



DATA
TERRA



CENTRES DE RÉFÉRENCE THÉMATIQUES

Systeme Terre-Environnement

Contribuer à la science ouverte

Par
Ghislaine Abbassi, responsable Communication Data Terra
et
Olivier Norvez, responsable Animation PNDB



Love Data Week

Nancy, Université de Lorraine, 16 mars 2023



SOMMAIRE



DATA
TERRA



Pôle National
de Données de Biodiversité

01

LES DÉFIS DE
DEMAIN

02

IRs Data Terra et PNDB
QUI SOMMES-NOUS ?

03

CENTRES DE RÉFÉRENCE
THÉMATIQUES



DATA
TERRA



01 LES DÉFIS DE DEMAIN

LA TERRE, UN SYSTÈME DYNAMIQUE COMPLEXE

- De nombreux processus géophysiques et environnementaux,
- A différentes échelles spatiales et temporelles,
- Interactions permanentes entre la Terre solide, les surfaces continentales, l'océan, les compartiments atmosphériques et anthroposphère.

ÉVOLUTION DES BESOINS

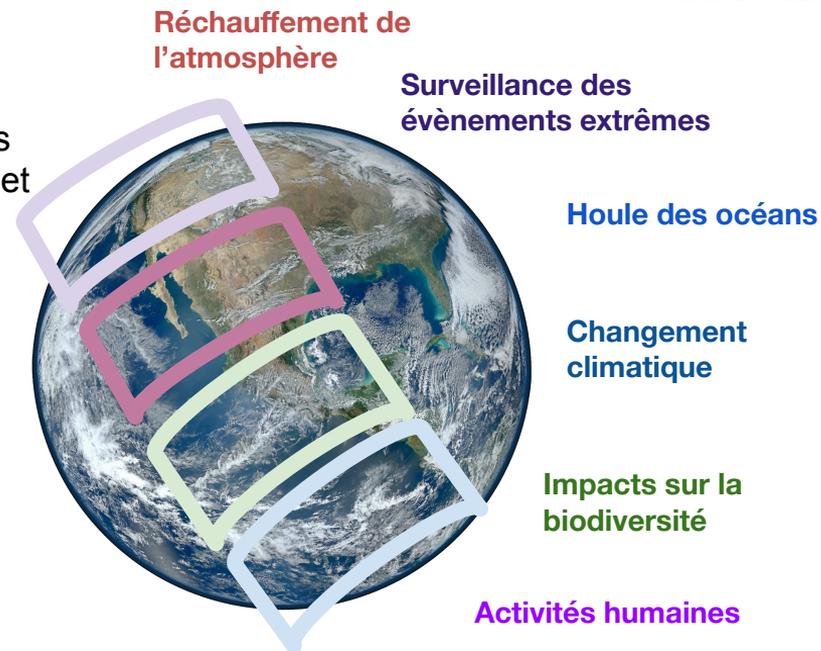
- **Approches plus intégrées de la complexité**
- **Données multi-sources, multi-capteurs permettant une multi-échelle (in situ, spatiale, spectrale et temporelle), séries long terme.**
- **Gestion de la diversité et volume : services cloud, Big Data et IA.**
- **Interdisciplinarité et transversalité des communautés scientifiques**



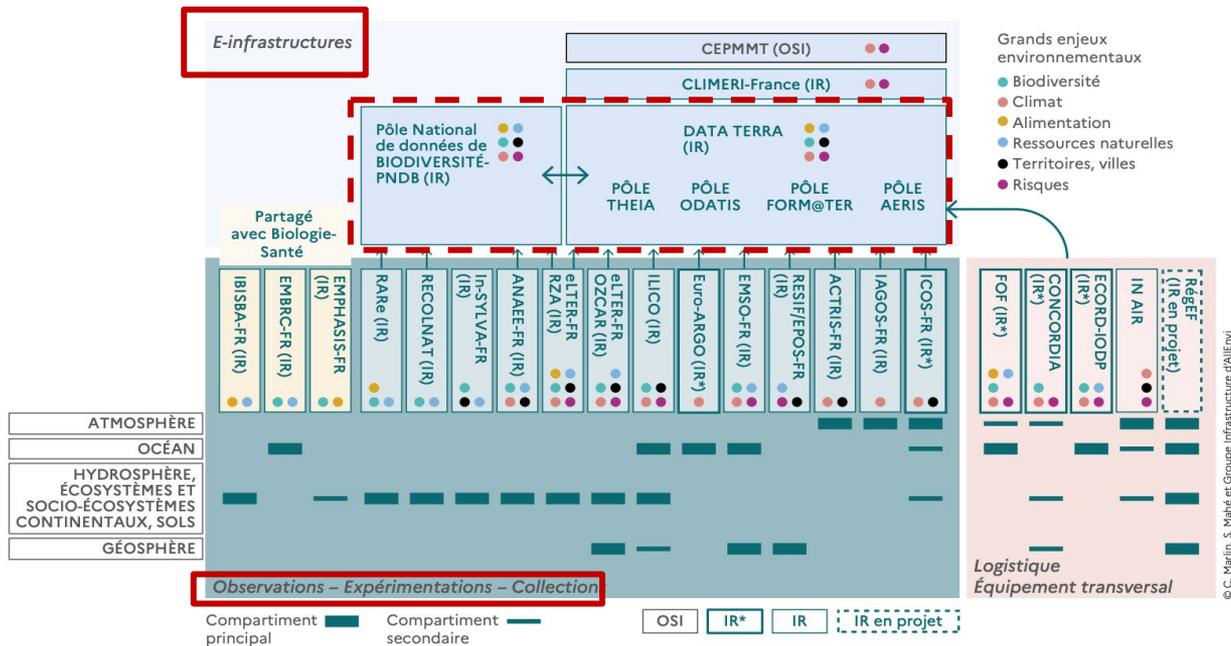
DATA
TERRA



Pôle National
de Données de Biodiversité



PAYSAGE ACTUEL DES INFRASTRUCTURES DE RECHERCHE



PNDB et DATA TERRA sont **deux E-Infrastructures de Recherche** par **Grands compartiments du système Terre** et **grands enjeux environnementaux** :

- PNDB : Biodiversité
- THEIA : Surf.continentales
- ODATIS : Ocean
- FORM@TER : Terre Solide
- AERIS : Atmosphere

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.

LA RÉPONSE DES IRs AUX DÉFIS SCIENTIFIQUES ET TECHNIQUES

Contribution aux initiatives internationales et européennes en **appui aux politiques publiques de développement durable**


constellation satellites
optique, radar...

Développement de services **sur le cycle complet de la donnée**

Accès simplifié aux données multi-sources


Observations long-terme In-Situ

Se baser sur les capacités, institutions, structures et **ressources existantes**

DONNÉES
ENVIRONNEMENT
/ CLIMAT / TERRE


Variables bio-geochimiques

Approches multidisciplinaires intégrées pour l'utilisation des données de recherche d'observation de la Terre

Interopérabilité des services


Modèles, algorithmes...

Comprendre les données par les {méta}données : définitions et contexte

"Les données de recherche sont définies comme des enregistrements factuels utilisés comme sources primaires pour la recherche scientifique, et qui sont généralement acceptés dans la communauté scientifique comme nécessaires pour valider les résultats de la recherche."

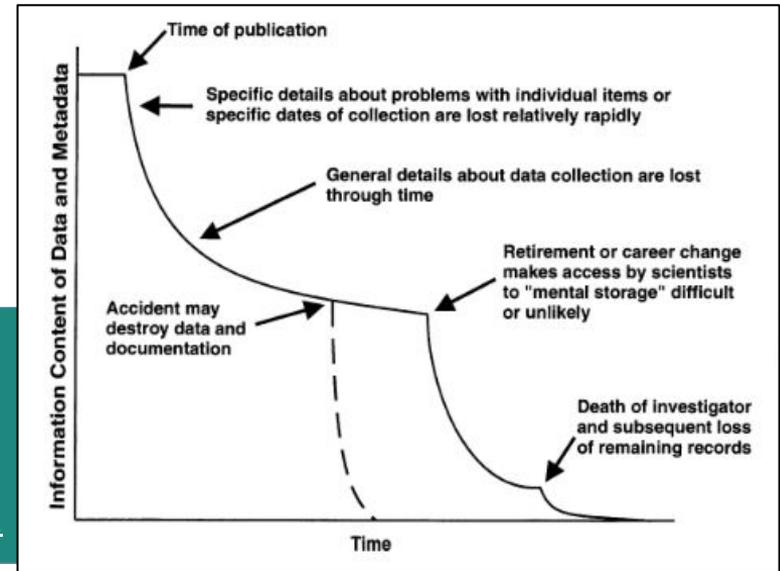
OCDE, 2007. <https://www.oecd.org/sti/inno/38500813.pdf>

OECD Principles and Guidelines for Access to Research Data from Public Funding



"Les métadonnées, que l'on peut définir simplement comme « des données sur les données », sont un moyen de nommer les choses et de représenter les données et leurs relations."

Christine L. Borgman, 2020. <https://books.openedition.org/oep/14692>



Principes FAIR et Cycle de la donnée

Planification

Gestion des données

Acquisition

Acquisition et sécurité des données

Stockage

Sauvegarde des données

Publication

Partage des données

Réutilisation

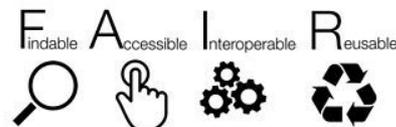
Découverte des données et citation

Pérennisation

Archivage des données

un ensemble de principes **directeurs** pour gérer les données de la recherche visant à les rendre **faciles à trouver, accessibles, interopérables et réutilisables** par l'homme et la machine.

Hétérogénéité dans les types de données, leur provenance, les standards utilisés
Diversité d'objets à relier entre eux
Perte d'informations dans le temps
Reproductibilité





DATA
TERRA



02 LES IRs DATA TERRA et PNDB

DATA TERRA, une e-Infrastructure de Recherche dédié au système Terre



€42m
(2020)



1000
produits &
services



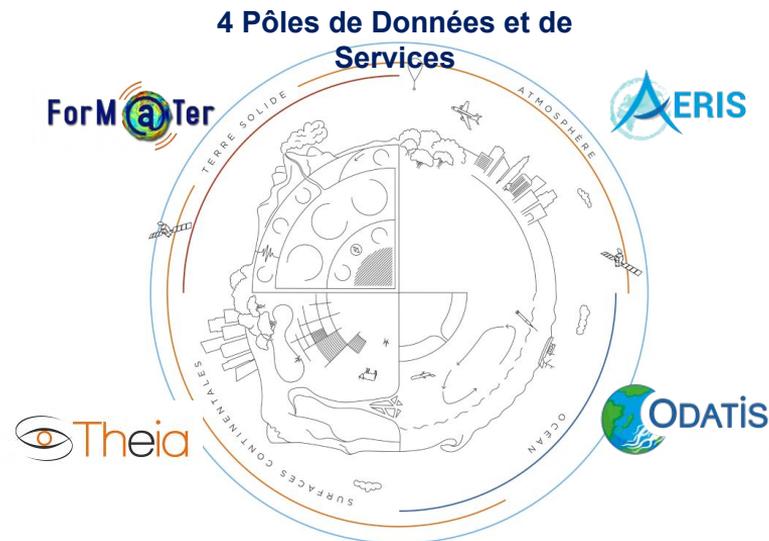
15,000
usagers



100,000 TB
(2022/2023)

La mission principale de l'IR DATA TERRA est de développer un **dispositif global d'accès et de traitement de données, produits et services** dédiés à l'**observation de la Terre**.

- 26 organismes de recherche et universités partenaires
- 30 Centre de Données et de Services
- 32 Conseils scientifiques
- 200 FTE / 450 scientifiques, ingénieurs et techniciens



2 Dispositifs
DATATERRA INTER-PÔLES
Imagerie spatiale THRS
FAIR Data Communautés



MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR
ET DE LA RECHERCHE
*Liberté
Égalité
Fraternité*



PNDB, le Pôle National de Données de Biodiversité



Pôle National
de Données de Biodiversité

2018, Inscription sur la **feuille de route des Infrastructures** nationales de recherche du Ministère de l'Enseignement Supérieur, de la Recherche (MESR), pilotée par Allenvi.

Portage par **18 établissements partenaires**

(+ Fondation pour la recherche sur la biodiversité, **FRB**).

=> Mise en œuvre confiée au **Muséum National d'Histoire Naturelle** via PatriNat (OFB-MNHN-CNRS-IRD)

et un co-portage de l'animation avec la FRB



9 organismes et 10 universités

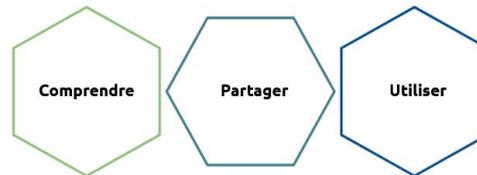


2022 : **Centre de référence thématique** "système Terre-Environnement" (MESR & Recherche Data Gov.) => Structuration des communautés autour des données

Horizon 2024 : 5ème pôle de l'IR Data Terra



=> Une e-infrastructure ***pour et par*** les communautés de Recherche pour accompagner et faciliter la compréhension, le partage et l'utilisation des {méta}données de biodiversité



www.pndb.fr



GAIA DATA, Equipex+ PIA 3 un projet structurant



DATA
TERRA



8

Années pour la réalisation du projet



21

Organismes partenaires



62 M €

Budget Total (dont 16, 2 millions d'euros de l'ANR)



400

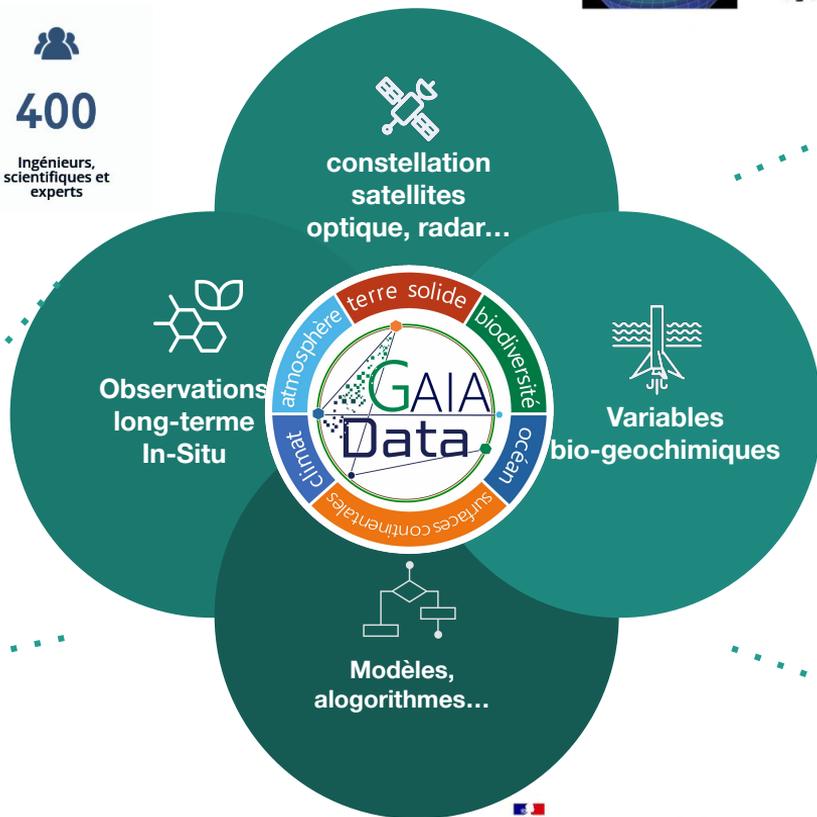
Ingénieurs, scientifiques et experts

Contribution aux initiatives internationales et européennes en **appui aux politiques publiques de développement durable**

Accès simplifié aux données multi-sources

Interopérabilité des services

www.gaia-data.org



Développement de services **sur le cycle complet de la donnée**

Se baser sur les capacités, institutions, structures et **ressources existantes**

Approches multidisciplinaires intégrées pour l'utilisation des données de recherche d'observation de la Terre



Ouvrir la science!





DATA
TERRA



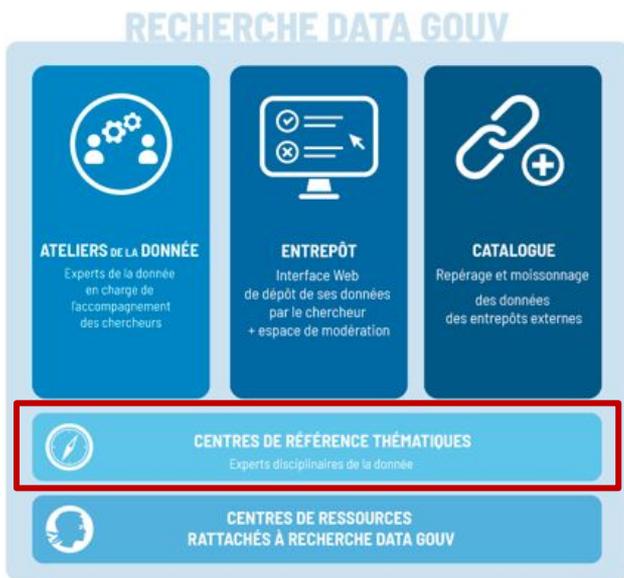
Pôle National
de Données de Biodiversité

03 CENTRES DE RÉFÉRENCE THÉMATIQUES SYSTÈME TERRE ET ENVIRONNEMENT

CENTRES DE RÉFÉRENCE THÉMATIQUES SYSTÈME TERRE ET ENVIRONNEMENT POUR CONTRIBUER A LA SCIENCE OUVERTE



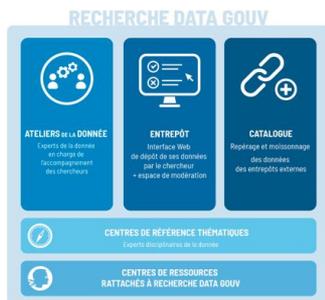
Un écosystème au service du partage et de l'ouverture des données de recherche **FÉDÉRER, ACCOMPAGNER, PARTAGER, OUVRIR, RÉUTILISER**



Structuration progressive des centres de références thématiques avec les organismes de Recherche : **DATA TERRA et PNDB** dans le domaine **Système Terre en Environnement**



CENTRES DE RÉFÉRENCE THÉMATIQUES SYSTÈME TERRE ET ENVIRONNEMENT POUR CONTRIBUER A LA SCIENCE OUVERTE



Définir

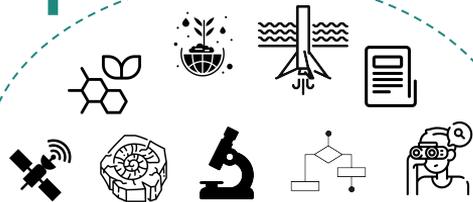
- les **normes internationales** de description et de diffusion des données (ouverture, période d'embargo, accès restreint).
- et **diffuser les bonnes pratiques** de collecte, documentation, traitement, et diffusion des données.
- la **liste des entrepôts de données de référence** (nationaux et/ou internationaux) vers lesquels orienter les chercheurs pour le dépôt des données et que *Recherche Data Gouv* moissonnera

Contribuer à la définition de l'**arborescence** des données de l'entrepôt *Recherche Data Gouv* **thématiques Système Terre et Environnement**

Soutenir l'**articulation** entre les dispositifs thématiques spécialisés et *Recherche Data Gouv* et l'**inscription** dans l'**écosystème international**



1



données Terre-Environnement : *in situ*, données satellitaires, modèles algorithmiques

Caractérisation de mon jeu de donnée (=métadonnées) via un accompagnement

- Qui est le "data owner" ?
- De quoi parle mon jeu de données ?
- Quel est le cadre d'acquisition (recherche, politiques Publiques, ...) ?
- Quel est le format et standard utilisé ?
- Publication d'un data paper en parallèle ?
- création d'un Plan de gestion de données

2



exemples de Systèmes d'Informations dont entrepôts institutionnels et catalogues de données des infrastructures

Données ouvertes : moissonnage, standardisation des métadonnées et enrichissement (description attributs et termes issus de ressources terminologiques) des métadonnées + liens directs de téléchargement

Centres de références Thématiques

Carte des ateliers de la donnée

PNDDB
Pôle National de Données de Biodiversité

DATA TERRA

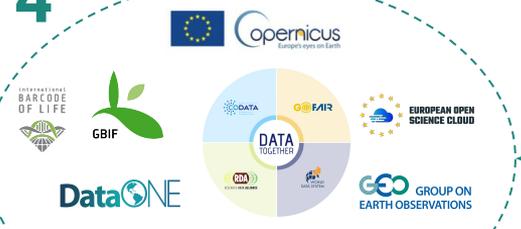
outils et services pour la description, l'accès, la validation, l'analyse et la réutilisation des données Terre-Environnement

3

Infrastructures de recherche

Métadonnées et workflows analytiques visibles sur catalogues/instances internationales

4



Systèmes d'informations et catalogues de données internationaux

PARTENAIRES DATA TERRA



PARTENAIRES PNDB





DATA
TERRA



PNDB

CENTRES DE RÉFÉRENCES THÉMATIQUES
Système Terre-Environnement

Merci ! Questions ?

Love Data Week
Nancy, Université de Lorraine, 16 mars 2023

Par
Ghislaine Abbassi, responsable Communication Data Terra
et
Olivier Norvez, responsable Animation du PNDB



Ou
Ouvrir
la science !

