



EaSy Data

Earth System Data Repository

Lancement officiel de l'Entrepôt des données de longue traîne du Système Terre et de l'Environnement

6 novembre 2023, Paris



Climat



Biodiversité



Océan



Surfaces
continentales



Terre solide



Atmosphère



OUVERTURE OFFICIELLE

Isabelle BLANC, Administratrice ministérielle des données, des algorithmes et des codes sources, Conseillère auprès de la Directrice Générale recherche et innovation et de la Directrice Générale enseignement supérieur et insertion professionnelle, DGRI-DGESIP/MESR

Laurent CROUZET, Chef de Département, Services et Infrastructures Numériques, DGRI/MESR

Nicolas ARNAUD, Directeur du CNRS-INSU (visioconference)

Michaël CHELLE, Directeur du programme scientifique « Donnée et numérique » BRGM

Frédéric HUYNH, Directeur IR Data Terra

Programme **Lundi**



14h - 14h20

Mots de bienvenue et discours d'ouverture



14h20-15h

Présentation de l'entrepôt EaSy Data : positionnement au sein des écosystèmes nationaux et européens, organisation et démonstration en ligne

Intervenantes :

Christelle PIERKOT, Ingénieure de Recherche IR Data Terra / CNRS, co-pilote EaSy Data

Helene BRESSAN, Ingénieure BRGM, co-pilote EaSy Data



15h - 15h20

Témoignages et exemples de dépôts



15h20-15h45

Échanges



15h45-16h

Conclusions et perspectives

Intervenants : Jean-Marie FLAUD, Expert "Sciences atmosphériques et climat" MESR

DATA TERRA, coordonne et structure l'ensemble des acteurs des données système Terre / Environnement de l'ESR

Data Terra permet au travers de portails unifiés l'accès et traitements de l'ensemble des données et produits qualifiés, standardisés et inter-opérables pour adresser des enjeux scientifiques et des défis sociétaux interdisciplinaires

- Producteurs de données d'observation multi-sources
- 30 Centre de Données et de Services
- 32 Centres d'expertise scientifiques
- 250 FTE / 500 scientifiques, ingénieurs et techniciens

- Coordination/articulation avec les OSU, SNO
- En France
- Outre-Mer
- Etranger



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

L'IR DATA TERRA propose des services aux données d'observation du système Terre et de l'environnement



En 2024, 5 pôles de données et de services

AERIS

Atmosphère

FORM@TER

Terre solide

ODATIS

Océan

THEIA

Surfaces
continentales

PNDB

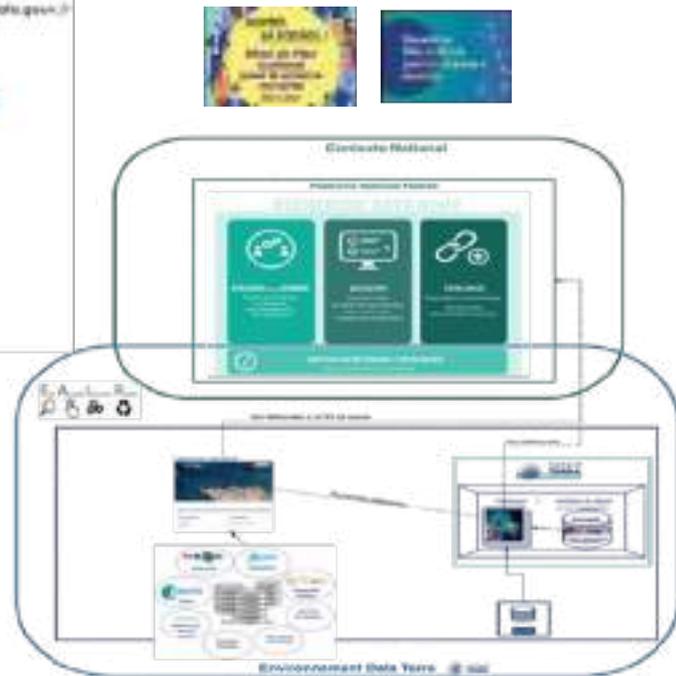
Biodiversité

dispositif
transversal

DINAMIS

Accès données
satellites THRS

DATA TERRA : accès unifiés pour accéder aux données et services d'observation du système Terre et de l'environnement



Service transversal

Recherche Data Gov.
Data Terra : Centre de Référence Thématique



Données "orphelines" longue traîne



Atmosphère



Terre solide



Océan



Surf.continentales

🕒 Catalogue de données et de services



🕒 Services associés

Support aux campagnes, services aux producteurs de données, cluster de calculs, logiciels et produits thématiques, ateliers thématiques, appel à projets, newsletter, adhésion en ligne, etc



EaSy Data

Earth System Data Repository

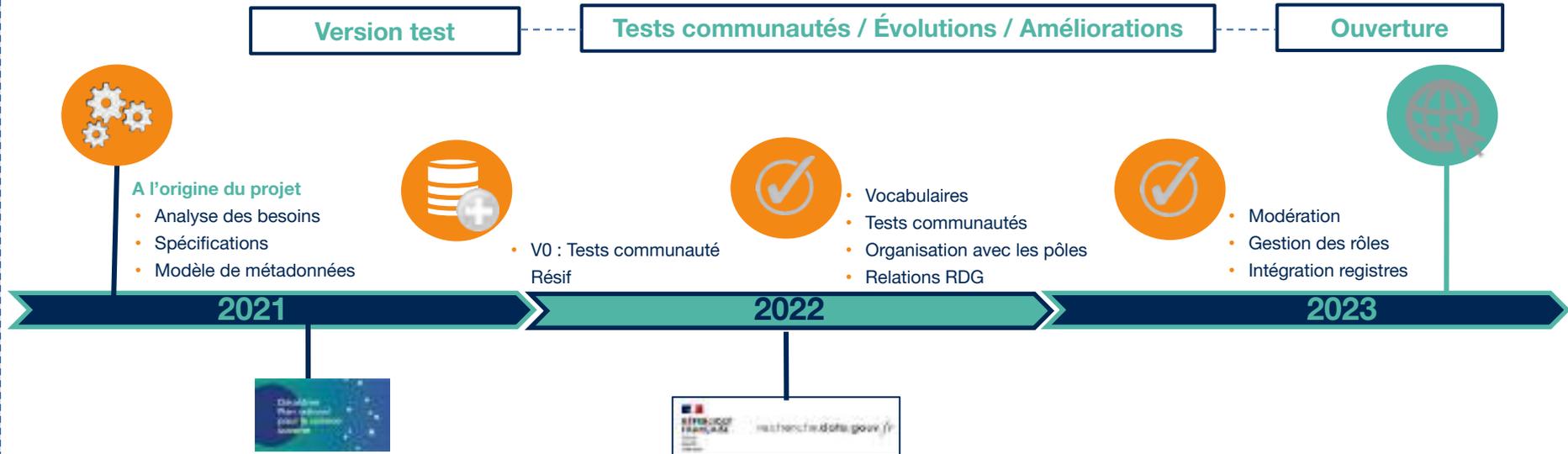
Entrepôt thématique des données de longue traîne de l'Environnement

*« Un service de dépôt et de diffusion pour les
données de la recherche du Système Terre et
de l'Environnement »*



H. Bressan – C. Pierkot
06 Novembre 2023





Équipe projet

- Véronique Bertrand – RESIF
- Hélène Bressan – BRGM
- Christelle Pierkot – CNRS/ Data Terra
- Marine Vernet – Ifremer / Data Terra

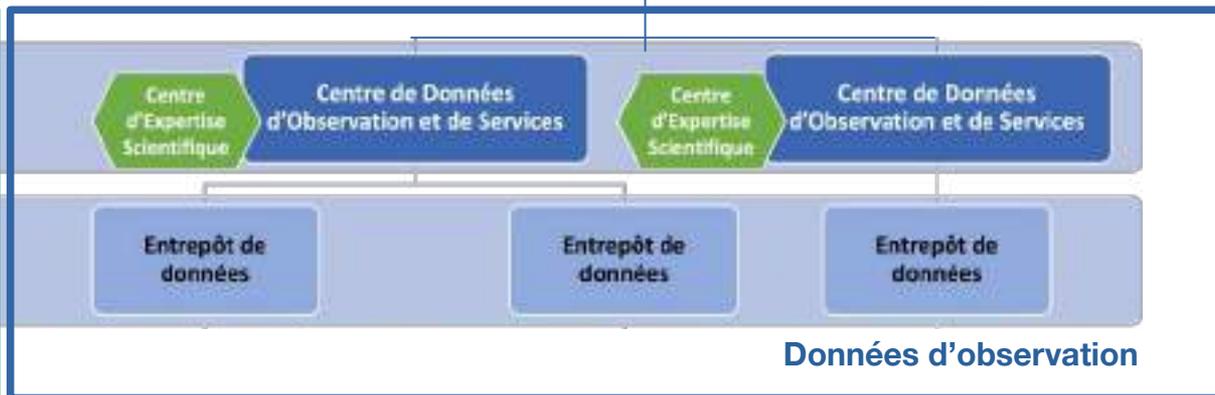
Avec le soutien de :

Sandrine Dalmar – IRD / Data Terra





Données de longue traîne



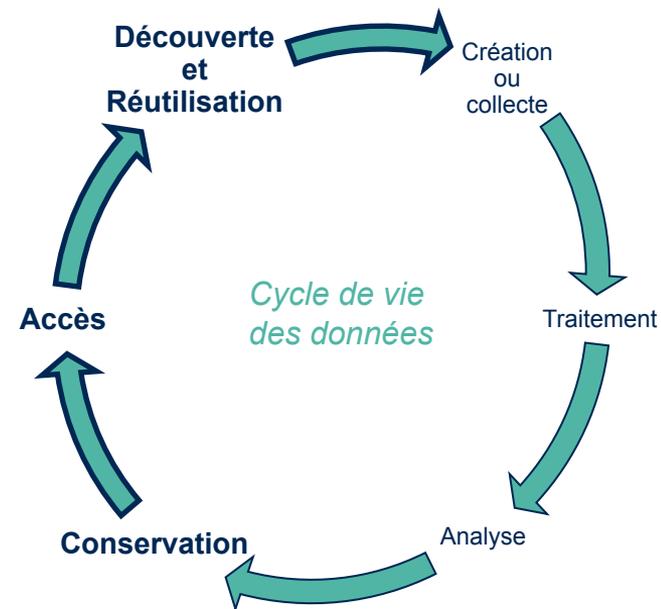
SEANOE

cupidon

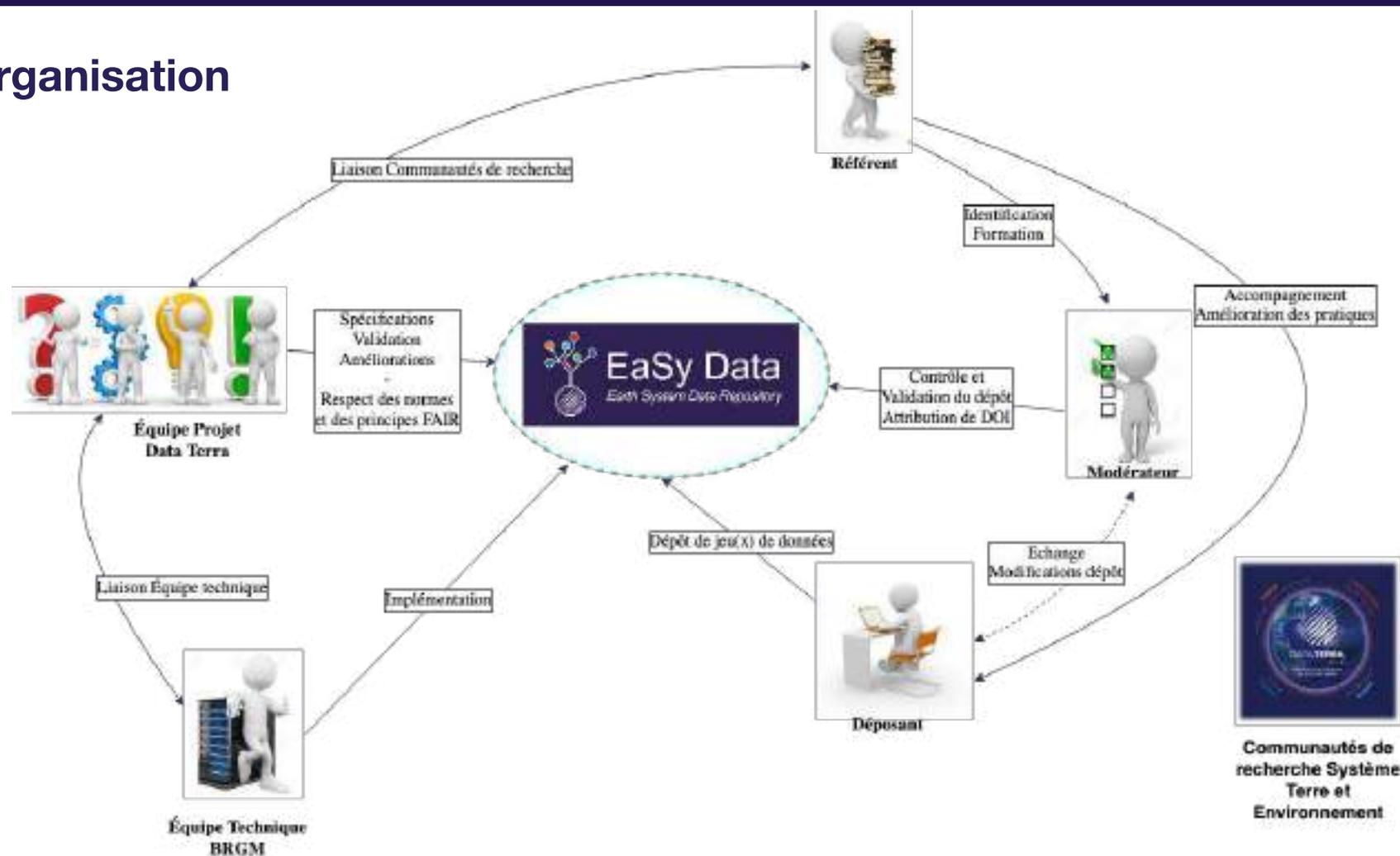


Capitaliser les données de la recherche et leur attribuer un DOI

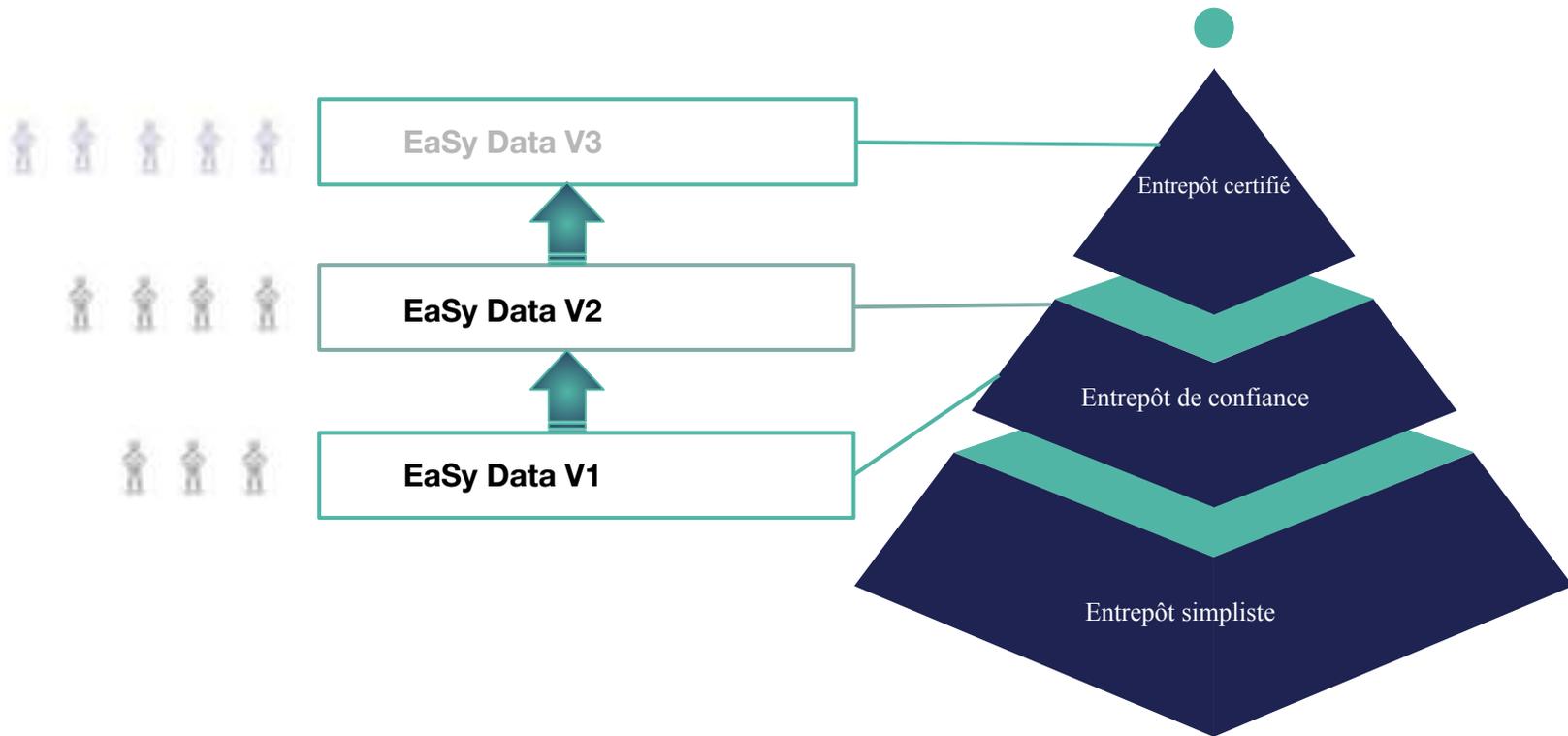
- ▶ une simplicité de dépôt
 - ▶ un respect des standards communautaires
 - ▶ un respect des principes FAIR
 - ▶ une pérennisation et visibilité des données
- ▶ les données de la Recherche du Système Terre et Environnement
 - ▶ pour les membres de la communauté de la recherche en science de la terre et de l'environnement



Organisation



Un entrepôt thématique de confiance...



... En route vers la certification



EaSy Data

Earth System Data Repository

Démonstration dépôt

<https://www.easydata.earth>



Interface - Niveaux de granularité

L'entrepôt Data Terra propose deux niveaux de granularité :

- Un niveau **Dépôt** qui englobe un ou plusieurs jeux de données

Un niveau **Jeu de données** qui est composé d'un ou plusieurs fichiers.

le déposant peut choisir, en fonction de ses besoins de citation, le niveau de granularité souhaité. Par exemple :

- 1 dépôt avec 1 seul jeu de données global associé
- 1 dépôt avec plusieurs JDD associés partageant des attributs communs (ex: type de données, capteur, localisation, etc.) mais ayant leur propre spécificité (ex: période de collecte)

Les DOI sont fournis uniquement pour les jeux de données.



EaSy Data
Earth System Data Repository

Démonstration modération

<https://www.easydata.earth>



EaSy Data

Earth System Data Repository

Publier, partager et découvrir les données du Système Terre et de l'Environnement



Déposer des données



Rechercher des données



Aide



EaSy Data
Earth System Data Repository

Merci

entrepot@data-terra.org

<https://www.easydata.earth>

Équipe projet

- Véronique Bertrand – Epos-France
- Hélène Bressan – BRGM
- Christelle Pierkot – CNRS/ Data Terra
- Marine Vernet – Ifremer / Data Terra

Avec le soutien de :

Sandrine Dalmar – IRD / Data Terra



L'ENTREPÔT DES DONNÉES DE LONGUE TRAÎNE DU SYSTÈME TERRE ET DE L'ENVIRONNEMENT

EaSy Data propose **une interface conviviale et simple d'utilisation** pour le dépôt et l'accès aux données longue traîne de recherche du domaine Système Terre et de l'Environnement. Il est basé sur des standards reconnus par les communautés, facilitant la découverte et la réutilisation des données déposées.

Les métadonnées respectent les principes FAIR conformément aux recommandations nationales et européennes. Des vocabulaires sont utilisés pour améliorer la qualité des descriptions des jeux de données et garantir un certain niveau d'interopérabilité entre les systèmes. Un dispositif de modération disciplinaire assure la qualité des dépôts effectués. EaSy Data rend visible les jeux de données d'autres entrepôts thématiques (e.g. Seanoe, Data Indores). Les données stockées dans EaSy Data seront référencées au niveau national dans le catalogue de Recherche Data Gov.



L'ENTREPÔT DES DONNÉES DE LONGUE TRAÎNE DU SYSTÈME TERRE ET DE L'ENVIRONNEMENT

Lancement de la démo



The screenshot shows the main interface of the EaSy Data website. At the top left is the EaSy Data logo and name. The background features a satellite-style map of Earth. In the center, there is a dark purple box with the EaSy Data logo and the text "EaSy Data Earth System Data Repository". Below this, the main heading reads "Publier, partager et découvrir les données du Système Terre et de l'Environnement". Two large teal buttons are positioned below the heading: "Déposer des données" (with a database and gear icon) and "Rechercher des données" (with a magnifying glass icon). At the bottom, there are three links: "À propos", "Contact", and "Mentions légales". The text "Service porté par IR Data Terra et développé par le BRGM" is visible at the bottom center.

L'ENTREPÔT DES DONNÉES DE LONGUE TRAÎNE DU SYSTÈME TERRE ET DE L'ENVIRONNEMENT

POUR QUI ?

- ✓ LES SCIENTIFIQUES
DU SYSTÈME TERRE ET ENVIRONNEMENT
- ✓ LES PARTENAIRES NATIONAUX, EUROPÉENS ET
INTERNATIONAUX

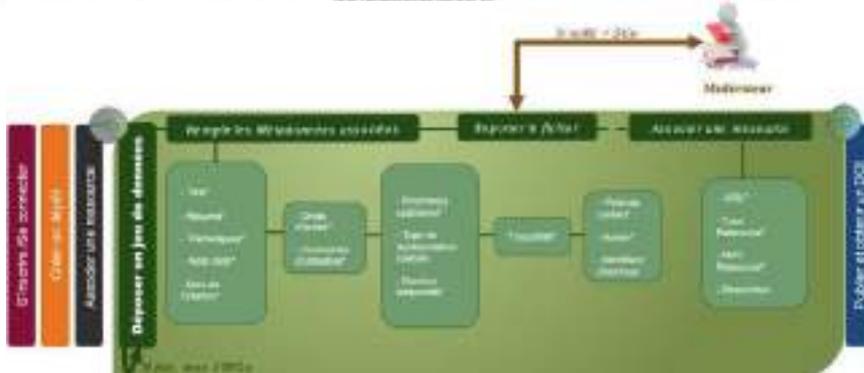
QUEL TYPE DE DONNÉES ?

- ✓ DONNÉES LONGUE TRAÎNE
- ✓ ISSUES DE RECHERCHE EN VUE DE PUBLICATION
- ✓ ISSUES DE PROJETS

Un procédé simple
en 5 étapes clés



Un modérateur
pour accompagner
le déposant



Pérennisation et visibilité des données :

- Stockage adapté et diffusion de données de qualité
- Attribution de DOI aux données déposées / citation
- Référencement des données déposées dans le catalogue de la plateforme nationale fédérée

Simplicité du dépôt :

- Authentification via EduGAIN/Renater, ORCID
- Prise en compte d'identifiants chercheurs,
- Dépôt en quelques étapes
- Modération et accompagnement par des spécialistes

Qualité des métadonnées et interopérabilité :

- Norme ISO 19115-3,
- Interface adaptée aux spécificités des données géo-référencées
- Thesaurus communautaires
- Services avec possibilité d'associer des ressources aux données (publication, code source, site web projet, ...)



EaSy Data
Earth System Data Repository

Témoignages et exemples de dépôt

Bref rappel du contexte

- ❑ Besoin exprimé par AERIS dès 2019 : **outil simple de dépôt et de fourniture de DOI** pour des **données associées à une publication** ou des données « orphelines »
- ❑ Avec l'idée de **privilégier une solution Data Terra**
- ➔ Nous sommes très satisfaits de voir aboutir ce projet et cette première version de l'entrepôt **répondra déjà bien à nos besoins**

Appel à projets AERIS

- ❑ Appel **ouvert en continu** sur notre site web
- ❑ Mais examen (instruction technique et scientifique) 3 ou 4 fois par an des demandes par notre Conseil Scientifique (CS)
- ❑ Exemples : base de données d'un projet, chaîne de traitement, outils d'analyse ou de visualisation de données, support aux campagnes de mesures, etc.
- ❑ Manque de **réactivité** pour répondre à certaines demandes plus "simples"
- ➔ L'entrepôt EaSy Data va nous aider à combler ce manque, nous permettra d'**offrir un meilleur service** et probablement de toucher de **nouveaux utilisateurs** (grâce à l'intégration dans l'écosystème Recherche Data Gouv)
- ➔ Un examen *a posteriori* des dépôts par notre CS nous permettra d'identifier des jeux d'intérêt qui pourraient être valorisés dans des services du pôle

Utilisation de EaSy Data par AERIS

- ❑ Principal cas d'utilisation : **fourniture rapide** (1 à 5 jours, sous réserve de la « qualité » des métadonnées saisies et des données déposées) **d'un DOI pour des données associées à une publication**
- ❑ Et aussi : prise en charge plus simple de **données de longue traîne** (métadonnées et curation minimales)
- ❑ Pour tous les autres cas (en particulier si demande de services associés aux données) : passage par notre appel à projets pour une prise en charge par un de nos CDS (Centre de Données et de Services) et un ajout au catalogue AERIS
- ➔ L'utilisation envisagée est **complémentaire des services plus avancés que nous pouvons offrir via l'Appel à projets**

Exemples

- ❑ Une demande concernant un jeu de données associé à un papier en cours de révision vient d'arriver (analyse de profils d'Ozone IASI) => premier jeu publié via EaSy Data
- ❑ Données de projets ou campagnes de mesures multi-disciplinaires : on ne sait pas toujours bien prendre en charge les composantes hors atmosphère dans le catalogue de données du pôle (ex : le dernier projet reçu à notre AAP concerne aussi des données de Végétation)
- ➔ L'entrepôt EaSy Data nous offre une meilleure solution pour les publier et un cadre pour **fluidifier la collaboration avec les autres pôles** sur ce type de demandes

EaSy Data : un service de proximité attendu et structurant

- **Historique** : phase de test menée depuis 2021 avec la communauté Résif-EPOS (sismologie, géodésie, gravimétrie), besoin identifié sur des données orphelines.
- **Intérêt** : mieux connaître et cartographier la communauté de nos **petits producteurs** de données. →
- **Objectif** : co-construire la phase de **modération** des dépôts pour favoriser la **réutilisation** des jeux de données (travail sur les vocabulaires/thesaurus).



Exemples de dépôts en Terre Solide

- Paramètres d'intensité des mouvements sismiques issus des vitesses et accélérations du sol mesurées pendant 20 ans en France métropolitaine par le réseau Résif (RLBP et RAP). Base observationnelle pour les méthodes du génie parasismique et pour les normes de construction.
- Modèle tomographique des Alpes.
- Catalogues de sismicité (essaims naturels ou induits, produits d'instrumentation ultra-dense et de nouvelles méthodes de détection [IA]...).
- Jeux de données de référence pour les méthodes IA de discrimination des événements sismiques (séismes, tirs de carrière...).
- ...

Les dispositifs de partage des données / productions THEIA

→ Pour les communautés structurées :

- Entrepôts des IR d'observation (OZCAR, ANAEE, RZA, etc) pour les **données d'observation long-terme** (principalement *in-situ*)
- Entrepôts des **Centres d'Expertise Scientifique (CES)** pour les **données (collection) / productions** (chaînes de traitement par exemple)



Dispositif de modération → par AAP THEIA

- Appel ouvert en continu sur notre site web
- Mais examen (instruction technique et scientifique) 3 ou 4 fois par an des demandes par notre Conseil Scientifique (CS)
- Objectif d'une prise en charge par un de nos **CDS (Centre de Données et de Services)** et un ajout au **catalogue THEIA**

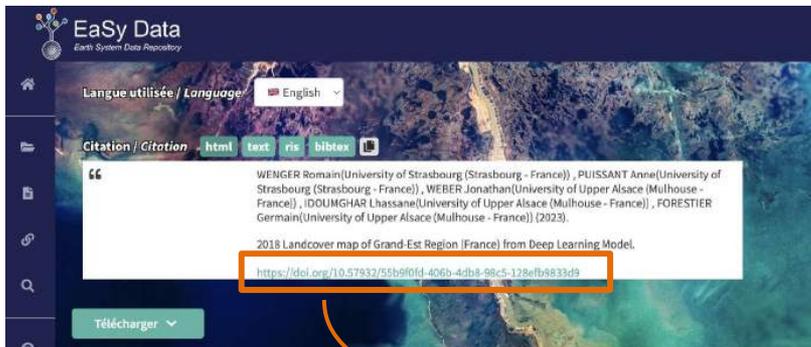
→ Dans tous les autres cas - EaSy Data :

- **Partager les productions scientifiques** dites « orphelines » issues de **projets de recherche (publications)** avec l'obtention d'un **DOI** ;
- **Deux conditions** : en dehors d'un cadre d'observation sur le long-terme, aucune autre solution communautaire (INRAE, IRD, etc) disponible.
- Prise en charge simple de **données de longue traîne** (métadonnées et curation minimales) mais **validées scientifiquement** → *basée sur une expertise scientifique adaptée* ;

→ jeux de données issus du traitement de données **d'observation par imagerie** (terrestre, aérienne ou spatiale) et/ou de données **d'observations de terrain ou in-situ** et/ou de leurs **combinaisons** ;

Exemples de la solution apportée par EaSy Data :

- une campagne d'**observations d'objets spécifiques** (arbres urbains, etc) dans la cadre d'un projet (inter)disciplinaire ;
- une **cartographie d'occupation des sols** sur un territoire réalisée à partir de données spatiales combinées avec des données acquises sur le terrain dans le cadre d'un projet de recherche (Thèse, ANR, etc)
- ... à laquelle on peut associer un jeu de données créé pour entraîner un modèle d'apprentissage.



EaSy Data
Earth System Data Repository

Langue utilisée / Language: English

Citation / Citation: html text ris bibtext

WENGER Romain (University of Strasbourg (Strasbourg - France)), PUISSANT Anne (University of Strasbourg (Strasbourg - France)), WEBER Jonathan (University of Upper Alsace (Mulhouse - France)), IDOUMGHAR Hassane (University of Upper Alsace (Mulhouse - France)), FORESTIER Germain (University of Upper Alsace (Mulhouse - France)) (2023).

2018 Landcover map of Grand-Est Region (France) from Deep Learning Model.

<https://doi.org/10.57932/55b9f0fd-406b-4db8-98c5-128efb8833d9>

Télécharger

Liens de téléchargement / Download links

[10_classes_style_2.qml](#)
[Grand_Est_2018_2020.tif](#)



Résumé / Abstract

2018 Landcover map at 10m with 10 thematic classes on the Grand-Est Region (France) from a deep learning model described in Wenger et al., 2022 (DOI: [10.1080/01431161.2022.2054295](https://doi.org/10.1080/01431161.2022.2054295)).

Traçabilité (origine) / Lineage (origin)

The deep learning model used to produce the landcover map is trained by using the MultiSensGE Dataset (doi.org/10.5381/zenodo.6375466) which contains multi-temporal patches of Sentinel1&2 imagery (year 2020) over the Grand-Est region (France) and a modified version of the OCSGE2 dataset for reference (Wenger et al., 2022 - doi.org/10.5194/isprs-annals-V-3-2022-635-2022, 2022). To explore the

<https://doi.org/10.57932/55b9f0fd-406b-4db8-98c5-128efb8833d9>

L'ENTREPÔT DES DONNÉES DE LONGUE TRAÎNE DU SYSTÈME TERRE ET DE L'ENVIRONNEMENT

Les communautés scientifiques concernées



Atmosphère

Données de longue traîne.

Exemples :

- Un besoin urgent d'un DOI sur un jeu de donnée pour publier.
- Des données anciennes du domaine atmosphère en local (ordinateur ancien, etc)



Océan

ne concernent que des jeux de données ayant été produits dans un contexte interdisciplinaires, mutualisés avec une autre communauté du Système Terre.



Terre solide

SISMOLOGIE

- Jeux de données sismologiques transformées
- Collection de données incluant des données d'entrepôts de données d'observation différents
- Modèle tomographique (après avis CES)
- Catalogue sismicité (si hors périmètre SNO Renass)
- Carte d'intensité macrosismique (si hors périmètre SNO BCSF)

GRAVIMÉTRIE

- Jeux de données de cartographie terrestre obtenues avec des gravimètres relatifs à ressort non rattachées à une station absolue
- Jeux de données de cartographie aéroportées (après discussion BGI)



Surfaces continentales

Données de longue traîne.

Deux catégories de données :

- Données de type "**observations de l'imagerie**" (remote observations) définie comme « Tout(e)s données issue(s) des traitements d'un ou de plusieurs capteurs permettant d'observer une variable à distance (i.e. en dehors de son environnement immédiat) »
- Données de type "**observations in-situ**" définie comme « Données d'une variable observée directement depuis son milieu, à l'endroit ou à proximité de là où se trouve l'objet et/ou le phénomène d'intérêt, ou combinaison de plusieurs variables de ce type »

entrepot@data-terra.org

<https://www.easydata.earth>



EaSy Data
Earth System Data Repository



SOUTENU PAR

Les pôles de données de l'IR DATA TERRA



Le projet GAIA Data Equipex+ PIA 3



**DATA
TERRA**



Pôle National
de Données & Services



anr



Les partenaires de l'IR DATA TERRA

