

IR DATA TERRA : L'E-Infrastructure du Système Terre et environnement Pôles de données et services pour le système Terre

Emmanuel Chaljub (Univ Grenoble Alpes, OSUG), directeur pôle **ForM@Ter**,
Frédéric Huynh (IRD), directeur **IR Data TERRA**

Patrice Henry (CNES), directeur pôle **AERIS** ; Erwan Quimbert (IFREMER), directeur pôle **ODATIS** ; Anne Puissant (UNISTRA), directrice pôle **THEIA** ; Richard Moreno (CNES), **directeur technique** ; Jean-François Faure (IRD), secrétaire exécutif **DINAMIS**

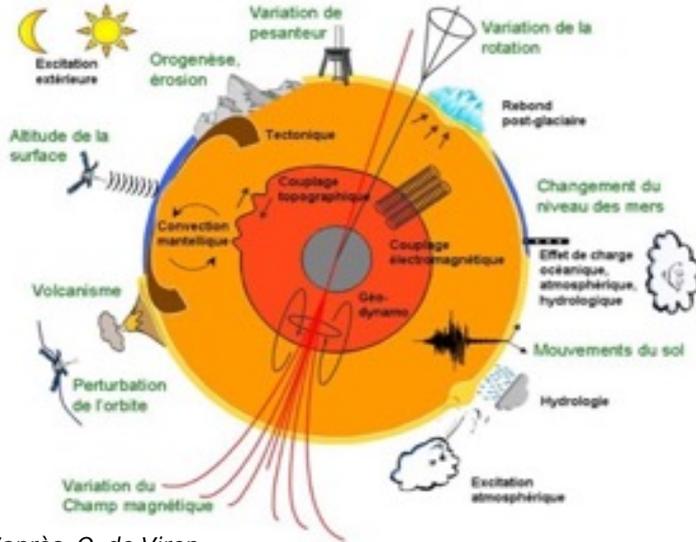
Christelle Pierkot (IR Data Terra), **Hélène Bressan** (BRGM)

Contexte et Enjeux



La Terre, un système complexe dynamique :

- processus géophysiques et environnementaux,
- différentes échelles spatiales et temporelles,
- interactions permanentes *entre les compartiments Terre solide, surfaces continentales, océan, atmosphère et l'anthroposphère.*

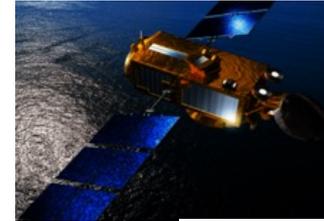


Comprendre ces processus nécessite l'**accès**, le **croisement** et l'**analyse** de **données complexes, hétérogènes** (nature, niveau, fréquence, volume, financement), d'**origines multiples** (*satellites, in-situ, campagnes d'observation, résultats d'expérimentation, sorties de modèles*), de plus en plus **massives** (missions satellites THR, nouveaux capteurs, réseaux denses, fibre optique, science citoyenne) issues de nombreux producteurs.

D'après O. de Viron

Développer un dispositif global d'accès, de traitement de **données multi-sources et de développement de services** permettant **d'observer, comprendre et prévoir** de manière **intégrée le fonctionnement et l'évolution du système Terre**

- Faciliter l'**accès** et l'**utilisation** des **données et produits** de qualité sur l'ensemble des compartiments du **système Terre**
=> **Données spatiales, aéroportées, sols, in-situ**
- Développer des **services de visualisation et de traitements** adaptés aux besoins, à l'accroissement de la volumétrie et aux avancées technologiques
- Favoriser la **mutualisation, interopérabilité**, émergence d'**approches multi- et inter-disciplinaires**
- Servir les communautés **scientifiques**, les acteurs de l'**action publique et de l'innovation**
- Mettre en œuvre une **stratégie nationale**, européenne et internationale



© Lison Bernet

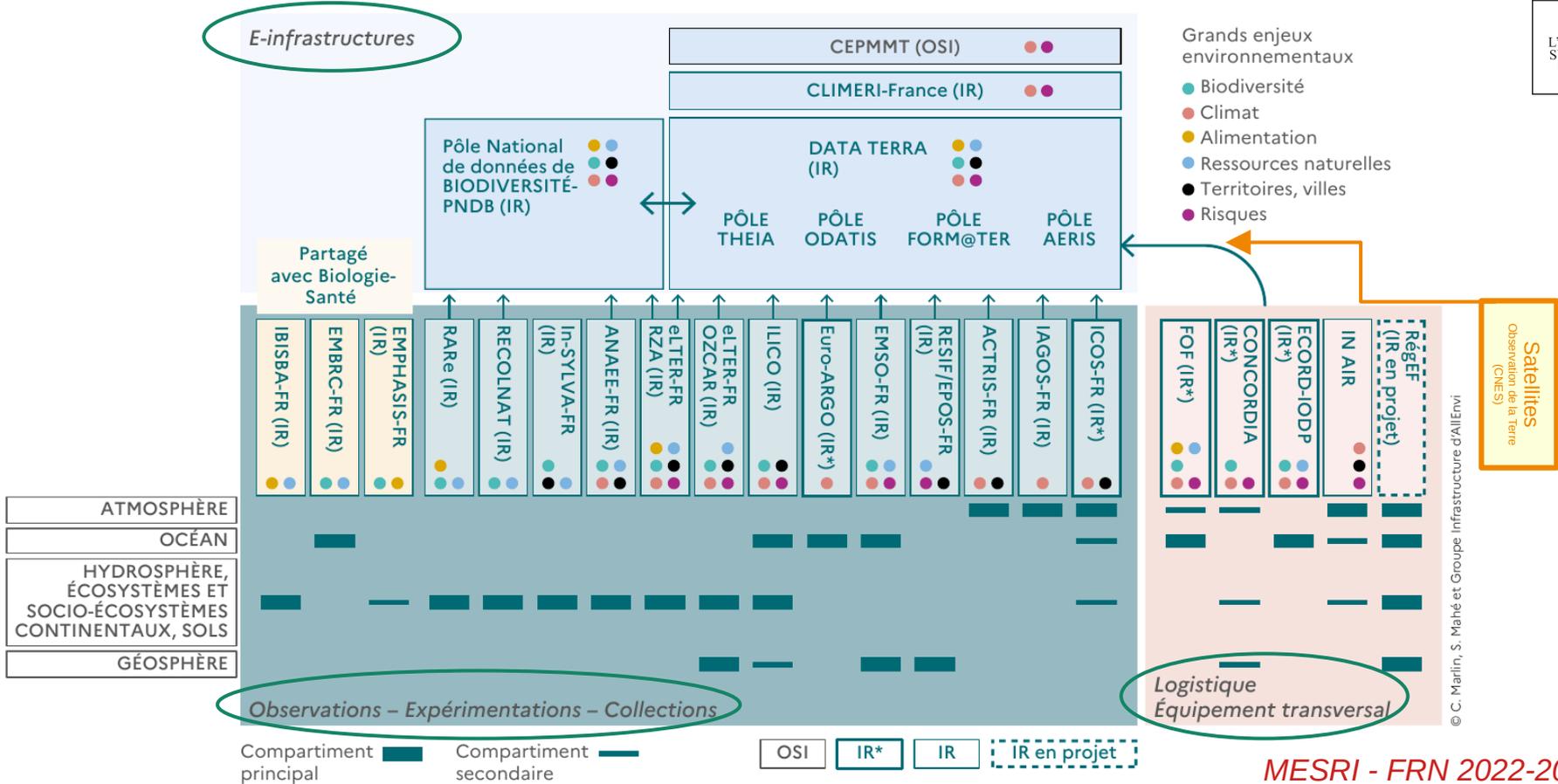


Data Terra en quelques chiffres

- 26 organismes et universités (34 d'ici 2025, avec l'intégration du PNDB !)
- **4 pôles de données** : AERIS, ForM@Ter, ODATIS, THEIA (+ PNDB d'ici 2025)
- Des services (DINAMIS) et groupes de travail transversaux (Science, Tech, Comm, International, ...)
- **30 Centres de Données (d'Observation) et de Services (CDoS) et Infrastructures de Données Spatiales (IDS)**
- 32 Consortium d'Expertise Scientifique
- 200 ETPT / 450 scientifiques, ingénieurs et techniciens
- 33 M€ (2016), 39 M€ (2017), 40 M€ (2019), 42 M€ (2020),...
- Plus de **500 produits et services**, plus de **15000 utilisateurs**
- 50 000 To (2018) ; 100 000 To (2022/2023) ; **150 Peta (2025)**

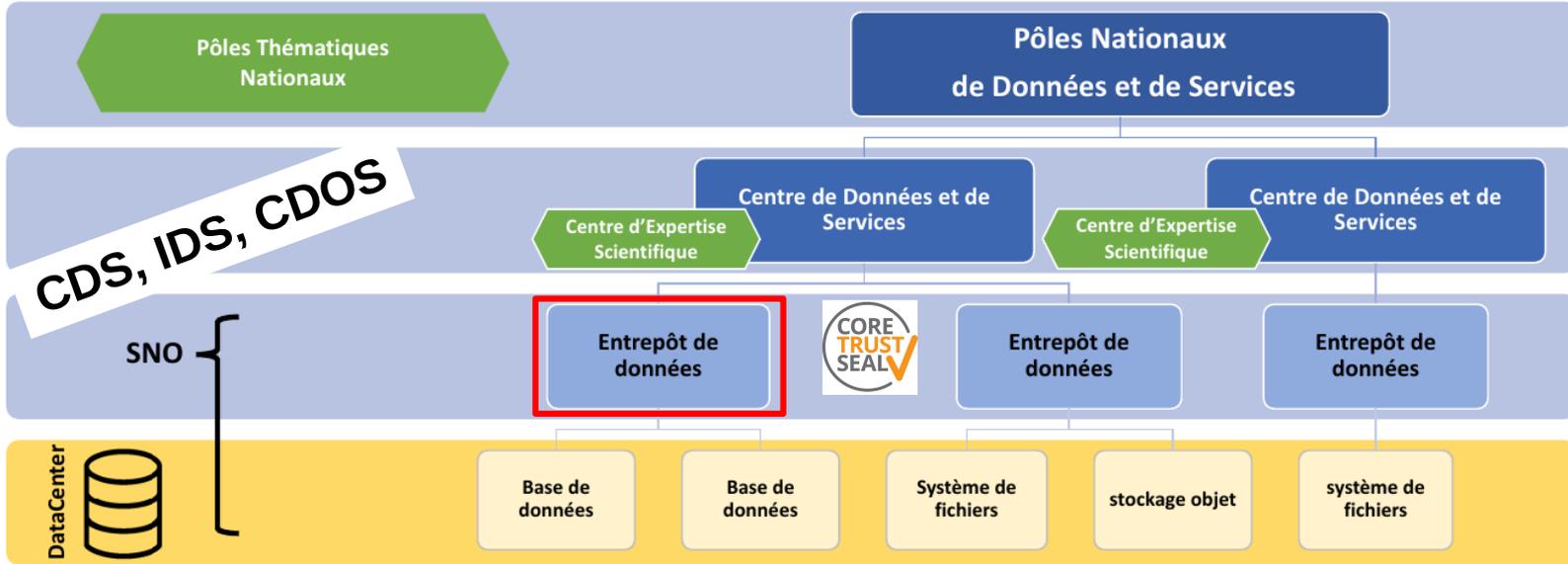


Le paysage actuel des infrastructures



MESRI - FRN 2022-2025

Organisation de la e-infrastructure des données d'observation



E-infra

 IR Observation

opérationnel / coordination / physique

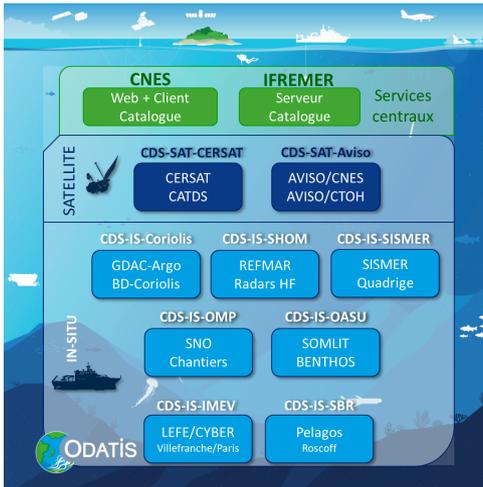
INSU 2022



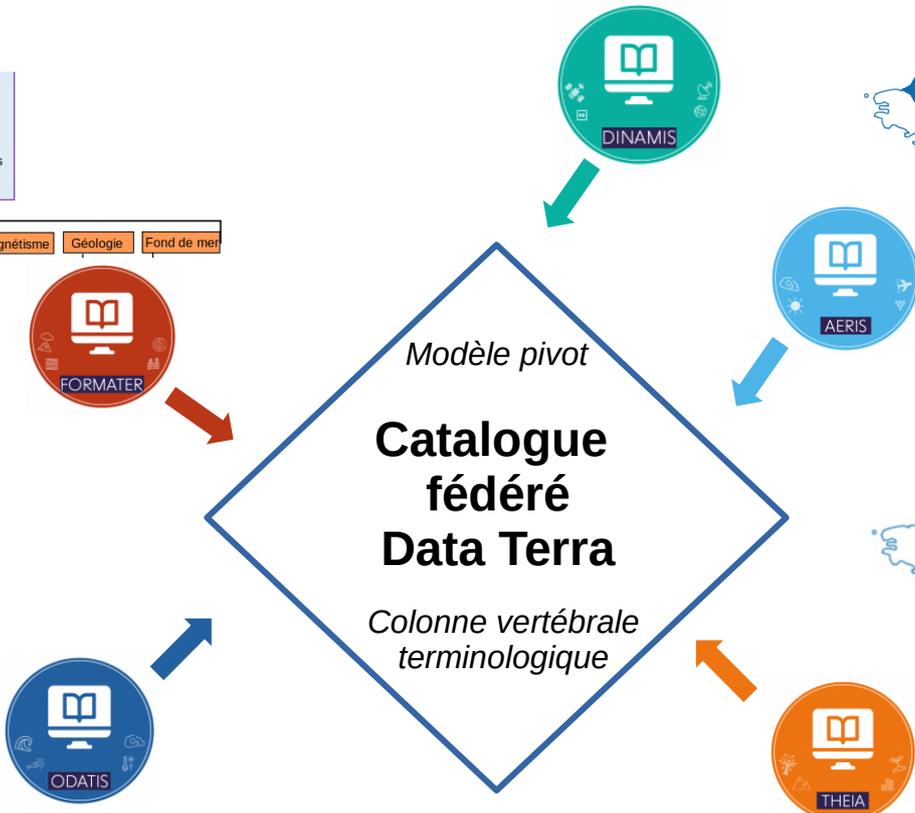
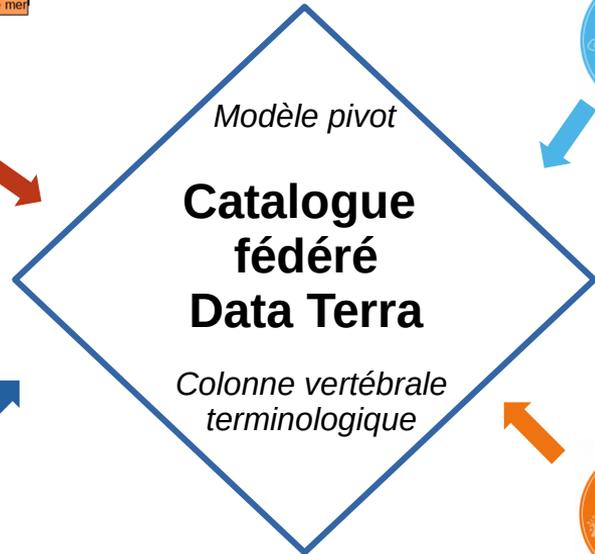
ForM@Ter



- Géochimie et Minéralogie
- Sismologie
- Gravimétrie
- Géodésie
- Volcanologie
- Déformations du sol
- Magnétisme
- Géologie
- Fond de mer



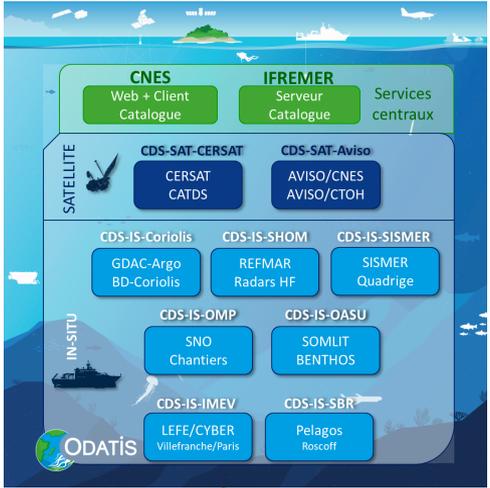
ODATIS



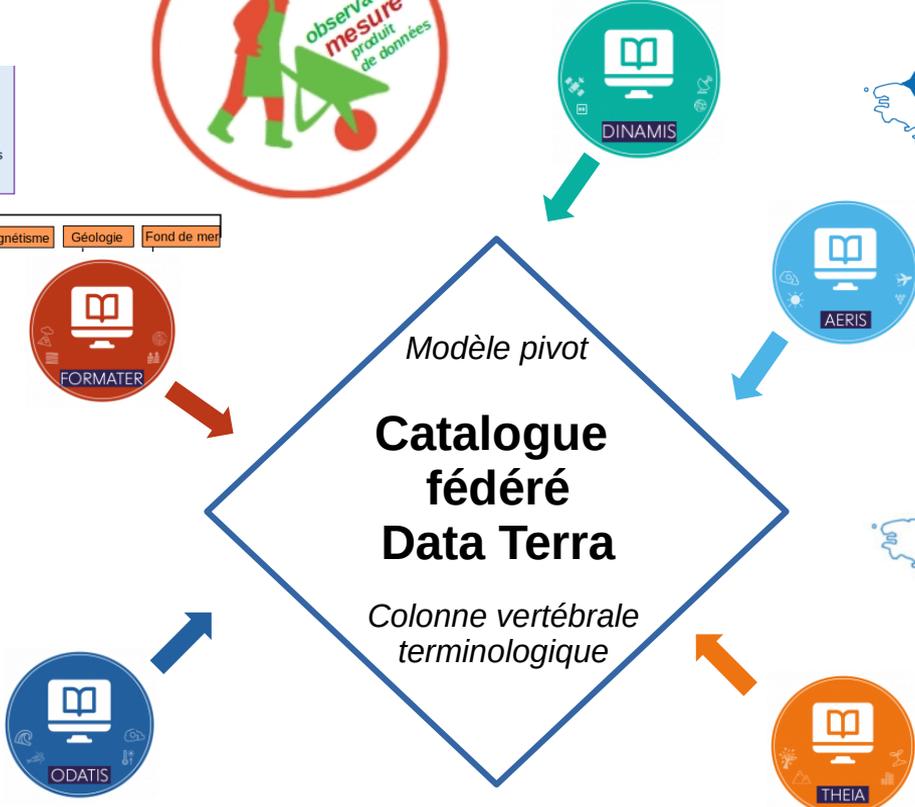
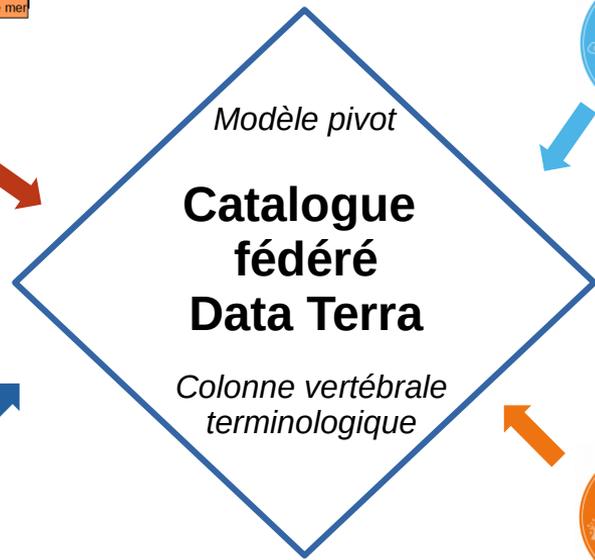
ForM@Ter



- Géochimie et Minéralogie
- Sismologie
- Gravimétrie
- Géodésie
- Volcanologie
- Déformations du sol
- Magnétisme
- Géologie
- Fond de mer



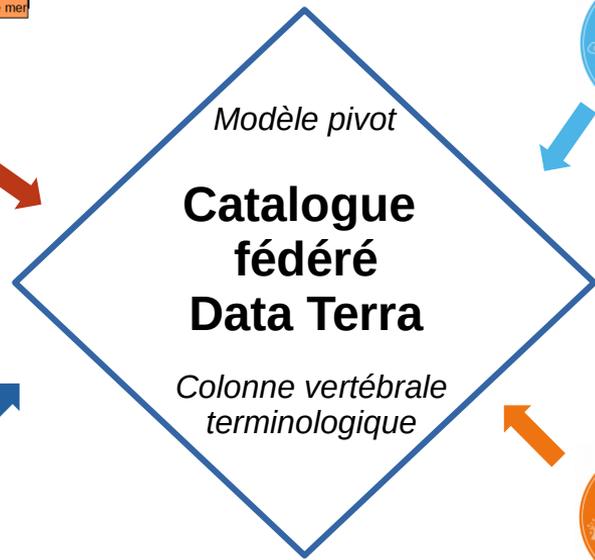
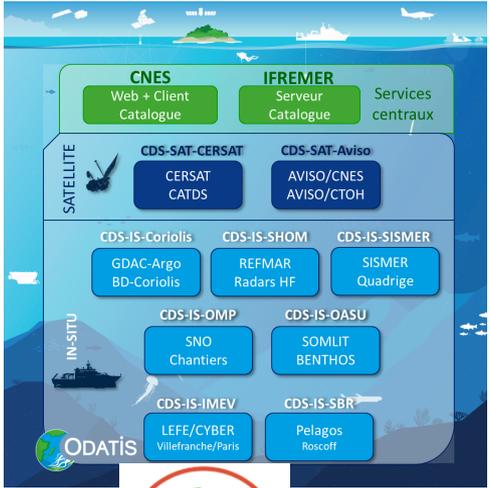
ODATIS



ForM@Ter



- Géochimie et Minéralogie
- Sismologie
- Gravimétrie
- Géodésie
- Volcanologie
- Déformations du sol
- Magnétisme
- Géologie
- Fond de mer



ODATIS

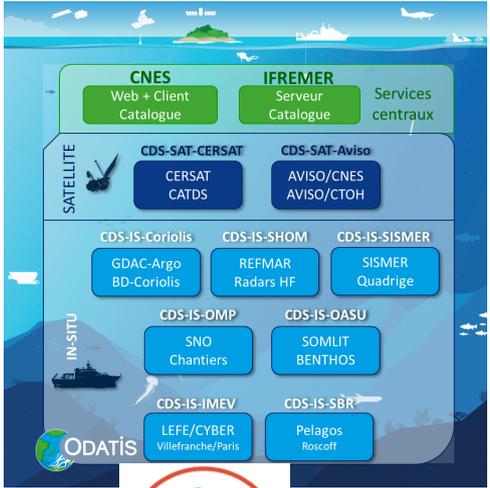
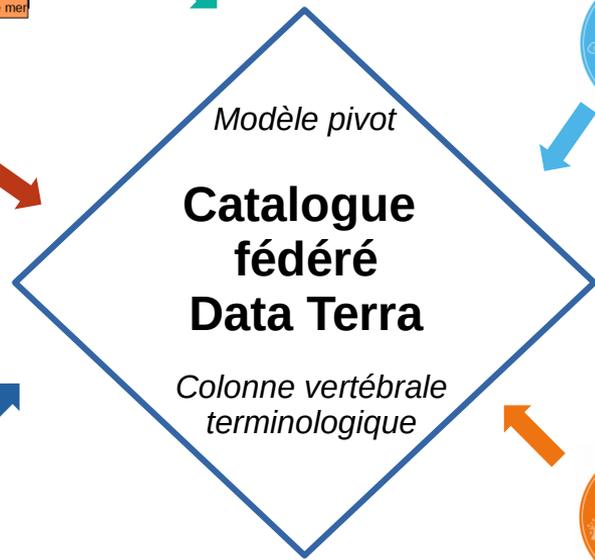


SEANOE

ForM@Ter



- Géochimie et Minéralogie
- Sismologie
- Gravimétrie
- Géodésie
- Volcanologie
- Déformations du sol
- Magnétisme
- Géologie
- Fond de mer



ODATIS



SEANOE



Centre de référence thématique du domaine Terre & Environnement

- Définition des normes de description des données
- Définition des normes de diffusion des données (ouverture, période d'embargo, accès restreint)
- Définition/diffusion des bonnes pratiques de collecte, documentation, traitement, et diffusion des données
- Définition de 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
- ...

RECHERCHE DATA GOUV

ATELIERS DE LA DONNÉE
Experts de la donnée en charge de l'accompagnement des chercheurs

ENTREPÔT
Interface Web de dépôt de ses données par le chercheur + espace de modération

CATALOGUE
Repérage et moissonnage des données des entrepôts externes

CENTRES DE RÉFÉRENCE THÉMATIQUES
Experts disciplinaires de la donnée

CENTRES DE RESSOURCES RATTACHÉS À RECHERCHE DATA GOUV

Mise en oeuvre de l'entrepôt de données du domaine Terre & Environnement

Besoin de la communauté scientifique & technique pour assurer la phase de modération !



Composante de EOSC-France

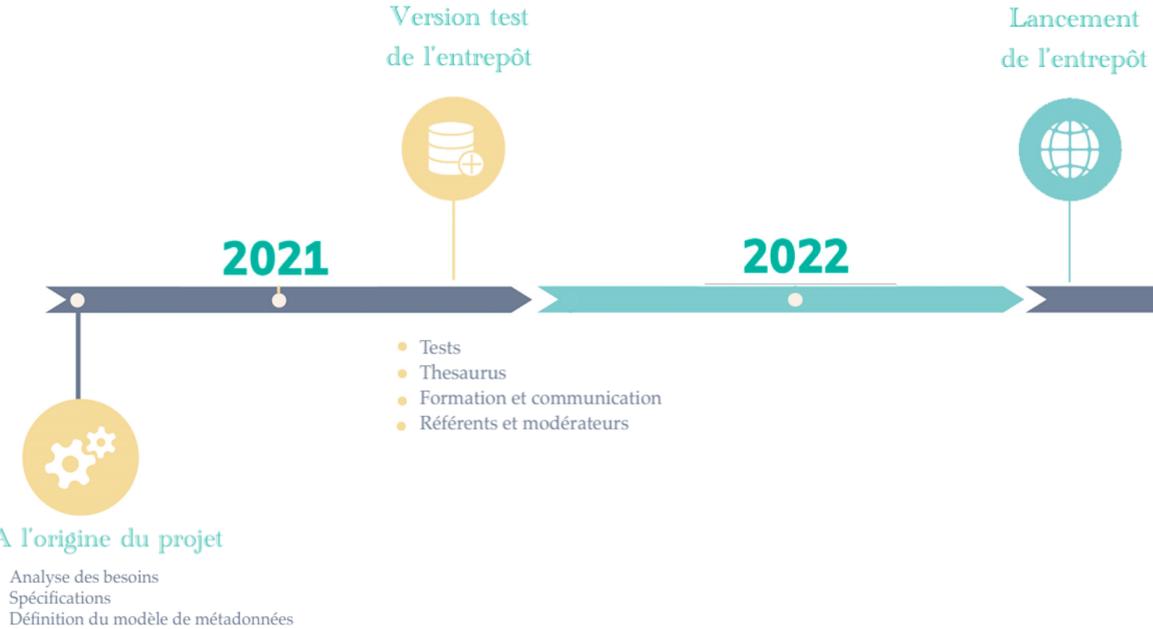
Représentation française dans GO-FAIR (mutualisation RDA)



Équipe projet

- Véronique Bertrand - RESIF
- Hélène Bressan - BRGM
- Christelle Pierkot - Data Terra
- Marine Vernet - Data Terra

Hébergement





Besoins des pôles

=> Gestion de données produites dans des entités qui ne disposent pas de moyens pour les FAIRiser et les stocker.

=> Produits à valeur ajoutée issus de données d'observation produites par les pôles, en vue de publication et l'obtention d'un DOI.

=> Données issues de projets nationaux et internationaux dont Data Terra est porteur.

Enquête communautés

- ⇒ **Données :**
 - ⇒ **Volumétrie : Fichier : entre 4Mo et 1To et Dépôts : entre 20 Mo et 10 Po**
 - ⇒ **Formats hétérogènes**
- ⇒ **Droits d'accès :**
 - ⇒ **Ouvert 76 %**
 - ⇒ **Embargo 50 % avec durée variable en moyenne 2 ans**
 - ⇒ **En demande d'accès : 45 %**
- ⇒ **Visibilité et accessibilité :**
 - ⇒ **Moissonnage des entrepôts existants**
 - ⇒ **Référencement dans des systèmes nationaux et/ou internationaux**
- ⇒ **Divers**
 - ⇒ **Facilité d'utilisation et simplicité du dépôt**
 - ⇒ **Gestion des versions**
 - ⇒ **Exposition des métadonnées selon différents formats**
 - ⇒ **Statistique de téléchargement et d'utilisation des dépôts**

Analyse des solutions existantes

- ⇒ **Retours d'expérience**
 - ⇒ **Seanoe**
 - ⇒ **DataSud**
- ⇒ **Analyse de l'existant**
 - ⇒ **Pangaea**
 - ⇒ **Zenodo**
 - ⇒ **Dataverse CIRAD**

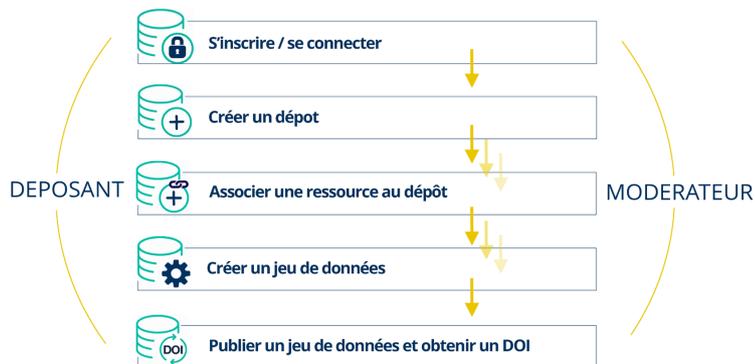
Choix technique GeoNetwork



Spécifications

- ⇒ **Données géo-référencées**
- ⇒ **Métadonnées normalisées (ISO 19115)**
- ⇒ **Gestion multilingue**
- ⇒ **DOI et Citation**
- ⇒ **Identification des auteurs**
- ⇒ **Niveaux de granularité**
- ⇒ **Vocabulaires**
- ⇒ **Licence CC BY par défaut**
- ⇒ **Volumétrie (5Go par fichier, 100Go par dépôt)**
- ⇒ **Contraintes d'accès et embargo**
- ⇒ **Formats de données non contraint**
- ⇒ **Authentification Renater**
- ⇒ **Recherche à facettes**
- ⇒ **Associer des ressources aux données**

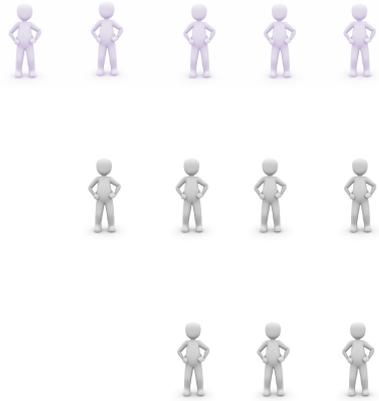
Workflow de dépôt



Améliorations prévues

- ⇒ **Connexion software Heritage - Hal**
- ⇒ **Gestion des versions**
- ⇒ **DataPaper**
- ⇒ **Analyse antivirus**
- ⇒ **Statistique d'utilisation**
- ⇒ **Prise en compte retours utilisateurs**

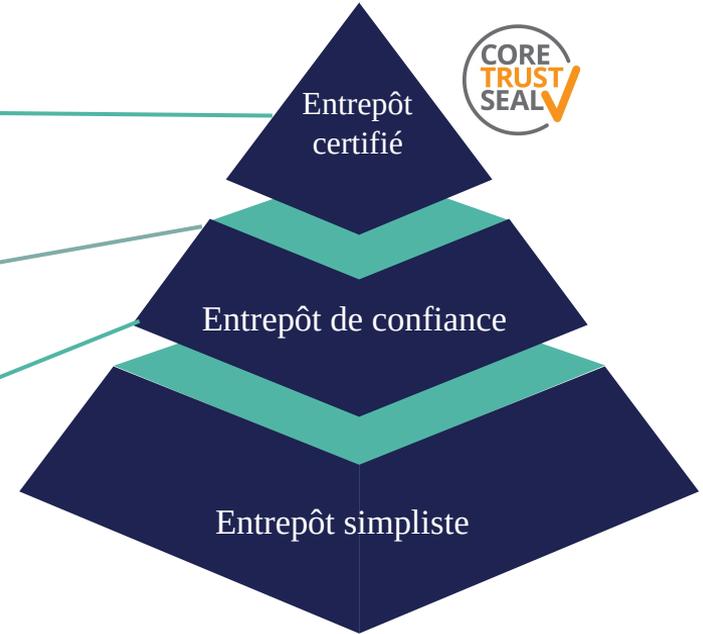




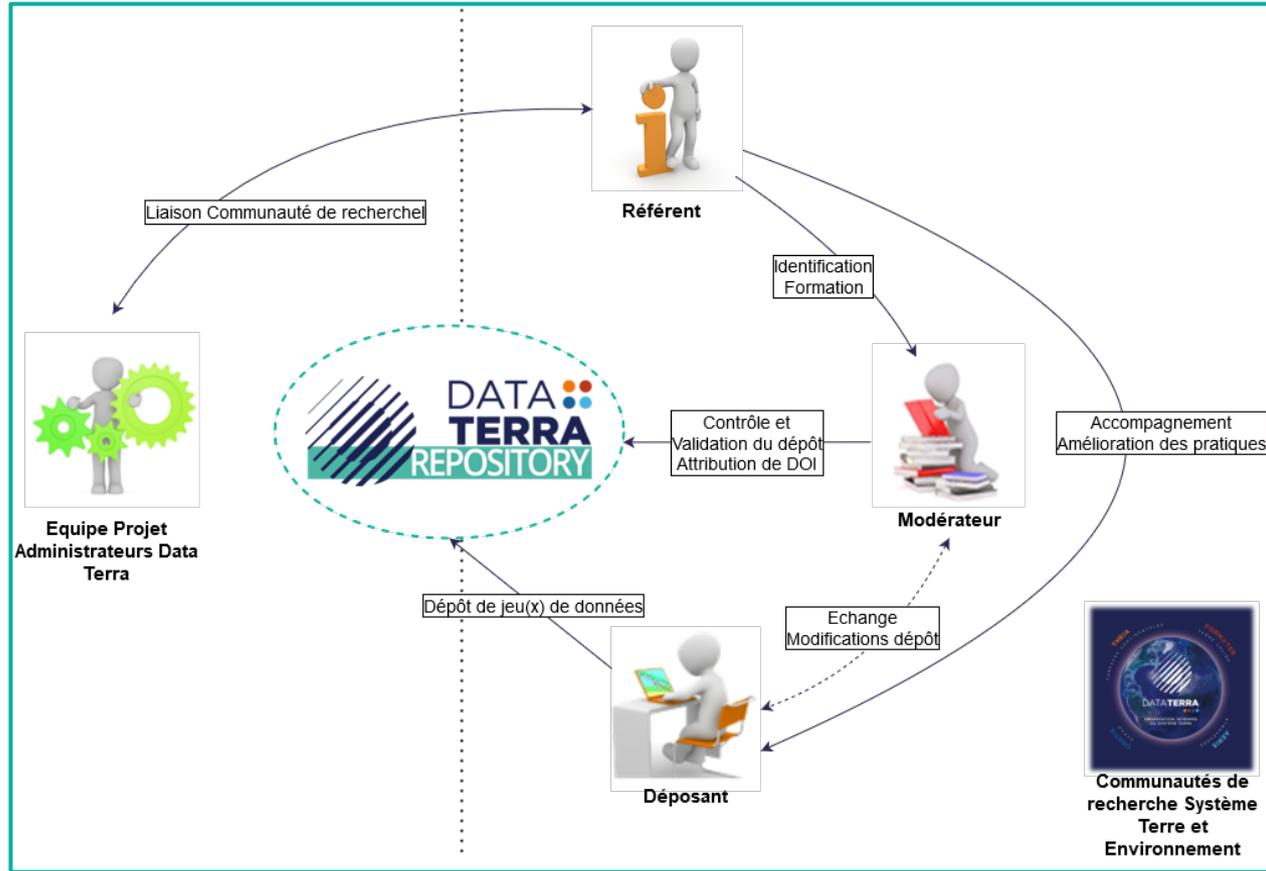
Entrepôt Data Terra V3

Entrepôt Data Terra V2

Entrepôt Data Terra V1



Organisation EDT



Support

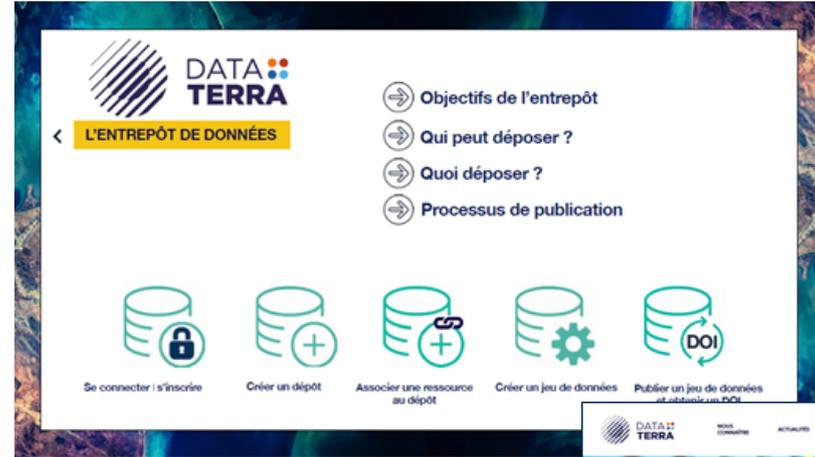
- Guide de dépôt
- Guide de bonnes pratiques
- Guide de modération
- Qualité attendue
- Périmètre

Promotion

- Page web : <https://www.data-terra.org/donnees-services/entrepot-de-donnees-data-terra/>
- Flyer

Ateliers

- Ateliers : sismologie (Résif), positionnement (GNSS)
- Université de Lille 12/05 - Fabrique de la Science Ouverte Session Entrepôts de données disciplinaires
- Ateliers de la données



Genèse

- Besoins identifiés pour entreposer les données de longue traîne
- Intégration de V. Bertrand (Résif) dans le GT entrepôt Data Terra
- Présentation du projet d'entrepôt de données Data Terra lors des Journées Scientifiques et Techniques Résif, novembre 2021.

Méthode

- Recherche de volontaires
- Choix de jeux de données ciblés
- Formation aux différentes rôles
- Beta-test de dépôt réalisé de bout en bout.

Retours

- Très instructif pour améliorer l'outil !
- Prise de conscience de l'intérêt de l'ingénierie de la donnée :
 - Identification des standards de description (méta)données
 - Format des (méta)données
 - Vocabulaire contrôlé, thesaurus, ontologie...
- Des problèmes de périmètre : diversité de la longue traîne

=> ODATIS / SEANOE

- ⇒ Moissonnage - Compatibilité des schémas ?
- ⇒ Retour d'expérience en terme d'organisation à considérer pour les autres pôles.

=> Théia et AERIS

- ⇒ Définition du périmètre
- ⇒ Mise en place de l'organisation
- ⇒ Vocabulaires : Mots clés et Thématiques

=> CLIMERI

- ⇒ Communauté structurée et internationale
- ⇒ Groupe de travail interne
- ⇒ Volumétrie importante

=> PNDB

- ⇒ Compatibilité des niveaux de granularité ?
- ⇒ Moissonnage d'entrepôts existants ?





- **Ouverture progressive en 2023**

- solution technique ok (V2)
- Facteur limitant : mise en oeuvre de la modération avec les communautés scientifiques

- **Encore des questions ouvertes**

- Préservation à long terme (> 10 ans), archivage pérenne
- Modèle économique
- Moyens humains pour assurer la continuité du service (0,2 ETP SEANOE)
- Missions associées au rôle de centre de référence thématique

