



Présentez vous en *quelques mots* :

- Sujet ?
- Début ou fin de thèse ?
- Projet de publication et/ou de diffusion des données ?
- Votre thèse relève-t-elle d'un projet ANR ? Européen ?
- Votre thèse traite-t-elle des données personnelles ou sensibles?
- Votre labo est-il en ZRR ?
- Avez-vous discuté de la diffusion des données liées à votre thèse avec vos encadrants ?
- Attentes/questions par rapport à la formation ?
- Qui a participé :
 - à des journées sur la gestion des données ?
 - à la formation RGPD



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

Formation doctorale 16/12/2025



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*

Lier ses productions scientifiques

Aide et ressources



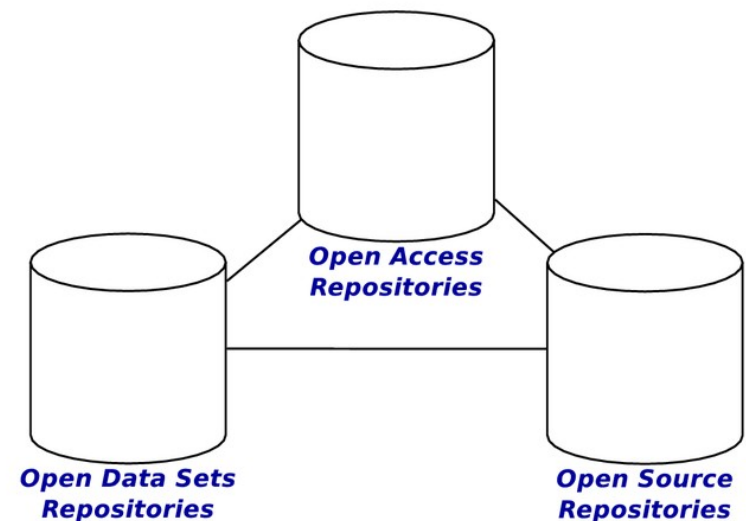
- **Processus de publication :**

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 ?

Votre avis ?

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 ?



Pourquoi partager (enjeux pour la recherche et enjeux pour vous) ?

Que partager (dans votre cas précis) ?

- **Selon quels critères ?**
- **Savez-vous déjà ce que vous pourriez diffuser ou non ?**
- **Avez-vous déjà des données prêtes à diffuser ?**
- **Avez-vous des réticences au partage ?**

Travail en binôme (20 mn)

Restitution incrémentale (10mn)

Pourquoi diffuser ?



- **Intégrité et ré-utilisabilité**

- **Éthique** et intégrité scientifique (voir code de la recherche, art D211-2 et l'art de l'OFIS)
- Retour vers les citoyens
- 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 ?



FAIR DATA PRINCIPLES

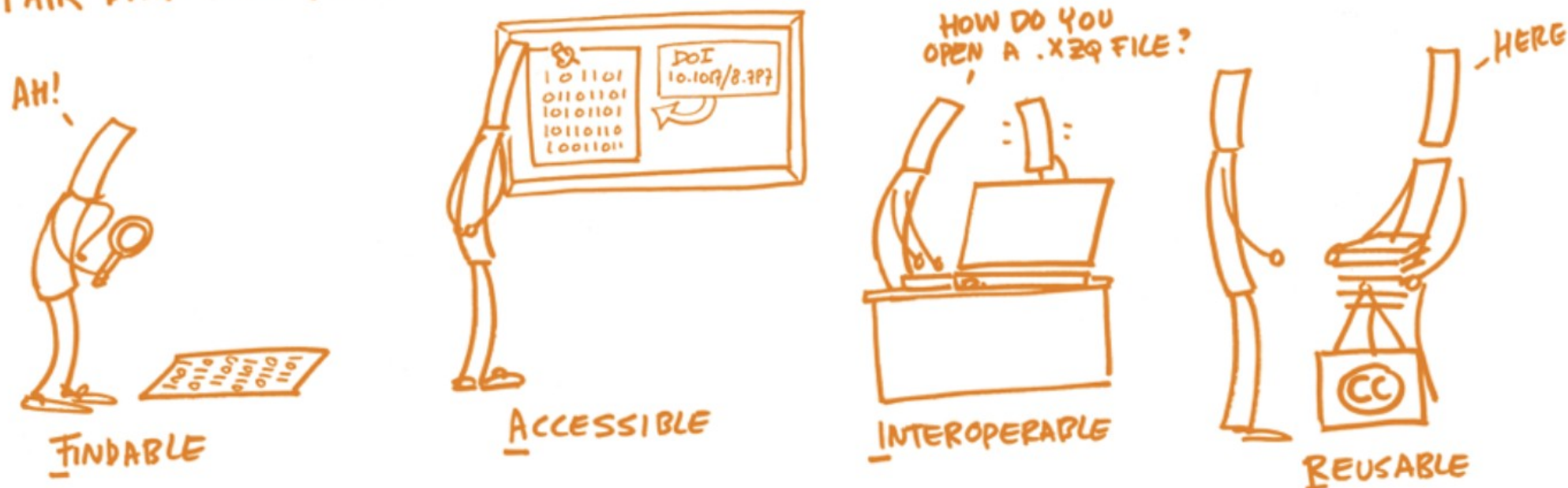


Image: <https://book.fosteropenscience.eu/>

Voir aussi la fiche : Le minimum à FAIR

Pourquoi diffuser ?



- **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 diffuser ?



- 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

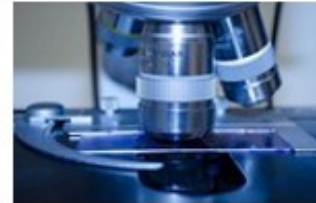
- France : [2e Plan National pour la Science Ouverte](#) (2021-2024)
 - Axe 2 : Structurer, partager et ouvrir les données de la recherche
 - Axe 3 : Ouvrir et promouvoir les codes sources produits par la recherche
- CNRS : [Plan données de la recherche](#) (nov 2020)
- [Charte du CEA pour la science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

Types de données de la recherche

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



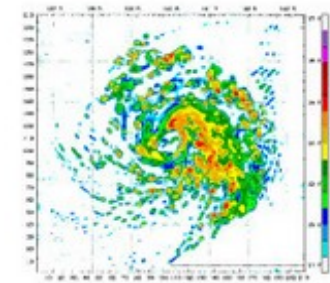
publicdomainpictures.net

Données expérimentales

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

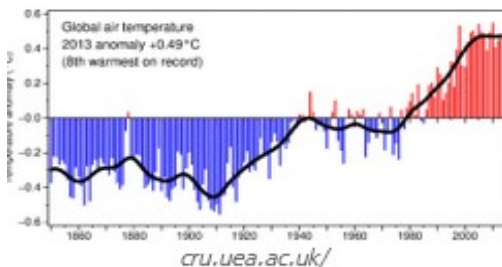
Données computationnelles ou de simulation

- 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: Atmos

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

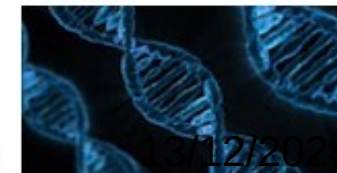


Données dérivées ou compilées

- 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.



12/12/2024



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



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



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 : le principal critère**

- 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 ?
- Quelles questions pratiques 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...)






















Comment diffuser ? focus licences



Exemples :

- 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**
- Licence ouverte Etalab (open licence, compatible cc-by)
- 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)



Définir l'usage des jeux de données que vous diffusez

- **Des ressources pour vous aider :**
 - Arbre Aide à la décision sur la diffusion des données de recherche (Cirad)
- **Ressources DoraNum**
 - Aspects juridiques et éthiques (RGPD, éthique, droit 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⟩](#)

Comment diffuser ?



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.

Comment diffuser ?



Quel accès ? Focus respect du RGPD

Le principe de l'anonymisation et de la pseudonymisation

PROCESSUS	PSEUDONYMISATION	ANONYMISATION
STATUT DES DONNÉES	Personnelles (restent indirectement identifiantes et donc soumises au RGPD et à la loi Informatique et Libertés)	Anonymes
RÉUTILISATION DES DONNÉES	Sous conditions	Sans restriction
UTILITÉ DES DONNÉES	Préservée car pas d'altération du niveau de détail des données	Plus ou moins altérée en fonction des objectifs poursuivis et des méthodes appliquées
MÉTHODES À METTRE EN OEUVRE	Compteur, générateur de nombres aléatoires, fonction de hachage , chiffrement à clé secrète, etc.	Dépend des objectifs poursuivis : confidentialité différentielle , randomisation, k-anonymat, l-diversité, t-proximité, etc.
COMPLEXITÉ DE LA MISE EN OEUVRE	Simple à moyenne	Dépend des objectifs poursuivis : simple dans certains cas comme l'agrégation ou le comptage et complexe dans d'autres

Exemples d'outils : [amnesia](#), [ARX](#)

Voir fiche pratique CNIL :

[enjeux et avantages de l'anonymisation et de la pseudonymisation \(2022\)](#)



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
- **Inconvénient** : de nouvelles expériences peuvent confirmer ou non la qualité de vos données
- Exigence des éditeurs : rendre accessible, au moins aux reviewers, les données liées à une publication
- 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

Quand diffuser ?



En Bref

À PRIVILÉGIER



Partager des données liées à une publication

Déposer ses données **avant** de publier son article : mentionner l'**identifiant pérenne** des données dans l'article et inversement, mentionner l'identifiant pérenne de l'article dans la description des données.

-
-
- **entrepôt de données** de confiance (disciplinaire ou institutionnel), indépendant de la revue.
-
-

À ÉVITER



Déposer ses données dans un entrepôt **après** avoir publié son article.

-
-
- Déposer les données associées à l'article dans l'entrepôt de la revue, ou sous la forme de « **supplementary data** » ou de « **supplementary materials** ».
-
-

Partager les données liées aux publications scientifiques. Guide pour les chercheurs. Comité pour la science ouverte. Collège des Données de la recherche. Décembre 2025.



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

Où diffuser ? Diffuser via un entrepôt



- 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

Où diffuser ? Diffuser via un entrepôt



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).

Où diffuser ? Diffuser via un entrepôt



Thématiques : quelques exemples

PANGAEA

ORTOLANG

Protein Data Bank in Europe

Nakala

PublicNeuro

Où diffuser ? via un entrepôt



- **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**
 - Vérifier les entrepôts cités dans les publications
- Recommandations du **financeur, de l'éditeur** (listes d'entrepôts)
- 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

Où diffuser ? via un entrepôt



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 ? Diffuser via un entrepôt



Des outils pour vous aider à choisir un entrepôt :

Le logigramme sur Recherche Data Gouv

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 liste mise à jour régulièrement (non exhaustive)

Sur Recherche Data Gouv, accès et description aux entrepôts thématiques

Où diffuser ? via un entrepôt



Des outils pour vous aider à choisir un entrepôt :

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)

Repositories for Sharing Scientific Data (National Institutes of Health)

Où diffuser ? via un entrepôt



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

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

Où diffuser ? via un entrepôt



Utilisation de **moteurs de recherche spécialisés** pour trouver des jeux de données et donc des entrepôts :

Datacite Commons

Dataset Search

OpenAire

Plus d'infos **ici**

Où diffuser ? via un entrepôt



A vous !

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

- Page des entrepôts de confiance sur Recherche Data Gouv
- CatOpidor
ou un des 3 annuaires :
 - re3data (Registry of Research Data Repositories)
 - Repository Finder(DataCite)
 - FAIRsharing



DEPOSER UN JEU DE DONNEES

Ex avec Recherche Data Gouv



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)



Je prépare mes données (suite)

S'être créé un compte et posséder des droits sur l'entrepôt de son choix (en faire la demande via l'adresse « support » en cas de besoin)

Avoir un **ensemble de fichiers prêts**

- Arborescence : 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 (identifiant hal, doi ..)



Je prépare mes données (suite)

Respecter les principes FAIR (Findable, Accessible, Interoperable, Reusable)

Attention à la charge de travail que représente tous ces aspects, prenez-la en compte !

- Anticiper (description, organisation des fichiers, etc)
- Préparer une arborescence
- Rédiger un Read me

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

Une plateforme nationale des données de recherche

- Entrepôt pour déposer et publier des données
- Catalogue pour signaler des données déposées dans des entrepôts externes (à venir)

Une fédération nationale des centres de compétences

- Géographique : centres de compétences géographiques - ateliers de la donnée
- Thématiques : centres de compétences disciplinaires
- Global : centres de ressources (DoraNum)

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 assurée

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

Deux exemples de jeu de données

ExpeSmarthouse - Données d'une maison connectée

Cross-referenced data on electoral disputes and French legislative election results



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

Déposons un jeu de données

Focus Codes

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

Fiche pratique

Webinaire ([support](#) et [vidéo](#))

Documentation HAL :
[déposer le code source d'un logiciel](#)

INIST (et al.) (2025). DoRANum-Accès et visualisation : Préserver et diffuser les codes sources des logiciels de recherche. DoRANum.
<https://doi.org/10.13143/657A-ZK37>

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 :

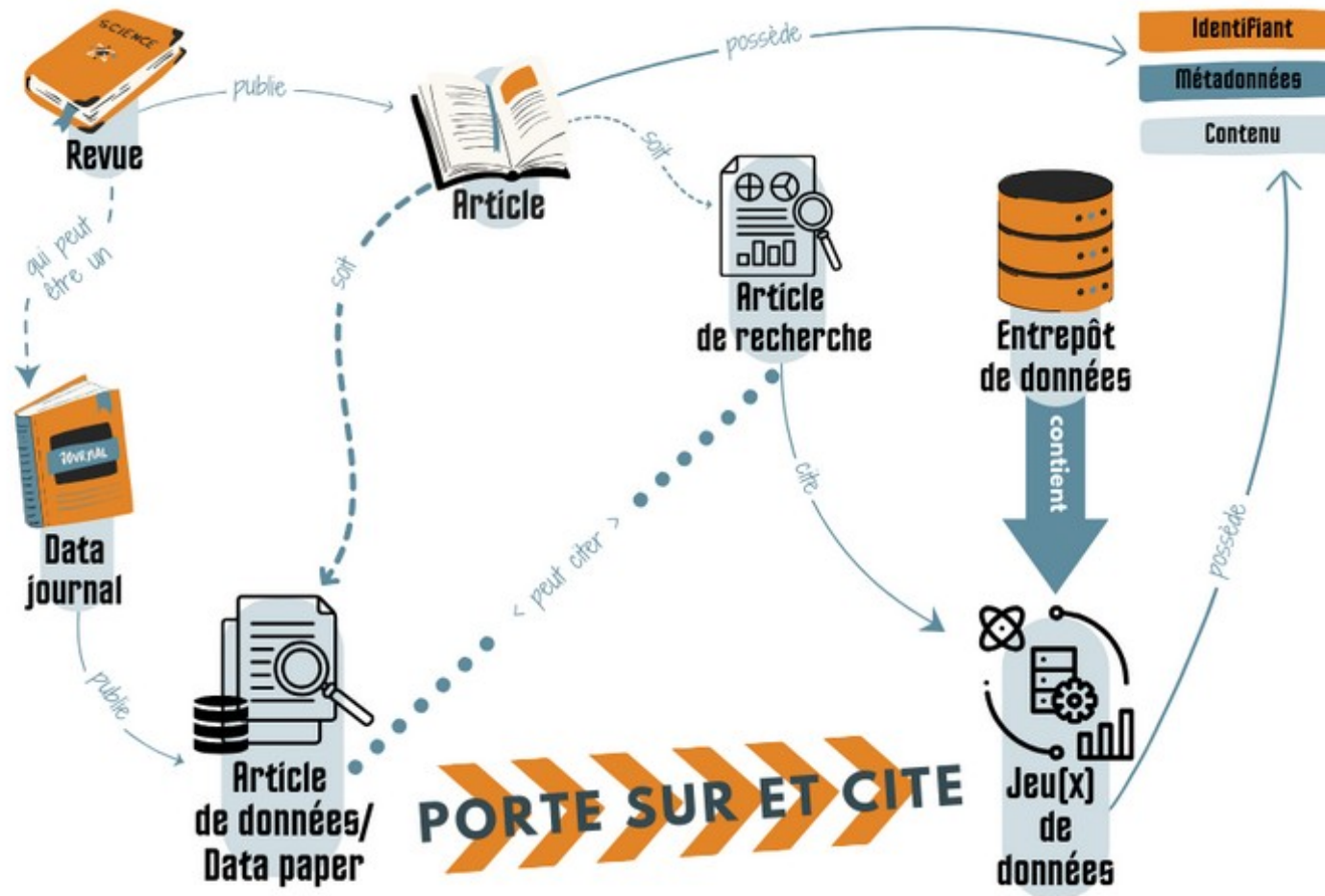
Déposez dans HAL le post-print de la publication

(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>)

Où diffuser ? via une publication



LE CONCEPT D'ARTICLE DE DONNÉES



Flodine Papin, pour ADOC Lorraine, d'après L. Candolia, B. Castellani, P. Moyni et A. Tseli, "Data Journals : A survey (preprint)", Journal Association for Information Science and Technology, 2014, p. 1-20, figure 6, <https://doi.org/10.1002/asit.23256>

Où diffuser ? via une publication



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)
- Data & Corpus (Univ Lorraine) - SHS

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 ?)

Où diffuser ? via une publication



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

Où diffuser ? via une publication



« **Le data paper est généralement structuré en 3 points :**

- description du jeu de données et de son contexte,
- description des méthodes d'obtention,
- explication du potentiel de réutilisation des données. »

(voir Dedieu, L. 2022. *Publier un Data paper, en 5 points*. Montpellier (FRA) : CIRAD, 5 p. <https://doi.org/10.18167/coopist/0057>)

Exemple :

<i>(Scientific Data)</i>	<i>(JOHD)</i>
Methods	1- overview
Data Records	2 Method
Data Overview (optional)	3 Dataset Description
Technical Validation	4 Reuse Potential
Usage Notes (optional)	
Data Availability Code Availability	

Où diffuser ? via une publication



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>

Où diffuser ? via une publication



Ressources

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

Publier un software paper

Software Sustainability Institute, 2021.

In which journals should I publish my software?

DoRANum, Data papers et Data journals
L'essentiel et « Pour aller plus loin »

Où diffuser ? via une publication



Diffusion en libre accès de ses articles :

- Dans tous les cas :
 - **déposer son post print dans une archive ouverte** (HAL, arXiv)
cf Loi pour une République Numérique, art 30 : Droit de déposer un article accepté pour publication dans une archive ouverte
voir Guide d'application et la FAQ
 - **si paiement de frais de publication** (APC, Article Processing Charges),
déposer la **version éditeur immédiatement dans une archive ouverte**
- Si possible, privilégier les revues :
 - **des revues sans frais pour les lecteurs et les auteurs**
 - avec frais de publication raisonnable (< 2000 €)
 - en **libre accès intégral**

Voir liste des revues sans APC du Directory of Open Access Journals (DOAJ)
et le Journal Checker Tool



Lier publications, données, et codes

Référencement

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

A noter : vous pouvez, depuis un jeu de données, associer également votre thèse déposée dans HAL

Comment lier des données à une publication lors du processus de *reviewing* de l'article (révision par les évaluateurs de la revue) et comment les référencer dans la notice HAL de l'article.



Comment lier des données à un article dans une revue scientifique ?

Les données ne sont pas encore dans un entrepôt

Les données sont déjà dans un entrepôt -> un PID leur est attribué

Une phase de *reviewing* des données est-elle nécessaire ?

Oui

Non

Non

Oui

Déposer les données dans un entrepôt sans les publier

Fournir aux reviewers de l'article une URL privée lui donnant accès aux données *

Publier le dépôt -> un PID est attribué aux données

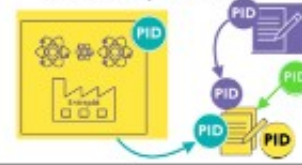
Déposer et publier vos données dans un entrepôt -> un PID est attribué aux données



Indiquer le PID des données (et le PID du code/logiciel le cas échéant) dans l'article à publier



Indiquer le PID du data paper et le PID des données (et le PID du code/logiciel le cas échéant) dans l'article à publier



Veiller à ce que soit indiqué le PID des données (et le PID du code/logiciel le cas échéant) dans la notice HAL de l'article en plus du PID de l'article



Veiller à ce que soient indiqués tous les PID dans la notice HAL de l'article



PID des données
PID du data paper
PID de l'article
PID du code ou logiciel ayant servi pour la production des données
 Identifiant HAL

PID ? Persistent Identifier
 Les identifiants persistants garantissent un lien stable vers la ressource en ligne (ex : DOI)

Partager les données liées aux publications scientifiques
 . Guide pour les chercheurs.
 Comité pour la science ouverte. Collège des Données de la recherche.
 Décembre 2025.

* Dans le cas où l'entrepôt ne fournit pas d'URL privée, vous pouvez communiquer le PID des données publiées dans l'entrepôt ou transmettre les données au reviewer par un autre moyen

Dernière recommandation !



Sur HAL

Ressources associées

Relation

Entrepôt

Cite (Cites) ▼

Recherche Data Gouv ▼

Rechercher un dépôt dans Recherche Data Gouv (identifiant, titre)

Commentaire

Annuler

Ajouter

Voir la [documentation](#) sur HAL

Ressources associées

Relation

Entrepôt

Cite (Cites) ▼

Autre source (autre entrepôt, archive, ...) ▼

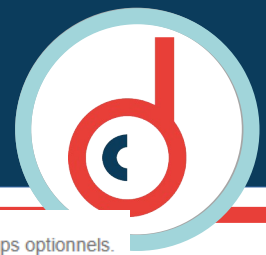
Identifiant DOI

Commentaire

Annuler

Ajouter

Dernière recommandation !



Sur
Recherche
Data Gouv

Publication associée ?

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

Type de relation ?

Est cité par



Citation ?

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

Type d'identifiant ?

Sélectionner...

Identifiant ?

ex. pour DOI : "10.15454/AEIOUY"

URL ?

Adresse URL, commençant par https://

Contenus associés ?



Jeu de données associé ?

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

Citation ?

Nom, Prénom (Année). Titre. Version. Editeur. Type de ressource. Identifiant



Type d'identifiant ?

Sélectionner...

Identifiant ?

ex. pour DOI : "10.15454/AEIOUY"

URL ?

Adresse URL, commençant par https://

Dernière recommandation !



Sur
Recherche
Data Gouv

Métadonnées de workflow de calcul ^

Type de workflow ?	<input type="text"/>	
URL vers le dépôt de code externe ?	<input type="text" value="https://..."/>	+
Documentation ?	<input type="text"/>	+

Pour aller plus loin :

[Partager les données liées aux publications scientifiques](#). Guide pour les chercheurs.
Comité pour la science ouverte. Collège des Données de la recherche. Décembre 2025.



Une adresse mail :

[sos-data\[at\]univ-grenoble-alpes.fr](mailto:sos-data[at]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[at]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
- Aide pour lier les productions scientifiques

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](#)