

Data sharing : trouver, réutiliser et citer des données de recherche : COURS

Sommaire

- [Sommaire](#)
- [1. Introduction](#)
- [2. Trouver des données de recherche avec les outils bibliographiques](#)
- [3. Trouver des données de recherche grâce aux entrepôts de données](#)
- [4. Trouver des données de recherche pour les réutiliser](#)
- [5. Citer des données de recherche](#)
- [6. Politiques de données des éditeurs et des organismes de financement de la recherche](#)
- [7. Conclusion](#)

0. Tour de table

- Dans quel laboratoire travaillez-vous?
- Quelles sont vos attentes pour le stage?

1. Introduction

Définition et typologie des données de recherche

On devrait plutôt utiliser le pluriel pour les termes "définition" et "typologie", car plusieurs approches sont possibles.

L'OCDE donne ainsi une définition très large des données de recherche.

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.

Source: Organisation de coopération et de développement économiques. (2007). Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. Organisation de coopération et de développement économiques. <https://doi.org/10.1787/9789264034020-en-fr>

On trouve aussi des définitions plus restrictives, variables en fonction des domaines scientifiques, du contexte, etc.

Ainsi le NIH (National Institute of Health) fournit-il la définition suivante dans le cadre de ses appels à projet.

"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."

Source: National Institutes of Health. (2021). NOT-OD-21-013: Final NIH Policy for Data Management and Sharing. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>

S'agissant des types de données, on retient les grandes catégories suivantes. A noter que le type de données peut aussi s'envisager du point de vue de l'achèvement.

- Données d'observation
- Données expérimentales
- Données de références
- Données de simulation numériques
- Données dérivées ou compilées

Fiche pratique d'1 page :

Bracco, L., Bouchet-Moneret, F., & Jouneau, T. (2021). Que sont les données de la recherche ? <https://doi.org/10.5281/zenodo.5495210>

Modes de partage des données de recherche

Voici quelques exemples de différents modes de partage des données de recherche.

- Répondre manuellement à des demandes.

Availability of data and materials

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

Afficher cet exemple en ligne : <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

[Explore Data](#)[About ▼](#)[Help ▼](#)[Login](#)

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

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

evesff@bham.ac.uk

Publication date: October 25, 2019

Publisher: Dryad

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

Data Files

[Download dataset](#)

> October 25, 2019

Related Works

Article

<https://doi.org/10.3390/ijerph16193782>

Metrics



25 views

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>

Afficher cet exemple en ligne : <https://doi.org/10.5061/dryad.qz612jm91>



Parmi les modes de partage des données ci-dessous, lequel vous semble-t-il le plus fréquent?

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



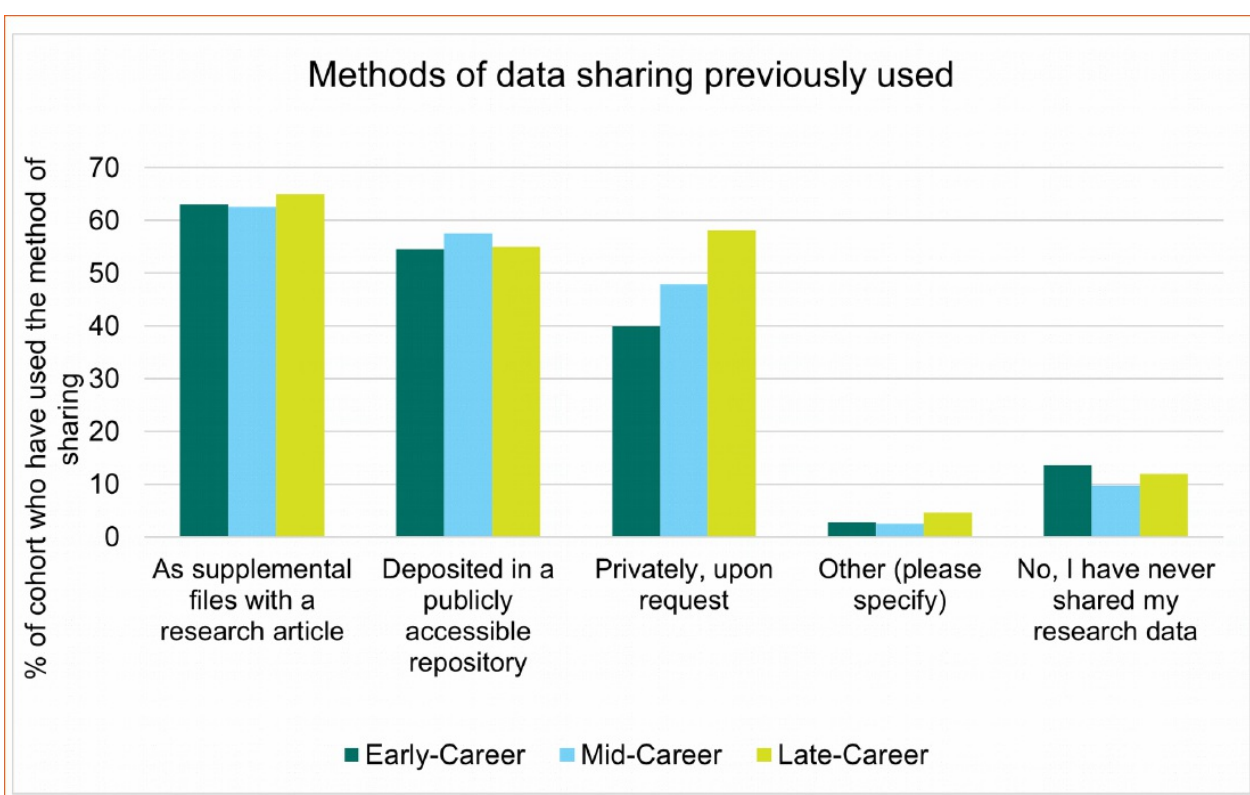
Quels avantages et quelles limites associez-vous à chacun d'entre eux?



Dans une enquête menée par l'éditeur PLOS, ce sont les **fichiers supplémentaires** qui sont les plus utilisés. Il convient toutefois de noter que les sciences humaines et sociales sont peu représentées dans les 1477 réponses reçues, qui se répartissent comme suit.

- 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>

Notez que 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 elle-même.

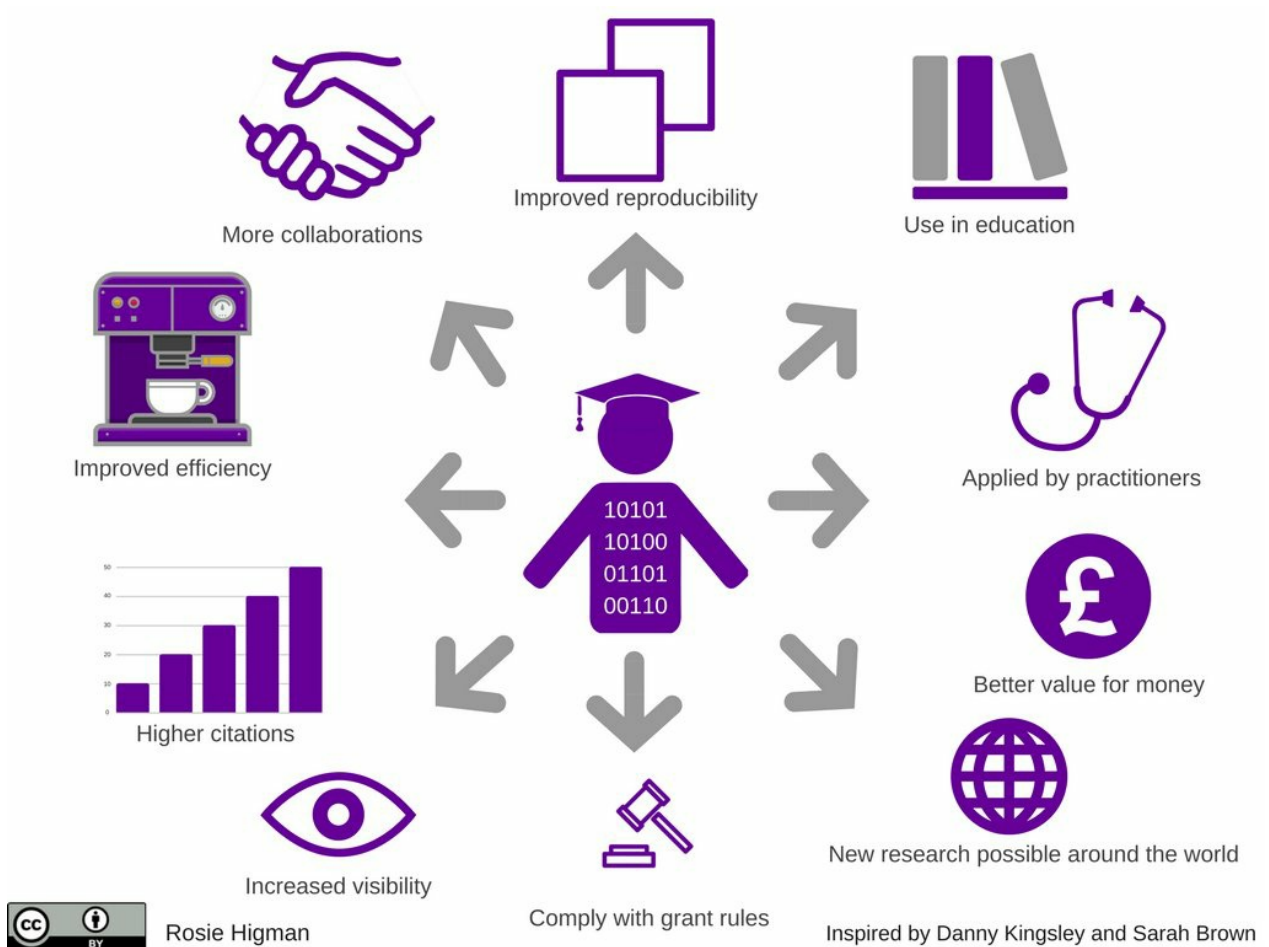
Exemple :

Article : Rousselot, N., Tombrey, T., Zongo, D., Mouillet, E., Joseph, J.-P., Gay, B., & Salmi, L. R. (2018). Development and pilot testing of a tool to assess evidence-based practice skills among French general practitioners. *BMC Medical Education*, 18(1), 254. <https://doi.org/10.1186/s12909-018-1368-y>

Fichiers déposés dans Figshare : Rousselot, N., Tombrey, T., Drissa Zongo, Mouillet, E., Jean-Philippe Joseph, Gay, B., & Salmi, L. (2018). Development and pilot testing of a tool to assess evidence-based practice skills among French general practitioners [Data set]. In Figshare. <https://doi.org/10.6084/M9.FIGSHARE.C.4297076.V1>

Enjeux

Les enjeux du partage des données sont multiples et variés, comme l'illustre la figure ci-dessous.



Source : Higman, R. (2017). Talking carrots and sticks for RDM at the #datastewards networking day, sharing the graphic on reasons to share we use in @UoMRDMSERVICE [trainingpic.twitter.com/bMsj6h57RG](https://twitter.com/bMsj6h57RG) [Tweet]. @RosieHLib. <https://twitter.com/RosieHLib/status/936587668607160320>



Par lesquels de ces enjeux vous sentez-vous concerné actuellement, compte tenu de votre sujet de recherche, du degré d'avancement de votre thèse, etc.?

Ces enjeux s'entendent également du point de vue du processus de publication.




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
	Distrust of published results	Data and code sharing	Can scientist obtain similar results from the same data?	Institutions and funding agencies to value it Policymakers to allow it
	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

Fig 4. A summary of our findings and proposed solutions

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>

Fiche pratique en 1 page :

Bracco, L., Bouchet-Moneret, F., & Jouneau, T. (2021). Pourquoi ouvrir ses données ? <https://doi.org/10.5281/zenodo.5495217>

2. Trouver des données de recherche avec les outils bibliographiques

On peut distinguer au moins 3 méthodes pour trouver des données de recherche. Dans cette partie nous abordons les 2 premières, la troisième fait l'objet de la partie suivante.

- Méthode 1 : utiliser les fonctionnalités intégrées aux **bases de données bibliographiques**, par exemple Scopus et PubMed.
- Méthode 2 : interroger les **outils de recherche spécialisés** qu'il s'agisse de moteurs de recherche comme DataCite Search ou d'index de base de données (Data Index du Web of Science).
- Méthode 3 : utiliser les **entrepôts de données, généralistes** (tels que Zenodo, Dryad ou Figshare) ou **spécialisés**.

Méthode 1, fonctionnalités intégrées aux bases bibliographiques

En plus du *data availability statement* (parfois non accessible sans abonnement à la revue), des outils intégrés aux bases de données bibliographiques permettent d'identifier les jeux de données **liés à des publications**.

Dans la nouvelle interface de **PubMed**, il s'agit de l'attribut *Associated data*, disponible sous la forme d'un filtre à partir d'une liste de résultats. Cet attribut filtre sur la présence dans l'article d'un lien vers une source de données comme [ClinicalTrials.gov](https://clinicaltrials.gov), GenBank, Figshare ou Dryad.

Dans **Scopus**, il s'agit de l'encart *Related Research Data*, accessible depuis la notice d'un article en particulier. Cette fonctionnalité est donc très limitée, puisqu'elle ne permet pas de sélectionner un ensemble d'articles ayant des données associées.

Le **type de publication** peut également être exploité comme critère de recherche ou de filtre. Il permet en effet d'identifier les articles de données, ou *data papers*. Les valeurs associées diffèrent en fonction des bases de données.

- Scopus -> choisir la valeur *data paper*
- PubMed -> choisir la valeur *dataset*

Méthode 2, les outils de recherche spécialisés

Certains des services suivants indexent uniquement des données de recherche, d'autres non.



DataCite Search

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

Fournisseur du service : l'organisation à but non lucratif DataCite - ses membres sont des organismes de recherche, des universités, des sociétés savantes, des fournisseurs de service, etc.

Couverture

DataCite Search permet de chercher parmi tous les objets auxquels un DOI a été attribué par DataCite. La couverture excède celle des données de recherche, car DataCite fournit à des organismes de recherche des DOI pour toutes leurs productions, qu'il s'agisse de données de recherche ou de thèses, de preprints, etc.

[Consulter la liste des entrepôts de données interrogeables par DataCite Search](#)

Autres caractéristiques

- Seule une recherche simple est possible.
- Les filtres disponibles sont limités à l'année d'enregistrement, au type de ressource et à l'affiliation.
- Depuis janvier 2020, DataCite Search affiche les liens (de citation, de référence, de relation) entre les objets.



Mendeley data search

URL : <https://data.mendeley.com/research-data>

Fournisseur du service : l'éditeur commercial Elsevier

Couverture

Le service de recherche de Mendeley Data interroge d'autres sources en plus de Mendeley Data.

Deux types de source sont distingués.

- *Data repositories* : entrepôts hébergeant uniquement des données
- *Article repositories* : on peut à nouveau distinguer plusieurs cas de figure.
 - Archives ouvertes institutionnelles hébergeant à la fois des publications et des données (University of British Columbia par exemple)
 - Serveur de preprints arXiv -> on trouve à la fois les figures et les tableaux insérés dans les articles et les fichiers de données supplémentaires
 - Plateforme ScienceDirect de l'éditeur Elsevier -> on trouve à la fois les figures et les tableaux insérés dans les articles et les fichiers de données supplémentaires

Autres caractéristiques

La syntaxe de recherche permet de construire des recherches avancées, elle est détaillée dans la [FAQ Mendeley Data > Data Search](#).

Exemple : liste de résultats pour la recherche sur le champ "Institution" contient "Bordeaux" :

[https://data.mendeley.com/research-data/?search=INSTITUTION\(Bordeaux\)](https://data.mendeley.com/research-data/?search=INSTITUTION(Bordeaux))



Google dataset search

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

Fournisseur du service : la société commerciale Google

Couverture

La couverture est limitée aux données, mais non aux données de recherche ; elle est indéfinie et potentiellement illimitée.

One thing hasn't changed however: anybody who publishes data can make their datasets discoverable in Dataset Search by using an open standard (schema.org) to describe the properties of their dataset on their own web page.

Noy, N. (2020, janvier 23). *Discovering millions of datasets on the web*. Google Blog. <https://blog.google/products/search/discovering-millions-datasets-web/>

Autres caractéristiques

Comme pour Google Scholar, les fonctionnalités de recherche sont très limitées.

- La syntaxe de recherche n'est pas définie.
- Les filtres sont peu nombreux.
- Lorsque le nombre de résultats excède 100, la même mention demeure affichée : "Plus de 100 ensembles de données trouvés".

La localisation est affichée sous la forme d'une carte lorsque des informations de géolocalisation sont associées aux données.



Dimensions

URL : <https://www.dimensions.ai/>

Fournisseur du service : la société commerciale Digital Science (propriétaire de Figshare, Altmetric, ReadCube, etc.)

Couverture

La couverture de Dimensions excède les données de recherche, puisque 6 types de contenu sont indexés : *publications, datasets, grants, patents, clinical trials, policy documents*.

Pour les données de recherche, seuls les objets de type *dataset* dans les sources sont indexés, ce qui exclut par exemple les vidéos, images, logiciels, etc. Les sources sont à la fois des entrepôts de données (Dryad, Pangaea, etc.) et des sites d'éditeurs, lorsque l'éditeur identifie par un DOI les **fichiers de données supplémentaires des articles**.

La restriction affichée aux seuls "jeux de données" n'est de fait pas entièrement exacte, car :

- les fichiers de données supplémentaires sont indexés à l'unité ;
- le type *dataset* n'est pas appliqué uniformément par tous les entrepôts de données. Certains peuvent l'utiliser pour des fichiers de données unitaires non organisés en un jeu de données - exemple :

Sordello, R., Flamerie, F., Livoreil, B., & Vanpeene, S. (2019). MOESM4 of Evidence of the environmental impact of noise pollution on biodiversity : A systematic map protocol [Data set]. In *Figshare*. <https://doi.org/10.6084/m9.figshare.7711562.v1>

NB d'autres fichiers supplémentaires à cet article, auxquels le type *document* a été assigné par Figshare, ne sont pas indexés par Dimensions, mais peuvent l'être par d'autres services précédemment mentionnés (Mendeley data search par exemple).

Autres caractéristiques

Un filtre *Fields of research* est disponible.

Notez que le filtre par institution n'est pas disponible dans la version gratuite ; la version gratuite permet par ailleurs d'interroger seulement 2 des 6 types de contenu : *publications* et *datasets*.



Autres services

D'autres services indexent des données et peuvent être utiles pour identifier des données existantes ; la liste suivante n'est pas exhaustive.

- [Web of Science Data Index](#) : sur abonnement - l'université de Bordeaux n'est pas abonnée à cette ressource.
- [B2FIND](#) : composante de l'infrastructure européenne EUDAT ; l'interface permet notamment d'effectuer des recherches par

couverture chronologique et géographique.

- **Isidore** : spécialisé pour les sciences humaines et sociales, le moteur de recherche d'HumNum permet de formuler des recherches ciblées sur les données, notamment grâce au critère "Par type de document" > "Données d'enquêtes", "Autres", etc.
- **Bielefeld academic search engine** : le moteur de recherche de l'université de Bielefeld permet de cibler des recherches sur les données, notamment grâce au critère "Type de document" > "Dataset", "Unknown", etc.

TP1, trouver des données avec les bases de données bibliographiques et les moteurs de recherche spécialisés



➡ Voir [Fiche TP](#)

🕒 Rendez-vous dans **10 minutes** pour une mise en commun.

3. Trouver des données de recherche grâce aux entrepôts de données

Cette méthode rejoint les astuces 2, 4 et 8 des 11 astuces recensées ci-dessous.

Gregory, K., Khalsa, S. J., Michener, W. K., Psomopoulos, F. E., Waard, A. de, & Wu, M. (2018). Eleven quick tips for finding research data. PLOS Computational Biology, 14(4), e1006038. <https://doi.org/10.1371/journal.pcbi.1006038>

Tip 1: Think about the data you need and why you need them.

Tip 2: Select the most appropriate resource.

Tip 3: Construct your query strategically.

Tip 4: Make the repository work for you.

Tip 5: Refine your search.

Tip 6: Assess data relevance and fitness -for -use.

Tip 7: Save your search and data- source details.

Tip 8: Look for data services, not just data.

Tip 9: Monitor the latest data.

Tip 10: Treat sensitive data responsibly.

Tip 11: Give back (cite and share data).

[lire l'astuce 2 sur le site de PLOS](#)

[lire l'astuce 4 sur le site de PLOS](#)

[lire l'astuce 8 sur le site de PLOS](#)

Interroger directement un entrepôt de données peut notamment permettre de bénéficier de fonctionnalités de recherche spécifiques.

👉 Cela implique au préalable d'identifier les entrepôts de données appropriés.

Entrepôts de données généralistes et spécialisés

Voici quelques exemples de jeux de données déposées dans des entrepôts de données généralistes et spécialisés.

Quelles différences observez-vous entre eux?

Exemple 1, agronomie et sciences sociales

Dépôt dans [CIRAD Dataverse](#)

Fok, M. (2021). Crédits formel et informel de trésorerie aux producteurs de coton au Bénin [Data set]. In CIRAD Dataverse.

Exemple 2, archéologie

Dépôt dans **Zenodo**

Lebrun, B., & Tribolo, C. (2020). Optically Stimulated Luminescence data from archaeological site Uzès (Gard, France) [Data set]. In Zenodo. <https://doi.org/10.5281/zenodo.4061220>

Exemple 3, psychologie

Dépôt dans **UK Data Archive ReShare**

Biggart, L. P. (2016). Emotional intelligence in social work [Data set]. In UK Data Service. <https://doi.org/10.5255/UKDA-SN-852431>

Exemple 4, écologie

Dépôt dans **ACEAS Data Portal (ACEAS = Australian Centre for Ecological Analysis and Synthesis)**

Haberle, S., Hopf, F., Tng, D., Johnston, F., & Bowman, D. (2014). Weekly pollen count data for the University of Tasmania, Hobart. [Data set]. In ACEAS. <https://doi.org/10.4227/05/5344E9A41A124>

Exemple 5, code informatique

Dépôt dans **Zenodo**

Althaus, C. (2018). Analyses of Ebola virus disease outbreaks [Data set]. In Zenodo. <https://doi.org/10.5281/zenodo.1285677>

Identifier un entrepôt de données spécialisé pertinent

La démarche pour identifier un entrepôt de données approprié pour **déposer** des données est détaillée dans le support de formation **Zenodo, Figshare, etc.: rechercher et partager des données de recherche grâce aux entrepôts de données**.

Pour un aperçu concernant les entrepôts de données **généralistes**, vous pouvez tester notre outil en ligne d'aide à la décision **Trouver un entrepôt de données**.

On se focalise ici sur la recherche d'un entrepôt de données dans la perspective de trouver des données existantes et non d'en déposer.

Aussi nous nous concentrons sur la recherche d'entrepôts **spécialisés**, pouvant offrir des fonctionnalités de recherche spécifiques, ou dont la couverture spécialisée permet de limiter le bruit. Parmi les outils mentionnés dans le support de formation précédemment cité, nous retenons l'annuaire **re3data**.

Trouver un entrepôt de données avec re3data

Registry of Research Data Repositories. <https://doi.org/10.17616/R3D> // citation simplifiée : re3data - <https://www.re3data.org/>

Il s'agit de la ressource de référence, indexant plus de 2000 entrepôts.

On peut naviguer dans re3data par :

- [sujet](#),
- [type de contenu](#),
- [pays](#).

 [Afficher la liste complète des entrepôts référencés dans 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** : plusieurs 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?

TP2, trouver un entrepôt de données spécialisé avec re3data



↔ Voir [Fiche TP](#)

🕒 Rendez-vous dans **15 minutes** pour une mise en commun.

4. Trouver des données de recherche pour les réutiliser

Nous allons aborder ici les astuces 1 et 6 de l'article "Eleven quick tips..." précédemment cité.

Tip 1: Think about the data you need and why you need them.

Tip 6: Assess data relevance and fitness -for -use.

[lire l'astuce 1 sur le site de PLOS](#)

[lire l'astuce 6 sur le site de PLOS](#)

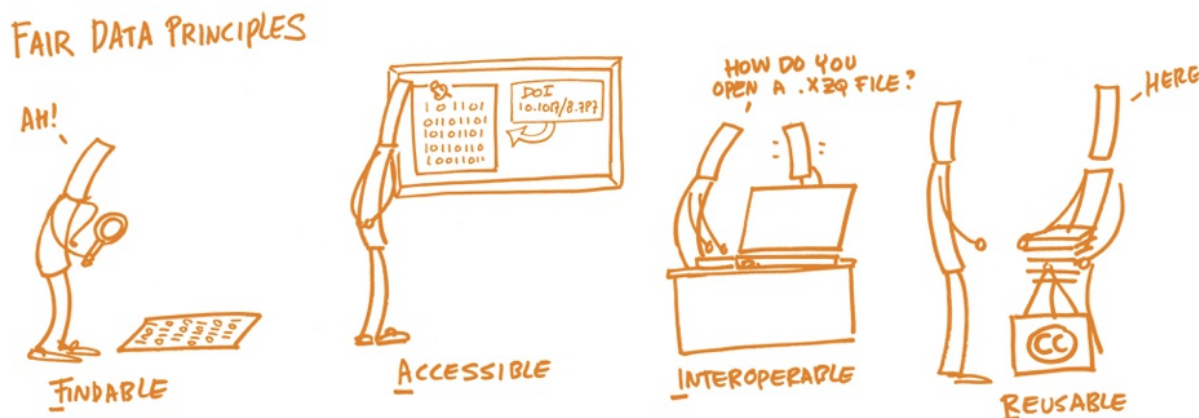
TP3, trouver et évaluer un jeu de données



↔ Voir [Fiche TP](#)

🕒 Rendez-vous dans **20 minutes** pour une mise en commun.

Focus sur les principes FAIR



Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P. L., Görögh, E., Helbig, K., Kramer, B., Labastida, I., Niemeyer, K., Psomopoulos, F., Ross-Hellauer, T., Schneider, R., Tennant, J., Verbakel, E., Brinken, H., & Heller, L. (2018). Open Science Training Handbook. FOSTER Plus Consortium. <https://book.fosteropenscience.eu/>

Les principes FAIR visent à assurer que les données sont partagées d'une manière qui permette et améliore leur réutilisation, à la fois par des humains et par des machines. A chaque principe sont associées des caractéristiques qui en précisent et spécifient l'application.

Les principes FAIR ont été développés initialement au sein [du groupement FORCE11](#), the Future of Research Communication and e-Scholarship, et présentés dans un article paru en 2016 dans *Scientific Data*.

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

Source : Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Santos, L. B. da S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, sdata201618. <https://doi.org/10.1038/sdata.2016.18>

Comprendre et appliquer les principes FAIR

L'infographie interactive ci-dessous vous permet d'explorer chacun des principes FAIR.

Urfist Méditerranée. (2019). Les principes FAIR [DoRANum]. <https://doi.org/10.13143/z7s6-ed26>

Aperçu statique de l'infographie FAIR

Les principes FAIR

Les chercheurs s'appuient sur les connaissances scientifiques antérieures, notamment sur les résultats publiés dans les articles scientifiques. La reproductibilité des résultats, ainsi que leur croisement, ne sont cependant envisageables qu'avec des données originelles et leurs conditions d'obtention. C'est pourquoi la science ouverte vise à faciliter l'accès aux publications scientifiques et aux données de la recherche. Cette facilitation s'accompagne d'un certain nombre de mesures pour rendre les données scientifiques facilement découvrables, accessibles, interopérables et réutilisables. Ce sont les principes FAIR : Findable, Accessible, Interoperable, Reusable.

URFIST

CC BY-NC-SA

Références ...

Le site de l'INRAE consacre des pages détaillées et en français aux principes FAIR, y compris des recommandations et des questions à se poser lors de la rédaction d'un plan de gestion de données. Ces pages sont accessibles à partir de la page suivante.

IST-Données de la Recherche. (2018, août). Produire des données FAIR. Datapartage. <https://datapartage.inrae.fr/Produire-des-donnees-FAIR>

Le Fonds national suisse de la recherche scientifique propose sous la forme d'un tableau récapitulatif une analyse pratique en 3 rubriques : *In other words, Researcher's responsibility, Requirements to be fulfilled by the repository.*

Fonds national suisse de la recherche scientifique. (2017). Explanation of the FAIR data principles. http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf

Données FAIR et ouvertes?

L'application des principes FAIR aux données implique que ces dernières doivent pouvoir être trouvées, comprises et réutilisées. Cela ne signifie pas qu'elles doivent être nécessairement en accès entièrement libre.

Ainsi des données peuvent être **FAIR mais non ouvertes**, si elles appliquent les principes FAIR, mais sont soumises à des restrictions d'accès.

Des données peuvent être **ouvertes mais non FAIR**, si elles sont mises à disposition publiquement, mais sont dépourvues d'une documentation suffisante ou d'une licence précisant clairement les conditions de réutilisation.

5. Citer des données de recherche

Cela correspond à une partie de l'astuce 11 de l'article "Eleven quick tips..." cité précédemment.

Tip 11: Give back (cite and share data).

[lire l'astuce 11 sur le site de PLOS](#)

Citation en bibliographie

De plus en plus d'éditeurs incitent les auteurs à citer les jeux de données au même titre que les articles dans leur bibliographie, dans une rubrique dédiée ou dans la bibliographie principale.

Scientific Data requiert ainsi que les données soient citées dans la bibliographie principale des articles ; la politique de citation des données de cette revue est détaillée dans l'éditorial suivant de 2019.

Data citation needed. (2019). Scientific Data, 6(1), 27. <https://doi.org/10.1038/s41597-019-0026-5>

Exemple de données citées dans la bibliographie principale, référence n° 21 :

Power, A. M., Merder, J., Browne, P., Freund, J. A., Fullbrook, L., Graham, C., Kennedy, R. J., O'Carroll, J. P. J., Wieczorek, A. M., & Johnson, M. P. (2019). Field-recorded data on habitat, density, growth and movement of *Nephrops norvegicus*. Scientific Data, 6(1), 7. <https://doi.org/10.1038/s41597-019-0013-x>

19. Merder, J. *et al.* Density-dependent growth in 'catch-and-wait' fisheries has implications for fisheries management and Marine Protected Areas. *Ambio*. <https://doi.org/10.1007/s13280-019-01158-1> (2019).

[Article](#) [PubMed](#) [Google Scholar](#)

20. Haynes, P. S. *et al.* Growth in *Nephrops norvegicus* from a tag-recapture experiment. *Sci. Rep.* **6**, 35143, <https://doi.org/10.1038/srep35143> (2016).

[ADS](#) [CAS](#) [Article](#) [PubMed](#) [PubMed Central](#) [Google Scholar](#)

21. Power, A. M. *et al.* Field-recorded data on habitat, density, growth and movement of *Nephrops norvegicus*. *figshare*. <https://doi.org/10.6084/m9.figshare.c.4249163>  (2019).

Comment citer un jeu de données?

Avec un logiciel de gestion bibliographique, *Zotero*

- Enregistrer les jeux de données sous le type de document "Article de revue"
- Spécifier : `type: dataset` dans le champ `Extra`

Info	Notes	Marqueurs	Connexe
Clé de citation: 2016_rusch			
Type de document	Article de revue		
Titre	Data from: Grape moth density in Bordeaux vineyards depends on local habitat management despite effects of landscape heterogeneity on their biological control		
Auteur	Rusch, Adrien		
(...) Résumé	1. Biological control of crop pests is a major ecosystem service affected by several variables acting at multiple spatial s...		
Publication	Dryad Digital Repository		
Volume			
Numéro			
Pages			
Date	2016-12-25		
Collection			
Titre de la coll.			
Texte de la coll.			
Abrév. de revue			
Langue			
DOI	10.5061/dryad.85096		
ISSN	0021-8901		
Titre abrégé			
URL			
Consulté le	15/11/2017 à 10:20:15		
Archive	Dryad Digital Repository		
Loc. dans l'archive			
Catalogue de bibl.	datadryad.org		
Cote			
Autorisations			
Extra	type: dataset		
Date d'ajout	15/11/2017 à 10:20:15		

- Choisir un style bibliographique qui prend en charge le type de document "Dataset", par exemple [APA](#) ou Vancouver

Avec l'outil *Citation Formatter* fourni par DataCite

Ce service en ligne est accessible à l'adresse : <https://citation.crosscite.org/>

⚠ A vérifier, non fonctionnel le 2022-01-07

A partir d'un DOI, puis du choix d'un style bibliographique et de la langue, il génère automatiquement une citation correctement mise en forme.

Data availability statement

La bibliographie ou la liste des références citées n'est pas la seule rubrique d'un article dans laquelle des données de recherche peuvent ou doivent être mentionnées.

Le *data availability statement* ou *data access statement* décrit où les données sous-jacentes à l'article peuvent être trouvées, et sous quelles conditions elles sont accessibles.

L'université de Manchester propose plusieurs exemples de formulation pour prendre en compte les cas suivants :

- données en accès ouvert,
- analyse secondaire de données existantes,
- contraintes éthiques,
- contraintes commerciales,
- contraintes de coût,
- données originales dans un format propriétaire,
- données non numériques,
- absence de création de données nouvelles,

- données disponibles de façon ouverte.

The University of Manchester Library. Data access statements. <https://www.library.manchester.ac.uk/using-the-library/staff/research/research-data-management/sharing/data-access-statements/>

Voir aussi les exemples par discipline mis à disposition par l'université de Bristol.

University of Bristol. Sharing research data. <http://www.bristol.ac.uk/staff/researchers/data/sharing-research-data/>

L'ordre et la nature des informations à mentionner peuvent varier en fonction des éditeurs, voir par exemple les recommandations de Springer Nature, incluant des liens vers des exemples dans des articles publiés.

Springer Nature. Data availability statements. <http://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880>

6. Politiques de données des éditeurs et des organismes de financement de la recherche



Nous l'avons mentionné plus haut de façon éparse, de plus en plus d'éditeurs scientifiques et d'organismes de financement de la recherche adoptent des politiques de données, le partage des données étant l'une des composantes de politiques couvrant un spectre plus large d'actions et de recommandations.

Editeurs

PLOS a été l'un des premiers éditeurs à définir une politique de données en 2014, obligeant notamment au dépôt en accès ouvert des données sous-jacentes aux articles.

L'**accès aux données sous-jacentes** aux articles et la **citation des données** constituent les deux points les plus évidents concernant le partage des données.

Une politique de données et plus largement s'entend toutefois plus largement et peut recouvrir bien plus d'aspects, comme le montre le tableau ci-dessous, qui représentent 14 caractéristiques, réparties en 6 niveaux de politique.

14 journal research data policy features arranged as 6 policy types (tiers)

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exceptions to policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Embargoes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Supplementary materials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data repositories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data citation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data licensing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Researcher/ author support	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data availability statements		<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data formats and standards				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (specific data types)				<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (all papers)				<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Peer review of data				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Data Management Plans (DMPs)				<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

☐ Provide information

The text for the policy feature will be included in the policy template but it is clear that the feature will not be enforced and checked as part of the publishing or peer review process

☒ Provide information and action

The text of the policy feature is included and makes clear where applicable that the feature will be checked and enforced in the publishing or peer-review process

Source : Hrynaszkiewicz, I., Simons, N., Hussain, A., Grant, R., & Goudie, S. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. Data Science Journal, 19(1), 5. <https://doi.org/10.5334/dsj-2020-005>

Si on considère plus largement le processus de publication, les données ne sont elles-mêmes qu'une constituante d'une démarche globale d'ouverture et de transparence. L'initiative **Transparency and Openness Promotion (TOP)** s'inscrit dans cette perspective, comme l'illustre le tableau synthétique des critères pris en compte et leur degré d'application.

TOP Guidelines summary table.

	Not Implemented	Level I	Level II	Level III
Citation Standards	No mention of data citation.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used consistent with journal's author guidelines.	Article is not published until providing appropriate citation for data and materials following journal's author guidelines.
Data Transparency	Journal encourages data sharing, or says nothing.	Article states whether data are available, and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Analytic Methods (Code) Transparency	Journal encourages code sharing, or says nothing.	Article states whether code is available, and, if so, where to access it.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Research Materials Transparency	Journal encourages materials sharing, or says nothing.	Article states whether materials are available, and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.
Design and Analysis Transparency	Journal encourages design and analysis transparency, or says nothing.	Journal articulates design transparency standards.	Journal requires adherence to design transparency standards for review and publication.	Journal requires and enforces adherence to design transparency standards for review and publication.
Study Preregistration	Journal says nothing.	Article states whether preregistration of study exists, and, if so, where to access it.	Article states whether preregistration of study exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies and provides link and badge in article to meeting requirements.
Analysis Plan Preregistration	Journal says nothing.	Article states whether preregistration of study with analysis plan exists, and, if so, where to access it.	Article states whether preregistration with analysis plan exists, and, if so, allows journal access during peer review for verification.	Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.
Replication	Journal discourages submission of replication studies, or says nothing.	Journal encourages submission of replication studies.	Journal encourages submission of replication studies and conducts results blind review.	Journal uses Registered Reports as a submission option for replication studies with peer review prior to observing the study outcomes.

More information available at cos.io/top

[Afficher le tableau sur le site TOP](#)

Notez que cette démarche n'est pas propre à un domaine disciplinaire. Les 16 revues recensées par TOP comme atteignant le degré 3 relèvent ainsi de domaines différents.

Voir [la rubrique "Data Reproducibility Policies" sur le site de TOP](#) :

The following **16** journals use shared data and code to reproduce the computational findings reported in articles prior to publication. **This represents level 3 of the TOP Guidelines for Data, Materials, and Code.**

- The nine journals of the [American Economic Association](#)
 - Author instructions and guidance available [here](#)
- [American Journal of Political Science](#)
- [Biostatistics](#)
- [Journal of Experimental Political Science](#)
- [Journal of Peace Research](#)
- [Meta-Psychology](#)
- [Organic Syntheses](#)
- [Quarterly Journal of Political Science](#)

TOP attribue un "**TOP factor**" aux revues, en se fondant sur le tableau présenté ci-dessus. En 2021, plus de 900 revues ont reçu un "TOP factor".



- Retrouvez-vous un titre que vous connaissez dans la liste des revues évaluées?
- Son score vous surprend-il? Pourquoi?

Organismes de financement

Deux aspects constituent des éléments centraux dans la politique des organismes de financement.

- La recherche de la conformité aux **principes FAIR** ; l'approche "*as open as possible, as close as necessary*" de l'Union européenne est ainsi reprise par l'ANR.
- La mise en oeuvre de bonnes pratiques de gestion des données, décrites dans un **plan de gestion de données** (DMP - data management plan).

Pour une présentation générale de la politique de science ouverte de l'Union européenne, voir le livret ci-dessous, duquel est extrait la figure concernant les impacts et résultats attendus.

Union européenne. (2021). Horizon Europe, science ouverte: Partage précoce des connaissances et des données, et collaboration ouverte. Office des publications de l'Union européenne. <https://doi.org/10.2777/838069>

Résultats et impacts attendus



Les obligations et recommandations concernant la gestion des données de recherche sont détaillées à la diapositive 8 de la présentation suivante.

López de San Román, A. (2021, avril 12). Open Science in Horizon Europe. LERU CE7 Seminar. <https://doi.org/10.5281/zenodo.4681073>

La politique de science ouverte de l'**ANR** est détaillée sur la page "[La science ouverte](#)" de son site internet .

7. Conclusion

Reprenons la conclusion de l'article

Gregory, K., Khalsa, S. J., Michener, W. K., Psomopoulos, F. E., Waard, A. de, & Wu, M. (2018). Eleven quick tips for finding research data. *PLOS Computational Biology*, 14(4), e1006038. <https://doi.org/10.1371/journal.pcbi.1006038>

Regardless of whether you are acting as a data seeker or a data creator, remember that 'data discovery and reuse are most easily accomplished when: (1) data are logically and clearly organized; (2) data quality is assured; (3) data are preserved and discoverable via an open data repository; (4) data are accompanied by comprehensive metadata; (5)

algorithms and code used to create data products are readily available; (6) data products can be uniquely identified and associated with specific data originator(s); and (7) the data originator(s) or data repository have provided recommendations for citation of the data product(s)'

Crédits



Ce document est mis à disposition selon les termes de la [Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 3.0 France](#).

Auteur

Frédérique Flamerie



orcid.org/0000-0001-6014-0134

Icônes

Bezjak, S., Clyburne-Sherin, A., Conzett, P., Fernandes, P. L., Görögh, E., Helbig, K., Kramer, B., Labastida, I., Niemeyer, K., Psomopoulos, F., Ross-Hellauer, T., Schneider, R., Tennant, J., Verbakel, E., Brinken, H., & Heller, L. (2018). *Open Science Training Handbook*. FOSTER Plus Consortium. <https://book.fosteropenscience.eu/>
[CC-Zero]