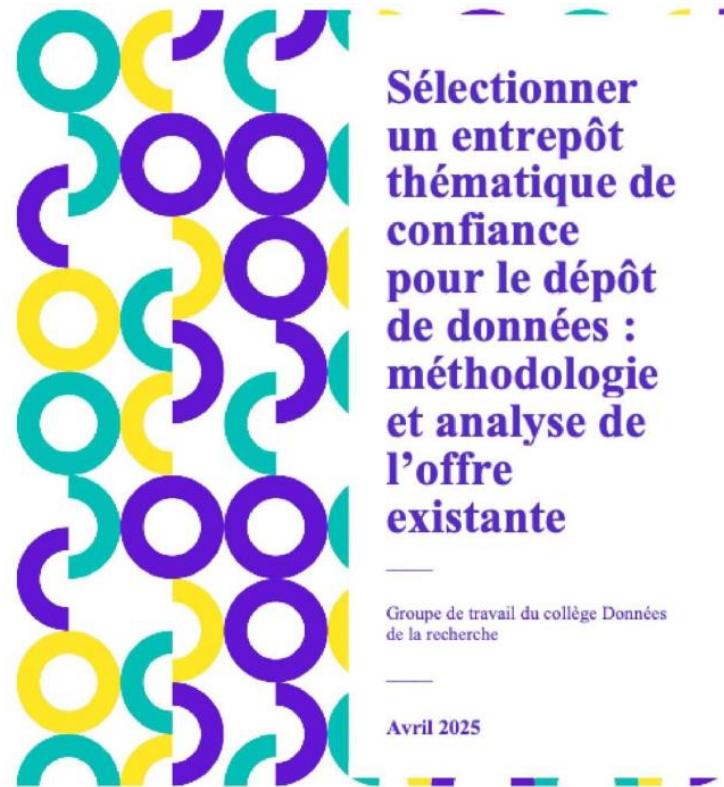


Les entrepôts de données de confiance

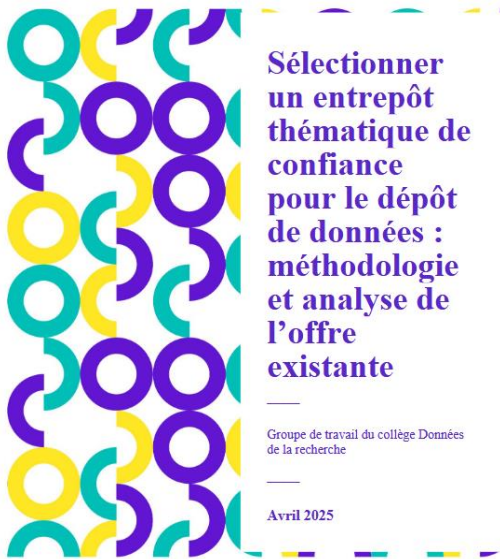
Curathon ADELE 2026



Les entrepôts de données de confiance : la liste du Comité pour la Science Ouverte



La liste des entrepôts thématiques de confiance du COSO



- ◆ **Commande du MESRE/ DGRI**
- ◆ **Mission confiée au Collège des données de la recherche du Comité pour la Science Ouverte (mars 2023)**

→ **Un GT dédié sur les entrepôts thématiques de confiance**

→ **3 axes de travail :**

- Établir la liste de critères de confiance permettant d'apprécier la qualité d'un entrepôt thématique
- Proposer une note méthodologique et une première liste d'entrepôts thématiques de confiance
- Proposer une stratégie de pérennisation de la liste d'entrepôts thématiques de confiance

Guider les équipes de recherche pour sélectionner un entrepôt de confiance

→ « L'**ambition** des politiques du ministère de l'Enseignement supérieur et de la Recherche concernant les données de la recherche est de **faire en sorte que ces données soient progressivement structurées en conformité avec le principes FAIR** (Facile à trouver, Accessible, Interopérable, Réutilisable), **préservées** et **partagées ou ouvertes** par des **entrepôts de données de confiance**.

Afin de **guider les équipes de recherche vers l'entrepôt le plus adapté** pour le partage et l'ouverture des données de leur **domaine thématique**, (...) il est **indispensable d'identifier** les **entrepôts thématiques nationaux et internationaux de confiance, certifiés Core Trust Seal ou non** »



Guider les équipes de recherche pour sélectionner un entrepôt de confiance

Plans nationaux pour la science ouverte

Développer une démarche proactive d'**ouverture des données associées aux articles** et publications sur des entrepôts de confiance thématiques ou sur Recherche Data Gov.



Exigences des financeurs

ANR

Les **données** doivent être:

- **déposées dans un entrepôt de données**
- Aussi vite que possible
- **Selon les modalités indiquées dans le DMP**
- En open access sauf restriction légale

HORIZON EUROPE

[La politique Science ouverte de l'ANR](#)

[Open Science in Horizon Europe](#)

Guider les équipes de recherche pour sélectionner un entrepôt de confiance

Constats:

- ◆ Certaines communautés scientifiques déjà structurées autour d'entrepôts de confiance bien identifiés (astrophysique, génomique, cristallographie), d'autres non
- ◆ Pas toujours de recommandations explicites de la part des sociétés savantes, communautés, financeurs: choix délégué au chercheur

Risques:

- ◆ Choix d'un entrepôt sans politique exigeante de description des données
- ◆ et/ ou entrepôt porté par un éditeur commercial



Comment sélectionner un entrepôt ?

« Mon éditeur/financeur demande de déposer des données dans un entrepôt »
« Je souhaite partager mes données »

1/ Se renseigner auprès de votre communauté de recherche (équipe de recherche, DU,...) : pour prendre connaissance des pratiques établies en matière d'entrepôts identifiés/ de confiance

2/ Chercher un entrepôt dans un catalogue

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

3473 Repositories

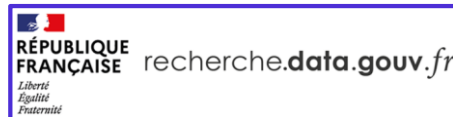


FAIRsharing.org
standards, databases, policies



Comment sélectionner un entrepôt ?

« Mon éditeur/financier demande de déposer des données dans un entrepôt »
« Je souhaite partager mes données »



1/ Se renseigner auprès de votre communauté de recherche (équipe de recherche, DU,...) : pour prendre connaissance des pratiques établies en matière d'entrepôts identifiés/ de confiance

2/ Consultez la liste des entrepôts de confiance

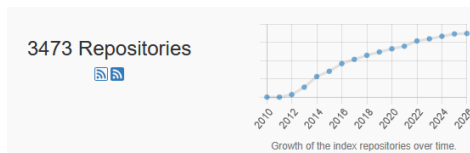
3/ Si aucun entrepôt de confiance n'existe ou ne semble remplir les critères « de confiance » suffisants, vous pouvez déposer dans Recherche Data Gouv : entrepôt national et multidisciplinaire pour les chercheurs qui n'ont pas de solution d'entrepôt disciplinaire

Comment sélectionner une entrepôt?

Chercher un entrepôt dans un catalogue



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

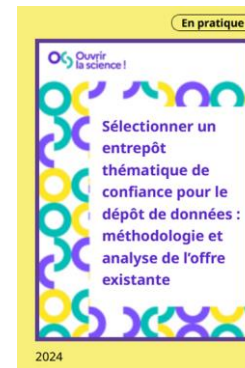


FAIRsharing.org
standards, databases, policies

2873 Databases

Repositories	1612
Knowledgebases	1042
Knowledgebase/Repositories	219
Institutional Repositories	199

Consulter la **liste des entrepôts thématiques « de confiance »** établie par le comité pour la Science Ouverte (COSO)

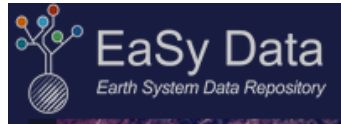


Recommandations au niveau national:

1) déposer ses données en priorité sur un **entrepôt disciplinaire** (ou entrepôt thématique)



2) Si difficultés à identifier un entrepôt disciplinaire : possibilité de se tourner vers un **entrepôt « généraliste », pluridisciplinaire**



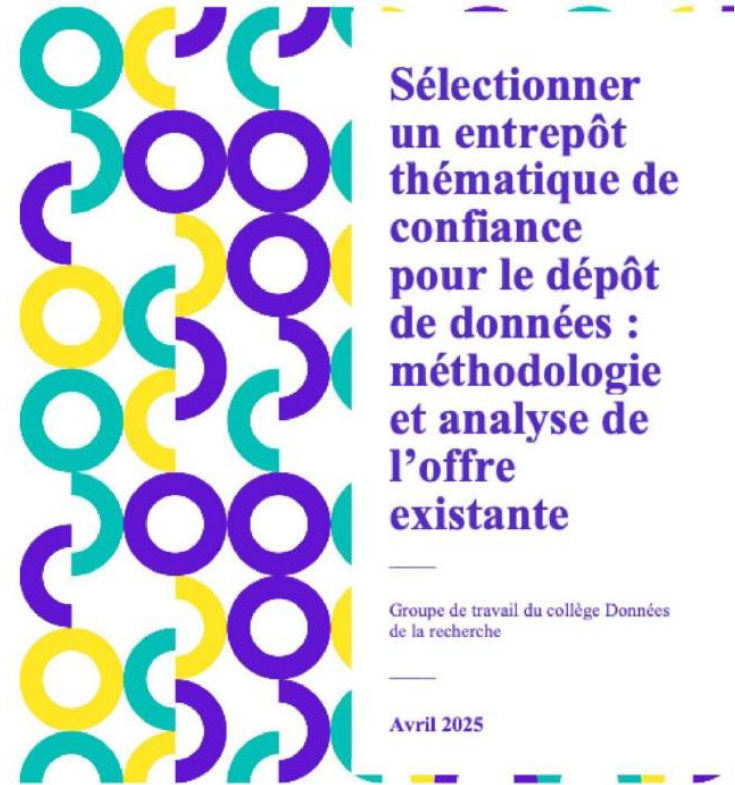
[Entrepôt Recherche Data Gouv](#)



Espace institutionnel
de l'Université de Strasbourg



Les critères retenus pour établir la liste



GT dédié sur les entrepôts thématiques de confiance

- Proposer une définition de critères de confiance permettant d'apprécier la qualité d'un entrepôt à l'aune des objectifs de «FAIRisation» des données
- Œuvrer à l'identification d'entrepôts thématiques de confiance

Sources d'identification des entrepôts

- Le **groupe de travail dédié** du Collège données de la recherche, composés de **membres du Collège** et d'**experts extérieurs**
- La **littérature scientifique et la littérature grise** (articles décrivant le fonctionnement des entrepôts)
- Les **annuaires d'entrepôts** (CatOpidor, Re3data, Fairsharing, Opendoar)
- Les **plateformes disciplinaires** dédiées à la gestion des données de recherche (comme le consortium allemand NFDI, Dataacc.org, Elixir...) proposant un premier recensement d'entrepôts thématiques
- Des **retours d'expérience** de la communauté scientifique

Entrepôt de données de confiance

Un **entrepôt de données** : « Infrastructure de stockage et de services facilitant le **dépôt, la description, le partage en accès ouvert, la découverte et la réutilisation, par des humains ou des machines**, de jeux de données propres à une communauté scientifique. Ces **jeux de données** sont **associés à des métadonnées** et sont **conservés à moyen ou long terme.** »

*Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante, **Collège Données de recherche, Comité pour la Science Ouverte***

Un **entrepôt** de données **thématique**:
structuré autour d'un certain type de données/ une
communauté scientifique

vs.

Un **entrepôt généraliste/pluridisciplinaire** :
peut accepter des données pluridisciplinaires - ne
s'adresse pas à une communauté spécifique

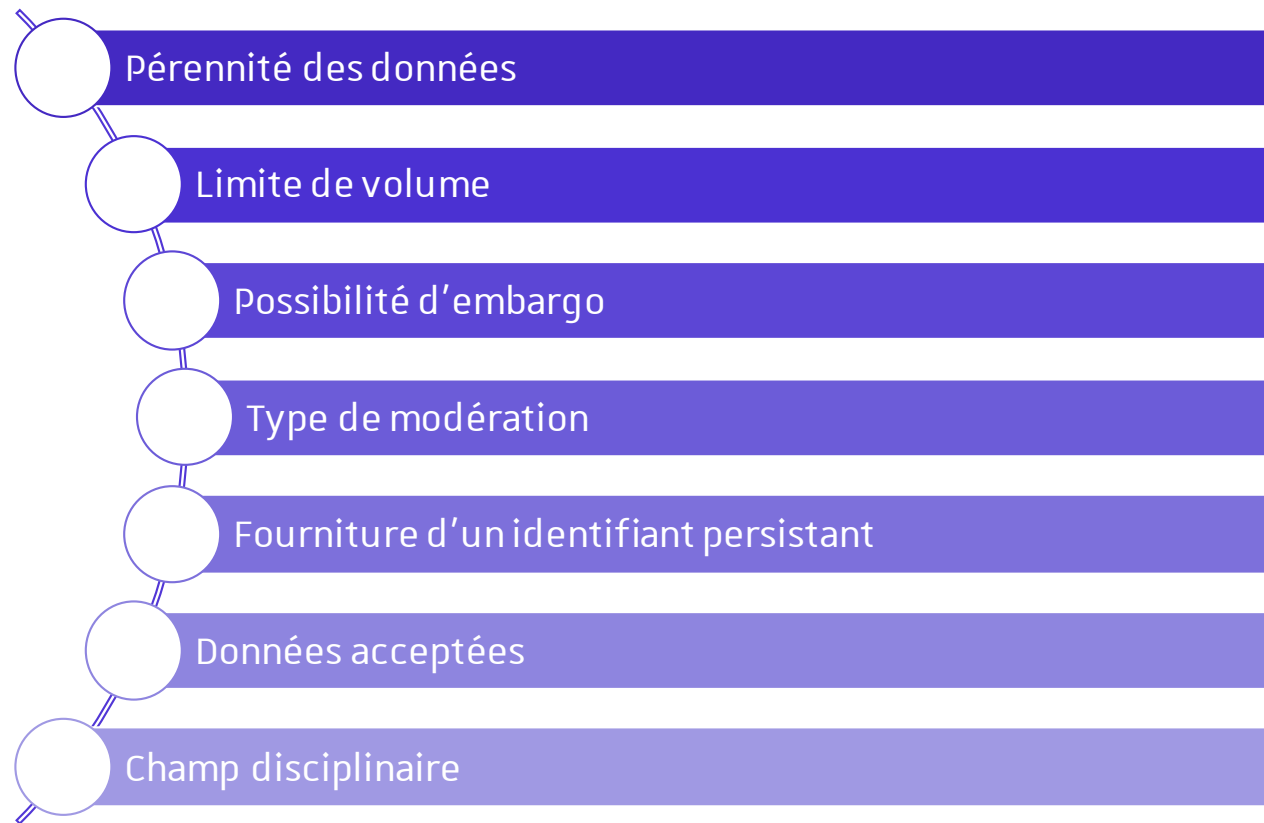
Un **entrepôt** de données thématique « **de confiance** »?

Capacité à répondre à certain nombre de critères:

- Données intelligibles et réutilisables grâce aux métadonnées et à la documentation
- Accès aux données dans la durée (pérennité de l'infrastructure)
- Politique de modération lors du dépôt

Liste de critères de confiance d'un entrepôt thématique

Critères retenus (critères de description)



Critères retenus (critères de description)

Pérennité des données

Engagement de l'entrepôt à **préserver** les données déposées pendant une **période de temps expressément définie**

Limite de volume

Information importante pour les disciplines générant des **volumes importants de données**. Permet **d'anticiper le coût à prévoir si participation financière demandée** à partir d'un certain volume

Possibilité d'embargo

Un embargo peut être souhaité par le chercheur ou demandée par le financeur et/ou l'éditeur. **Tous les entrepôts n'offrent pas cette option**

Type de modération

Vérification des métadonnées, intervention humaine ou automatisée, contrôle scientifique des données,...

Fourniture d'un identifiant persistant

Tout identifiant persistant (DOI, Handle,...) contribuant à faciliter la « découvrabilité » des jeux de données selon les principes FAIR

Données acceptées

Description des types de données acceptées par l'entrepôt (emploi de la **terminologie spécifique propre à chaque discipline** afin de faciliter le choix du déposant) *Exemple: spectres RMN au format constructeur*

Champ disciplinaire

S'appuie sur la **nomenclature développée par HAL** – offre jusqu'à 3 niveaux de granularité. Permet de mieux décrire les entrepôts sélectionnés

Critères d'exclusion



Critères d'exclusion

Absence de modération des dépôts

Exclusion si aucune modération humaine ou automatisée

Absence d'identifiant persistant

Exclusion si pas proposé. identifiant persistant = PID (exemple: un DOI) ou identifiant interne

Absence de garantie sur le pérennité de l'infrastructure

L'entrepôt doit être maintenu et proposer une durée de préservation des données d'au moins 5 ans. (critère parfois difficile à définir)

Entrepôt pratiquant la cession des droits

Exclusion car ne correspond pas au cadre légal français

Politique tarifaire excessive

Entrepôts conditionnant le moindre dépôt de faible volume à des frais. *Exemple: des entrepôts qui facturent à partir de tranches de 10 Go.*

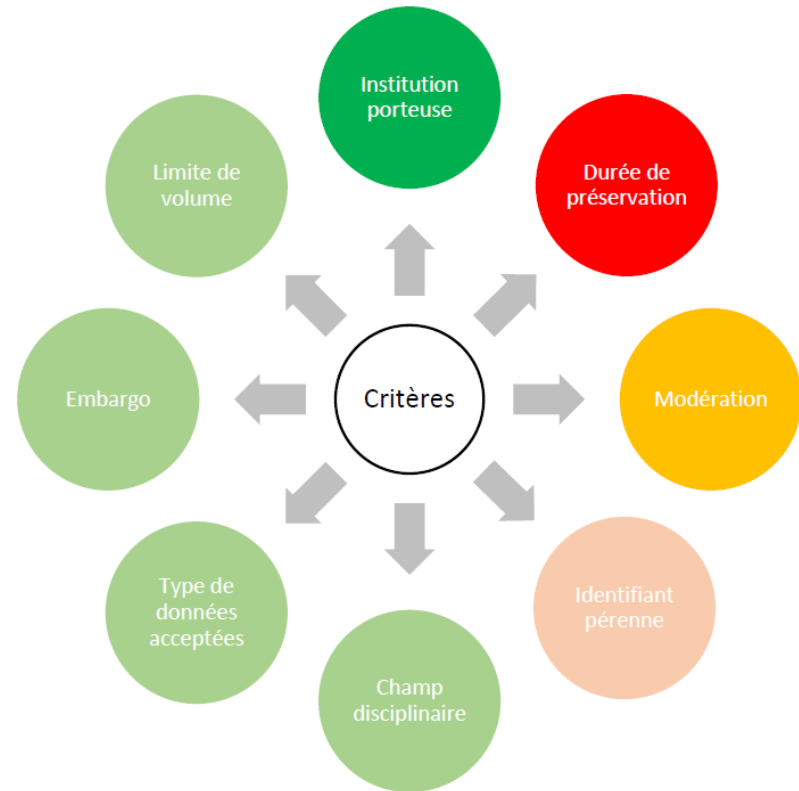
Localisation des données hors UE pour certains types de données

Point de vigilance pour les données de santé. Exclusion des pays ne respectant pas le RGPD + des entrepôts dont les dépôts ne sont pas totalement anonymisables

Dépôt restreint par l'affiliation institutionnelle

Possibilité pour le chercheur de déposer indépendamment de son affiliation institutionnelle (Pratique recherchée: auto-dépôt libre)

Difficultés d'application des critères



Difficultés d'application des critères

« La **modération n'est pas appréhendée de la même façon dans toutes les disciplines**, certaines d'entre elles souhaitant garantir une grande liberté pour le déposant quand d'autres contrôlent les métadonnées du jeu de données, les métadonnées intrinsèques et les données elles-mêmes. Dans ce contexte, il n'a pas été aisé de trouver un juste équilibre et **certains entrepôts recensés ne répondent pas complètement au critère de modération. Ils ont néanmoins été retenus pour leur conformité aux autres critères et pour leur importance au sein de leur discipline.** »

Extrait de Sélectionner un entrepôt thématique de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante

Liste des entrepôts de confiance du COSO : remarques

- ◆ Liste **ouverte et non exhaustive**
- ◆ Liste **mise à jour régulièrement**
Exemple : retrait de plusieurs entrepôts en raison du contexte géopolitique (ne veut pas dire qu'ils ne pourront pas réintégrer la liste)
- ◆ Critères de sélection: **exclusion de critères très techniques** (type interopérabilité, API, OAI-PMH: pas inclus pour l'instant)
- ◆ Pas d'analyse technique des protocoles de moissonnage et API des entrepôts concernés
- ◆ Pas d'analyse exhaustive du **degré de FAIRisation** des données déposées sur les entrepôts
- ◆ Si un entrepôt n'apparaît pas sur la liste : pas nécessairement un « mauvais » entrepôt
- ◆ La liste des entrepôts assure une **couverture française/ européenne/ internationale**
- ◆ Certains entrepôts : en cours de réflexion sur les aspects à améliorer pour mieux répondre aux critères : **échange avec les entrepôts**

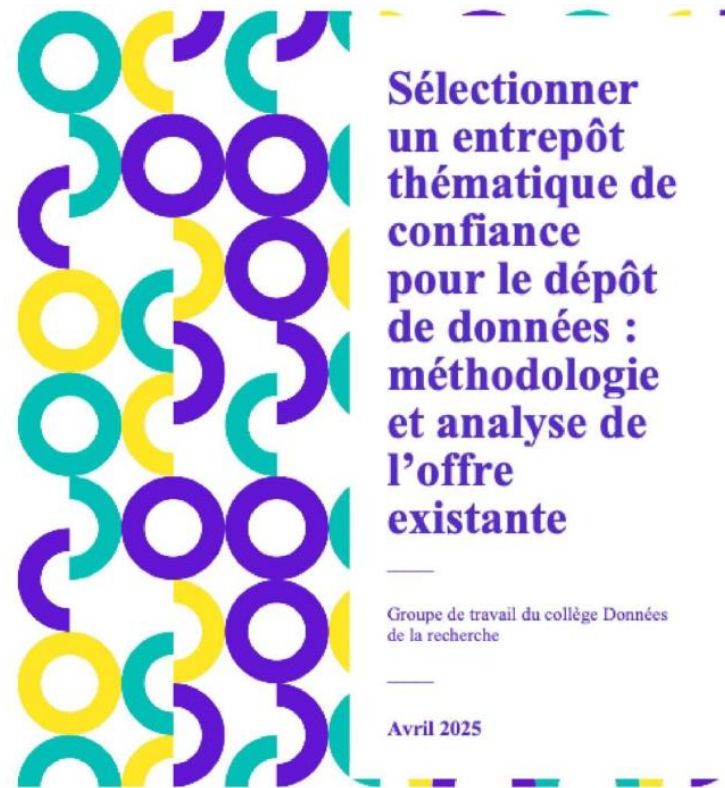
 **Contactez le groupe du COSO en charge des entrepôts de confiance:**
entrepots-confiance@groupe.renater.fr

Répartition disciplinaire des entrepôts sélectionnés

Domaine	Nombre d'entrepôts
Astronomie	2
Biologie	13
Chimie	9
Physique	2
Sciences de l'environnement	7
Sciences humaines et sociales	16 (dont 6 en archéologie et 3 en linguistique)

[Liste des entrepôts de confiance](#)
(mise à jour avril 2025)

Les entrepôts thématiques de confiance: quelques exemples



Les entrepôts de confiance: liste disponible sur Recherche Data Gouv

<https://recherche.data.gouv.fr/fr/entrepots>

ArkeOpen	Archéologie	Electron Microscopy Public Image Archive (EMPIAR)	Biologie	MorphoSource	Archéologie
Array Express	Biologie	Endangered LAnguages Repository (ELAR)	Sciences humaines et sociales	Nakala	Sciences humaines et sociales
Cambridge Structural Database (CSD)	Chimie	European genome-phenome archive (EGA)	Biologie	NeuroMorpho	Biologie
Centre de données astronomiques de Strasbourg (CDS)	Astronomie	European Nucleotide Archive (ENA)	Biologie	nmrXiv	Chimie
Centre de Données Socio-Politiques (CDSP)	Sciences humaines et sociales	Geovistory	Sciences humaines et sociales	NOMAD	Physique
Chemotion	Chimie	German Neuroinformatics Node (G-Node)	Biologie	Open Context	Archéologie
Collections de CORpus Oraux Numériques (CoCoON)	Sciences humaines et sociales	Global Biodiversity Information Facility (GBIF)	Sciences de l'environnement	Open Reaction Database (ORD)	Chimie
Conservatoire National des Données 3D (CND3D)	Archéologie	GnPLS	Biologie	Outils et Ressources pour un Traitement Optimisé de la LANGUE (ORTOLANG)	Sciences humaines et sociales
Crystallography Open Database (COD)	Chimie	HEP DATA	Physique	PANGAEA	Sciences de l'environnement
Data Station Archaeology (DANS)	Archéologie	Image Data Resource (IDR)	Biologie	Paris Astronomical Data Center (PADC)	Astronomie
Didómena	Sciences humaines et sociales	Inventaire des données de recherche en environnement et sociétés (data.InDoRES)	Sciences de l'environnement	Plant Trait Database (TRY)	Biologie
EaSy Data	Sciences de l'environnement	ioChem-BD	Chimie	Protein Data Bank Europe (PDBe)	Biologie
ebrains	Biologie	IsoArcH	Archéologie	PRoteomics IDentifications (PRIDE)	Biologie
EELS Data Base	Chimie	Long-term Access and Usage of Deeply Annotated Information (Laudatio)	Sciences humaines et sociales	Public-nEUro	Biologie
		MassBank	Chimie	Qualidata (UK data service)	Sciences humaines et sociales
				Qualitative Data Repository (QDR)	Sciences humaines et sociales
				Quetelet-Progedo-Diffusion	Sciences humaines et sociales
				SEANOE	Sciences de l'environnement
				Sextant	Sciences de l'environnement
				ShareLoc.XYZ - An open platform for sharing single-molecule localization microscopy data and tools	Biologie
				StrendaDB	Chimie
				SupraBank	Chimie
				World Data Center of Climate (WDC)	Sciences de l'environnement

Entrepôt de confiance en biologie: PRIDE

<https://recherche.data.gouv.fr/fr/entrepots>

PRoteomics IDentifications (PRIDe)



[Aller sur l'entrepôt](#) →

Biologie

Protéomique

Données acceptées

Données de spectrométrie de masse

Institutions porteuses

European Molecular Biology
Laboratory European
Bioinformatics Institute (EMBL-EBI)

Conseil pour la recherche en
biotechnologie et sciences
biologiques (BBRSC)

Identifiants pérennes

DOI Interne (PDX)

Limite de volume

Limite de volume par fichier : Pas de limite

Limite de volume par jeu de données : Pas de limite

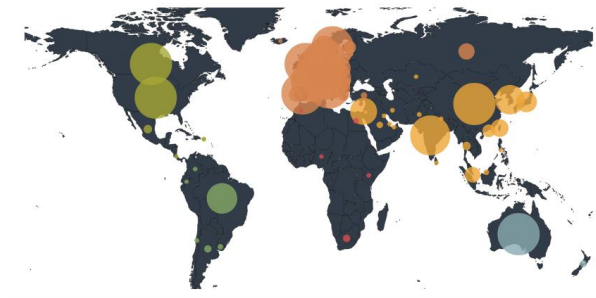
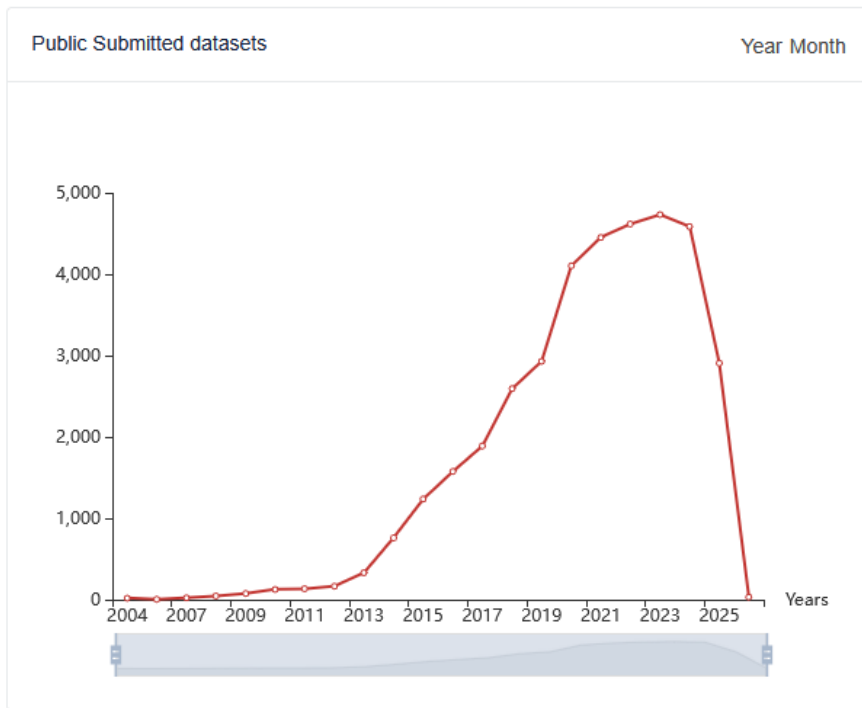
Modération assurée par l'entrepôt

Modération automatique lors du téléchargement des fichiers, puis validation par des experts du protocole expérimental, des paramètres de spectrométrie de masse et de la complétude des données. Voir https://www.ebi.ac.uk/pride/markdownpage/pridesubmissiontool#submission_details.

Remarques

Les jeux de données sont régis par la licence CC0 depuis 2018. Fait partie de l'infrastructure ELIXIR. <https://elixir-europe.org/platforms/data/elixir-deposition-databases>

Entrepôt de confiance en biologie: PRIDE



Jeu de données déposés par pays :

France : 1786

Etats-Unis : 7684

Chine : 4287

Allemagne : 5614

Inventaire des données de recherche en environnement et sociétés (data.InDoRES)



Bioarchéologie

Ecologie

Environnement

Données acceptées

Données environnementales, Données écologiques, Données géographiques

Institutions porteuses

CNRS Écologie & Environnement (CNRS-INEE)

Muséum national d'Histoire naturelle (MNHN)

Identifiants pérennes

DOI

Limite de volume

Limite de volume par fichier : Pas de limite précisée

Limite de volume par jeu de données : 2 Go par jeu de données mais passage prévu à 4 ou 5 Go prochainement.

Modération assurée par l'entrepôt

La modération est opérée soit par l'équipe data.InDoRES soit par le référent local du réseau OpenDoRES de l'unité de recherche quand il existe. La complétude des métadonnées ainsi que la compréhension des contenus sont vérifiés par le modérateur. Le contenu scientifique est quant à lui laissé à la responsabilité du déposant.

Remarques

Les données déposées dans cet entrepôt qui rentrent dans le périmètre d'EaSy Data seront moissonnées pour être visibles dans EaSy Data. Le périmètre des données déposables inclut les disciplines au carrefour entre les SHS et l'écologie.



[Aller sur l'entrepôt](#) →

Sciences de l'environnement

Sciences de la Terre

Données acceptées

Données de longue traîne du système terre et de l'environnement (exemple : issues de projet), hors données d'observation prises en charge par les poles de données. Pas de format imposé mais des préconisations.

Institutions porteuses

Data Terra

Identifiants pérennes

DOI

Limite de volume

Limite de volume par fichier : 5 Go

Limite de volume par jeu de données : 100 Go par dépôt. Possibilité de formuler une demande en cas de volumétrie plus importante.

Modération assurée par l'entrepôt

Modération sur les métadonnées et le périmètre. Une prise en charge de la modération est effectuée par des experts de domaine dans chaque pôle de données de Data Terra. Une interface de modération est fournie. Modération organisée par thématiques (liste thématique ici : https://terra-vocabulary.org/ncl/DataTerraRepositoryFairIncubator/_Voc_thematique).

Remarques

Licence Creative Commons avec CC-BY-4.0 par défaut. Mise en place d'une organisation interne pour modération/curation (modérateurs i.e. experts thématiques des poles de données) et pour amélioration des pratiques (référents des poles de données + ateliers de la donnée). Possibilité de déposer en Fr ou en En. Principes FAIR respectés. Certification CTS sera demandée



[Aller sur l'entrepôt](#) →

Sciences humaines et sociales

Données acceptées

Données de recherche issues de projets scientifiques en Sciences Humaines et Sociales

Institutions porteuses

CNRS

Huma-Num

Ministère de l'Enseignement
supérieur et de la Recherche

Identifiants pérennes

DOI

Limite de volume

Limite de volume par fichier : Il n'y a pas formellement de limitation pour la taille des données déposées. Cependant, s'il est envisagé de déposer de gros volumes (supérieurs à 10 Go par fichier), il est nécessaire de prendre contact avec l'équipe NAKALA au préalable. Voir :

Limite de volume par jeu de données : Pas de limite formelle

Modération assurée par l'entrepôt

La modération générale est gérée par le contrôle des accès des déposants et la vérification automatique d'une série de métadonnées. Une modération humaine est réalisée sur demande, le jeu de données déposé ayant alors un label de qualité "Données publiques modérées." Voir : <https://documentation.huma-num.fr/nakala-qualite/>.

Pourquoi déposer ses données dans un entrepôt?

- ◆ Visibilité
- ◆ Reproductibilité, transparence, intégrité scientifique
- ◆ Favoriser le partage et la réutilisation des données pour de nouvelles questions de recherche
- ◆ Eviter le gaspillage de ressources (répéter les mêmes expériences)
- ◆ Bonne pratique de gestion des données (principes FAIR)

Pour la première fois à l'échelle d'un pays entier, la France, une étude montre que les pratiques de science ouverte sont associées à une augmentation des citations



<https://doi.org/10.48550/arXiv.2508.20747>

The citation advantage of linking publications to research data

<https://doi.org/10.1371/journal.pone.0230416>

Pourquoi déposer ses données dans un entrepôt?

Pratique déconseillée : livrer ses données en tant que « Supplementary material » à l'éditeur scientifique

Data Availability Statement et Supplementary Materials:

- ♦ Très souvent les données ne sont pas
 - « disponibles »
 - « interopérables »
 - « réutilisables »
- ♦ Le DAS associé à un article n'est pas une garantie nécessairement suffisante (cf vidéo)
- ♦ Risque de transfert des droits vers l'éditeur
- ♦ Pour que les données soient utiles, puissent être analysées, réinterprétées, elles doivent être organisées, intelligibles - principes FAIR



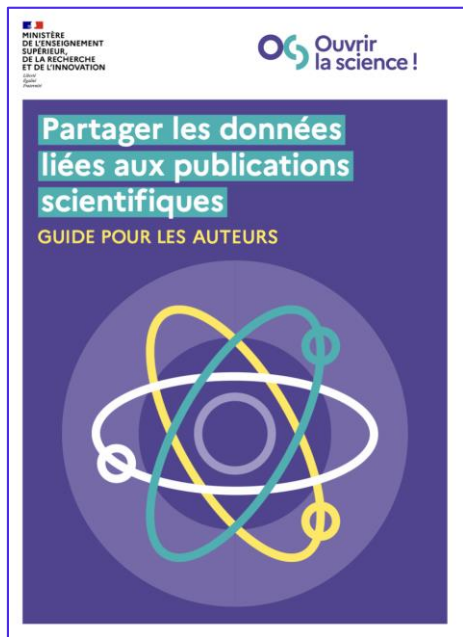
**Data Sharing and Management Snafu in
3 Short Acts**

[NYU Health Sciences Library](https://www.youtube.com/watch?v=66oNv_DJwDc)

https://www.youtube.com/watch?v=66oNv_DJwDc

Comment procéder pour sélectionner un entrepôt ?

Pratique déconseillée : livrer ses données en tant que « Supplementary material » à l'éditeur scientifique



En Bref

À PRIVILÉGIER



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.

À ÉVITER



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

Partager des données liées à une publication

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

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

Le Collège Données de la recherche du Comité pour la science ouverte

Missions Projets associés Productions associées L'équipe

Principaux travaux

- ♦ Proposer une méthodologie pour sélectionner des entrepôts de confiance
- ♦ Etablir et maintenir la liste des entrepôts de confiance
- ♦ Elaborer des recommandations sur les PIDs
- ♦ Soutenir la certification CTS pour les entrepôts français
- ♦ Faciliter l'appropriation des plans de gestion des données
- ♦ Faciliter l'appropriation des principes FAIR
- ♦ Evaluer la qualité d'une donnée et d'une métadonnée
- ♦ Gérer les données de recherche dans les projets de science participative
- ♦ Contribution au passeport Science ouverte et livret sur les données

<https://www.ouvrirlascience.fr/college-donnees-de-la-recherche/>



En conclusion : Comment sélectionner un entrepôt pour ses données de recherche?

- ◆ Sonder sa communauté de recherche
- ◆ Evaluer le degré de fiabilité de l'entrepôt à l'aune des critères du COSO
- ◆ Contacter son atelier de la donnée de proximité :

ADELE Helpdesk

