

amU | MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ANDOR2024

CEDRE | Centre de formation et de soutien aux Données de la REcherche présente

Assises Nationales des DONNÉES de la Recherche

26 & 27 novembre 2024
Mucem, Marseille

- Temps forts :
- Remise des prix Science Ouverte données de la recherche et logiciel libre
 - Restitution du groupe de travail sur la "Transparence et publicité des résultats de la recherche en santé"



IR DATA TERRA : E-Infrastructure pour des données (spatiales, in-situ) et services FAIR au service d'une observation intégrée du Système Terre

Frédéric Huynh (IRD), directeur IR Data TERRA

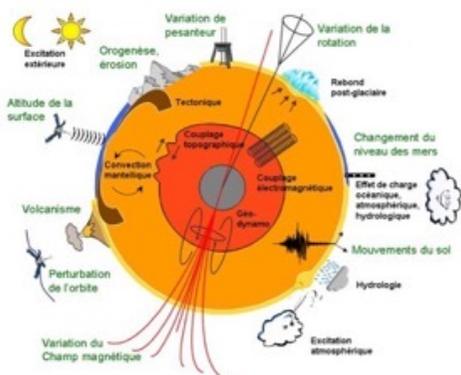
Emmanuel Chaljub (UGA-CNRS), directeur pôle Form@Ter ; Patrice Henry (CNES), directeur pôle AERIS ; Erwann Qumbert (IFREMER), directeur pôle ODATIS ; Anne Puissant (UNISTRA), directrice pôle THEIA ; Jean-Denis Vigne (MNHN) coordinateur PNDB, Richard Moreno (CNES), directeur technique ; Jean-François Faure (IRD), secrétaire exécutif DINAMIS



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.*



D'après O. de Viron

Comprendre ces processus nécessite l'accès et l'analyse de données complexes, nombreuses, volumineuses et d'origines multiples (satellites, in situ, campagnes, observations à long terme, résultats d'expérimentation, modèles, etc ...)

Dispositif d'accès à des **données, produits et services** permettant **d'observer, comprendre et prévoir** de manière **intégrée** l'histoire, fonctionnement et évolution du système Terre soumis aux changements globaux



Enjeux

Atmosphère

Surfaces continentales

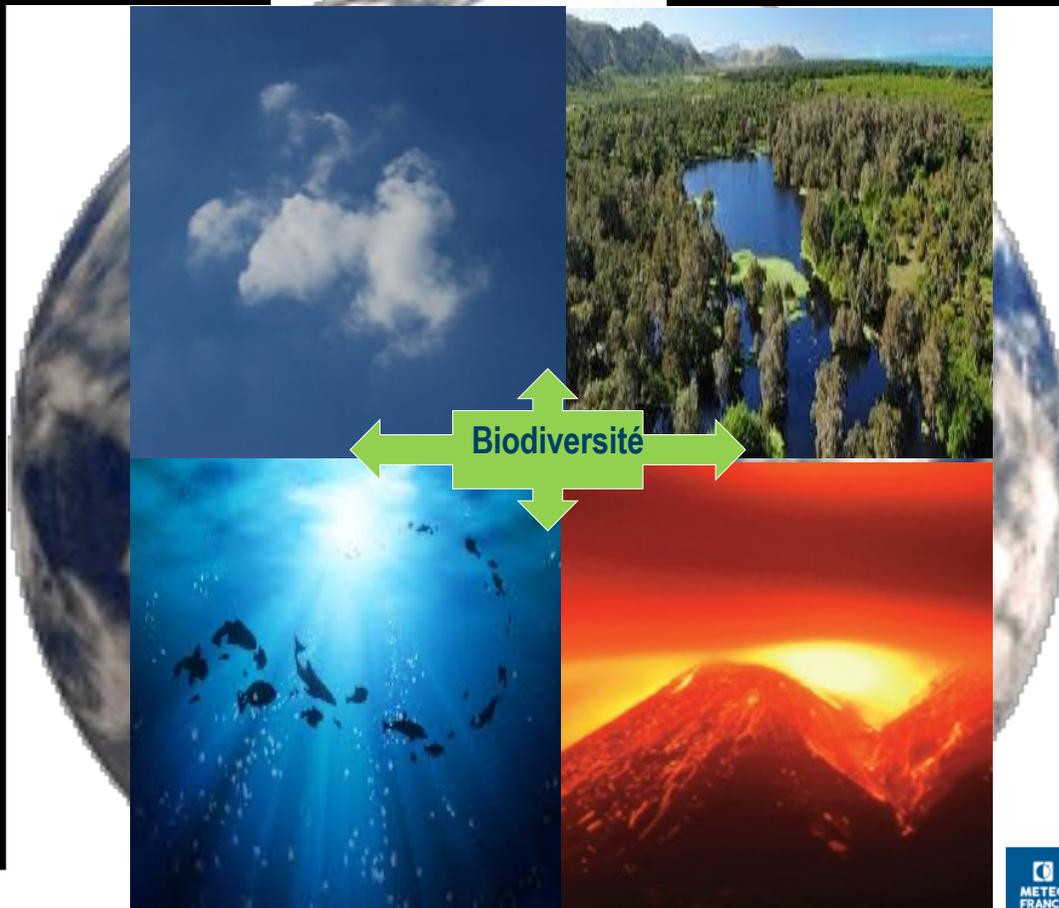
Réchauffement
climatique
Erosion de la
biodiversité
Pollution de l'air
Usage et qualité de l'eau
Dégradation des sols
Submersion
Risque hydro-climatique
Tension sur les
ressources biologiques
et minérales
Acidification de l'océan

Alimentation
Santé
Energie & Matériaux
Risques naturels
Sécurité
Migrations

Biodiversité

Océan

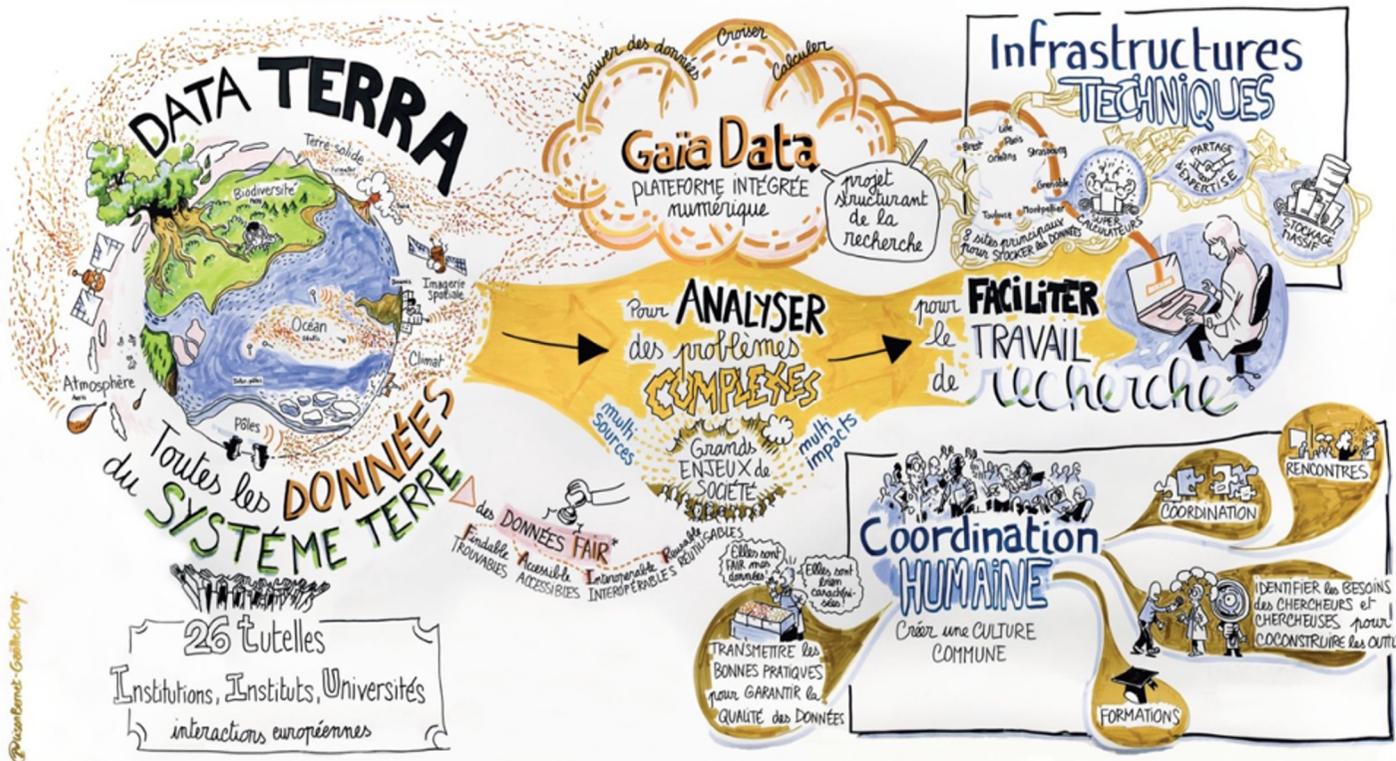
Terre interne



Elisabeth Vergès, MENESR-DGRI--SSRI-A1



AMBITIONS

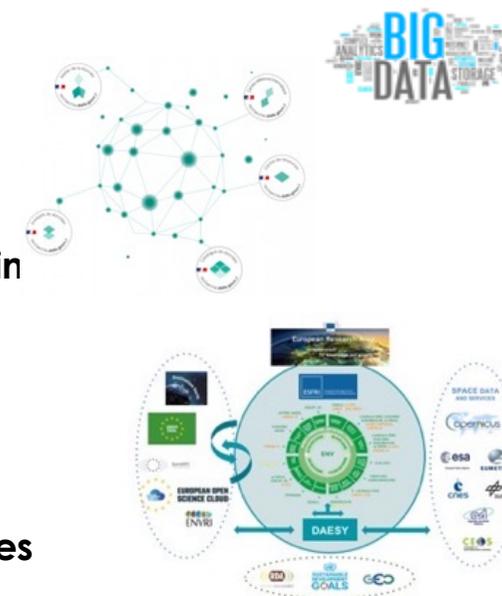
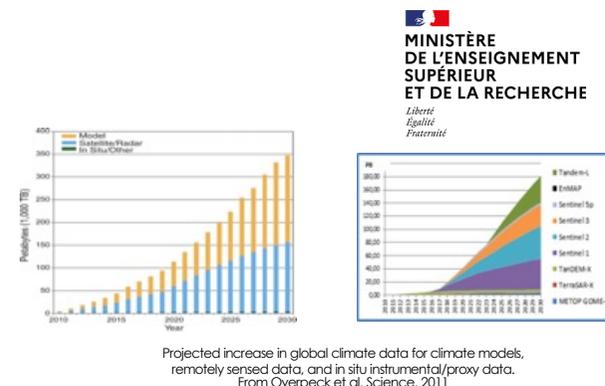


Développer un dispositif global d'accès, de traitement de **données multi-sources** et de développement de **services** permettant d'**observer, comprendre et prévoir** de manière **intégrée** le **fonctionnement et l'évolution** du système Terre

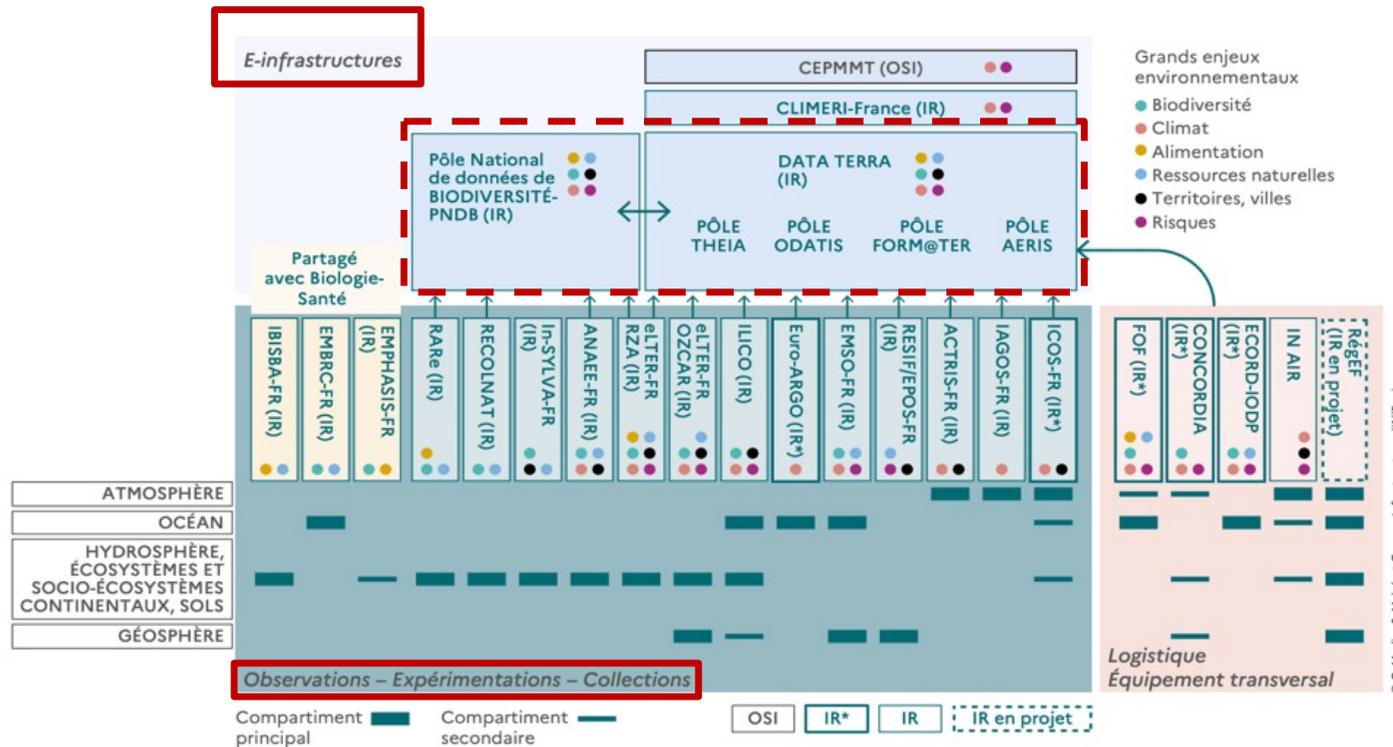
Accéder, partager, traiter et combiner les données spatiales et in-situ produites par les sciences du système Terre pour répondre aux grands défis scientifiques et accompagner les transitions

Besoins et verrous

- **Evolution des besoins des communautés scientifiques**
données **multi-sources**, **multi-capteurs**, services d'accès aux données, traitements, analyse/modélisation, IA, approches intégrées des interactions et complexité, observation à long terme
- **Augmentation exponentielle du nombre de données, diversités des sources, complexités, ...**
Spatiales, in-situ, modèles => besoins d'analyse/réanalyse, traitements intelligents
- **Méthodes et outils de FAIRisation des données spécifiques des domaines**
Découverte, méta-données enrichies, vocabulaires partagées, ...
=> écosystèmes d'acteurs
- **Co-construire une infrastructure distribuées de services spécifiques du domain système Terre et Environnement**
=> enjeux scientifiques et données transverses
- **Concilier structuration nationale, européenne et internationale**
=> modèle d'IR Data Terra en Europe et à l'international
- **Concilier recherche d'excellence et développement de partenariats avec les acteurs publics et économiques**



Positionnement de l'IR Data Terra - Feuille de route des IR/IR* E-Infrastructure du domaine environnement / système Terre

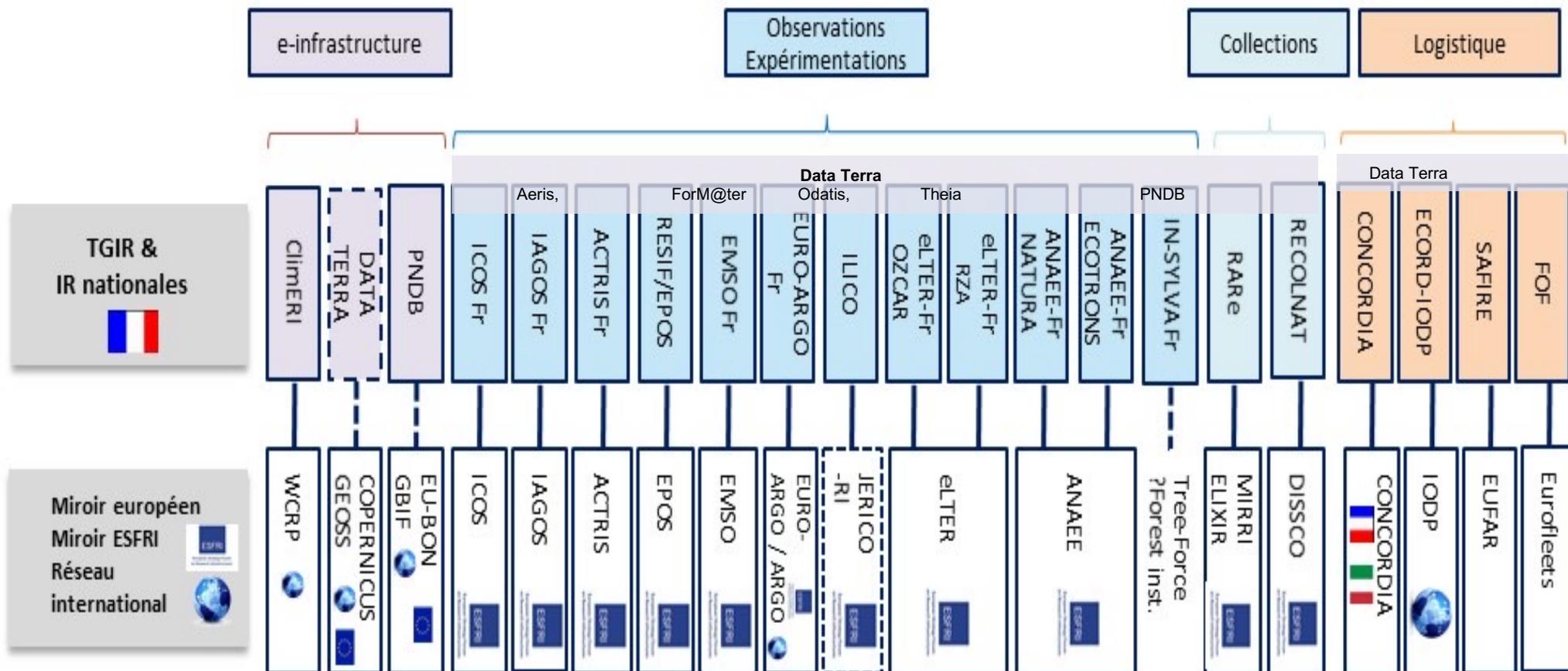


DATA TERRA E-Infrastructure de Recherche organisée à partir de pôles de données et services Grands compartiments du système Terre:

- 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.

Positionnement singulier de l'IR Data Terra dans la Feuille de route ESFRI des IR européens => vers une E-infra thématique env. en Europe (2025-2027)

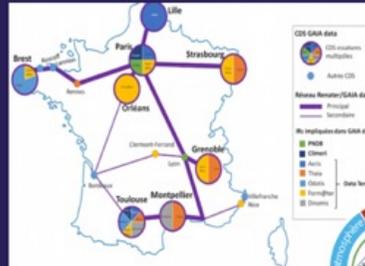




Integrated platform for Earth system sciences data and services

Develop a global system of access to data, products and treatment services to analyse, understand, model and predict history, functioning and evolution of the Earth system subject to global changes

- 26 Research Organizations and Universities
- 30 Data and Services Centers (CDOS)
- 32 Consortium of Scientific Expertise (CES)
- 230 FTE / 450 scientists, engineers



DATA & SERVICES HUBS

- AERIS** Atmosphere
- ODATIS** Ocean, coastal
- Theia** Land Surfaces
- ForM@Ter** Solid Earth
- PND B** Biodiversity
- CLIMER** Climat models, data simulation

CROSS-CUTTING SERVICES

- DATA TERRA DINAMIS** HR Spatial Imagery
- DATA TERRA INTER-PÔLES** FAIR Data and Services Networking Communities

Long term observation data



constellation satellites optic, radar...



In situ : physical and biological variables



Data from models Climat, water, ecosystem



€42m/year



1000 products & services



15,000 permanents users



> 100 PB (2022/2023)



DATA TERRA : 5 PÔLES nationaux de données et services - compartiments du système Terre et dispositifs transversaux



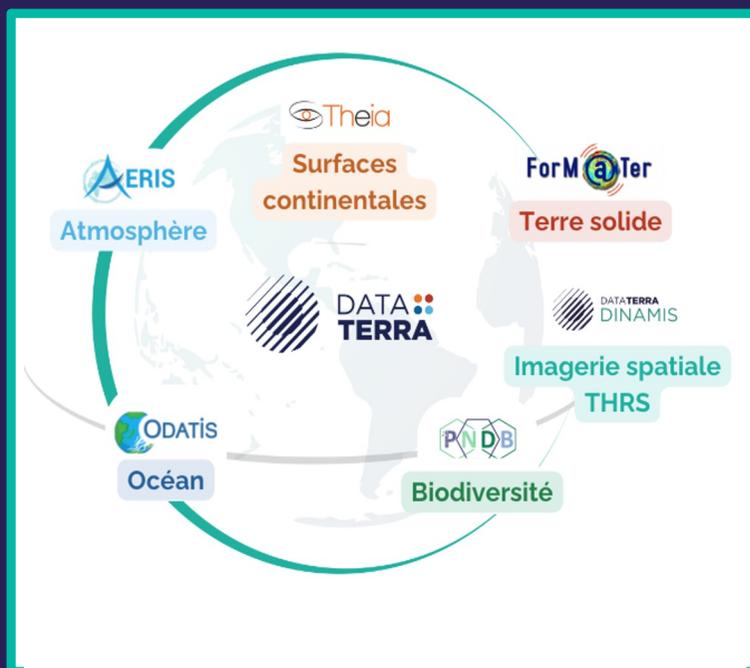
Faciliter l'accès et l'utilisation des données et produits de qualité sur l'ensemble des compartiments du système Terre (Données spatiales, aéroportées, sols, in-situ)

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



PROJETS SCIENTIFIQUES ET TECHNIQUES

Nationaux, Européens et Internationaux

- EOSC : FAIR-EASE, ...
- GAIA DATA
- PEPR ...



GRUPE DE TRAVAIL
SCIENCE, TECH, FORMATION,
EUROPE & INTERNATIONAL,
COMMUNICATION,
COOPÉRATION Pays du Sud





Services découverte, Accès et Gestion de données

Découverte (s'y retrouver parmi des milliers de jeux de données)

Catalogues (métadonnées, vocabulaires, ontologies)

Consultation et accès aux données via des web services interopérables

Visualisation avancée

Accompagnement des communautés pour la **FAIRisation**



Services transversaux pour faciliter les travaux transdisciplinaires

Grille de données, cloud, portail connaissances, SSO, métriques, support utilisateurs & formation – animation communautés

Support aux campagnes

Analysis Ready Data
Datacubes, ...



Earth Analytics Lab, VRE exploration de la donnée, bac à sable

Virtual Analysis Platform - VAP : écosystème Notebook/PANGEO/STAC

Datacubes

Capacité à se connecter directement sur les centres Traitements à la demande (WPS)

NoCode/Low Code: Galaxy-E, FG/VIP, ~Matlab/Simulink



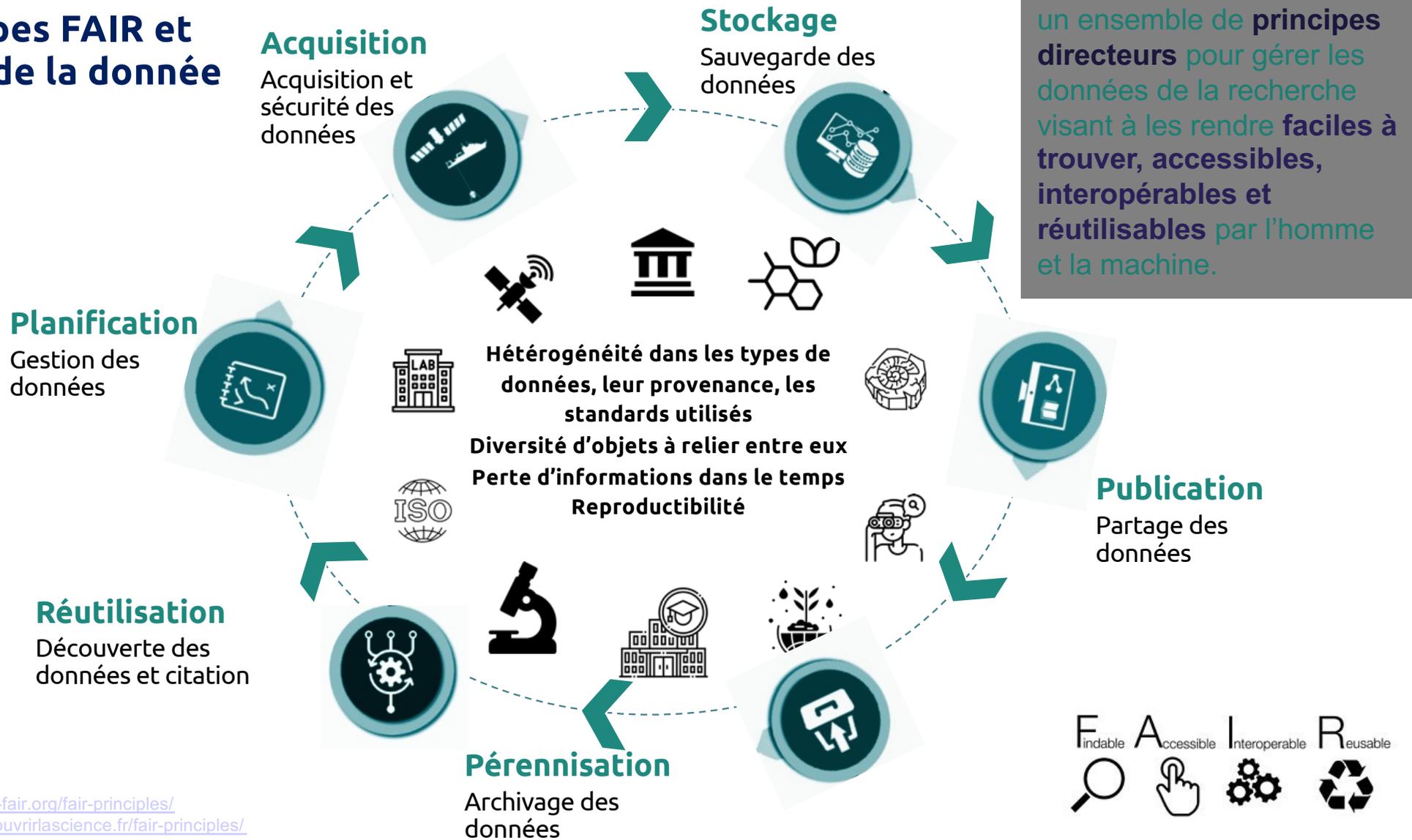
Services de production réguliers

Optimisation des traitements (outils orchestration) et formats de données (Zarr, CoG, Dask, ...)

Supporté sur un continuum d'infrastructures partagées



Principes FAIR et Cycle de la donnée



<https://www.go-fair.org/fair-principles/>
& <https://www.ouvrirlascience.fr/fair-principles/>

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

Services autour des données d'observation du système Terre interopérables et interdisciplinaires



En 2024, 5 pôles de données et de services

AERIS	FORM@TER	ODATIS	THEIA	PNDB
Atmosphère	Terre solide	Océan	Surfaces continentales	Biodiversité

Dispositif transversal
DINAMIS
Accès données satellites THRS

Dispositif transversal
ART
Animation Régional Thématiques

Dispositif transversal
Inter-pôles
Expertise collective FAIR Data and Services



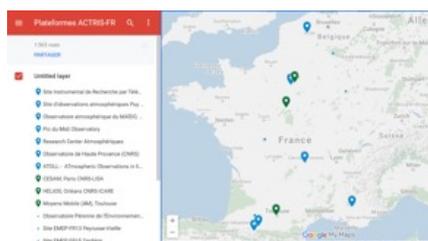
Caractériser et analyser l'atmosphère

dynamique, physique et chimie atmosphérique
évolution du climat, qualité de l'air, ...

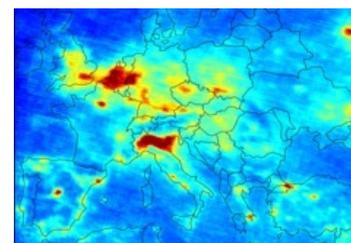
Le pôle de données et de services pour l'atmosphère AERIS fédère au niveau national des activités de gestion de données et d'expertise scientifique en atmosphère.

Les données proviennent de plateformes instrumentées, de réseaux d'observation, d'instruments à bords de satellites, ballons, avions, de mesures de laboratoires, inventaires, modèles, données de campagnes scientifiques.

Quelles thématiques pour les territoires ? Exemple de la qualité de l'air

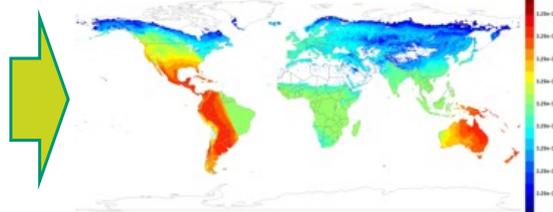


Mutualiser la surveillance de la qualité de l'air

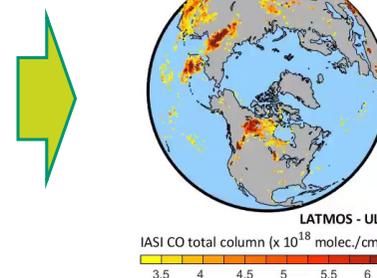


Créer un réseau européen de surveillance de la qualité de l'air

Estimer et quantifier les émissions de carbone et de particules



Suivre les effets des feux de forêts pour la qualité de l'air





ForM@Ter Connaitre la structure et la composition de la Terre

les risques telluriques et les ressources



www.poleterresolide.fr

Le pôle de données et de services pour la terre solide Form@ter fédère au niveau national des activités de gestion de données et d'expertise scientifique en matière de terre solide.

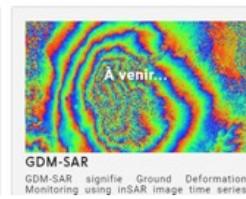
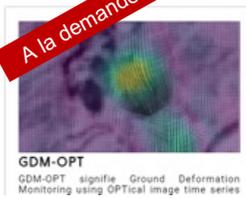
Les données proviennent de missions satellitaires, d'instruments in situ et de résultats d'expérimentations, de modélisations et de simulations.

Des services de calcul sont également disponibles.

Quelles thématiques pour les territoires ?

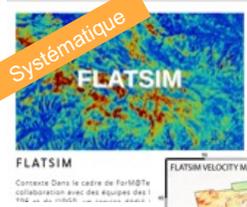
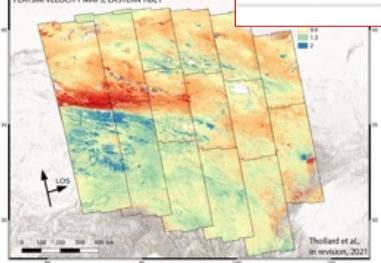
Merci de vous connecter avant de commencer.
(Lors de la première connexion, suivre les instructions)

A la demande



Calcul de déformation du sol par imagerie spatiale

Systématique

Applications

Activation de l'initiative CREST²

ERUPTION DE LA SOUFRIERE SAINT-VINCENT: EVALUER LES ALEAS VOLCANIQUES EN TEMPS REEL GRACE A L'IMAGERIE OPTIQUE HAUTE-RESOLUTION

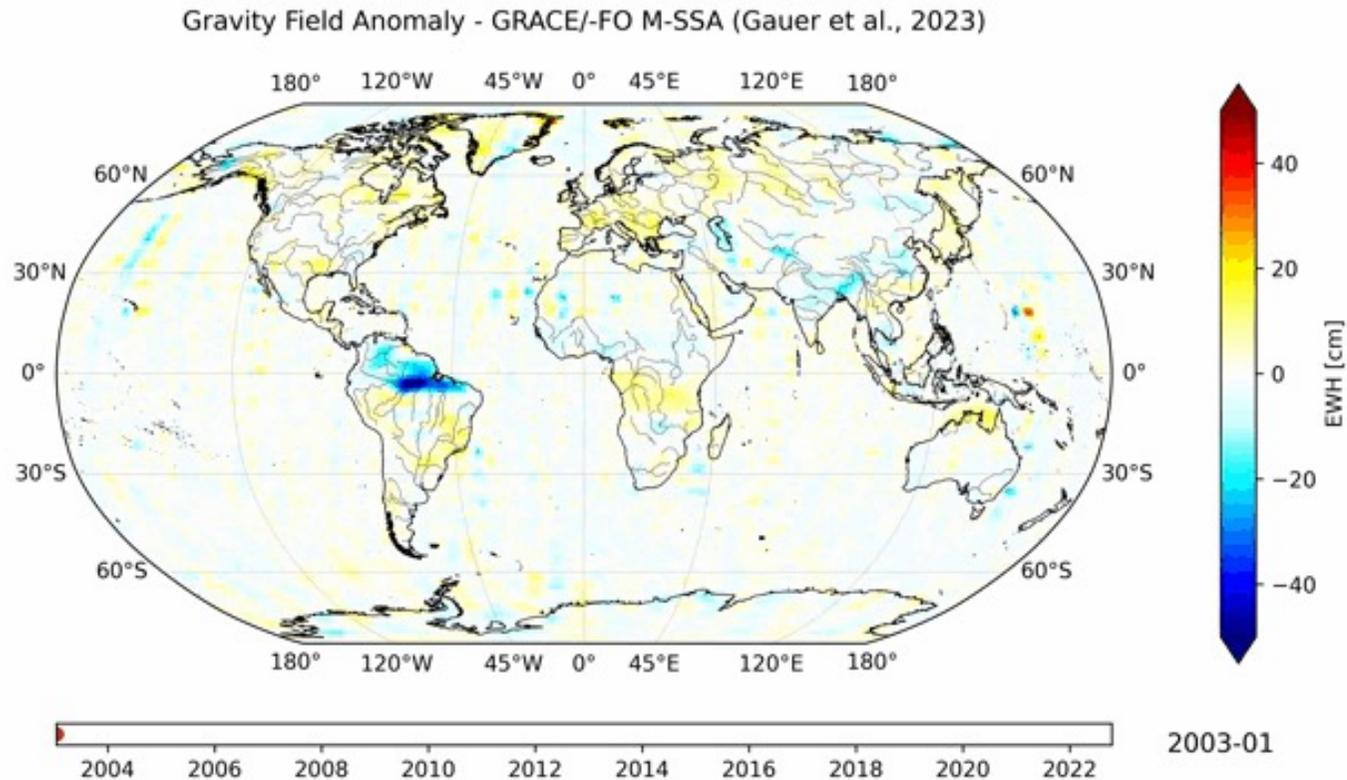
Suivi spatio-temporel de régions critiques

Applic

Déformations de la surface de la Terre à l'échelle continentale
Suivi spatio-temporel de régions critiques
→ grandes zones de failles actives et systèmes magmatiques, glissements de terrain et déformation lithosphérique à l'échelle de grands massifs...



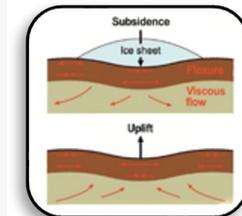
Terre interne – dynamique externe



Mission **GRACE** => carte du champ de gravité de la Terre chaque mois

- redistribution des masses
- Les principaux signaux hydrologiques:
 - L'alternance des saisons entre l'Amazone sud et nord et la mousson en Inde et en Afrique
 - La fonte des glaces du Groenland et de l'Alaska
 - Mais aussi des signaux plus profonds comme la signature du séisme du Japon qui apparaît en 2011

Le champ de gravité de la Terre mesuré est une variable essentielle de l'observation du climat car c'est un moyen unique pour accéder à une mesure **globale** des stocks d'eau.





www.odatis-ocean.fr

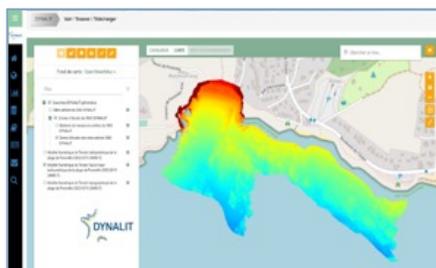
Observer et comprendre les dynamiques des océans

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

Le pôle de données et de services pour l'océan **Odatis** fédère au niveau national des activités de gestion de données et d'expertise scientifique pour les océans.

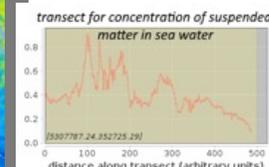
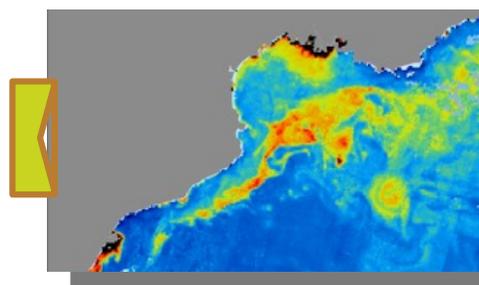
Les **données** proviennent de missions satellitaires, d'instruments in situ (plateformes fixes, flotteurs, gliders, radars, campagnes en mer, mesures en laboratoires, ...)

Quelles thématiques pour les territoires ?



Suivi du trait de côte

Qualité de l'eau



Oxygénation des eaux



Bathymétrie LASER et LIDAR

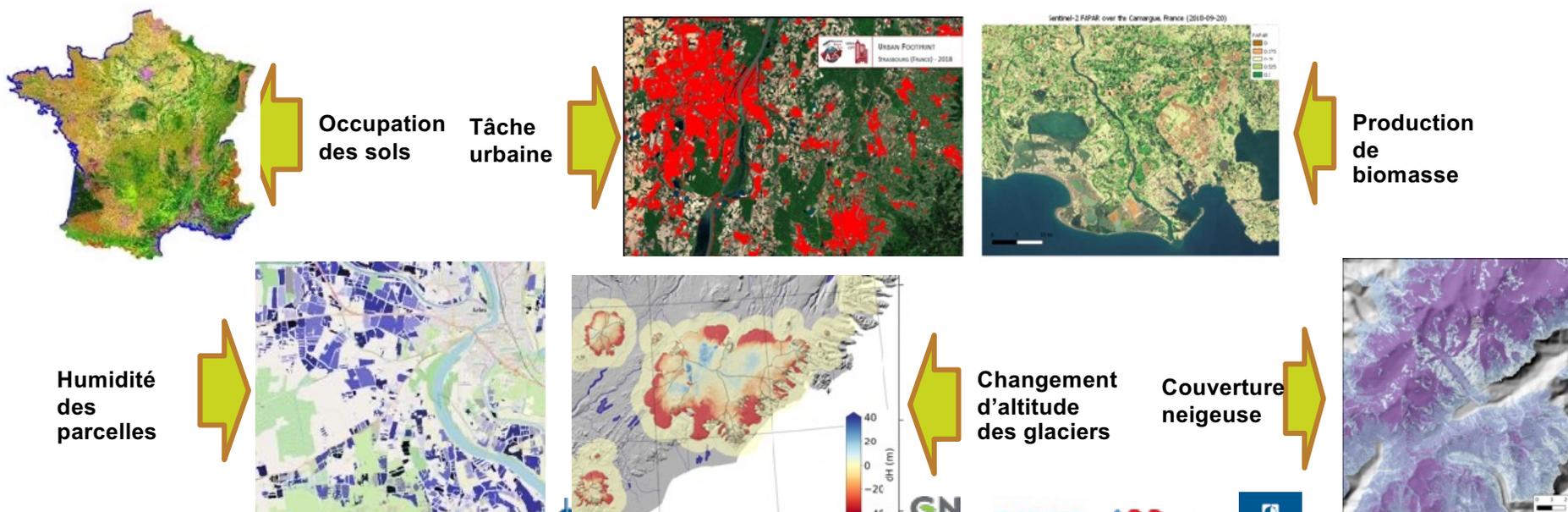


Observer les dynamiques des surfaces continentales et ressources naturelles

Le pôle de données et de services pour les surfaces continentales Theia fédère au niveau national des activités de suivi des écosystèmes et des ressources environnementales.

Les données proviennent de missions satellitaires et d'instruments in situ. Elles soutiennent le développement de produits, modèles et logiciels à valeur ajoutée dans dix thématiques : agriculture, biodiversité, climat, eau, forêt, littoral, neige & glace, risques naturels, santé, urbain.

Quelles thématiques pour les territoires ?



LAND SURFACES

Observe the dynamics of land surfaces and natural resources

The Theia land surfaces data and services hub federates **ecosystem** and **environmental resource monitoring activities** at the national level.

The data comes from **satellite missions and in situ instruments**. They support the development of value-added **products, models and software in ten themes**: agriculture, biodiversity, climate, water, forest, coast, snow & ice, natural risks, health, urban.

STUDY CASE

SOILS OCCUPATION



URBAN PLOTS



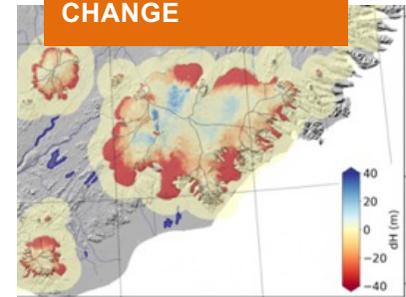
BIOMASS PRODUCTION



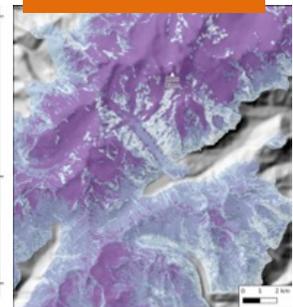
HUMIDITY PLOTS



GLACIER ALTITUDE CHANGE



SNOW COVER



www.theia-land.fr



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION

Labex
Agatex
Intermed



Var. modélisées

- Humidité du sol THRS
- Humidité du sol HRS

Humidité superficielle basse résolution
 Humidité superficielle basse résolution IB
 Humidité spatiale haute résolution
 Humidité en zone racinaire

- Couleurs/qualité des eaux

Concentration en matières en suspension
 Turbidité
 Concentration en Chlorophylle-a
 Coefficient d'abs. de la matière organ. dissoute
 Présence d'algues toxique
 Débit sédimentaires et d'éléments

Eaux de surface à continuum Terre-Mer



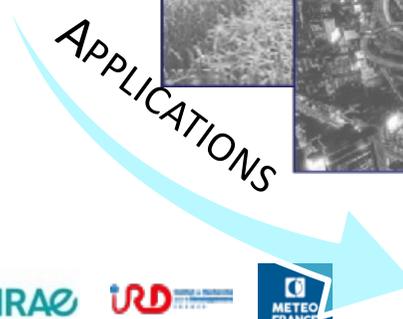
Var. thématiques (descriptives)

- Surfaces en eau (y compris surfaces inondées)
- Hauteur des lacs et des rivières
- Volumes d'eau (y compris des surfaces inondées)
- Irrigation (parcelle irriguée)
- Mangrove
- Zones humides

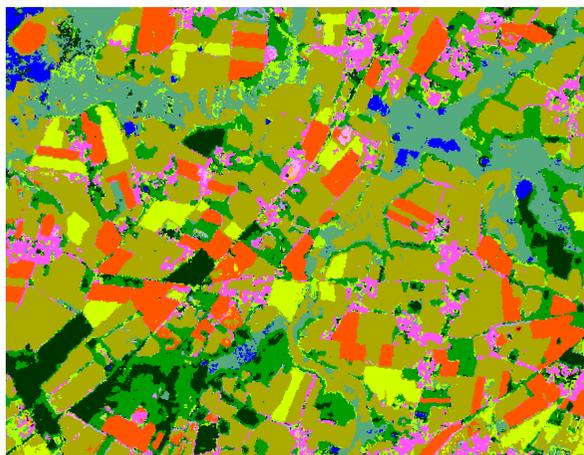
Var. modélisées (suite) :

- *Evapotranspiration*

flux de chaleur de surface
rayonnement net
evapotranspiration journalière
evapotranspiration décadaire
indice de stress



- Nomenclature de 23 classes d'occupation des sols depuis 2018, compatibles avec les 17 classes antérieures (2016 & 2017)
- A partir de Sentinel-2 et de données exogènes (climat, registre parcellaire, etc.)
- Chaîne de traitement : iota2 (open source)
- Production par Theia : annuelle sur la métropole. Productions saisonnières en réflexion.
- Echelle : 10 ~20m
- Applications : Agriculture, aménagement, risque, environnement, etc.

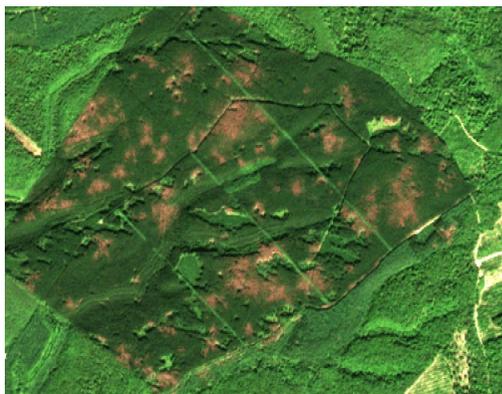


(5 millésimes)

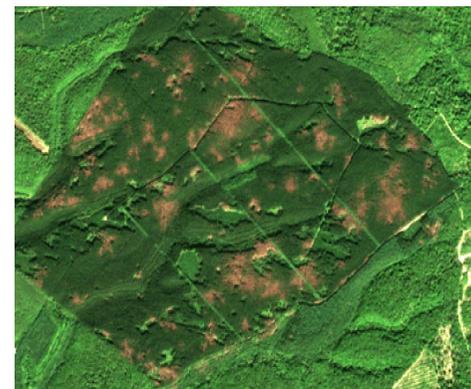
- | | |
|-------------------------------------|-------------------------------|
| Urbain dense | Tubercules / Racines |
| Urbain diffus | Prairies |
| Zones industrielles et commerciales | Vergers |
| Routes | Vignes |
| Colza | Forêts de feuillus |
| Céréales à paille | Forêts de conifères |
| Protéagineux | Pelouse |
| Soja | Landes |
| Tournesol | Surfaces minérales |
| Maïs | Plages et Dunes |
| Riz | Glaciers et neiges éternelles |
| | Eau |

- [biodivMapR](#) | Package R permettant de produire des cartes d'indicateurs de biodiversité à partir de l'information issue de données d'imagerie optique
- [FragScape](#) | PlugIn QGIS permettant de calculer les indicateurs de fragmentation du paysage de Jaeger
- [BioDispersal](#) | PlugIn QGIS permettant de modéliser les continuités écologiques
- [Fordead](#) | un package python pour la détection d'anomalies de végétation à partir d'images SENTINEL-2

Période de
détection



Classe de
confiance

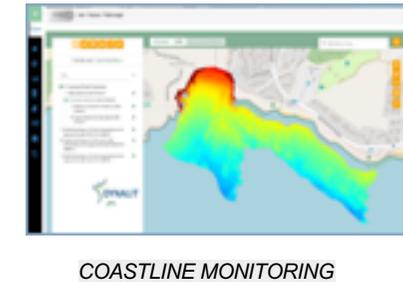
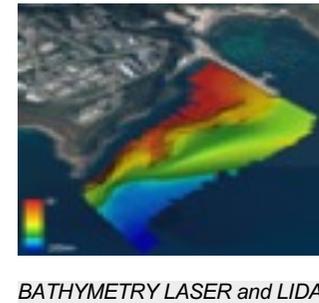
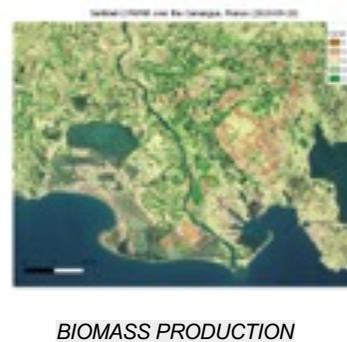
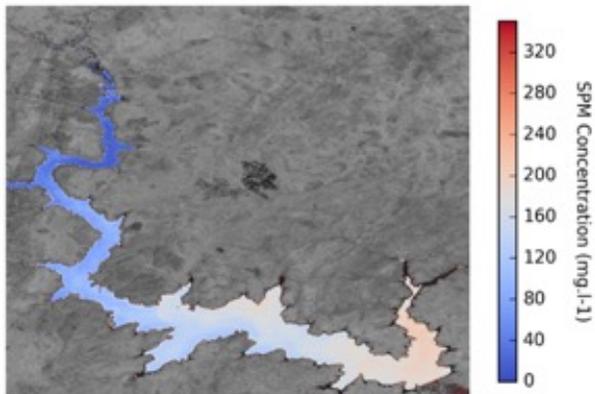
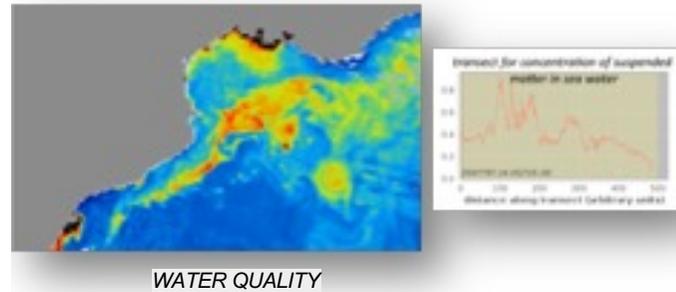




CONTINENTAL WATER COLOR

Distribution of the first water quality products (SPM suspended particulate matter) in 2018 through THEIA

We expect to cover the 30 foremost watershed (in terms of discharge) and 20 lakes in the world





dinamis.data-terra.org

Permettre et faciliter l'accès et les usages des images satellitaires de très haute résolution spatiale



Le Dispositif Institutionnel National d'Approvisionnement Mutualisé en Imagerie Satellites, DINAMIS, s'adresse à 3 types d'utilisateurs :

Utilisateurs institutionnels autorisés français

- Pour la recherche scientifique, l'enseignement supérieur et les acteurs des territoires

Utilisateurs institutionnels autorisés scientifiques étrangers

- Pour la coopération au Sud, dans l'Union Européenne et internationale

Utilisateurs privés (France)

- Pour les porteurs de projets R&D et les prestataires de commandes publiques



Couverture métropolitaine 2020. Copyright ADS - IRD, INRAE, IGN



Littoral Sénégal. Pléiades Copyright CNES, ADS



Montpellier. Pléiades © CNES 2020, distribution Airbus DS

Une offre institutionnelle

Couvertures spécifiques Pléiades littoraux sableux métropolitains et guyanais, DOM-TOM...

Couvertures millésimées SPOT 6-7 France métropolitaine

Acquisitions ad-hoc Pléiades et SPOT 6-7 Monde entier - Exprimées par les UIA, quel que soit leur lieu d'affectation.

Imageries complémentaires haute résolution

Relais vers le programme CNES Spot World Heritage (Spot 1-5), plateforme CNES PEPS (Sentinel 2), Geosud, programme CNES Kalideos : RapidEye, CosmoSkyMed, TerraSar,-X, Aster...



ÉRUPTION VOLCAN TONGA

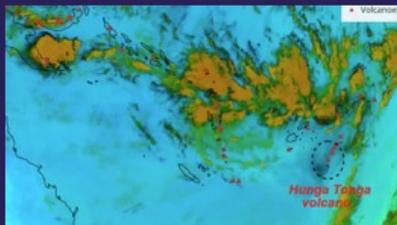
14 janv. 2022 l'éruption volcanique sous-marine a entraîné des vagues qui ont frappé l'Australie, la Nouvelle-Zélande, le Japon ainsi que les côtes ouest de l'Amérique du Nord et du Sud durant plusieurs jours. Toutes les informations recueillies par les scientifiques **facilitent ainsi les interventions d'urgence** et aident à **mieux comprendre et anticiper ce type de catastrophes naturelles à l'avenir.**



Les îles Tonga sont composées d'environ 170 îles

AERIS

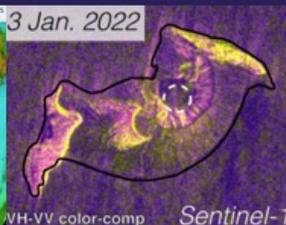
Atmosphère



Vues par les satellites géostationnaires

FORM@TER

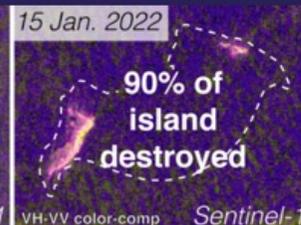
Terre solide



Évolution de la morphologie de l'île volcanique Hunga Tonga vue par imagerie satellite

ODATIS

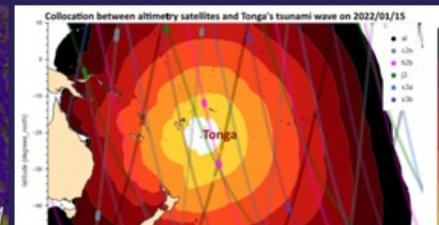
Océans



Mesures de la hauteur du tsunami au large, par les satellites altimètres et via les marégraphes côtiers

THEIA

Surfaces continentales



Cartographier les dommages pour guider les secours

DINAMIS

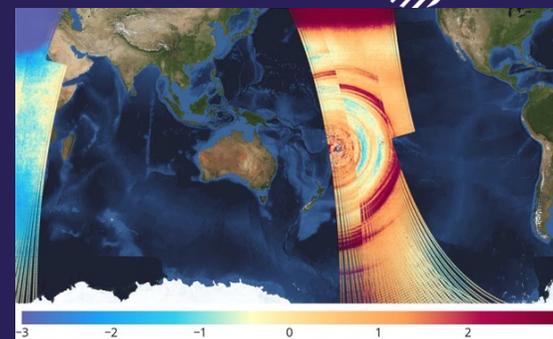
Images satellites HD



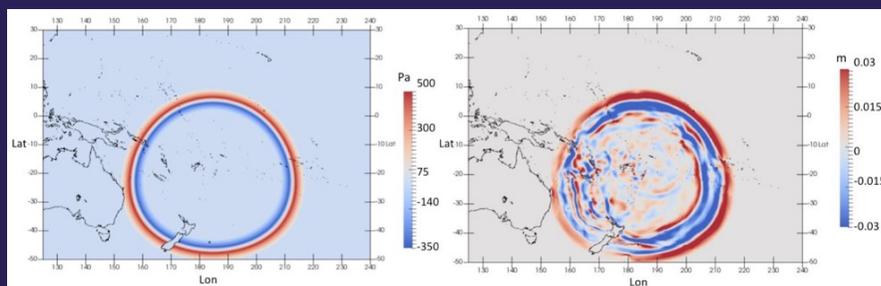


Images Pléiades-HR post-événement (à droite) et des images WorldView pré-événement (à gauche) sur les îles

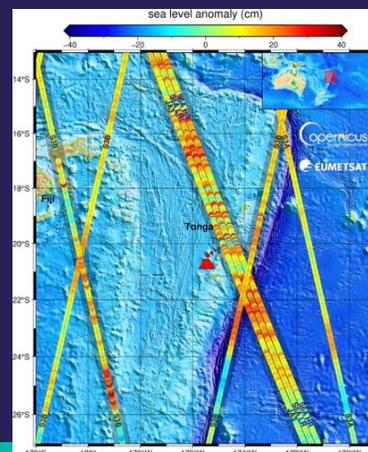
ERUPTION TONGA



Perturbations du champ de température dues aux ondes de gravité se propageant dans la stratosphère mesurées par le sondeur IASI

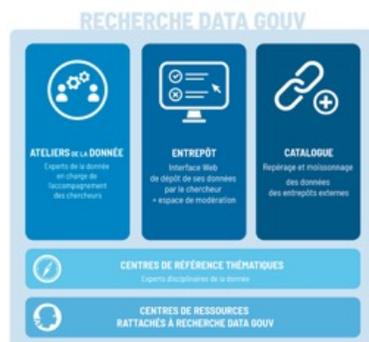


Propagation de l'onde de Lamb et propagation du tsunami associé, 3 heures après l'explosion. A gauche : champ de pression en surface (en Pa) modélisé par une onde sinusoïdale se propageant à la vitesse du son. A droite : élévation de la surface de l'eau (en m) générée par l'onde de Lamb et simulée



Anomalies du niveau de la mer (hauteur du niveau de la mer par rapport à une moyenne à long terme) mesurées par Sentinel-3A et -B, Sentinel-6 et Jason-3, les 14 et 15 Janvier à 04:15 UTC6

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



EaSy Data, entrepôt de données longue traîne



POUR QUI ?

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

QUEL TYPE DE DONNÉES ?

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

DONNÉES LONGUE TRAINE

Données ou produits issues de travaux de recherche de durée finie (projets ou publications) dont l'acquisition/l'élaboration, la préservation et la diffusion ne sont pas organisées de façon pérenne ou communautaire. Une (grande) partie des données de longue traîne sont dites « orphelines », au sens où elles ne bénéficient pas d'un service d'entrepôt de données susceptible de les préserver et de les rendre accessibles à la communauté scientifique".





Data Terra in the european landscape

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

Data Tera contributes to an increasing number of EOSC projects, ESFRI, DestinE, Copernicus

FAIR EASE (01/09/2022, 5M€, 26 partners, coord: Data Terra/CNRS A. Rizzo, IRD). Objectif: develop et leverage within EOSC the distributed and integrated observation and modelling for the Earth System, Environment and Biodiversity, with the user communities and the European Science Research Infrastructures. 143 PM/1,2M€

FAIR IMPACT (01/06/2022, 10M€, coord: DANS). Objectif: develop a FAIR EOSC « EOSC de données et de services FAIR », by supporting the practical implementation of the FAIR principles in the scientific communities and for research results at the national, European and international levels. 14PM

GEO INQUIRE (01/10/2022, 15M€, coord: GFZ). Objectif: Access to observations, data products and services for monitoring and simulating the geosphere dynamic processes, at unprecedented levels of detail and accuracy. 8PM

DT GEO (01/09/2022, 10M€, coord: BSC). Objectif: Deploy a digital twin prototype for geophysical and human-made environmental extreme events: Coupled Digital Twin Components (DTC) for risk associated to earthquakes and induced seismicity, volcanic eruptions, and triggered tsunamis (earthquakes, landslides). 4PM

others: BioDT, InterTwin

Data Terra: integrated French contribution to ESFRIs and Infrastructure integrati

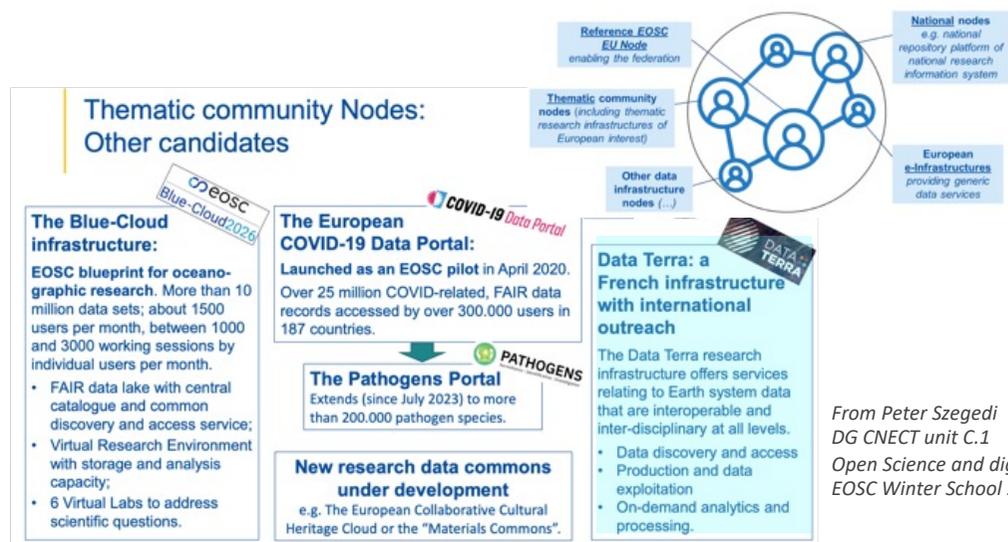
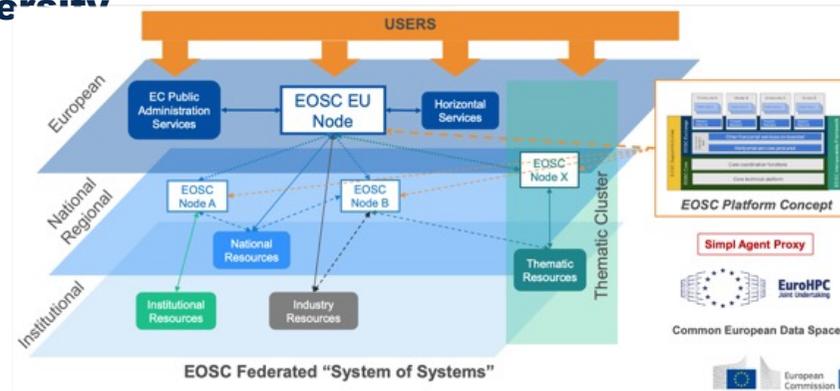
- > ACTRIS, ICOS, IAGOS, EU-FAR, Euro-ARGO, EMSO, SeaDataCloud, JERICO, eLTER, EPOS
- > ENVRI-FAIR



Data Terra: The French EOSC thematic community Resources/node for earth system, environment and biodiversity

Operate the French resources contribution to an Earth System, Environment and Biodiversity Thematic EOSC Cluster

- ▶ In synergy with ESFRI Clusters (ENVRI Fair) and other **European national thematic cluster initiatives**: e.g., NFDI@Earth and NFDI@Biodiversity (Germany), ITINERIS (Italy), UK environmental services ...
- ▶ **Sustainable science-driven thematic resources and services** integrated into EOSC portal and leveraging EOSC core services
- ▶ **Open and trusted FAIR long-term high-quality curated multi-domain and multi-component data**: observation (space, airborne, in-situ), simulation and experiments
- ▶ **Thematic reference data models and standards (data, metadata)**, machine readable catalogs, certified thematic repositories
- ▶ **Science-driven expertise and resources in-and-across integrated sub-domain**: AERIS, ODATIS, Form@Terre, Theia, PNDB, CLIMERI, DINAMIS
- ▶ **High-quality and trusted open data corpus** with standardised thesaurus and ontology, and knowledge-based evaluation and certification methodologies:
 - **ML/DL models for multi-modal data cross-analysis**
 - **Foundation models** (unimodal and multimodal) for Earth science, environment, biodiversity discovery,
 - **Science-based Digital Twins** in Earth science, environment and biodiversity



From Peter Szegedi
DG CNECT unit C.1
Open Science and digital mo
EOSC Winter School 2024

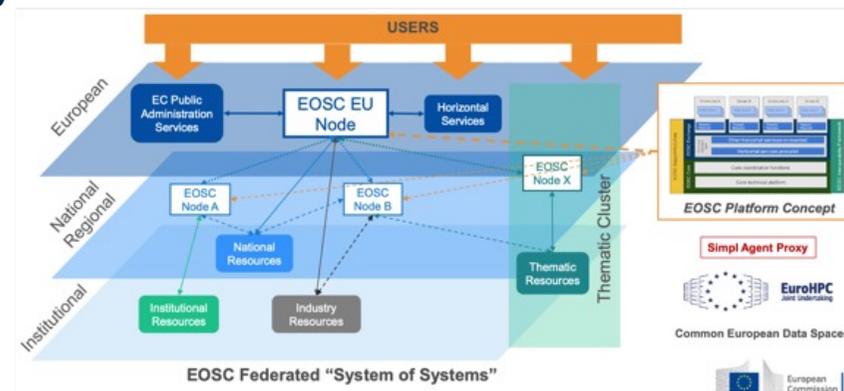


Data Terra: The French EOSC thematic community Resources/node for earth system, environment and biodiversity



Operate the French resources contribution to an Earth System, Environment and Biodiversity Thematic EOSC Cluster

- ▶ In synergy with ESFRI Clusters (ENVRI Fair) and other **European national thematic cluster initiatives**: e.g., NFDI@Earth and NFDI@Biodiversity (Germany), ITINERIS (Italy), UK environmental services ...
- ▶ **Sustainable science-driven thematic resources and services** integrated into EOSC portal and leveraging EOSC core services
- ▶ **Open and trusted FAIR long-term high-quality curated multi-domain and multi-component data**: observation (space, airborne, in-situ), simulation and experiments
- ▶ **Thematic reference data models and standards (data, metadata)**, machine readable catalogs, certified thematic repositories
- ▶ **Science-driven expertise and resources in-and-across integrated sub-domain**: AERIS, ODATIS, Form@Terre, Theia, PNDB, CLIMERI, DINAMIS
- ▶ **High-quality and trusted open data corpus** with standardised thesaurus and ontology, and knowledge-based evaluation and certification methodologies:
 - **ML/DL models for multi-modal data cross-analysis**
 - **Foundation models** (unimodal and multimodal) for Earth science, environment, biodiversity discovery,
 - **Science-based Digital Twins** in Earth science, environment and biodiversity



Proposal for the potential candidate Nodes

Life sciences

BBMRI ERIC
Elixir Hub
EMBL
Euro-BioImaging ERIC
Instruct ERIC

Social Sciences & Humanities

ARIADNE RI
CLARIN ERIC

Environment

ACTRIS ERIC
CNR (Blue-Cloud)

Photon & Neutron

ESRF (PANOSC)

(Astro)particle physics

CERN
CNRS-LAPP (ESCAPE)

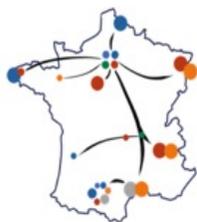
Multinational e-infra

EUDAT

National scope/mandate

AT ACONET / EOSC Support Office
DE NFDI
DK DelC
FI CSC – IT Center for Science
FR CNRS (Data Terra)
HR SRCE
HU HUN REN
IT Foundation ICSC
LU Lux NDS
NL SURF
NO NRIS
PL NCN
SE NDS
SI ARNES
SK CVTI SR
UA BITP

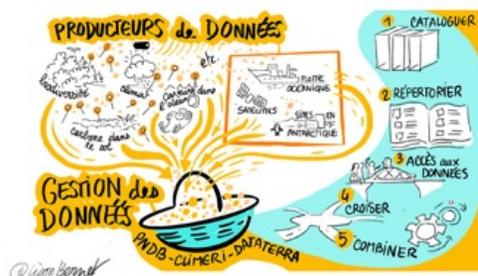
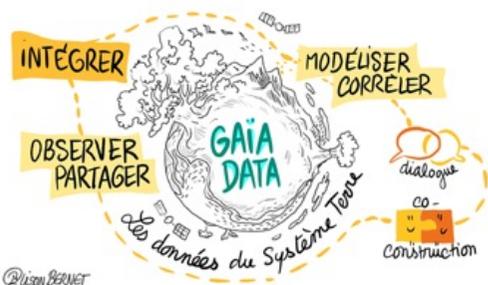




EQUIPEX+ /PIA3

OBJECTIF : Développer et mettre en œuvre **une infrastructure/plateforme intégrée de données FAIR et de services distribués pour l'observation, la modélisation et la compréhension du Système Terre, de la Biodiversité et de l'Environnement**

- **sur l'ensemble du cycle de la donnée**, de son acquisition (spatiale, sols, in-situ, simulation) jusqu'à ses **multi-usages** (qualification/validation, stockage, accès, traitements/croisements de données multi- sources/extraction de connaissances, produits/services)
- **pour la communauté scientifique** contribuant à la connaissance du système Terre, de la biodiversité et de l'environnement ; **acteurs publics et privés**



Equipement Structurant pour la Recherche / EQUIPEX+



Budget total : 62 M€ (coûts complets)

Demande ANR-EQUIPEX+PIA3 : 16,2 M€

Durée : 8 ans (06/2021 -> 06/2029)

Forte contribution RH : 339 ETP (soit 4066 p.m.) personnels permanents + 59 ETP (711 p.m.) cdd

Apports additionnels des organismes : 25 postes (recrutements cdi, postes permanents, mobilités)

3 Infrastructures de Recherche : Data Terra, Climeri-France, PNDB

21 Partenaires : CNRS, CNES, IFREMER, IRD, BRGM, IGN, INRAE, Météo-France, MNHN, CEA, IPGP, CINES, Sorbonne Univ., Univ. Grenoble-Alpes, Univ. Lille, Univ. F. Toulouse, UNISTRA, SHOM, OCA, FRB, CERFACS

Equippe Projet :

Responsable Scientifique et Technique : Frédéric Huynh (IRD/Data Terra)

Coordination Technique : Joel Sudre (CNRS/Data Terra)

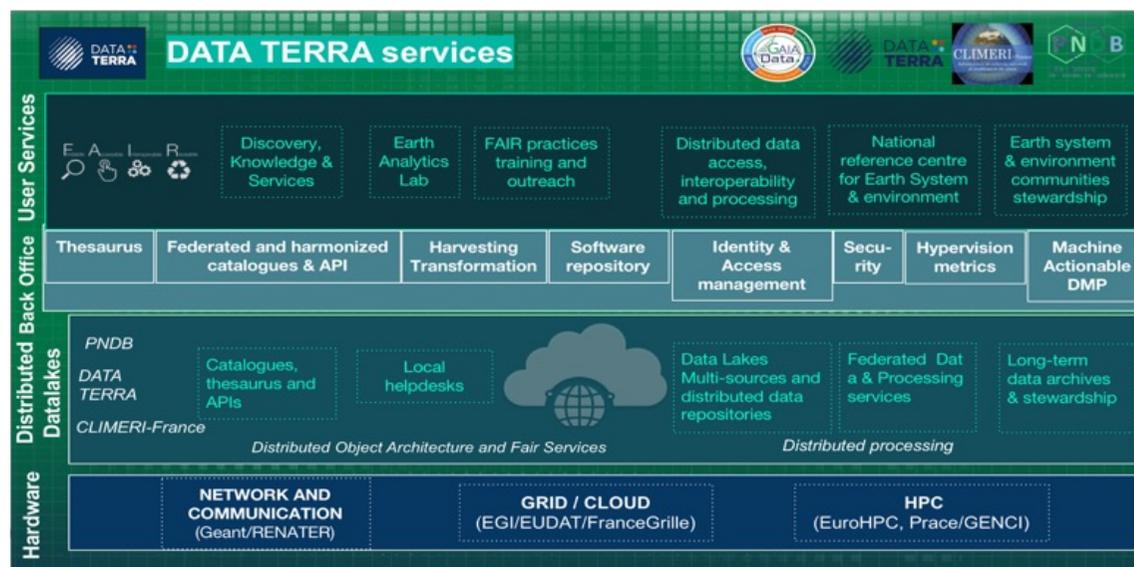
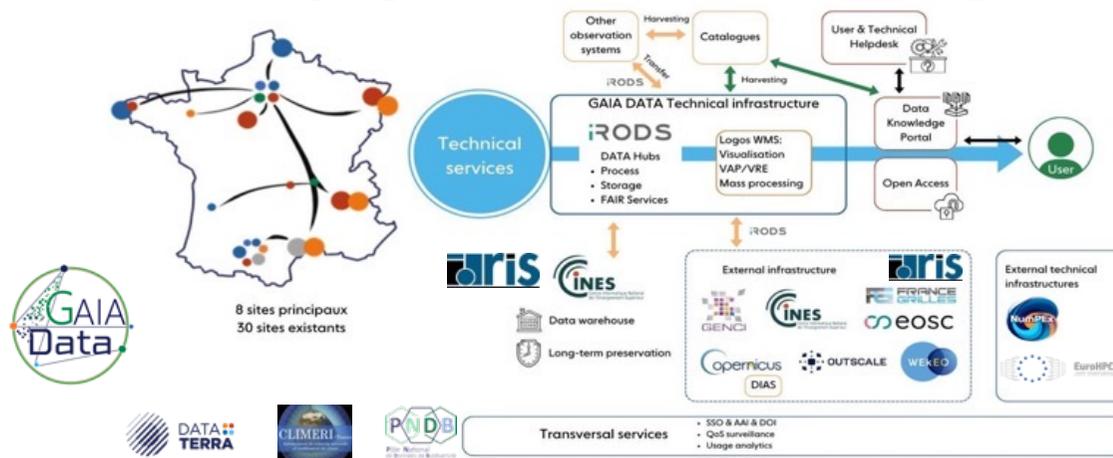
Responsables WP : K. Ramage(CNRS)/E. Chaljub(UGA) (WP2), R. Moreno (CNES)/G. Levasseur (SU) (WP3), JF Piolle (Ifremer) (WP4), C. Blanke (WP5)





Data Terra: The national E-infrastructure operating services developed through the national GAIA Data project

- ◆ Interoperable FAIR multi-domain data and compute services across a continuum of infrastructures
- ◆ Open AI-powered models for Earth system and environment science: e.g.,
 - ▶ foundation models for science, image analysis, event detection, data assimilation, physics-based models, Graph GNNs, surrogate models ...
- ◆ Science-driven Digital Twins projects: e.g.,
 - ▶ Extreme events, Climate, Digital Ocean, Hydrology, Geo-hazards, Territory
- ◆ Science-driven applications: e.g.,
 - ▶ connecting Copernicus, EO data, socio-economic information, IT information, scenarios ...
 - ▶ for decision making and economic activities, risk mitigation, sustainable urban areas, energy transition, water ...
- ◆ Support to National Priority Programs: e.g.,
 - ▶ One Water: "Water, Common Good",
 - ▶ FairCarbonN: Carbon in continental ecosystems, trajectories for carbon neutrality
 - ▶ Sous-Sol: géothermie, H2 and CO2 storage, water
 - ▶ TRACCS: Transforming Climate modelling for climatic services
 - ▶ Digital Ocean: toward a digital twin of the ocean system
 - ▶ NumPEX: numerics for exascale



Données, IA et souveraineté

L'utilisation des données, avec l'IA, exige un cadre de confiance, permettant d'accéder à des données certifiées :

- ✓ **Intègres** : qualité et traçabilité
- ✓ **Vraies** : les données sont-elles pertinentes pour l'utilisation, avec une expertise scientifique ?
- ✓ **Certifiées** : quel est le degré de confiance dans les résultats de l'utilisation des données

Les données ne sont rien sans la confiance que leur accorde un tiers crédible

Par son expertise scientifique, ses méthodes de validation scientifique et ses méthodologies, la recherche est au cœur de la production et de l'utilisation fiable des données.



Accompagner les transformations pour des bénéfices durables

L'IA va transformer les pratiques de recherche

- ✓ Favorise les interactions entre les domaines
- ✓ Accélère le décloisonnement nécessaire aux transitions
- ✓ Faire évoluer les organisations et leurs instincts de conservation

Conditions du succès pour des usages au service de la société

- ✓ Lier données de qualité, capacité de calcul/stockage et IA
- ✓ Garantir un continuum données, expertise scientifique pluridisciplinaire, modèles
- ✓ Capacité à superviser les usages de l'IA
- ✓ Disposer d'un cadre de confiance et éthique pour que les résultats soient adoptés pour la gestion des territoires

Disposer d'entité souveraine de recherche et de formation « IA, science et société »

Préconisation G1 Comité National Pilote d'Éthique du Numérique

Au vu de la complexité des enjeux de l'IA générative/fondation et de ses impacts à moyen et long termes, il est nécessaire de créer une entité souveraine (centre de compétence) consacrée à la recherche et à la formation relatives aux questions éthiques des systèmes d'IA en lien avec leurs enjeux scientifiques, techniques, sociétaux et environnementaux. (CNPEN)

EVOLUTIONS RECENTES ET PERSPECTIVES

Data Terra est l'infrastructure de recherche dédiée aux services numériques du domaine environnement et système Terre. Créée en 2016/2017, elle est inscrite sur la feuille de route nationale 2016, 2018, 2021 des IR/IR*.

Forte évolution de l'IR Data Terra :

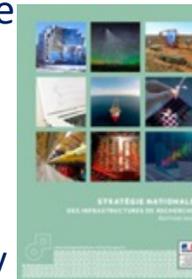
- périmètre scientifique sur l'ensemble des thématiques du système Terre
- missions élargies : Centre de Référence Thématique dans le cadre de Recherche Data Govv
- Statut de E-Infrastructure de recherche du domaine environnement et système Terre réaffirmée dans la FRN 2021-2025
- Innovation transfert, numérique-réseau d'infra distribuée-Gaia Data, ...
- ambitions européennes (EOSC, JN-DestinE, Copernicus, ...) et internationales (Pays du Sud
- valoriser pour partager et influencer sur le modèle Data Terra : souveraineté des données, continuum données-connaissances, spatiale-in-situ modèle, place et mobilisation de la science au travers des données pour les défis de société (transitions numériques et systémiques)

Un modèle (fédération, mutualisation, engagements inter-organismes, projets collaboratifs) à renforcer, partager et à développer pour s'adapter et être à l'échelle des nouvelles missions, enjeux et défis



e-infrastructures

Le Pôle de données et services du système Terre DATA TERRA est une infrastructure qui se structure pour offrir un portail unique vers les quatre pôles (AERIS pour l'atmosphère, ODATIS pour l'océan, FORM@TER pour la Terre solide et THEIA pour les surfaces et interfaces continentales) ainsi que des services et outils. Il permettra de contribuer aux grands programmes européens comme COPERNICUS et internationaux comme GEOSS



E-infrastructures

Une seule infrastructure de données et de services du système Terre (composantes biologiques et physiques) issue de l'intégration des IR DATA-TERRA et PNDB associées dans le projet GAIA DATA (PIA 3/Equipex+) est attendue à court/moyen terme.

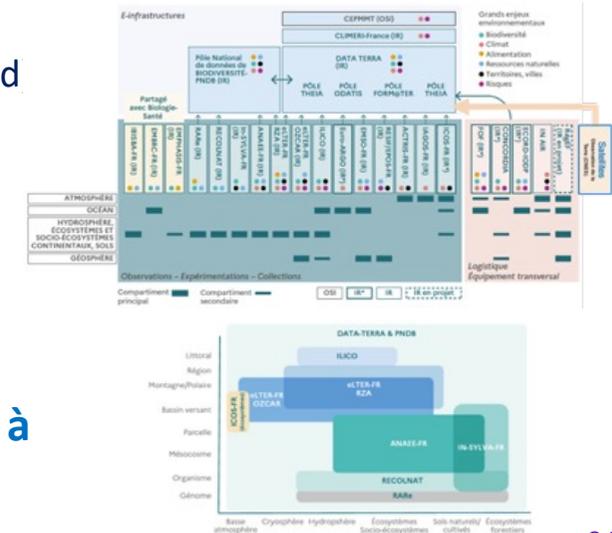


Figure 5: Infrastructures de recherche françaises dans le domaine « système Terre », regroupées et mises en perspective continentales et globales. Les infrastructures de logistique pour l'ensemble du domaine ne figurent pas sur cette figure.



DATATERRA



contact@data-terra.org
+33 (0)4 67 54 87 08

www.data-terra.org