



Comment bien diffuser ses données à l'issue de sa thèse

Formation doctorale 13/12/2024



- Présentez vous en *quelques mots* :
 - Sujet ?
 - Début ou fin de thèse ?
 - Projet de publication et/ou de diffusion des données ?
 - Attentes/questions par rapport à la formation ?
 - Qui a participé
 - aux deux journées sur la gestion des données ?
 - à la formation RGPD



Pourquoi diffuser ?

Que diffuser ?

Comment diffuser ?

- *A quelles conditions d'accès ?*
- *Licences*

Quand diffuser ?

Où diffuser ?

- *Entrepôts et critères de sélection*
- *Supplementary materials et datapapers*

Aide et ressources



- **Processus de publication :**

1 - Article scientifique : principaux résultats d'un travail scientifique

2 - Diffusion des données : reproductibilité – réutilisabilité- citabilité

3 - Diffusion des codes : reproductibilité – réutilisabilité - citabilité

4 - Diffusion thèses : sites institutionnels dédiés - plus de détails scientifiques et techniques

Pourquoi diffuser ?



- Pourquoi ne pas partager ?
 - Données **personnelles** et données **sensibles**
 - Par ex, environnementales, santé,
 - Procédures spécifiques (anonymisation...)
 - Données présentant des risques
 - défense nationale
 - sécurité publique, des Etats, établissements
 - Données liées à une zone à régime restrictif (ZRR) >>> voir plus loin
 - Données liées à des secrets professionnels

Pourquoi diffuser ?



- **Pourquoi ne pas partager ?**
 - Données comportant une **valeur économique**
 - Brevet
 - Convention/contrat avec des entreprises
 - En vérifier les termes
 - Discussions nécessaires
 - **Publication en cours** de préparation
 - Possibilité de déposer ses données associées et de mettre un embargo avant la diffusion

Pourquoi diffuser ?



- **Intégrité et ré-utilisabilité**

- **Éthique** et intégrité scientifique (voir code de la recherche, art D211-2 et l'art de l'OFIS)
- Garantir la **reproductibilité** des résultats
 - Fiabilité
 - Transparence
- Assurer la **re-utilisabilité** des données
 - Intérêt pour d'autres projets scientifiques
 - Favoriser le progrès de la recherche et l'émergence de nouvelles recherches
 - Mettre en oeuvre les principes **FAIR** pour les codes et les données

FAIR = Findable, Accessible, Interoperable, Reusable

Pourquoi diffuser ?



Voir aussi la fiche :
Le minimum à FAIR

DoraNum, les principes FAIR, 2024



- **Pourquoi partager ? En réalité**

Certaines données ne sont produites ou traitées que par le biais de coopérations nationales et internationales

- Des données **précieuses ou uniques**
- Dont le coût de production ou de traitement
 - mobilise des moyens très importants tout au long du cycle de vie au niveau des Etats
 - et nécessite une exploitation scientifique commune ou partagée
 - Ex : données astronomiques, physique des hautes particules, données génomiques, données de sismologie...
 - Ex: boson de Higgs



- **Pourquoi partager ? En réalité**

Augmenter la visibilité

- Données accessibles et citables indépendamment de l'article
- Lier les données à ses publications
- Augmenter la visibilité de ses recherches,
- Donc la citation de sa recherche !

Pourquoi diffuser ?



Pourquoi partager ? En réalité

Répondre aux exigences des **éditeurs**:

Préconisation des éditeurs sur l'accès aux données liées aux publications (data sharing)

Quelques exemples

EDP Sciences

Plos One

Elsevier

Springer / Nature

Taylor and Francis

Sondage : avez-vous été soumis à cette exigence éditoriale ?

Pourquoi diffuser ?



Pourquoi partager ? En réalité

Répondre aux exigences des financeurs

- [Horizon Europe 2021-2027](#) :

La «science ouverte» deviendra le mode opératoire d'Horizon Europe. Il exigera donc un accès ouvert aux publications et aux données.»

- ANR (contribuer à l'[ouverture des données](#) quand c'est possible)
- NIH [Data Management and Sharing Policy](#)

Répondre aux exigences des Etats et établissements

- France : [2e Plan National pour la Science Ouverte](#) (2021-2024)
 - Axe 2 : Structurer, partager et ouvrir les données de la recherche
 - Axe 3 : Ouvrir et promouvoir les codes sources produits par la recherche
- CNRS : [Plan données de la recherche](#) (nov 2020)
- [Charte du CEA pour la science ouverte](#) (2021)
- [Charte UGA](#)



Quelle est votre situation, votre point de vue sur la diffusion ?

- Des restrictions ?
- Des réticences ?
- Quelle importance, quel intérêt, pour vous de diffuser vos jeux de données ?

En avez vous discuté avec votre directeur de thèse ? Des collègues ?



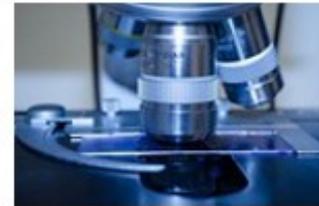
- **Analyse des données : des exemples en Science, Technologie et Médecine**
- **Est-ce que les données brutes suffisent à assurer la reproductibilité ou la réutilisation**
- **Est-ce que les données ont nécessité un pré-traitement / analyses?**
 - coûteux en temps ou en ressources
 - Si oui, diffusion du processus de pré-traitement / analyses
- **Est-ce que des outils/codes/logiciels sont nécessaires pour exploiter les données (pour pouvoir utiliser les analyses par exemple)**
 - indiquer les outils nécessaires pour l'exploitation des données
 - Si oui, diffuser si possible les outils/codes/logiciels liés

Impact environnemental : limiter le volume de données

Il existe différents types de données de la recherche qui diffèrent selon la manière dont les données sont produites et selon leur valeur supposée.

Données d'observation

- capturées en temps réel ;
- habituellement uniques et donc impossibles à reproduire ;
 - ➔ Exemples : neuroimagerie, photographie astronomique, données d'enquête.



publicdomainpictures.net

Données expérimentales

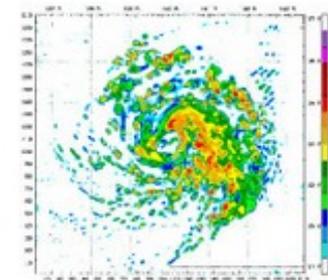
- obtenues à partir d'équipements de laboratoire ;
- souvent reproductibles mais parfois coûteuses ;
 - ➔ Exemples : chromatogrammes, puces à ADN.



publicdomainpictures.net

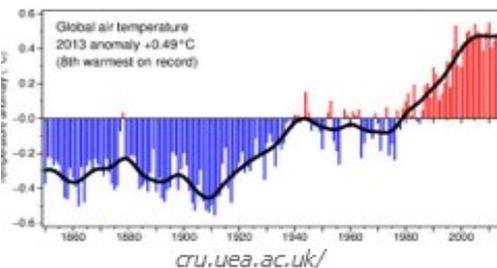
Données computationnelles ou de simulation

- générées par des modèles informatiques ou de simulation ;
- souvent reproductibles si le modèle est correctement documenté ;
 - ➔ Exemples : modèle météorologique, modèle de simulations sismiques, modèle économique.



Typhoon_Mawar_2005_computer_simulation.gif: Atmoz

Données de la recherche et publications scientifiques dans *Une introduction à la gestion et au partage des données de la recherche*, INIST



Données dérivées ou compilées

- issues du traitement ou de la combinaison de données "brutes" ;
- souvent reproductibles mais coûteuses ;
 - ➔ Exemples : fouille de texte, bases de données compilées.

Données de référence

- collection ou accumulation de petits jeux de données qui ont été revus par les pairs, annotés et mis à disposition ;
 - ➔ Exemples : GenBank, base de données de cristallographie, collection de lettres ou archive d'images historiques.



publicdomainpictures.net



Analyse des données : des exemples en Sciences humaines et sociales

- **Observations (par ex, observations urbaines)**
 - Captation unique
- **Enquêtes - questionnaires**
 - Administration unique
 - Données initiales (réponses brutes)
 - Données issues du dépouillement et des croisements
 - >>> diffuser les données initiales permettra de nouveaux croisements
- **Sources historiques, littéraires, juridiques, linguistiques**
 - Sont-elles accessibles en ligne (par ex sur Gallica)
 - Si oui, pas la peine de les déposer
 - La constitution du corpus est-elle originale ?
 - Si oui, le diffuser
- **Point de vigilance :**
 - Respect du RGPD
 - Impact environnemental : limiter le volume de données



Analyse des données : des exemples en Sciences humaines et sociales

- **Captation et numérisation**
 - **Numérisation**
 - Conserver les images brutes
 - Y-a-t-il des questions de droits d'exploitation ?
 - **Entretiens et transcriptions**
 - Y-a-t-il des questions de respect du RGPD
 - Diffuser les données initiales (enregistrement audio) ?
 - Diffuser uniquement la transcription ?
 - **Captation vidéo et transcription/analyse**
 - Y-a-t-il des questions de respect du RGPD ?
- **Bases de dépouillement – Exploitation - Visualisation**
 - Y a-t-il production d'une base de données originale (ex : base de données relationnelle) ?
 - Y-a-t-il des visualisations (par ex, graphes) originales ?

Impact environnemental : limiter le volume de données



- **Critère d'utilisabilité**

- Quelles données répondent aux objectifs de votre thèse ?
 - Pensez à la soutenance !
- Quelles données peuvent représenter un intérêt pour la communauté?

- **Statut des données**

- Est-ce que ces données existent déjà ?
 - Si oui, vérifier la pérennité de la source des données
- Est-ce que ces données sont uniques ?
 - Si oui, les diffuser
 - Si non, déterminer s'il vaut mieux diffuser les données ou le processus les ayant fournies (par exemple simulation, code source)



Diffuser des données « négatives » ou « non concluantes » ?

- Problématique dans toutes les disciplines
 - Enjeu de reproductibilité
 - Biais des résultats
-
- Voir Plan National Science Ouverte 2021-2024 : « Réduire le biais de publication, qui est la tendance à ne publier que les études ayant obtenu un résultat positif, au détriment des résultats peu concluants ou négatifs. »

Votre avis ? Votre expérience ?



Et pour vos propres données ?

- Avez-vous un projet de publication impliquant vos données ?
- Savez-vous déjà ce que vous pourriez diffuser ou non ?
- Avez-vous déjà des données prêtes à diffuser ?
- Quelles questions vous posez-vous ?



Définir l'usage des jeux de données que vous diffusez

- **Condition indispensable** pour la réutilisation de vos données
- Nécessité de choisir une licence d'usage
- **Licence** : contrat par lequel un titulaire d'un droit de propriété intellectuelle concède en tout ou partie la jouissance de ce droit (droits de reproduction, de représentation, droit d'autoriser les œuvres dérivées...)
- Licences **Creative Commons**
 - BY : Attribution
 - SA : Share Alike (partage à l'identique)
 - NC : Non Commercial
 - ND : Non Derivative (pas de modification)

Il suffit d'ajouter les logos de chaque famille souhaitée au contenu auquel on veut appliquer la licence

- Consulter la liste de toutes les licences possibles selon les objets.

Les licences Creative Commons

		Utilisation Partage	Adaptation Modification	Utilisation commerciale	Modification de licence	
TRÈS LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
LIBRE						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix autorisés Les versions dérivées de l'œuvre doivent conserver la licence originale ou compatible
NON LIBRE						<ul style="list-style-type: none"> Utilisation commerciale autorisée Modifications ou remix NON permis
						<ul style="list-style-type: none"> Utilisation commerciale NON permise Modifications ou remix NON permis



BY

ATTRIBUTION

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

L'auteur doit être cité



SA

PARTAGE DANS LES MÊMES CONDITIONS

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Partage sous licence compatible



NC

POUR USAGE NON COMMERCIAL

Vous pouvez retenir, réutiliser, réviser, remixer et redistribuer.

Pour usage non commercial



ND

PAS DE MODIFICATION

Vous pouvez retenir, réutiliser et redistribuer.

Aucune modification permise



Définir l'usage des jeux de données que vous diffusez

Choix de licences

Des **outils de sélection de licences** pour les dépôts de données ou de codes :

- License Chooser (Creative Commons)
- License Selector (codes et données)
- Licentia by Inria



Définir l'usage des jeux de données que vous diffusez

- **Des ressources pour vous aider :**
 - Contexte juridique de l'Espace chercheurs ENPC et son logigramme dynamique (à qui appartiennent les données / peut-on les diffuser)
 - Arbre Aide à la décision sur la diffusion des données de recherche (Cirad)
- **Ressources DoraNum**
 - Aspects juridiques et éthiques (RGPD, éthique, doit et open data)



Définir l'usage des jeux de données que vous diffusez /

Pour aller plus loin, des guides :

- Nicolas Becard, Céline Castets-Renard, Gauthier Chassang, Martin Dantant, Laurence Freyt-Caffin, et al.. Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France. [Rapport de recherche] Comité pour la science ouverte. 2017, 45 p. [⟨hal-02791224⟩](#)
- INSHS, Les sciences humaines et sociales et la protection des données à caractère personnel dans le contexte de la science ouverte – Guide pour la recherche, v2, 2021.
- Véronique Ginouvès, Isabelle Gras. La diffusion numérique des données en SHS – Guide des bonnes pratiques éthiques et juridiques, Presses universitaires de Provence, 2018, Digitales, [hal-⟨01903040⟩](#)



Quel accès ? Plusieurs choix possibles :

Libre accès : Immédiat ? Différé ? Sur demande ?

Focus sur les Zones à Régime Restrictif (ZRR) et dispositifs de protection du potentiel scientifique et technique de la nation (PPST)

Enjeu : Protection des « savoirs et savoir-faire stratégiques » et technologies sensibles

Protection juridique et administrative

Contrôle des accès physiques et numériques

Pour la diffusion des données, c'est le chercheur qui décide

La diffusion de données doit au préalable avoir été expressément autorisée par le responsable de la ZRR

Le fonctionnaire de sécurité défense ou d'autres services peuvent être sollicités.



Quand décide-t-on de rendre ses données publiques ?

Il n'y a pas de règles, le mieux est d'ouvrir les données le plus tôt possible

- **Avantages** : vous êtes le premier à produire de nouvelles données
- **Inconvénients** : de nouvelles expériences peuvent confirmer ou non la qualité de vos données
- Un embargo peut aussi être appliqué afin de permettre un délai d'exploitation des données
- Les données sont souvent publiées au moment de la publication des résultats
- Diffuser ses données peut être une justification pour les financeurs

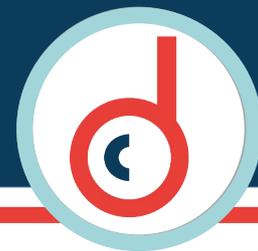


Quelles possibilités connaissez-vous pour diffuser vos données ?



Quel mode de diffusion ?

- Dépôt dans un entrepôt de données
- Publication données intégrées dans un article classique
- Publication de supplementary materials
- Publication d'un data paper



- Service en ligne permettant le dépôt, la description, la conservation, la recherche et la diffusion des jeux de données en vue de leur réutilisation.
- A ne pas confondre avec des plateformes de stockage ou d'archivage.
- Il existe des milliers d'entrepôts !
- Différents types d'entrepôts de données
voir la minute « entrepôts » (DoRaNum)

Où diffuser ? Diffuser via un entrepôt



Différents **types d'entrepôts** :

Généralistes : Zenodo (CERN)

Nationaux : Recherche Data Gouv

Institutionnel : ESRF Data Portal



Liés à des éditeurs : Mendeley

voir la note : AKDOGAN, M., Bracco, L., DU PASQUIER, D., Gauvrit, G., Heude, C., Laillier, B., León y Barella, A., Rousselot, C., & Sadowska, J. (2024). Les entrepôts de données d'éditeurs commerciaux : quelle stratégie adopter ?. GTSO Données de Couperin.
<https://doi.org/10.5281/zenodo.14382823>

« les éditeurs d'écrits scientifiques ne peuvent valablement obtenir de cessions exclusives de droits sur des données de recherche liées à la publication, qu'elles soient déposées sur un entrepôt de données ou qu'elles figurent comme supplementary materials de l'article ».

Cécile Arènes, Lionel Maurel, Stephanie Rennes. Guide d'application de la Loi pour une République numérique pour les données de la recherche. Comité pour la science ouverte. 2022. (hal-03968218).



Thématiques : ex : PANGAEA

Voir :
Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante
Comité pour la Science Ouverte. 2024.

Une première liste évolutive d'entrepôts (non exhaustive)

Sur Recherche Data Gouv, accès et description aux 1^{ers} entrepôts thématiques



- **Comment choisir un entrepôt approprié ?**
- **Quels seraient vos critères ?**



Quel entrepôt choisir ?

- En premier lieu s'il en existe, dans un **entrepôt disciplinaire**
 - **Pratique communautaire**
- Recommandations du **financeur, de l'éditeur**
- Recommandations du **projet de recherche, des partenaires**
- Recommandations de **l'établissement ou l'organisme de rattachement**



Liste de critères d'exclusion d'un entrepôt

(cf note méthodologique du Comité Science Ouverte « Sélectionner un entrepôt thématique de confiance pour diffuser ses données »)

- Absence de modération des dépôts
- Absence d'identifiant pérenne
- Absence de garantie sur la pérennité de l'infrastructure
- Entrepôts pratiquant la cession de droits
- Politique tarifaire excessive
- Localisation des données hors Union européenne pour certains types de données
- Dépôt restreint par l'affiliation institutionnelle



Les critères de choix

- **Statut et politique de l'entrepôt**
 - › Le statut public / privé de l'entrepôt?
 - › Lieu d'hébergement du serveur?
 - › Certification ? Reconnu ?
 - › Institution responsable de l'entrepôt ?
 - › La préservation sur le long terme des données ?

Modalités de dépôt

- › Types de données
- › Type de formats
- › Volume accepté
- › Gestion des versions
- › Qualité de la description (qualité des métadonnées) >>> standards ?
- › Type d'accès possibles aux données (Embargo / restriction de l'accès)
- › Licences
- › Lien avec la publication

En tout cas, c'est à vous de décider ...

Où diffuser ? via un entrepôt



Des outils pour vous aider à choisir un entrepôt :

Le logigramme sur Recherche Data Gouv

CAT OPIDOR – catalogue des services et entrepôts pour les données de la recherche. (Inist-CNRS) ,

DATAACC' – dispositif d'accompagnement à la gestion des données de la recherche en physique et en chimie. (UGA et Lyon 1)

La liste d'entrepôts dans le domaine biomédical (CeRIS)

Scientific Data Sharing – (National Institutes of Health)

Exemples d'entrepôts de données en SHS (DoraNum)



Utilisation d'un **annuaire pour identifier un entrepôt dans sa discipline** :

- re3data (Registry of Research Data Repositories)
- Repository Finder(DataCite)
- FAIRsharing



A vous de jouer !

10 mn pour trouver un entrepôt de données pertinent dans votre discipline

- Page des entrepôts de confiance sur Recherche Data Gouv
- CatOpidor
- re3data (Registry of Research Data Repositories)
- Repository Finder(DataCite)
- FAIRsharing



Je prépare mes données

- Choix des **données pertinentes** (brutes, traitées, analysées)
 - Avec un point de vigilance concernant les données personnelles ou confidentielles
 - En s'assurant que les volumes sont en adéquation avec ce que permet l'entrepôt choisi
 - En veillant à mettre les éléments nécessaires pour les utiliser (codes, logiciels, etc.)
- Choix des **formats** (ouverts)
- Choix des **éléments de description** : métadonnées générales et disciplinaires, Readme ...
- **Organisation et nommage** des fichiers
- Choix de la **licence et des modalités d'accès** (ouvert, restreint, avec embargo)
- **Respecter les principes FAIR (Findable, Accessible, Interoperable, Reusable)**

Où diffuser ? via un entrepôt



Je prépare mes données (suite)

S'être créé un compte et posséder des droits sur la collection Dataverse de son laboratoire (en faire la demande via l'onglet « support » en cas de besoin)

Avoir un ensemble de fichiers prêts (organisation logique, nommage clair et sans accents), présence d'un README (fichier texte permettant d'expliquer le contenu du jeu de données et d'éventuelles consignes d'utilisation)

Optionnel : Avoir à portée de main la référence de la publication à laquelle lier le jeu de données

Où déposer ses données quand on est dans un laboratoire UGA ?



Où diffuser ses données à l'UGA s'il n'existe pas d'entrepôt thématique pertinent ?

La plateforme nationale Recherche Data Gouv

Objectif :

Proposer une solution de dépôt dans un entrepôt national de confiance

3 volets :

- Entrepôt pour déposer et publier des données
- Catalogue pour signaler des données déposées dans des entrepôts externes (à venir)
- Accompagnement à la gestion des données : ateliers de la donnée, centres de ressources (INIST) et de référence thématiques

Où déposer ses données quand on est dans un laboratoire UGA ?



Espace UGA sur Recherche Data Gouv : **Data Repository Grenoble Alpes**

Organisation en collections par laboratoires, projets de recherche, etc

Tous les types de données sont acceptés

Taille max des jeux de données : 50 Go par fichier.

Accès restreint possible pour certains ou tous les fichiers d'un jeu de données.

Possible de créer une **url privée** pour un jeu de données non publié, par exemple pour de la relecture par les pairs

Modération

Rappel : **privilégier un entrepôt thématique** reconnu par sa communauté



Dépôt dans Recherche Data Gouv (suite)

Pour vous aider :

- **classes virtuelles**
- **tutoriels**
- **guides**
- **FAQ**

Un modèle de Readme



**Le bac à sable de Recherche Data Gouv :
<https://demo.recherche.data.gouv.fr/>**

A vous de jouer!

Déposer un jeu de données

Focus Codes

- Archivage (automatique) dans **Software Heritage**
- Signalement dans HAL

Fiche pratique
Webinaire
(support et vidéo)

J'ouvre mon code en 4 étapes



1

J'intègre 4 fichiers

README.md	→	Description et liens vers les documentations
AUTHORS.md	→	Les auteurs et les contributeurs du logiciel
LICENSE.txt	→	Je veux un copyleft = GNU GPL Je veux une licence permissive = MIT license
CODEMETA.md	→	À générer automatiquement avec le CodeMetagenerator de github

2

Quand je code

Je **documente et commente mon code** pour moi et pour les autres
J'utilise une **forge** comme le Gitlab de Gricad

3

J'archive mon code sur Software Heritage



4

Je signale mon code sur HAL grâce au SWHID



Mon code est ouvert

Contacts

Pour plus d'informations sur la diffusion des codes, contacter la Cellule Data Université Grenoble Alpes :

sos-codes-recherche@univ-grenoble-alpes.fr





3 manières de publier des données

- Les inclure dans un article (données intégrées ou embedded data)
- Les assembler en annexe dans un matériel supplémentaire (« supplementary materials »)
- Les publier dans un data paper (data article, data descriptor).



Focus sur les datapapers

Définition :

Les data papers (data articles / data descriptors) sont des articles qui ont pour but de décrire un ou plusieurs jeux de données, plutôt que des résultats d'analyse.

Les data papers peuvent paraître dans des revues classiques ou dans des revues spécifiques « data journals ».

Pratique :

Les données peuvent être déposées dans un entrepôt, recommandé par l'éditeur ou au choix de l'auteur

Voir la minute « Publier un data paper », DoRaNum

Où diffuser ? via une publication



Focus sur les datapapers

Enjeu :

Le datapaper valorise les données en exposant leur potentiel pour de nouveaux usages

Il facilite la réutilisation des données en mettant en évidence la qualité des données et des procédures, ainsi que la rigueur scientifique de l'étude.

Il apporte de la visibilité aux données, les rend plus facilement repérables et citables par d'autres études.

Le datapaper est examiné par les pairs

Recommandation :

Déposez dans HAL le post-print de la publication

(voir Dedieu, L. 2014. Rédiger et publier un data paper dans une revue scientifique, en 5 points. Montpellier (FRA) : CIRAD, 7 p. <https://doi.org/10.18167/coopist/0057>)



Focus sur les datapapers

Quelques exemples de datajournals :

- Data Science Journal – codata
- Scientific Data (Nature)
- Journal of Open Humanities Data (JOHD)
- Data In Brief (Elsevier)

Quelques listes :

La base « Où publier » du Cirad (sélectionner data papers)

liste « Revues publiant des Data papers et des Software papers » (CIRAD)

Dataacc (Physique-Chimie)

Possible dans de nombreuses revues « classiques » (PlosOne, CyberGeo ...)

Comment choisir ? S'informer sur :

Son modèle économique (accès ouvert ? APC?)

Ses modalités de diffusion (dépôt dans un entrepôt possible?)

Ses contraintes juridiques (cession de droit exclusive? Possibilité de mettre des licences CC ?)



Focus sur les datapapers

Dans tous les cas, se référer :

- aux guides/instructions des éditeurs à destination des auteurs
- aux templates mis à disposition par les éditeurs

Exemples :

- Template for data descriptor pour Scientific Data (Nature Publishing Group – Overleaf)
- Le template de Data in Brief
- Journal of Open Humanities Data (JOHD) data paper template (open humanities data)
- L'outil arpha Writing Tool.
- L'outil de génération de datapaper à partir du doi d'un jeu de donnée sur Recherche Data Gouv



5 exemples de datapapers :

- Rodríguez-Pérez, Q. and Zúñiga, F. R.: An earthquake focal mechanism catalog for source and tectonic studies in Mexico from February 1928 to July 2022, *Earth Syst. Sci. Data*, 15, 4781–4801, <http://doi.org/10.5194/essd-15-4781-2023>, 2023.
- Hosseini, K., Beelen, K., Colavizza, G., & Ardanuy, M. C. (2021). Neural Language Models for Nineteenth-Century English. *Journal of Open Humanities Data*, 7, 22. DOI: <http://doi.org/10.5334/johd.48>
- Kovylyaeva, A., Astapov, I., Dmitrieva, A., Borog, V., Osetrova, N., & Yashin, I. (2020). Experimental Data of Muon Hodoscope URAGAN for Investigations of Geoeffective Processes in the Heliosphere. *Data Science Journal*, 19(1), 11. DOI: <http://doi.org/10.5334/dsj-2020-011>
- Testolini, V. (2021). Data from “Ceramic Technology and Cultural Change in Sicily from the 6th to the 11th Century AD.” PhD Thesis. *Journal of Open Archaeology Data*, 9, 11. DOI: <http://doi.org/10.5334/joad.77>
- Selvam, R. M. et al. (2015). Data set for the mass spectrometry based exoproteome analysis of *Aspergillus flavus* isolates. *Data in brief*, 2, 42-47. <http://dx.doi.org/10.1016/j.dib.2014.12.001>



Ressources

CNRS, 2021. Guide de bonnes pratiques sur la gestion des données de recherche. Publier un Datapaper pour valoriser et expliciter les données.

INRAE, Pubier un datapaper,
<https://datapartage.inrae.fr/Partager-Publier/Publier-un-Data-Pape>
voir la FAQ

Software Sustainability Institute, 2021.
In which journals should I publish my software?

DoRANum, Data papers et Data journals

DoRaNum, 2018. La minute Publier un Data paper.

DoRANum, 2020. Webinaire Data paper - Une incitation à la qualification et à la réutilisation des jeux de données.

DoRaNum, 2022, une série de 3 vidéos sur les datapapers



Lier publications, données, et codes

Référencement

- Des publications sur HAL (et dépôt du texte intégral)
- Des données sur Recherche Data Gouv / autres entrepôts
- Des codes sur Software Heritage et HAL
- Lien entre toutes ces productions via leurs identifiants (doi)

Dernière recommandation !



Identifiants

Sur HAL

Identifiants

Ajoutez l'identifiant DOI, arXiv, PubMed, ADS, etc pour lier votre dépôt aux autres bases.

SWHID ▾



Données associées

Ajoutez l'identifiant DOI fourni par l'entrepôt où vos données sont archivées.



Publication associée ?

Un ou plusieurs des champs suivants pourraient devenir requis si vous complétez l'un de ces champs optionnels.

Sur Recherche Data
Gouv

Citation ?

Nom, Prénom (Année). Titre, Editeur. DOI



Type d'identifiant ?

Sélectionner...

Identifiant ?

ex. pour DOI : "10.15454/AEIOUY"

URL ?

Adresse URL, commençant par https://



Une adresse mail :

[sos-data\[at\]univ-grenoble-alpes.fr](mailto:sos-data@univ-grenoble-alpes.fr)

La [cellule data Grenoble Alpes](#) répond concrètement à toutes les demandes des communautés scientifiques de Grenoble sur les données.

- Aide à la diffusion des données et des codes
- Aide à la description des données
- Lien publications/données/codes
- Aide juridique
- Diffusion des bonnes pratiques

Et pour les codes : [sos-codes-recherche\[at\]univ-grenoble-alpes.fr](mailto:sos-codes-recherche@univ-grenoble-alpes.fr)



Equipe de la la BU : traitement des thèses et accompagnement des doctorants

Rôle :

- Signalement et diffusion de la thèse sur theses.fr, via [HAL thèses](https://hal.archives-ouvertes.fr/)

Quelques exemples de services :

- Accompagnement individuel au dépôt de la thèse
- Aide juridique

Une adresse : bu-theses@univ-grenoble-alpes.fr



- Le site science ouverte de l'UGA
- DoraNum
 - Dépôt et entrepôts
 - Data papers et data journals
- Site du CIRAD, Coopist, Gérer des données
- Passeport science ouverte
 - Codes
 - Données de la recherche