

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

Sommaire

- [Sommaire](#)
- [1. Introduction](#)
- [2. Critères de choix d'un entrepôt de données](#)
- [3. Trouver un entrepôt de données généraliste](#)
- [4. Trouver un entrepôt de données spécialisé](#)
- [5. Trouver des jeux de données existants](#)
- [6. Citer les jeux de données](#)
- [7. Conclusion](#)

1. Introduction

Que sont pour vous les "données de recherche" et comment les partagez-vous?



Nous prenons quelques minutes pour préparer ce tour de table.

Pouvez-vous réfléchir à propos du dernier article que vous avez lu/écrit : quel était le matériel supplémentaire (tableaux, figures, etc.)?

Pouvez-vous noter des **exemples** et des types de données de recherche pertinents pour votre domaine de recherche?

Quels **modes de partage** des fichiers de données sont-ils utilisés dans votre domaine?

🔊 Si vous n'avez pas de micro ou ne souhaitez pas prendre la parole, répondez dans la conversation Zoom.

Modes de partage des données de recherche

Méthode	Avantage	Inconvénient
Publier les données sur un site web (d'un projet, d'un laboratoire, etc.)	Les données sont facilement accessibles pour les autres Dissémination plus large de la recherche	Requiert une maintenance de la part du groupe de recherche Pas de contrôle sur qui accède aux données Incapacité à attribuer un DOI ou un autre identifiant pérenne au jeu de données
Soumettre les données à une revue ou publier un article de données ou <i>data paper</i>	Les données sont associées à l'article publié au sujet de ces données Les données sont partagées avec les pairs	Les données peuvent ne pas donner lieu à un article mais devoir quand même être partagées En fonction de la revue, l'accès peut être restreint aux chercheurs bénéficiant d'un abonnement à la revue
Répondre aux demandes de données	Conserver le contrôle sur qui utilise les données	Accès très limité aux données Peut ne pas être une méthode acceptable pour les agences de financement Peut être chronophage
Déposer les données dans un entrepôt de données	En fonction des entrepôts, accès paramétrable : fermé >> restreint >> sous embargo >> ouvert Ne requiert aucune maintenance par le groupe de recherche Capacité à attribuer un DOI ou un autre identifiant pérenne au jeu de données Les données sont partagées avec les pairs	Les limitations peuvent varier en fonction des entrepôts : soumission limitée à certains utilisateurs, soumission payante, nécessité de respecter un format de données et/ou de métadonnées, etc.

Adapté de : Llebot, C. Research Data Services : Sharing Your Data. *Oregon State University Libraries*. Repéré à <https://guides.library.oregonstate.edu/research-data-services/data-sharing>

Terminologie

Données de recherche

Data

Data in the sense used here are all digitally available objects (simple or complex) that emerge or are the result of the research process.

On entend ici par données tous les objets (simples ou complexes) disponibles sous forme numérique qui émergent ou sont le résultat du processus de recherche.

Entrepôt de données

Repository

Repository is defined as the infrastructure and corresponding service that allows for the persistent, efficient and sustainable storage of digital objects (such as documents, data and code).

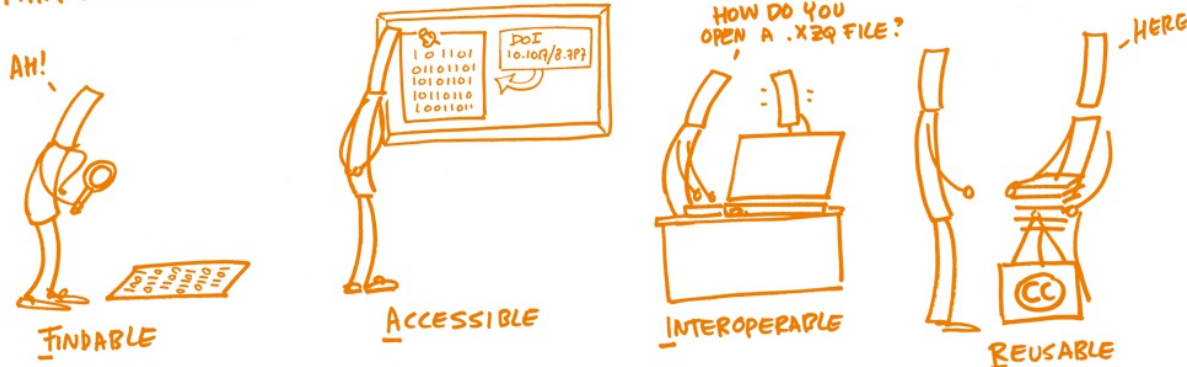
Un entrepôt est défini comme l'infrastructure et le service correspondant qui permet le stockage persistant, fiable et durable des objets numériques (tels que les documents, les données et le code).

Ces deux définitions sont tirées du [glossaire](#) du livre suivant :

Bezjak, S. et al. (2018). *Open Science Training Handbook*. Hannover : FOSTER Plus Consortium. Repéré à <https://book.fosteropenscience.eu/>

2. Critères de choix d'un entrepôt de données

FAIR DATA PRINCIPLES



En fonction du contexte, les critères de choix peuvent différer et la démarche d'identification et de choix d'un

entrepôt de données suivre différentes étapes.

Exemples de jeux de données dans des entrepôts de données

Quels critères pouvons-nous inférer des exemples ci-dessous?

Exemple 1

Dépôt dans **Open Science Framework**

Reynolds, N., & Green, C. (2019). Spatiotemporal modelling of radiocarbon dates using linear regression does not indicate a vector of demic dispersal associated with the earliest Gravettian assemblages in Europe [Data set]. *Open Science Framework*. <https://doi.org/10.17605/OSF.IO/6XRTS>

🔗 Il s'agit des données liées à l'article suivant :

Reynolds, N., & Green, C. (2019). Spatiotemporal modelling of radiocarbon dates using linear regression does not indicate a vector of demic dispersal associated with the earliest Gravettian assemblages in Europe. *Journal of Archaeological Science: Reports*, 27, 101958. <https://doi.org/10.1016/j.jasrep.2019.101958>

Depuis la page de cet article sur ScienceDirect, pouvez-vous accéder à ces données dans OSF? Comment?

Exemple 2

Dépôt dans **Zenodo**

Revers, N., Aline, R., & Vandekerckhove, R. (2017). Dutch Audio Description Corpus [Data set]. *Zenodo*. <https://doi.org/10.5281/zenodo.1035175>

Exemple 3 en spectrométrie

Dépôt dans **4TU.Centre for Research Data**

Plomp, E., Von Holstein, I., Koornneef, J., & Davies, G. (2019). Neodymium and strontium isotope analysis of modern human dental enamel using Thermal Ionization Mass Spectrometry (TIMS) [Data set]. *4TU.Centre for Research Data*. <https://doi.org/10.4121/uuid:d541a402-2701-47b2-ac6a-eaaa14c8c111>

Exemple 3 en psychologie

Dépôt dans **UK Data Archive ReShare**

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

Exemple 4 en écologie

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

Haberle, S. et al. (2014). Weekly pollen count data for the University of Tasmania, Hobart [Data set]. *ACEAS*. <https://doi.org/10.4227/05/5344E9A41A124>

Exemple 4 en linguistique

Dépôt dans **TROLLing** (TROLLing = The Tromsø Repository of Language and Linguistics)

Rainsford, T. (2020). Table of transcribed forms for: Syllable Structure and Prosodic Words in Old French [Data set]. *The Tromsø Repository of Language and Linguistics (TROLLing)*. <https://doi.org/10.18710/ymxpc>

Identifier et hiérarchiser vos critères

Voici quelques exemples de critère qui peuvent être pris en compte.

Le renvoi vers les exemples n'est pas exhaustif

- 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.rda-alliance.org/system/files/documents/CoretrustsealFR.pdf>

🔗 exemples 4 en linguistique, 3 en spectrométrie

- L'entrepôt gère-t-il différents **types d'accès**? Permet-il par exemple un accès restreint ou sous **embargo**?

🔗 exemple 2

- L'entrepôt prend-il en charge un **format de données** particulier?

📁 exemple 3 en spectométrie

- 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?

📁 exemples 4

Comme nous l'avons vu dans les exemples, les entrepôts de données fournissent des services variables et peuvent avoir différents niveaux d'exigence, concernant par exemple la description des données déposées. Cette description peut rester très sommaire ou au contraire devoir être très complète et structurée.

Critères et ressources complémentaires

Suivre les pratiques de votre communauté

Privilégier un entrepôt spécialisé pour votre discipline, ou l'entrepôt généraliste utilisé par votre communauté.

Suivre les recommandations ou obligations

Ces recommandations ou ces obligations peuvent émaner de différentes parties prenantes.

D'un organisme de financement

- [ERC - European Research Council](#) : liste commentée d'entrepôts spécialisés par discipline - p. 9 et suiv.
- [Commission européenne](#) : liste d'entrepôts généralistes et spécialisés par discipline - ces *Data Guidelines* comportent en outre des recommandations pour la préparation des données (notamment tabulaires).
- [Fonds national suisse pour la recherche scientifique](#) : liste d'entrepôts généralistes et spécialisés par discipline fréquemment utilisés par la communauté scientifique suisse et qui remplissent les critères Open Research Data (ORD) du FNS.

D'un éditeur

- [PLOS](#) : liste d'entrepôts généralistes et spécialisés par discipline (peu adapté pour les SHS)
- [Springer-Nature](#) : liste d'entrepôts généralistes et spécialisés par discipline (pas de rubriques *Humanities*, rubrique *Social sciences* plus étoffée que celle de PLoS)
- [Cambridge University Press](#) : sélection d'entrepôts généralistes, renvoi vers politique de chaque revue

Ressources complémentaires

DoRANum. (2018). Du choix de l'entrepôt au dépôt des données. *DoRANum*. Repéré à <https://doranum.fr/depot-entrepots/choix-entrepot-depot-donnees/>

*Cette infographie interactive guide la démarche de choix d'un entrepôt en regroupant les critères en trois sous-ensembles de caractéristiques, celles liées **aux données, au partage et à l'entrepôt**.*

Whyte, A. (2016). *Where to keep research data* (Rapport No. v. 1.1). Edinburgh : Digital Curation Centre. Repéré à <http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data/where-keep-research-data#3>

Pour chacune des questions, le guide définit 3 niveaux de capacité de service. Il met en avant également des points à considérer pour chacune des questions.

The checklist that follows addresses the five key questions posed in this guide:

1. is the repository reputable?
2. will it take the data you want to deposit?
3. will it be safe in legal terms?
4. will the repository sustain the data value?
5. will it support analysis and track data usage?

3. Trouver un entrepôt de données généraliste

Vous souhaitez en savoir plus concernant un entrepôt de données généraliste recommandé par une revue ou un financeur?
Vous souhaitez avoir une vue plus complète concernant les entrepôts de données généraliste?

L'outil en ligne d'aide à la décision [Trouver un entrepôt de données](#) vous aide à identifier l'entrepôt de données généraliste qui répond le mieux à vos 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 : Figsare, B2SHARE, et Harvard Dataverse.

• Entrepôts de données généralistes utilisés par ce service

- [4TU.ResearchData](#)
- [Figshare](#)
- [Harvard Dataverse](#)
- [Dryad](#)
- [Mendeley Data](#)
- [Open Science Framework \(OSF\)](#)
- [B2SHARE](#)
- [Zenodo](#)



Nous prenons quelques minutes pour que vous puissiez tester cet outil.

4. Trouver un entrepôt de données spécialisé

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** : 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?



Nous prenons quelques minutes pour la recherche suivante.

En filtrant la recherche sur les entrepôts intégrant la fonctionnalité d'**attribution de DOI**, trouvez-vous un entrepôt spécialisé dans votre domaine?

- 🔊 Si vous n'avez pas de micro ou ne souhaitez pas prendre la parole, répondez dans la conversation Zoom.

Outils complémentaires

Des services d'aide au choix et à la décision ont été développés, principalement à partir des données de re3data, et proposent des fonctionnalités supplémentaires.

FAIRsharing



[FAIRsharing.org](https://fairsharing.org) 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

En savoir plus concernant FAIRsharing :

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>

Data Deposit Recommendation Service



[Data Deposit Recommendation Service](#) est développé par l'infrastructure [DARIAH](#) et spécialisé pour les **sciences humaines**. Il utilise les données de re3data. 2 critères peuvent être spécifiés:

- le pays - > choisir *European Union* pour obtenir des réponses satisfaisantes,
- la discipline.

Pour certains entrepôts, il est possible de soumettre une demande de dépôt à partir de la fiche de cet entrepôt, par exemple [CLARIN-ERIC](#).

Repository Finder



[Repository Finder \(DataCite\)](#) est développé dans le cadre du [projet Enabling FAIR Data](#), porté par la *Coalition for Publishing Data in the Earth and Space Sciences*. Il permet d'identifier un entrepôt référencé dans re3data ou parmi les 208 entrepôts sélectionnés par le projet.

Ce service diffère des 2 précédents dans la mesure où on ne spécifie pas de critères de choix.

5. Trouver des jeux de données existants

Interroger directement un entrepôt de données

Il s'agit de l'une des approches possibles : interroger directement un entrepôt disciplinaire de son domaine de recherche, ce qui peut permettre de bénéficier de fonctionnalités de recherche spécifiques.

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

Gregory, K. et al. (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).

Trouver des jeux de données en lien avec les publications

En plus du *data availability statement* (parfois non accessible sans abonnement à la revue), des outils intégrés aux bases de données bibliographiques sont disponibles.

Dans le nouveau **PubMed**, il s'agit de l'attribut *Associated data*, disponible sous la forme d'un filtre à partir d'une liste de résultats.

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*

Les moteurs de recherche spécialisés

Le principal moteur de recherche spécialisé est [DataCite Search](#).

■ Vous trouverez une présentation détaillée de DataCite et d'autres moteurs spécialisés dans le support de la formation [Data sharing : trouver, réutiliser et citer des données de recherche](#) (formation du Collège des écoles doctorales de l'université de Bordeaux).

6. Citer les jeux de données

Où les données sont-elles citées?

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.

Voir par exemple pour une politique d'éditeur :

STM Publishing News. (2017, 14 septembre). Wiley announces new Data Sharing and Citation policies to improve transparency in research. *STM Publishing News*. Repéré à <http://www.stm-publishing.com/wiley-announces-new-data-sharing-and-citation-policies-to-improve-transparency-in-research/>

La politique de citation des données dans *Scientific Data* est détaillée dans cet éditorial :

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

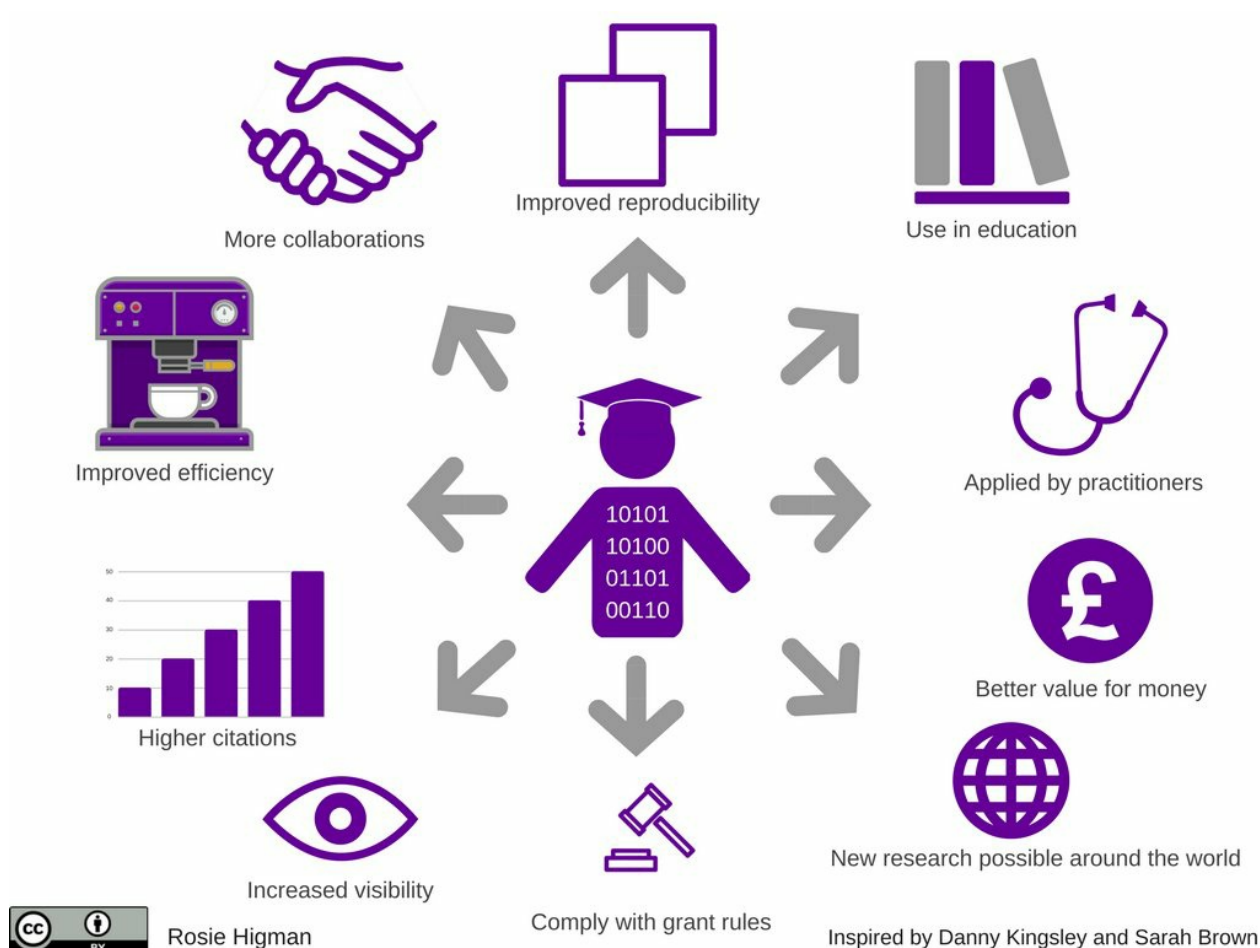
7. Conclusion

Reprenons la conclusion de l'article

Gregory, K. et al. (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)'

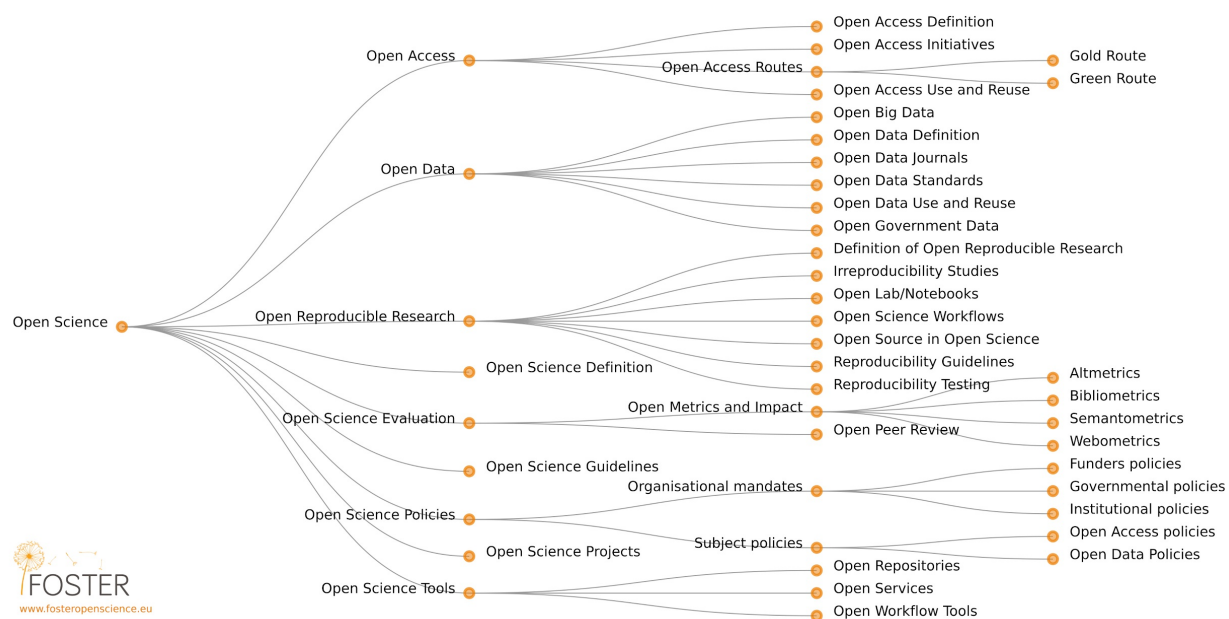
Les enjeux en image



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/bMJs6h57RG](https://twitter.com/bMJs6h57RG) [Tweet]. @RosieHLib. Repéré à <https://twitter.com/RosieHLib/status/936587668607160320>

La grande image de la science ouverte

Open Science Taxonomy



Source : Bezjak, S. et al. (2018). *Open Science Training Handbook*. Hannover : FOSTER Plus Consortium. Repéré à

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

Visuels et icônes

Bezjak, S. et al. (2018). *Open Science Training Handbook*. Hannover : FOSTER Plus Consortium. Repéré à <https://book.fosteropenscience.eu/> - [CC-Zero]