31. https://doi.org/10.5334/dsj-2021-031

NB ces modes de partage ne sont pas exclusifs les uns des autres, des fichiers supplémentaires à un article peuvent être déposés dans un entrepôt de données, parfois par la revue ellemême.

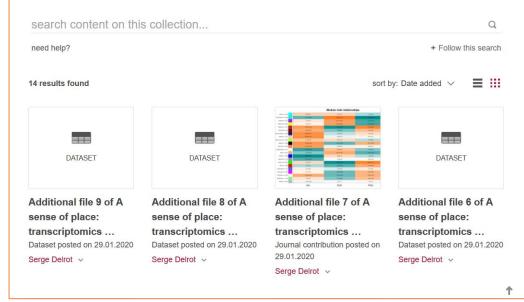
Exemple:

- → Article: Cramer, G. R., Cochetel, N., Ghan, R., Destrac-Irvine, A., & Delrot, S. (2020). A sense of place: Transcriptomics identifies environmental signatures in Cabernet Sauvignon berry skins in the late stages of ripening. BMC Plant Biology, 20(1), 41. https://doi.org/10.1186/s12870-020-2251-7
- → Fichiers déposés dans Figshare: Cramer, G. R., Cochetel, N., Ghan, R., Destrac-Irvine, A., & Delrot, S.(2020). A sense of place: Transcriptomics identifies environmental signatures in Cabernet Sauvignon berry skins in the late stages of ripening [Data set]. Figshare. https://doi.org/10.6084/mg.figshare.c.4834590.v1



Q

Search on figshare...





Log in Sign up

Enjeux



F. Flamerie - Données de recherche : entrepôts - ST - 2021-11

Enjeux

Stage	Problem	Solution	Helps to know	Agents
Data collection and analysis	Waste of research (duplication)	Preregistration	Who is doing what, how and when?	Researchers to adopt it Journals to enforce it
	Ethical concerns	Preregistration	Is the study in line with good research practices?	
	Flawed studies	Registered reports	Is the study designed appropriately to answer the question of interest?	
Publication process	Expedite reviewing	Open reviews	Are the findings verified by independent researchers?	Researchers to adopt it Journals to enforce it Institutions and funding agencies to value it Policymakers to allow it
	Distrust of published results	Data and code sharing	Can scientist obtain similar results from the same data?	
	Conflict of interests	Disclosure of conflict of interests, including editorial roles	Are the authors incentivized to publish these findings and to publish in this journal?	Researchers to adopt it Journals to enforce it
Communication	Misuse of preprints	Collaboration between journalists and scientists	Can the public trust scientific news in the media?	Journalists and news editors Institutions
	Misleading headlines, exaggerations			
	Paywalled manuscripts	Open Access on all manuscripts	The exact content of manuscripts used as a source. Can accelerate research through universal access to scientific findings	Researchers to favor it Policymakers and institutions to enforce it

Source: Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M., & Leyrat, C. (2021). Open science saves lives: Lessons from the COVID-19 pandemic. *BMC Medical Research Methodology*, 21(1), 117. https://doi.org/10.1186/s12874-021-01304-y

Enjeux - exemple politique d'éditeurs

"The Editorial team of *Behavioural Ecology and Sociobiology* and the Springer Nature Research Data team are delighted to report the implementation of a new research data sharing policy at *Behavioural Ecology and Sociobiology*, making it the first Springer journal to require that all published authors share their research data openly in appropriate data repositories."

Source : Grant, R. (2020, décembre 9). *Stronger data policies, more data sharing : Policy change at Behavioural Ecology and Sociobiology*. Research Data at Springer Nature.

http://researchdata.springernature.com/posts/stronger-data-policies-more-data-sharing-policy-change-at-behavioural-ecology-and-sociobiology

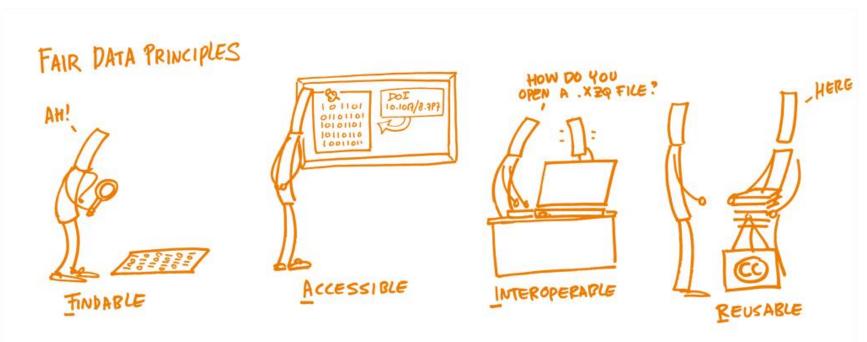
Entrepôts de données et principes FAIR

Principes FAIR -> les données doivent pouvoir être trouvées, comprises et réutilisées.

Cela ne signifie pas qu'elles doivent être nécessairement en accès entièrement libre.

Ces principes s'appliquent également aux métadonnées associées aux données.

Principes FAIR en détail -> Page web « FAIR et DMP » sur le site de la BU



Source image: The Open Science Training Handbook

Entrepôts de données et principes FAIR

DOI et citation :

Findable - Accessible

Documentation: Reusable

Une documentation plus structurée et standardisée permettrait d'augmenter les scores Interoperable - Reusable

Exemple: https://doi.org/10.5061/DRYAD.QZ612JM91

DRYAD

Explore Data | About ▼ | Help ▼

Signage interventions for stair climbing at work: more than 700,000 reasons for caution

Eves, Frank, University of Birmingham, Dhttps://orcid.org/0000-0002-4921-7169

evesff@bham.ac.uk

Publication date: October 25, 2019

Publisher: Dryad

https://doi.org/10.5061/dryad.qz612jm91

Citation

Eves, Frank (2019), Signage interventions for stair climbing at work: more than 700,000 reasons for caution, Dryad, Dataset, https://doi.org/10.5061/dryad.qz612jm91

Abstract

Increased stair climbing reduces cardiovascular disease risk. While signage interventions for workplace stair climbing offer a low-cost tool to improve population health, inconsistent effects of intervention occur. Pedestrian movement within the built environment has major effects on stair use, independent of any health initiative. This paper used pooled data from UK and Spanish workplaces to test the effects of signage interventions when pedestrian movement was controlled for in analyses. Automated counters measured stair and elevator usage at the ground floor throughout the working day. Signage interventions employed previously successful campaigns. In the UK, minute-by-minute stair/elevator choices measured effects of momentary pedestrian traffic at the choice-point (n = 426,605). In Spain, aggregated pedestrian traffic every 30 min measured effects for 'busyness' of the building (n = 293,300). Intervention effects on stair descent (3 of 4 analyses) were more frequent than effects on stair climbing, the behavior with proven health benefits (1 of 4 analyses). Any intervention effects were of small magnitude relative to the influence of pedestrian movement. Failure to control for pedestrian movement compromises any estimate for signage effectiveness. These pooled data provide limited evidence that signage interventions for stair climbing at work will enhance population health.

Observations from automated counters followed by bootstrapped regressions

Usage Notes

Read me file for UK interventions

Up from ground floor

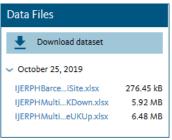
The coding used to allow analysis of the file with logistic regression is explained below. The dependent variable is stair and escalator choice. If you have any queries, please do not hesitate to contact me directly (evesff@bham.ac.uk).

For ascent, file IJERPHMultiSiteUKUp, the coding is as follows below.

INTSTAGE: 1 =baseline, 2=first intervention, 3=second intervention

BASEOINT1; the coding allows comparison of the baseline (0) with the first intervention (1).

PH2ASE1; the coding allows comparison of the first intervention (0) with the second (1).



Related Works

https://doi.org/10.3390/ijerph16193782





Licence CC: Reusable



Critères de choix d'un entrepôt de données

- → L'entrepôt est-il certifié?
 - > En savoir plus sur la certification *CoreTrustSeal* : RDA France. (2019). Entrepôts de données de confiance : Critères de conformité. Repéré à https://www.rd-alliance.org/system/files/documents/CoretrustsealFR.pdf
- → Gère-t-il différents **types d'accès**? Permet-il par exemple un accès restreint ou sous **embargo**?
 - > Y compris la gestion des demandes d'accès pour les données en accès restreint?
 - > Y compris la possibilité de définir le type d'accès fichier par fichier? [cf. exemple ICPSR ci-dessous]
- → Prend-il en charge un **format de données** particulier?
- → Quelles sont les exigences ou possibilités en termes de précision et structuration de la description des données déposées? Devrez-vous ou pourrez-vous fournir des métadonnées riches?
- → Propose-t-il d'autres services particulièrement adaptés à mes besoins? [cf. exemple connexion Zenodo-Github ci-dessous]

Dataverse gestion fine des accès



Spatiotemporal survey of major rice diseases in western Burkina Faso

Version 1.0



Barro, Mariam; Tollenaere, Charlotte; Wonni, Issa; Kassankogno, Itolou Abalo; Sereme, Drissa, 2021, "Spatiotemporal survey of major rice diseases in western Burkina Faso", https://doi.org/10.23708/8FDWIE, DataSuds, V1, UNF:6:cOi1B77EUelpAGcUzEjhog== [fileUNF]

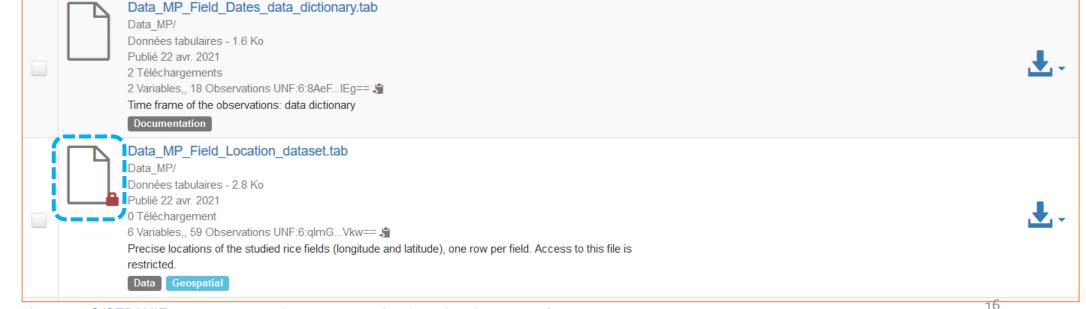
Citer le jeu de données -

Pour en savoir plus, consultez le document Data Citation Standards.

Modalités d'accès au jeu de données Contacter le Partager propriétaire Statistiques du jeu de données 🕢 27 Téléchargements 🕣

Description (

The study area is located in western Burkina Faso, in a 100x100kms region within the three provinces Houet, Kénédougou and Comoé, in the sudanian bioclimatic area. Major study sites correspond to three geographical zone (Bama, Banzon and Karfiguela), each of these zones presenting both an



Zenodo connexion GitHub



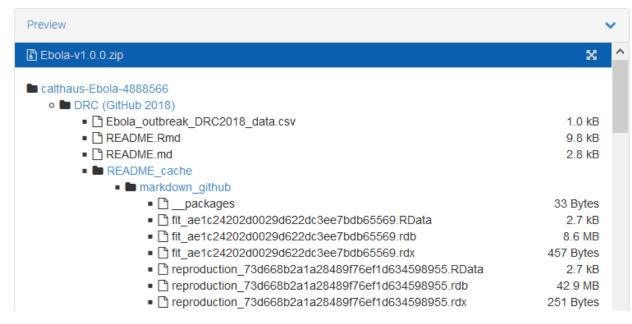
June 8, 2018 Open Access

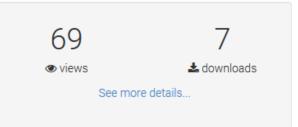
Analyses of Ebola virus disease outbreaks

(b) Christian Althaus

This release contains data and R code files from the following analyses of Ebola virus disease (EVD) outbreaks:

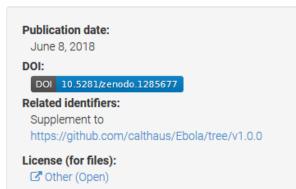
- Althaus CL. (2014) Estimating the reproduction number of Ebola virus (EVOB) during the 2014 outbreak in West Africa, PLoS Curr. 6.
- Althaus CL, Low N, Musa EO, Shuaib F, Gsteiger S. (2015) Ebola virus disease outbreak in Nigeria: Transmission dynamics and rapid control. Epidemics, 11:80-4.
- Althaus CL. (2015) Rapid drop in the reproduction number during the Ebola outbreak in the Democratic Republic of Congo. PeerJ, 3:e1418.
- 4. Althaus CL. (2015) Ebola superspreading70135-0). Lancet Infect Dis, 15:507-8.
- 5. Abbate JL, Murall CL, Richner H, Althaus CL. (2016) Potential impact of sexual transmission on Ebola virus epidemiology: Sierra Leone as a case study. PLOS Negl Trop Dis, 10:e0004676.
- 6. Althaus CL. (2018) Real-time analysis of the 2018 Ebola outbreak in the Democratic Republic of Congo). GitHub.





◆3 Log in





Exemple:

https://doi.org/10.5281/zenodo.1285676



Trouver un entrepôt de données généraliste

Outil d'aide à la décision en ligne UBx : « Trouver un entrepôt de données »

Trouver un entrepôt de données généraliste

Outil d'aide à la décision en ligne UBx : <u>Trouver un entrepôt de données</u>

- →Pour identifier l'entrepôt de données généraliste qui réponde le mieux aux besoins, à partir de la réponse aux 4 questions suivantes :
 - > Allez-vous publier des données dont l'accès doit être restreint?
 - > Recherchez-vous un entrepôt avec des options de dépôt gratuit?
 - > Souhaitez-vous pouvoir définir librement les conditions d'utilisation de vos données?
 - > Souhaitez-vous que votre dépôt de données soit relu avant d'être mis en ligne?
- →Une brève fiche descriptive accompagne chacun des huit entrepôts de données comparés : <u>4TU.ResearchData</u>, <u>Figshare</u>, <u>Harvard Dataverse</u>, <u>Dryad</u>, <u>Mendeley Data</u>, <u>Open Science Framework (OSF)</u>, <u>B2SHARE</u>, <u>Zenodo</u>

Trouver un entrepôt de données généraliste

Outil d'aide à la décision en ligne : <u>Trouver un entrepôt de données</u>

Aide au choix d'un entrepot

Trouver un entrepôt de données

Cet outil peut vous aider à trouver l'entrepôt qui réponde le spécialisé pour votre domaine de recherche, intern Pour plus d'informations concernant les données de recherch	pour rendre accessibles vos données de recherche ? mieux à vos besoins. Pour identifier un entrepôt de données ogez l'annuaire re3data ou contactez-nous. ne, consultez les pages de notre site web Les données de	
	ercne. loc-soutienrecherche@u-bordeaux.fr	
Cet outil est adapté du Data Repository Finder développé par l'université d'Utrecht. Il est en	cours de paramétrage et de finalisation, les contenus et les fonctionnalités sont susceptibles d'évoluer.	
1. Allez-vous publier des données dont l'accès doit être restreint ?	2. Recherchez-vous un entrepôt avec des options de dépôt gratuit ?	
Oui Pas nécessairement	○ Oui ○ Pas nécessairement	
3. Souhaitez-vous pouvoir définir librement les conditions d'utilisation de vos données ?	4. Souhaitez-vous que votre dépôt de données soit relu avant d'être mis en ligne	
Oui Pas nécessairement	Oui Pas nécessairement	
_		



Trouver un entrepôt de données spécialisé

Recommandations des éditeurs et des organismes de financement

Annuaire re3data

Service FAIRsharing

Recommandations

- → Editeurs : exemples <u>PLoS</u> et <u>Springer-Nature</u>
- → Organismes de financement
 - > <u>ERC European Research Council</u> : liste commentée d'entrepôts spécialisés par discipline p. 12 et suiv.
 - > <u>Commission européenne</u> : liste d'entrepôts généralistes et spécialisés par discipline ces *Data Guidelines* comportent en outre des recommandations pour la préparation des données (notamment tabulaires).

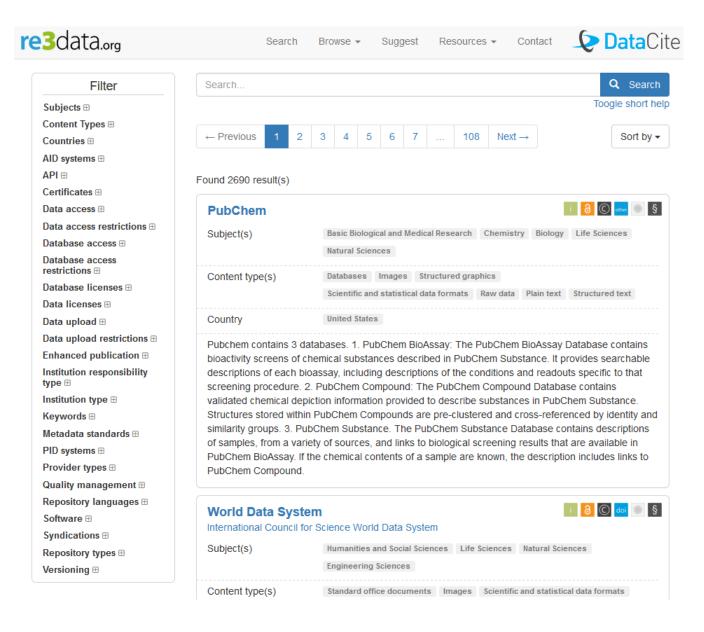
Annuaire re3data

Registry of Research Data Repositories.

https://www.re3data.org/

Il s'agit de la ressource de référence, indexant plus de 2000 entrepôts. On peut rechercher par mot-clé et naviguer dans re3data par :

- \rightarrow sujet,
- →type de contenu,
- \rightarrow pays.



Annuaire re3data

De nombreux filtres permettent ensuite d'affiner les listes de résultats, en fonction par exemple des critères suivants.

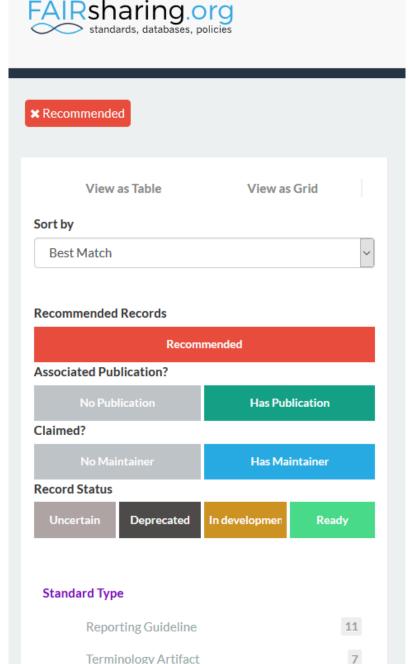
- → Certificates : par quelle certification l'entrepôt est-il qualifié?
- → Accès : gradient de valeurs de fermé à ouvert se décompose en 3 types d'accès :
 - > **Database access** : accès à l'entrepôt de données lui-même : sous quelles conditions un utilisateur peut-il accéder à la base de données en général?
 - Data access : accès aux jeux de données déposés dans un entrepôt de données spécifique : sous quelles conditions un utilisateur peut-il accéder à un jeu de données?
 - > **Data upload** : accès à la soumission de données : sous quelles conditions un utilisateur peut-il soumettre des données ?
- → **Versioning** : les jeux de données peuvent-ils être versionnés?

Service FAIRsharing

<u>FAIRsharing.org</u> répertorie non seulement des entrepôts mais également des standards, des méthodes, des vocabulaires, etc. Pour chaque ressource sont notamment spécifiés les critères suivants.

- Le statut : en développement / opérationnel / incertain / déprécié
- La recommandation : nom de l'éditeur, de la revue, etc. qui recommande

Voir: Sansone, S.-A. et al. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology*, 37(4), 358-367. https://doi.org/10.1038/s41587-019-0080-8





Merci pour votre attention

Des questions?

Contact: frederique.flamerie-de-lachapelle@u-bordeaux.fr