

# Séminaire SIST 24 - 6 juin 2024 - Marseille

## Session Entrepôts et écosystème de la science ouverte



# Panorama général des entrepôts de données de confiance au sein de l'écosystème de la science ouverte



Emilie Lerigoleur

**GEODE**  
GÉOGRAPHIE DE L'ENVIRONNEMENT



 emilie-lerigoleur

 0000-0002-0864-659X

IdRef : 253124867

 emilie.lerigoleur@univ-tlse2.fr



# Entrepôt de données de recherche / entrepôt de confiance

Source : <https://doranum.fr/glossaire-donnees-recherche/>

« Un **entrepôt de données de recherche** (Research Data Repository ou Data Repository) est une plateforme destinée à accueillir, conserver, rendre visibles et accessibles des données de recherche. Son rôle est de permettre le dépôt ou la collecte de données, leur description, leur accès, et leur partage en vue de leur réutilisation. Chaque entrepôt dispose généralement d'une politique de dépôt, de description et de diffusion des données. »

Il peut être disciplinaire ou thématique, être institutionnel ou centralisé.

# data warehouse

Un **entrepôt de confiance** (Trusted repository) répond aux critères de qualité exigés pour obtenir une certification (format des données, qualité des métadonnées, conditions d'accès et de réutilisation, identifiant pérenne, archivage à long terme...).

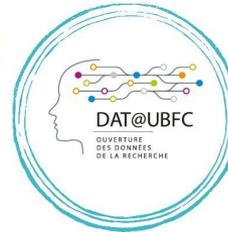
De nombreux entrepôts n'ont pas de certification mais sont cependant largement reconnus par la communauté scientifique et offrent des garanties de conservation à long terme.

Rem : certaines infrastructures de données ou infrastructures de recherche ou plateformes ou portails intègrent un **service d'entrepôt**



# Il existe une multitude d'entrepôts de données généralistes, institutionnels, multidisciplinaires, thématiques, certifiés, de confiance

## Quelques exemples



# L'écosystème de la science ouverte en France :

## Recherche Data Gouv

**14 ateliers de la donnée** (5 en cours)  
de proximité, 1er niveau expertise gestion raisonnée



**6 centres de référence thématiques**



**3 centres de ressources**  
DORANum, GIS URFIST, Centre de ressources entrepôt-catalogue RDG



**1 catalogue de données** (en cours)



# L'écosystème de la science ouverte en France :

## Les infrastructures de recherche de la feuille de route 2021

### Domaines principaux et secondaires

Sciences Humaines et Sociales 4

Sciences du Système Terre et de l'Environnement 22

Énergie 4

Biologie et Santé 28

Sciences de la matière et ingénierie 13

Astronomie et Astrophysique 11

Physique nucléaire et des hautes énergies 12

Recherche en Sciences du Numérique et Mathématiques 3

Information scientifique et technique 6

Services numériques (calcul et réseau) 5

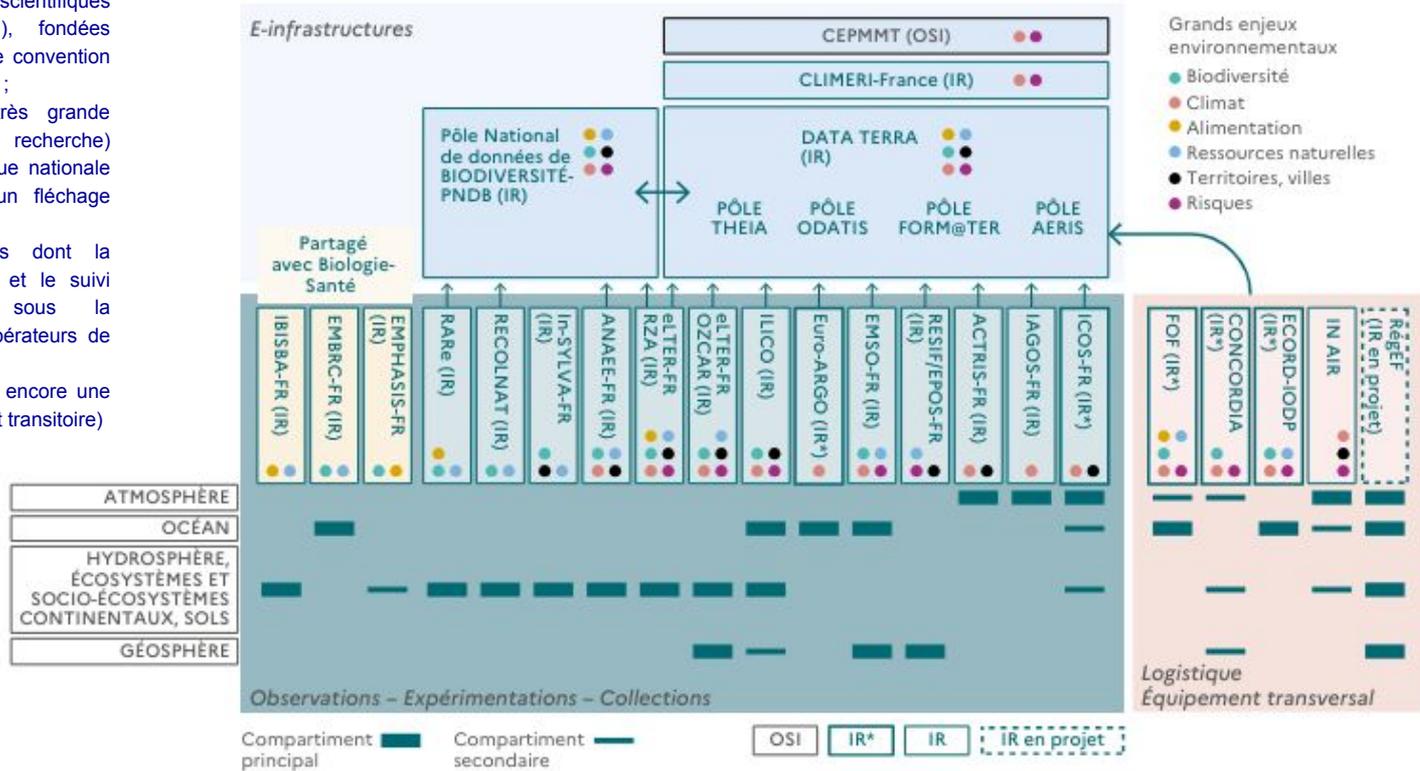


<https://data.enseignementsup-recherche.gouv.fr/pages/feuille-de-route-2018/?sort=acronyme>





**OSI** = organisations scientifiques internationales (OSI), fondées juridiquement sur une convention intergouvernementale ;  
**IR\*** = ex TGIR (très grande infrastructures de recherche) relevant d'une politique nationale et faisant l'objet d'un fléchage budgétaire du MESR ;  
**IR** = infrastructures dont la stratégie scientifique et le suivi budgétaire sont sous la responsabilité des opérateurs de recherche ;  
**projet** = n'ayant pas encore une pleine maturité ( statut transitoire)



© C. Marfin, S. Mahé et Groupe Infrastructure d'AlEnn

Figure 1 : OSI/IR\*/IR du domaine SST & ENV par grands types (observations – expérimentaux – collections, logistiques et e-infrastructures), par grands compartiments du système Terre (atmosphère, océan, hydrosphère–écosystèmes – socio–écosystèmes continentaux – sols et géosphère) et par grands enjeux environnementaux (biodiversité, climat, alimentation, ressources naturelles, territoires-villes, risques). En jaune figurent trois infrastructures partagées avec le domaine Biologie – Santé.



DOMAINE SCIENTIFIQUE PRINCIPAL	DOMAINE(S) SCIENTIFIQUE(S) SECONDAIRE(S)	ACRONYME	TITRE COMPLET DE L'INFRASTRUCTURE	TYPE	STATUT ESFRI
Sciences du système Terre et de l'environnement		CEPMMT	Centre Européen pour les Prévisions Météorologiques à Moyen Terme	OSI	
		CONCORDIA	CONCORDIA – station de recherche antarctique franco-italienne	IR*	
		ECORD/IODP	Programme international de forage profond en mer/European Consortium for Ocean Drilling Research/International Ocean Discovery Program	IR*	
		EURO-ARGO-France	Réseau in-situ global d'observation des océans/ European contribution to Argo program	IR*	EURO-ARGO : ESFRI landmark
		FOF	Flotte Océanographique Française	IR*	
		ICOS-France	Système Intégré d'Observation du Carbone/ Integrated Carbon Observation System	IR*	ICOS : ESFRI landmark
		ACTRIS-France	Aerosol, Cloud and Trace Gases Research Infrastructure – France	IR	ACTRIS : ESFRI landmark
		AnaEE-France	Analyse et Expérimentation sur les Écosystèmes – France	IR	ANAEE : ESFRI landmark
		CLIMERI-France	Infrastructure de recherche nationale de modélisation du système climatique de la Terre	IR	
		DATA TERRA	Pôles de données et services pour le système Terre	IR	
	eLTER-France OZCAR	Observatoires de la Zone Critique, Applications et Recherche	IR	eLTER : ESFRI project	
	eLTER-France RZA	Réseau des Zones Ateliers – Infrastructure des Socio-écosystèmes	IR	eLTER : ESFRI project	

Rem : certaines de ces infrastructures ont un **service d'entrepôt**





DOMAINE SCIENTIFIQUE PRINCIPAL	DOMAINE(S) SCIENTIFIQUE(S) SECONDAIRE(S)	ACRONYME	TITRE COMPLET DE L'INFRASTRUCTURE	TYPE	STATUT ESFRI
Sciences du système Terre et de l'environnement		EMSO-France	European Multidisciplinary Seafloor and water column Observatory – France	IR	EMSO : ESFRI landmark
		IAGOS-France	Instruments de mesure embarqués sur avions pour l'observation globale/ In-service Aircraft for Global Observing System	IR	IAGOS : ESFRI landmark
		ILICO	Infrastructure de Recherche Littorale et COTière	IR	
		IN AIR	Infrastructure Nationale des Aéronefs Instrumentés pour la Recherche	IR	
		In-Sylva-France	In-Sylva France Infrastructure Nationale de recherche pour la gestion adaptative des forêts	IR	
		PNDB	Pôle National de Données de Biodiversité	IR	
		RARe	Ressources Agronomiques pour la Recherche	IR	MIRRI : ESFRI landmark
		RECOLNAT	Réseau national des collections naturalistes	IR	DiSSCo : ESFRI project
		RESIF/EPOS	Réseau sismologique et géodésique français/ European Plate Observing System	IR	EPOS : ESFRI landmark
	RÉGEF	Réseau Géochimique et Expérimental Français	Projet		

Rem : certaines de ces infrastructures ont un **service d'entrepôt**



# Les critères pour choisir un entrepôt de données





**Vous ne savez pas quel  
entrepôt choisir**



**Faites une  
recherche  
dans un  
annuaire**

**Vous avez une idée de l'entrepôt  
qui pourrait vous convenir et  
souhaitez en savoir plus pour  
consolider votre choix**



**Consultez la  
politique de  
l'entrepôt**



**Naviguez sur  
le site de  
l'entrepôt**



**Il est important de choisir l'entrepôt avec soin et de  
vérifier s'il permet l'application des principes FAIR**

[https://doranum.fr/depot-entrepots/criteres-pour-choisir-entrepot\\_de\\_donnees\\_10\\_13143\\_zqpb-9449/](https://doranum.fr/depot-entrepots/criteres-pour-choisir-entrepot_de_donnees_10_13143_zqpb-9449/)



# Les annuaires recensant les entrepôts, dont les certifiés

## ● Annuaires internationaux

- [re3data.org](https://re3data.org) en recense 3218 dont [132](#) dirigés par une institution française avec possibilité de filtrer les certifiés ( ⚠ mise à jour !)
- [fairsharing.org](https://fairsharing.org) dénombre 2126 “databases”
- [CoreTrustSeal](https://www.coretrustseal.com) a certifié [126](#) entrepôts dont 8 français, gage de qualité :

Strasbourg Astronomical Data Centre < 2027	RESIF seismological Data Repository < 2027	IPSL, Institut Pierre-Simon Laplace < 2027	IFREMER-SISMER < 2026
Center for Socio-Political Data (CDSP) < 2026	ORTOLANG < 2026	European Synchrotron Radiation Facility < 2026	IDOC - IDOC-DATA Paris-Saclay < 2025

Certificates ☰
CLARIN certificate B (27)
CoreTrustSeal (151)
DIN 31644 (1)
DINI Certificate (11)
DSA (15)
ISO 16363 (1)
RatSWD (41)
Trusted Digital Repository (3)
WDS (45)
other (28)

## ● Annuaires nationaux

- [cat.opidor](https://cat.opidor.org) recense [177](#) entrepôts de données



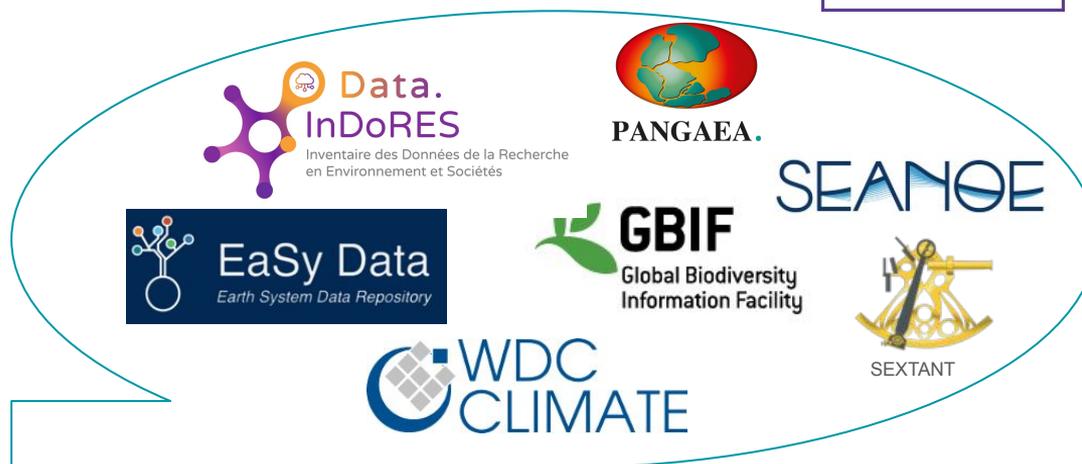
# La liste des 49 entrepôts de confiance produite par le collège “Données de la recherche” du Comité pour la Science Ouverte (CoSO)

## 2 livrables produits :

- La [note méthodologique](#) propose des critères d'exclusion et des critères de description pour orienter les communautés scientifiques dans leur choix d'un entrepôt pour l'ouverture ou le partage de leurs données. <[hal-04534321](#)>
- Liste des 49 entrepôts de confiance validée par le ministère MESR (I. Blanc)  
→ liste évolutive ! cf. <[tableur téléchargeable](#)> et <[lien sur site Recherche Data Gov](#)>



- archéologie **6**
- astronomie **2**
- biologie **10**
- chimie **9**
- neurosciences **3**
- physique **2**
- sciences de l'environnement **7**
- SHS **10**



## En bref...

- Un écosystème de la science ouverte national et international en évolution, avec une multitude d'entrepôts et d'infrastructures disposant ou non d'un service d'entrepôt
- Plusieurs annuaires listent ces entrepôts, dont les certifiés
- Choisir le bon entrepôt selon différents critères, en fonction de sa discipline, de son organisme d'appartenance, de ses tutelles
- Des listes d'entrepôts de confiance sont publiées et seront actualisées

### → Focus sur des entrepôts et infrastructures en lien avec les données d'observation :

- **L'entrepôt Easy Data**
  - Christelle Pierkot
- **Dat@UBFC : un exemple d'un atelier de la donnée labellisé par Recherche data gov**
  - Julien Pergaud, Léna Bouillard, Maelys Liseron-Monfils, Raphaël Melior, Rachel PROST, Nadir TABBOU, Hélène Tisserand, Catherine BERTRAND, Jean-François Couchot
- **Réseau des Zones Ateliers : La place du chercheur dans le dépôt des données**
  - Jérôme Dutroncy, Cécile Pignol, David Sarramia, Mathias Rouan, Wilfried Heintz, Françoise Le MOAL, isabelle charpentier

