



## IR DATA TERRA : l'E-Infrastructure du Système Terre et environnement

### Pôles de données et services pour le système Terre



**Frédéric Huynh (IRD), directeur IR Data TERRA**

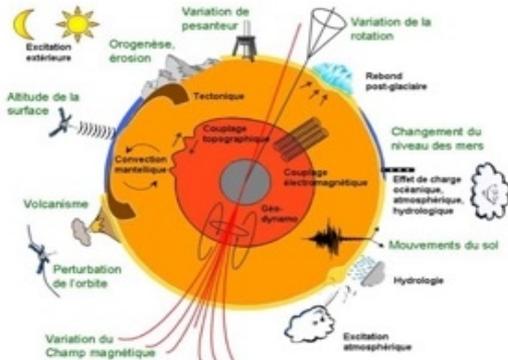
Emmanuel Chaljub (CNRS), directeur pôle **Form@Ter** ; Patrice Henry (CNES), directeur pôle **AERIS** ; Gilbert Maudire (IFREMER), directeur pôle **ODATIS** ; Anne Puissant (UNISTRA), directrice pôle **THEIA** ; Richard Moreno (CNES), **directeur technique** ; Jean-François Faure (IRD), secrétaire exécutif **DINAMIS**



# 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.*



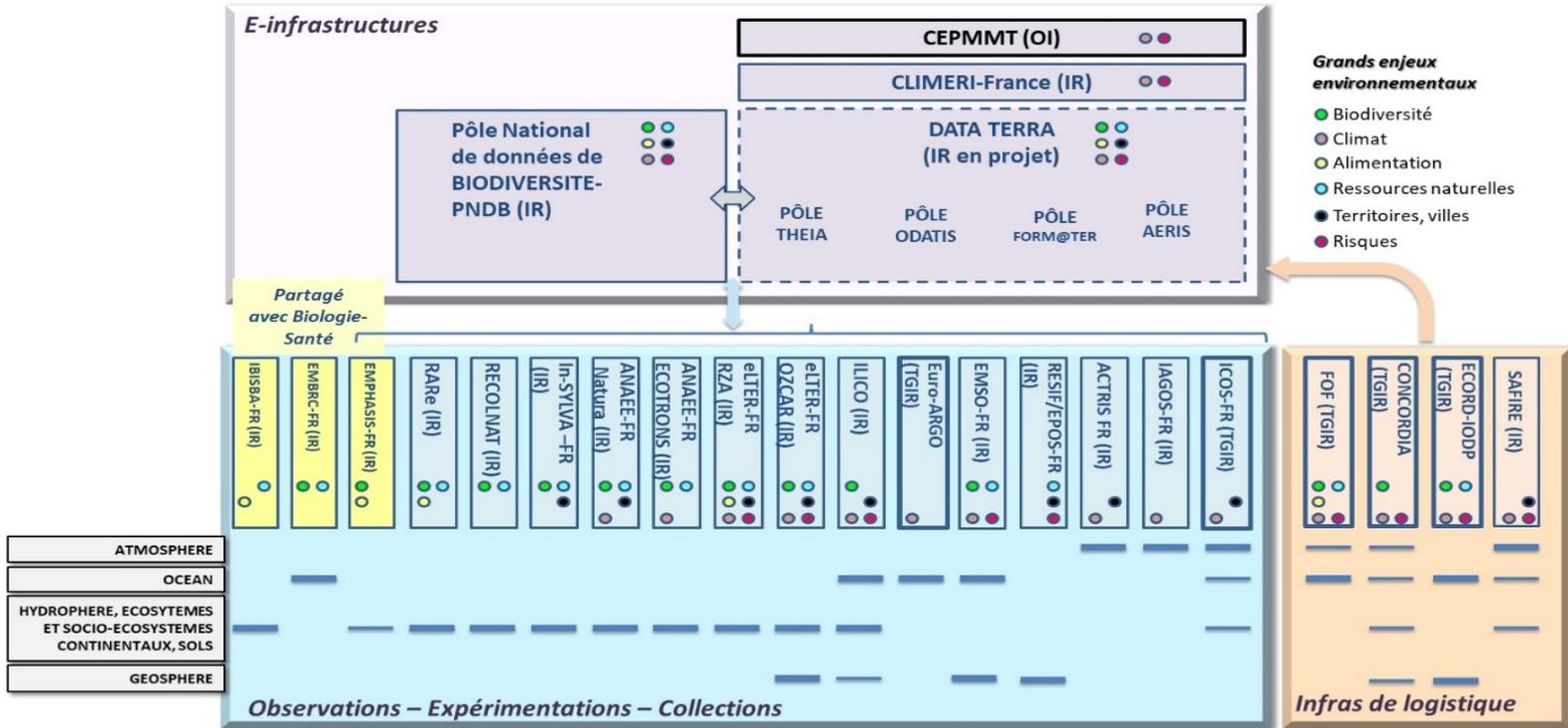
D'après O. de Viron

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 ainsi que des résultats d'expérimentation, modèles, etc ...*)

- **Groupe de réflexion « pôles thématiques en Observation de la Terre » 2012-2014** – CNES-CNRS/INSU, ... -  
*Création de pôles de données et d'une IR système Terre*
- **Feuilles de route nationale IR/TGIR – MESRI** contribution domaine « système Terre et Env. » 2016
- **Open science / open data**
- **Européanisation des activités : infrastructures de recherche (ESFRI...), services d'accès aux données spatiales et in-situ (EOSC, COPERNICUS, ...)**
- **Internationalisation des dispositifs de partage des données et services (GEO, GO FAIR, ...)**



# Le paysage actuel des infrastructures

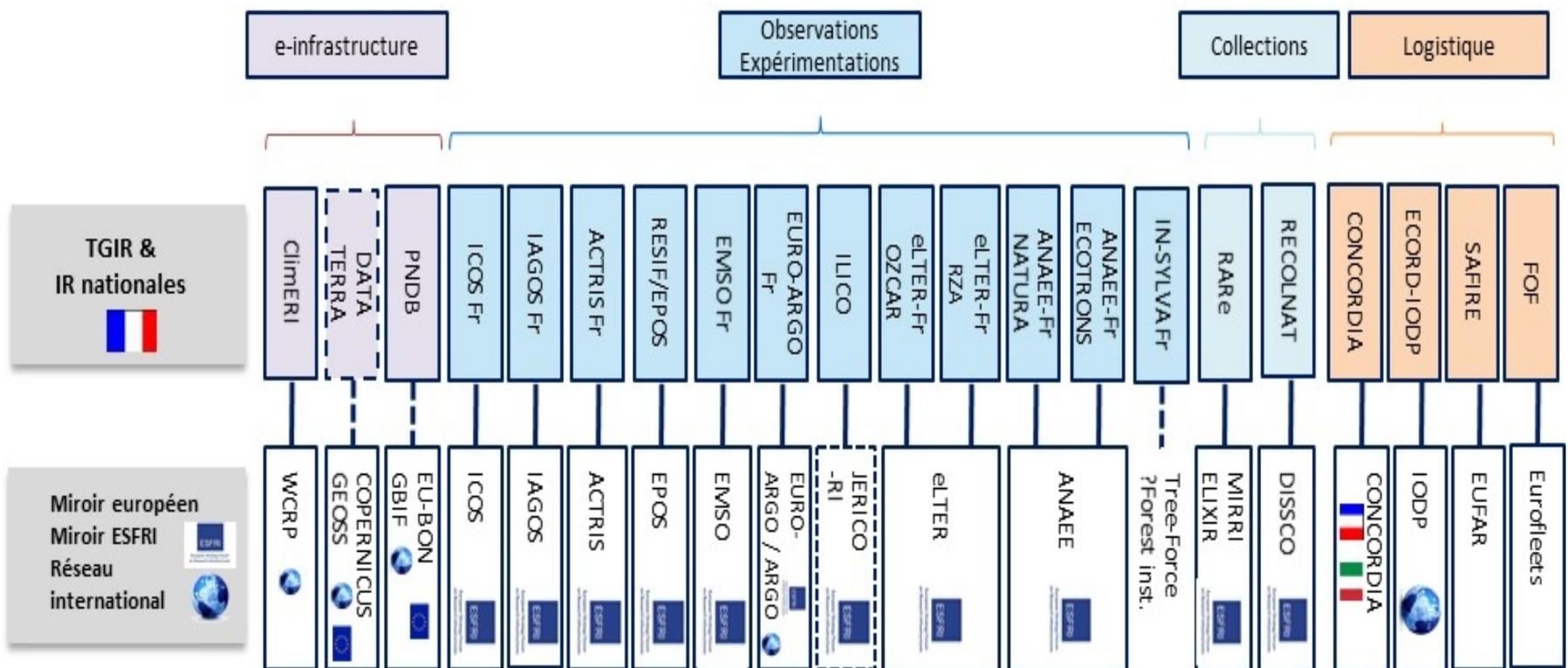


Compartiment principal — Compartiment secondaire —

OI TGIR IR IR en projet

MESRI - FRN 2022-2025 - HC-IR, mai 2021

# Le paysage actuel des infrastructures

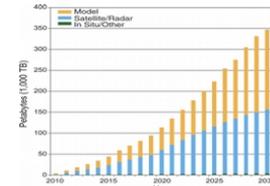


TGIR & IR nationales

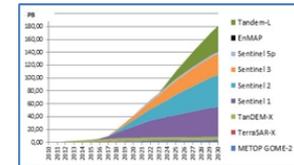
Miroir européen  
Miroir ESFRI  
Réseau international

# 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*
- **Augmentation exponentielle du nombre de données, diversités des sources, complexités, ...**  
*Spatiales, in-situ, modèles => besoins d'analyse/réanalyse, traitements intelligents*
- **Co-construire une infrastructure distribuées de services spécifiques du domaine système Terre et Environnement**  
**=> enjeux scientifiques et données transverses**
- **Concilier structuration nationale transversale à l'Europe**  
**=> 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**



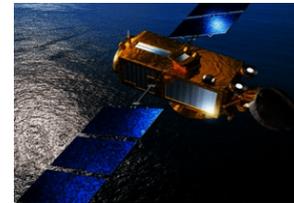
Projected increase in global climate data for climate models, remotely sensed data, and in situ instrumental/proxy data.  
 From Overpeck et al. Science, 2011



# Objectif général

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**



# Défis scientifiques transversaux - Verroux scientifiques et techniques associés aux données

CONTINUUM ET INTERFACES TERRE - MER

APPROCHE INTÉGRÉE DE LA ZONE CRITIQUE

DYNAMIQUE DES MILIEUX/ENVIRONNEMENTS  
URBAINS

CLIMAT ET CYCLE DE L'EAU

ALEAS ET IMPACTS SUR LA SOCIÉTÉ

CONTINUUM ET INTERFACES TERRE - MER

OBSERVATION ET MODÉLISATION DES IMPACTS DES CHANGEMENTS GLOBAUX

→ développement d'une approche intégrée

DYNAMIQUE DES ÉCOSYSTÈMES MARINS

Echelles spatiales fines (méso- et submésoéchelle)

→ accéder aux données in situ et à des produits spatiaux  
(ex. couleur de l'eau sur des zones géographiques méso-échelles)

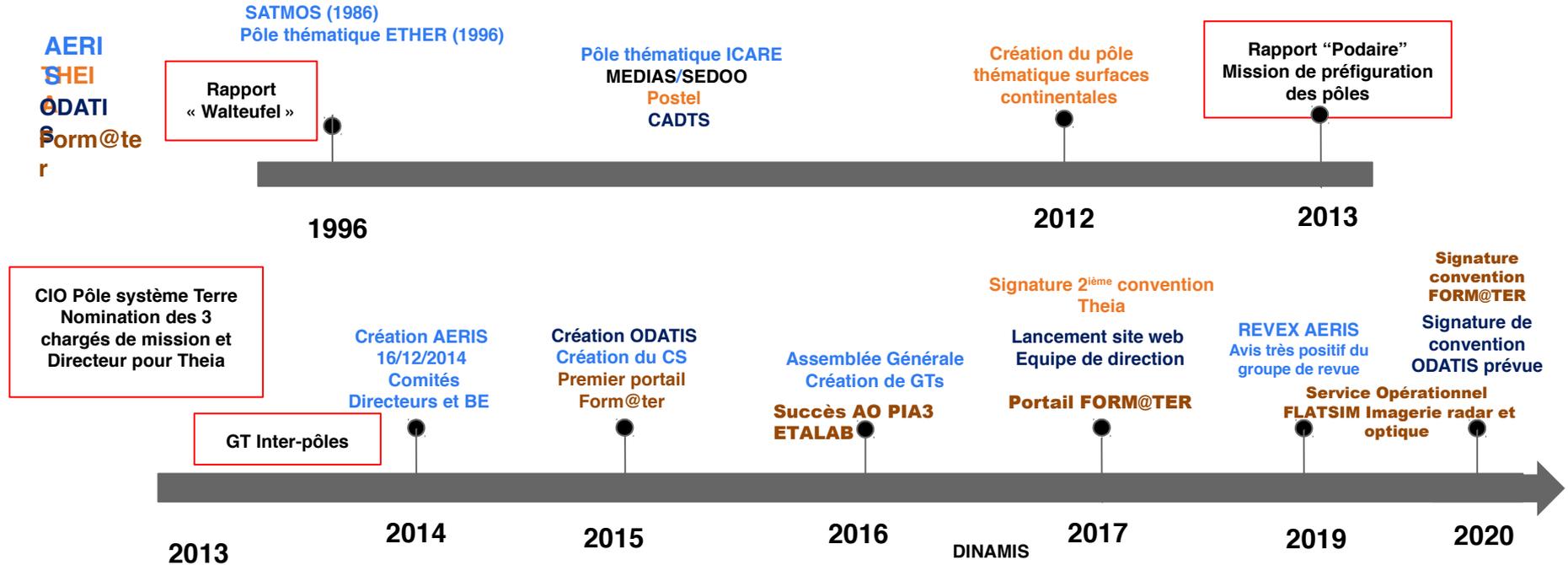
GESTION DURABLE DES RESSOURCES

CLIMATOLOGIE DE RÉFÉRENCE DU MILIEU CÔTIER

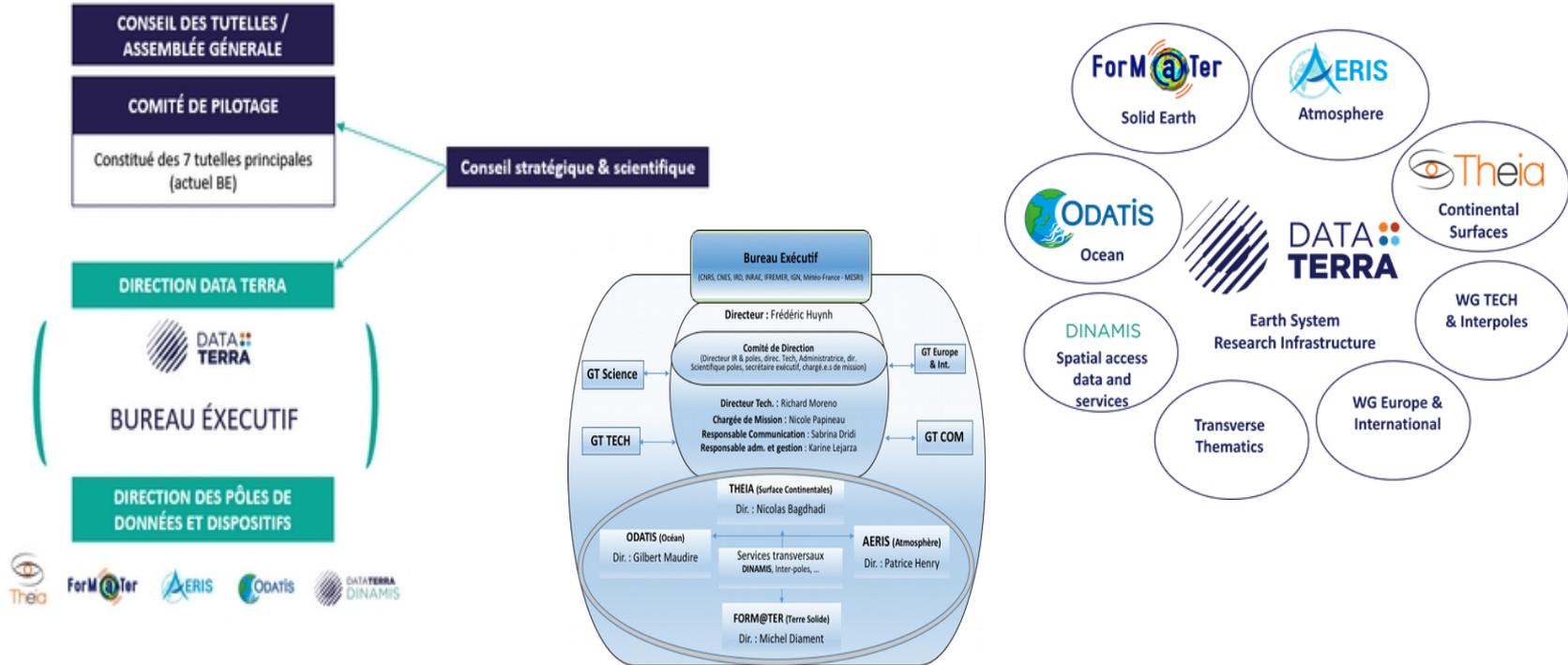
ÉVOLUTION DU NIVEAU DE LA MER

Sources : GT Science en lien avec les CS des pôles – priorités et prospectives CNRS/INSU, CNES (CPS, TOSCA, ..), IRD, INRAE, IGN, ...

# Structuration IR Data Terra et pôles : dates clés



# Organisation et gouvernance



# Data Terra en quelques chiffres

- 26 organismes et universités (34 d'ici 2025)
- 4 pôles de données : AERIS, ForM@Ter, ODATIS, THEIA (+ PNDB d'ici 2025)
- 6 services (DINAMIS) et groupes de travail transversaux
- 30 Centres de Données et de Services (CDS) et Infrastructures de données spatiales (IDS)
- 32 Consortium d'Expertise Scientifique
- 200 ETPT / 450 scientifiques, ingénieurs et techniciens appartenant à plus de
- 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)

# Caractériser et analyser l'atmosphère

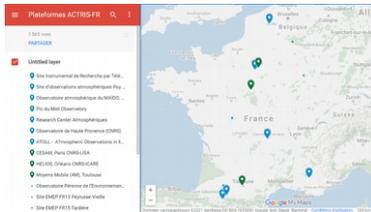
## dynamique, physique et chimie atmosphérique

### évolution du climat, qualité de l'air, ...

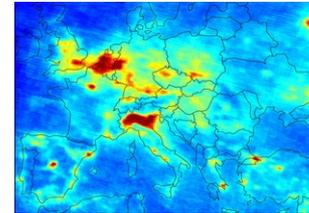
Le pôle de données et de services pour l'atmosphère AERIS fédère au niveau national des activités de gestion de données et d'expertise scientifique en atmosphère.

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

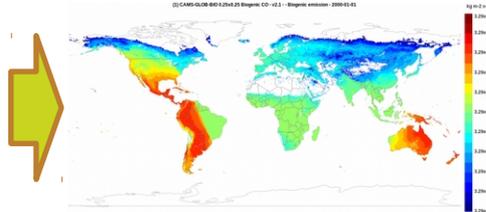


Mutualiser la surveillance de la qualité de l'air

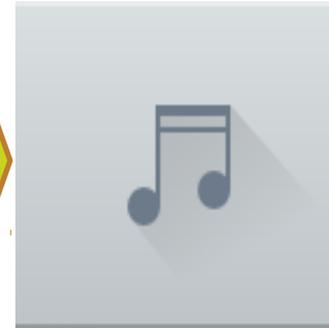


Créer un réseau européen de surveillance de la qualité de l'air

Estimer et quantifier les émissions de carbone et de particules



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



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 ?

**A la demande**

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

**GDM-OPT**  
GDM-OPT signifie Ground Deformation Monitoring using OPTical image time series

**DSM-OPT**  
DSM-OPT signifie Digital Surface Models from OPTical stereoscopic very-high

**GDM-SAR**  
GDM-SAR signifie Ground Deformation Monitoring using InSAR Image time series

Calcul de déformation du sol par imagerie spatiale

**Applications**

Activation de l'initiative **CIEST<sup>2</sup>**

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

**Systématique**

**FLATSIM**

Contenu Dans le cadre de ForM@Ter, le CNES développe, en collaboration avec des équipes des laboratoires ISTerre, IGL, ITRF et du IGN, un service dédié aux cartes de vitesse de déformation.

FLATSIM VELOCITY MAPS, EASTERN TIBET

Legend: 0.5 velocity (m/yr)  
0.5, 1, 1.5, 2

Thollard et al., in revision, 2021

**Applications**

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...

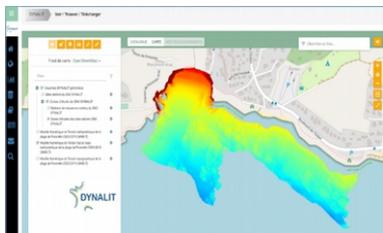
Suivi spatio-temporel de régions critiques

# Observer et comprendre les dynamiques des océans

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.

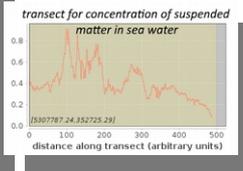
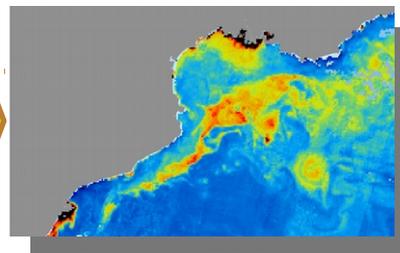
Les données proviennent de missions satellitaires, d'instruments in situ (plateformes fixes, flotteurs, gliders, radars, campagnes en mer, mesures en laboratoires, ...)

## Quelles thématiques pour les territoires ?



Suivi du trait de côte

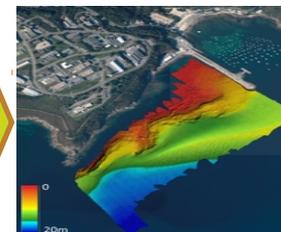
Qualité de l'eau



Oxygénation des eaux



Bathymétrie LASER et LIDAR

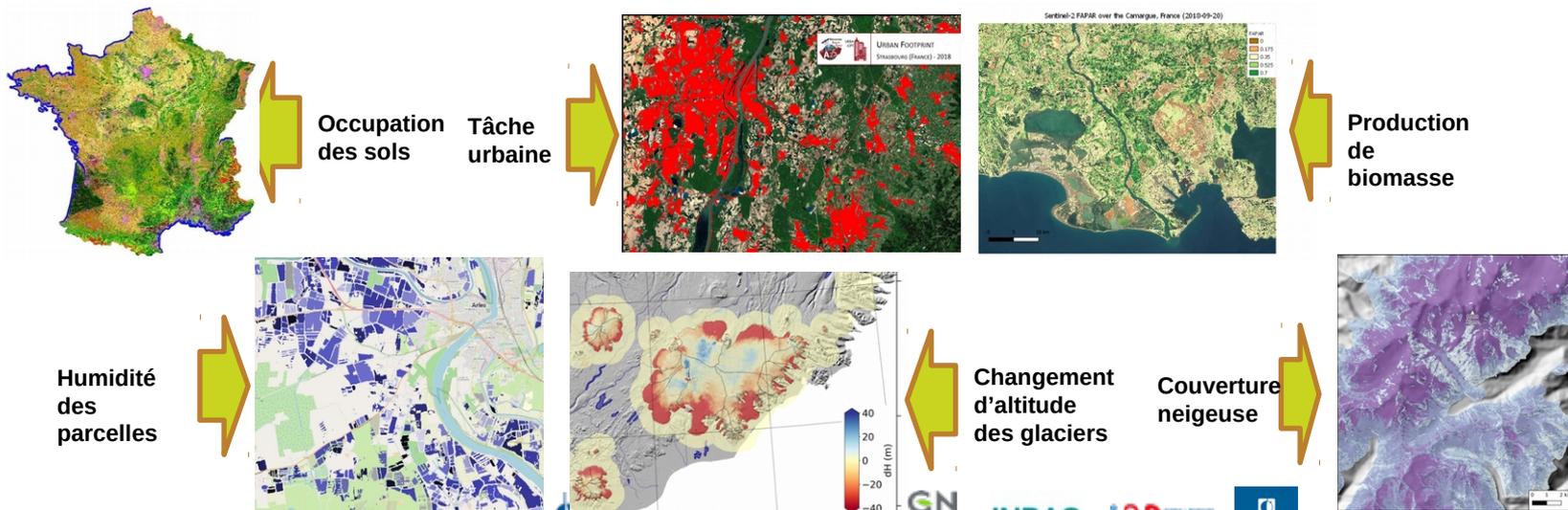


# Observer les dynamiques des surfaces continentales et ressources naturelles

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 ?



# 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 Satellitaire, 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...

# Participations à des projets nationaux, européens et internationaux

## ENVRI – FAIR : FAIRisation multi-domaines

Implication au travers des IR/TGIR ESFRI : ACTRIS, IAGOS, EPSO, EURO-ARGO, ANAEE ; 37 part., 19M€ - 16% Fr (CNRS, Ifremer, BRGM, INRA)

CEF – PHIDIAS (CEF – DG Connect, 3 M€) : service à la demande « données spatiales et environnementales » et HPC : CINES, CNRS, CSC, IFREMER, IRD, MARIS, Univ. Louvain, ...

EOSC-Pillar : gouvernance nationale et Eur. EOSC (INFRA-EOSC, 12 M€), coordination WP6 : 10 études de cas (FAIR Data and Services)

Seadatanet - SeadataCloud, projet d'appui à l'Infra (10 M€)

Blue Cloud : EOSC ocean (10 M€) Ifremer/Odatis, CNRS, Sorbonne Univ. Mercator-Ocean, ....

FAIR EASE (INFRA-EOSC 01-04) : 5 M€, 25 partenaires, coordination Data Terra

FAIR IMPACT (INFRA-EOSC 01-05)

Copernicus : services C3S, CAMS, Land, Marine, ...

EURO-HPC (CEA, GENCI, ...) : 2 cas d'usages (Hydrologie, Urbain)

CLUSSTER (ATOS, GENCI, CNRS, CS-Group, OVH,...) : partenaire associé

ANR Flash (open science) : 4 projets : FAIRisation, certification

GO-FAIR IN : GAIA Data In (global integrated Data Impl. Net.)

GEO : programme Board, Working Group, initiatives, ...

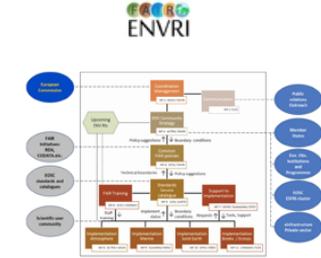


Figure 3: Work package structure and work flow of EOSC-EASE.



2022-2029

### 3.3 LE PROJET GAIA DATA

EQUIPEX+ /PIA3

Porté par 3 Infrastructures de Recherche

OBJECTIF : Développer et mettre en œuvre **une infrastructure/plate-forme intégrée de données FAIR** et de services distribuées 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) 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**



2022-2029

### 3.3 LE PROJET GAIA DATA

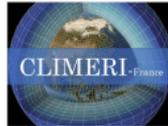
**Budget total : 62 M€ (coûts complets)**

**Demande ANR-EQUIPEX+PIA3 : 16,2 M€**

**Forte contribution RH : 339 ETP** (soit 4066 p.m.)  
personnels permanents + **59 ETP** (711 p.m.) cdd (1 pour  
5,6 permanents)

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

**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



# INFRASTRUCTURE GAIA DATA

## Grille de données et de services : 8 principaux centres en réseau

### Projets Equipex+ ou PIA4 infra

- FITS
- MesoNet
- Clusster

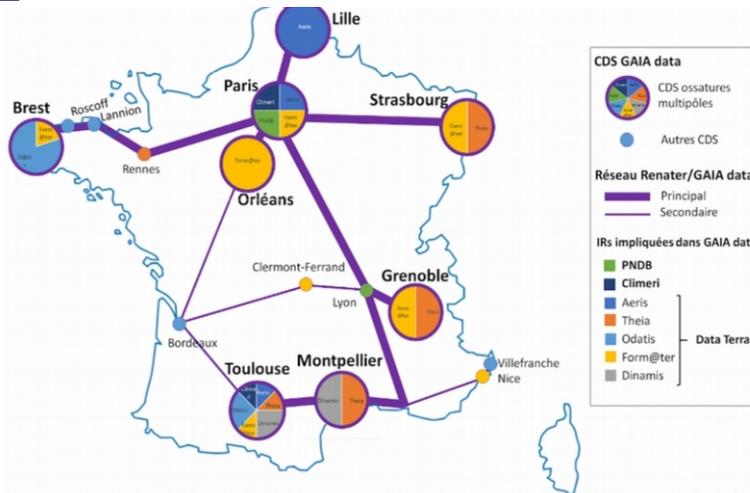
### Projets Equipex+ ou PEPR thématiques

- Obs4Clim
- TerraForma
- Marmor
- OneWater

### Projets H2020 – Horizon Europe

- IS-ENES
- PHIDIAS
- EOSC-Pillar
- FAIR EASE
- FAIR IMPACT

### Projets CPER en région



DESTINATION EARTH



- Mise en place d'un réseau dédié haut-débit et sécurisé
- Déploiement d'une grille de données (système iRODS AC) / S3 pour permettre un accès distant aux données et le transfert rapide et automatique de grands ensembles de données d'un centre vers un autre
- Interopérabilité des traitements entre les 8 centres de Gaia Data, avec les centres HPC en France et avec les clouds commerciaux (GAIA-X - DIAS)

## GAIA DATA Web Portal



Access, discovery, Knowledge & Services

Earth Analytics Lab

FAIR Workshop

Routine Processing

User helpdesk

Communities animation

## User Services

Thesaurus

Federated and harmonized catalogues & API

Harvesting Transformation

Software repository

Identity & Access management

Security

Hypervision metrics

Machine Actionable DMP

## Back Office Services



Individual catalogues, thesaurus and APIs

Local helpdesks



Existing services

Data Repository



Distributed processing

Data & Service Centers



Distributed datalake

Long-term archives

NETWORK



RENATER

GRID



HARDWARE



## Hardware

# SERVICES



## Services découverte, Accès et Gestion de données

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

**Catalogue** (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 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



## Longue vie au CDS

*poussière d'étoile*

