

Cartographie des solutions institutionnelles et des infrastructures

Damien Boulanger, SEDOO, AERIS

École thématique 2019 – e-Envir – Gif-sur-Yvette



DATA
TERRA

Plan

- Présentation de Data Terra
- Présentation des Pôles de Données et Services
- Dispositifs et activités transverses
- Solutions implémentées dans les Pôles



DATA 
TERRA

IR Data Terra

Contexte de création de l'IR Data Terra

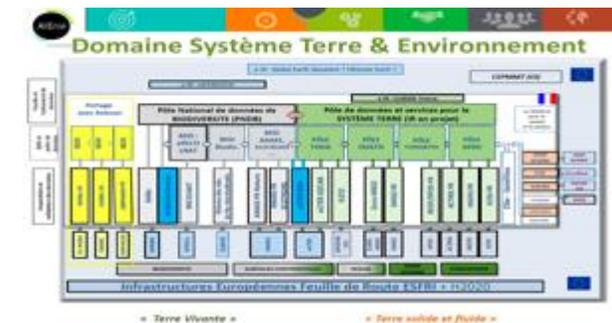
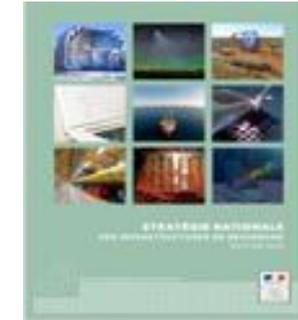
- **Groupe de réflexion** « Pôles thématiques en Observation de la Terre » 2012-2014 – CNES, CNRS/INSU

Nouveaux besoins et enjeux scientifiques ; fédérer et intégrer les pôles thématiques en pôles de données et de services au sein d'une infrastructure de recherche nationale

- **Feuilles de route nationale IR/TGIR – MESRI**

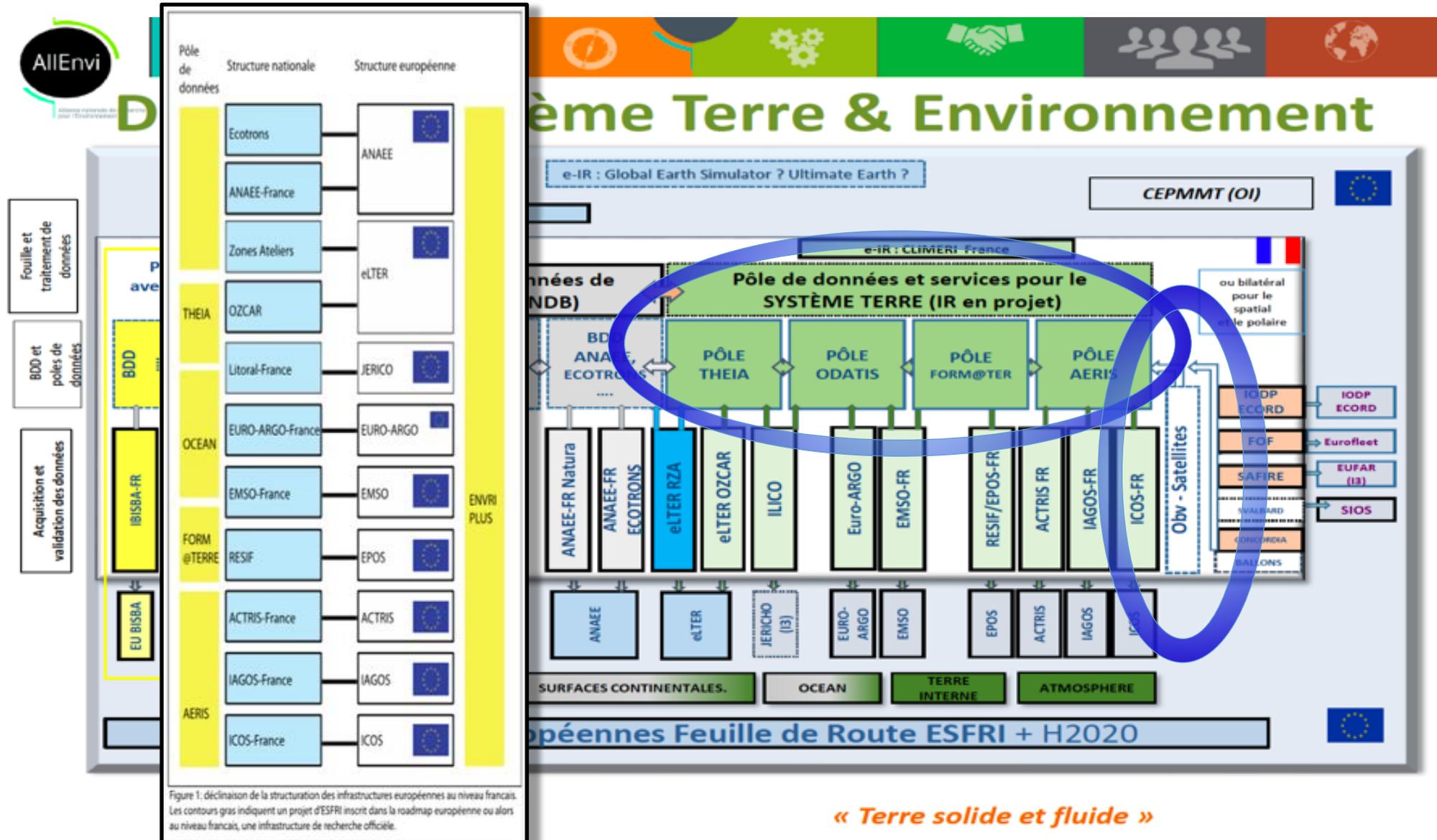
Contribution domaine « système Terre et Environnement »

- **Open science et open data**
- **Européanisation des infrastructures de recherche** (ESFRI, etc.) et des **services d'accès** aux données spatiales et in situ (EOSC, GO FAIR, COPERNICUS, etc.)
- **Internationalisation des dispositifs de partage des données et services** (GEO, etc.) autour des enjeux Changements Globaux et du Développement Durable



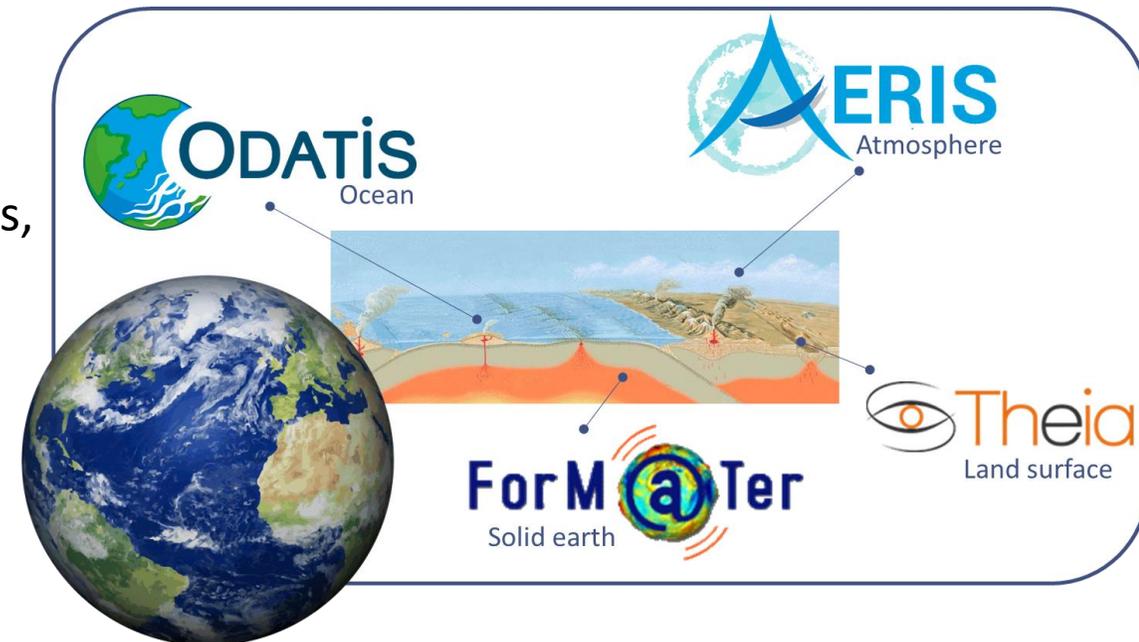
EUROPEAN OPEN SCIENCE CLOUD

Positionnement par rapport aux grands enjeux scientifiques, économiques et sociétaux liés aux infrastructures de recherche du domaine Système Terre et Environnement



Etat des lieux des Pôles

- 4 Pôles de données et de services (AERIS, FORMATER, ODATIS, THEIA)
 - **Très différents**
 - Pas le même niveau de maturité
- Lien fort avec le PNDB
- Hébergent et traitent
 - Des données des satellites français et européens, mais aussi d'autres pays
 - Des données in situ
 - Des données issues de modèles
- Chaque pôle est distribué géographiquement sur plusieurs Centres de Données et Services (CDS)
- Forte augmentation du volume des données
 - Spatiales : Copernicus, etc.
 - Mais probablement encore plus pour les données in situ



Contextes et enjeux autour des données

- Gérer et utiliser des données multi-sources, multi-domaines, multi-capteurs, multi-échelles spatiales, spectrales et temporelles
- « FAIRisation » des données et des services
- Open data et Open Science
- Intégration du potentiel de l'intelligence artificielle
- Développer des approches plus intégrées et systémiques orientées par les recherches

➔ Objectifs pour Data Terra et les Pôles

Associer l'expertise scientifique aux techniques de gestion des données

Partager les informations et les bonnes pratiques

Coordonner et fédérer au sein d'une même IR, l'ensemble des institutions, dispositifs et moyens existants

Servir la recherche nationale et internationale sur la planète, l'environnement, le climat, les risques, etc.

Quelques chiffres (2018)

- 4 pôles de données
- 20-30 Centres / Infrastructures de Données et de Services (CDS)
- 30 Centres / Consortium d'expertise scientifique (CES)
- 170 ETP / 350 scientifiques, ingénieurs et techniciens issus de plus de 30 organismes et universités
- 33 M€ (2016), 39 M€ (2017), 42 M€ (2018)
- 50 000 To (2017) - 100 000 To (2022)



DATA
TERRA

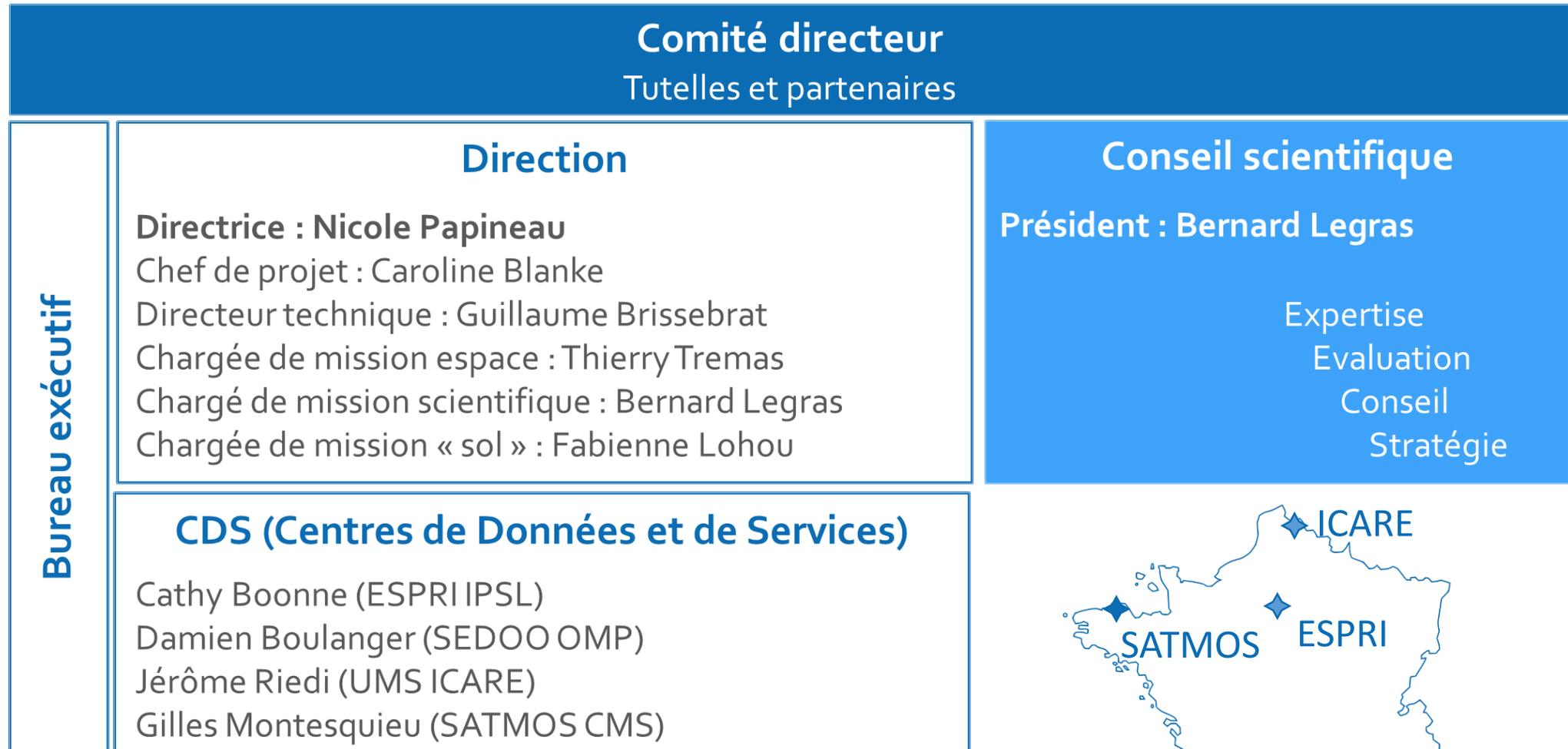
Pôles de Données et de Services

AERIS

- Pôle de Données et Services pour l'Atmosphère
- <https://www.aeris-data.fr>
- Objectifs
 - Fournir à la communauté scientifique un service et des outils performants pour collecter, traiter et distribuer les données atmosphériques en leur assurant une grande visibilité
 - Être un acteur reconnu sur la fourniture de données et de services au niveau français et européen

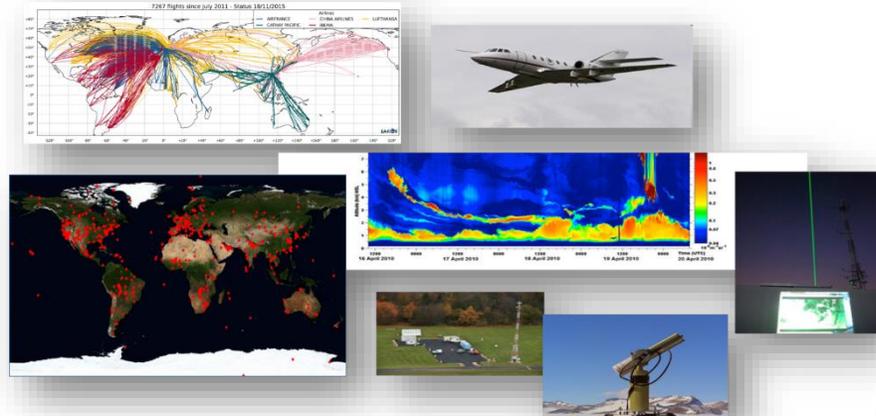


AERIS - Organisation



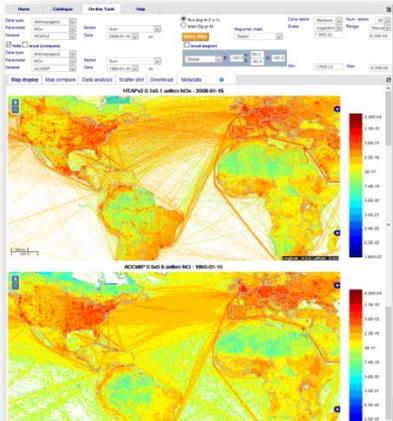
AERIS - Données et Services

Données sol et aéroportées



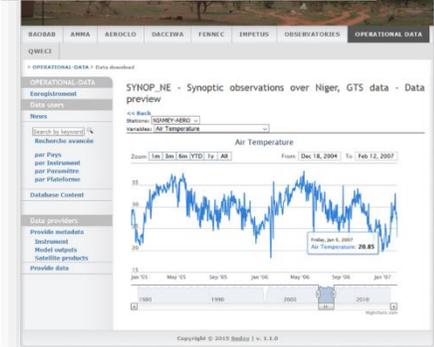
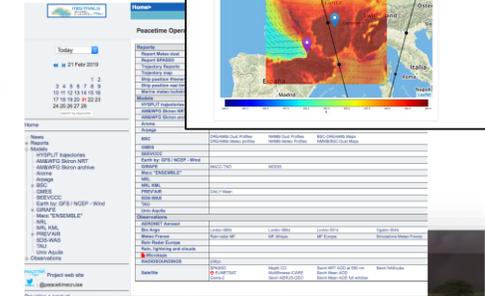
Bases de données

- ECCAD
- IUPAC
- GEISA

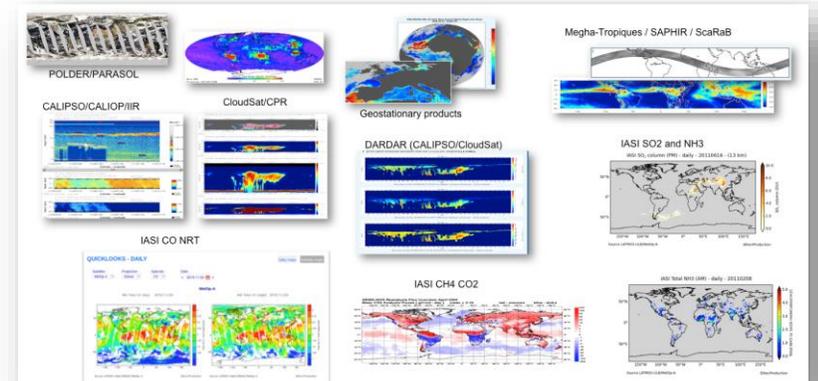


Outils

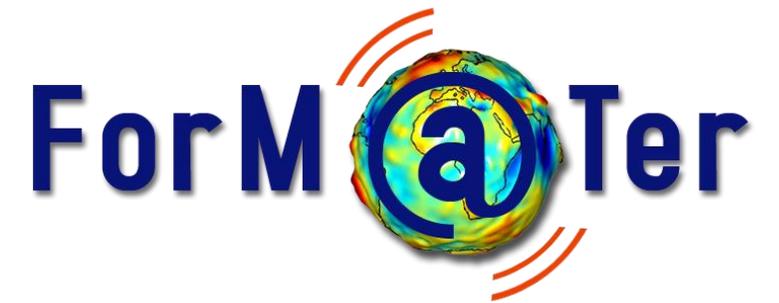
- Co-localisation sol satellite
- Outils de visualisation
- Outils de support aux campagnes



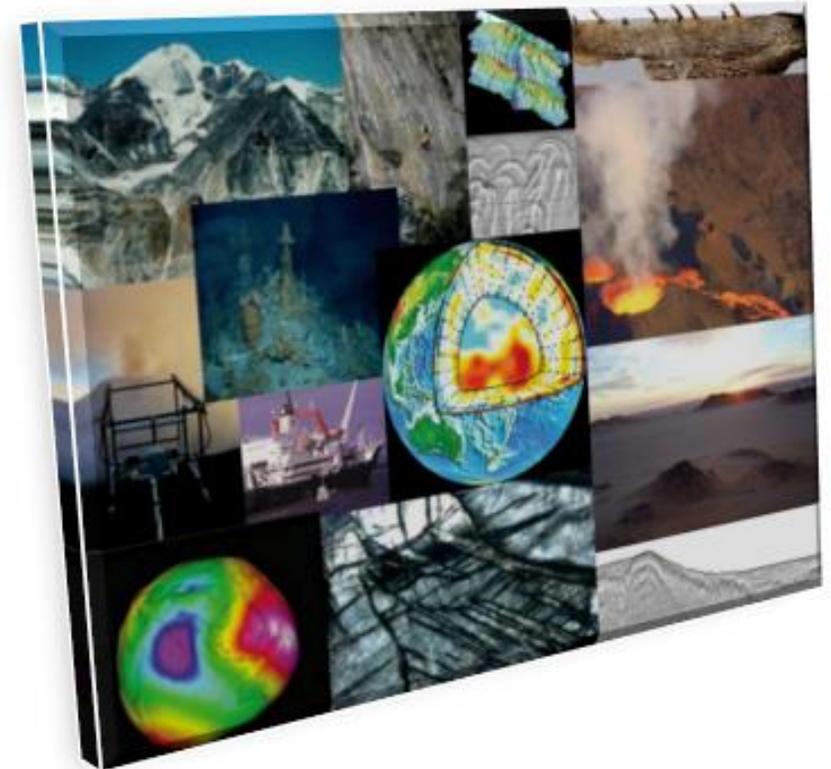
Données satellites



Formater



- Pôle de données et de services pour la Terre Solide
- <https://www.poleterresolide.fr>
- Objectif : faciliter l'accès aux données et contribuer à la création de nouveaux produits et services en apportant de la valeur ajoutée aux données spatiales et in situ



Formater – Objectifs

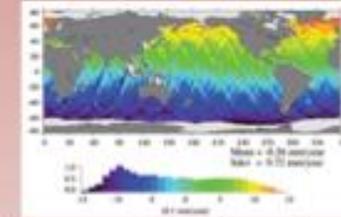
La Terre solide des enjeux scientifiques et sociétaux...

La formation de la Terre, sa structure



Les ressources

... et transversaux

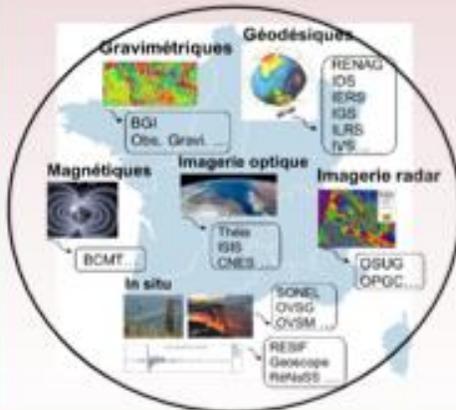


Exemple: le niveau des mers

Les risques telluriques

Les objectifs du pôle

Fédérer les centres de données et d'expertise existants



Positionnement dans le paysage des infrastructures en Europe



Des services pour la communauté et non spécialistes



Accès aux données



Outils et logiciels



Animation de la communauté

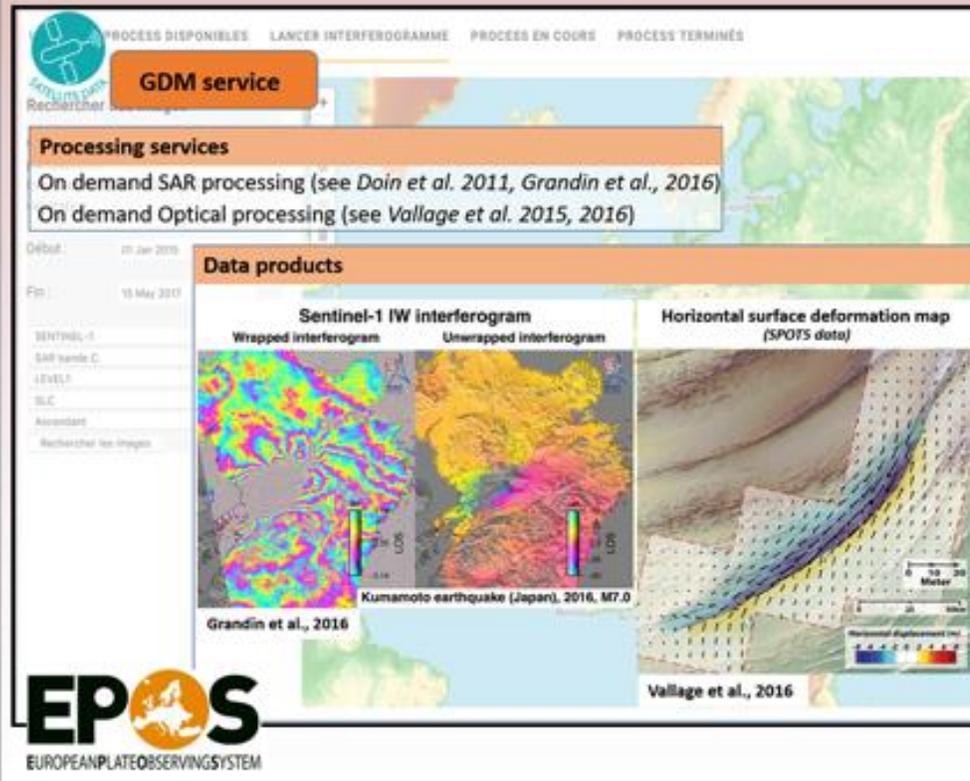


Support aux laboratoires

Formater – Services en construction

Ground Deformation Monitoring

Service de calcul sur demande
d'interférogrammes dans le cadre de
l'infrastructure de recherche Européenne
EPOS



Projet Etalab

Initiation d'une
plateforme pour la mise
à disposition de
produits à partir de
données et de services
de calcul répartis.



MUSCATE-SAR

Production systématique de séries
d'interférogrammes sur quelques
zones, à partir des données Sentinel-1
via la
plateforme de
calcul MUSCATE
développée par le
CNES



ODATIS



- Pôle de données et de services pour l'océan
- <https://www.odatis-ocean.fr>
- Objectifs
 - Fédérer au niveau national les activités de gestion de données et d'expertise scientifique en océanographie
 - Promouvoir et faciliter l'utilisation des données d'observations réalisées dans l'océan ou à son interface avec les autres milieux, à partir de mesures in situ et de télédétection



ODATIS - Organisation

Comité directeur (CD)

Représentants des 6 partenaires:

CNES, CNRS-INSU, Ifremer, IRD, SHOM, U. Marines

Bureau exécutif

Equipe de direction

Direction: G.Maudire

Direction technique: G.Dibarboure

Direction scientifique: S.Schmidt

2 chargés de mission : in-situ : J.Sudre, Satellite:
T.Guinle

Représentants des centres de données
et de services

Conseil scientifique

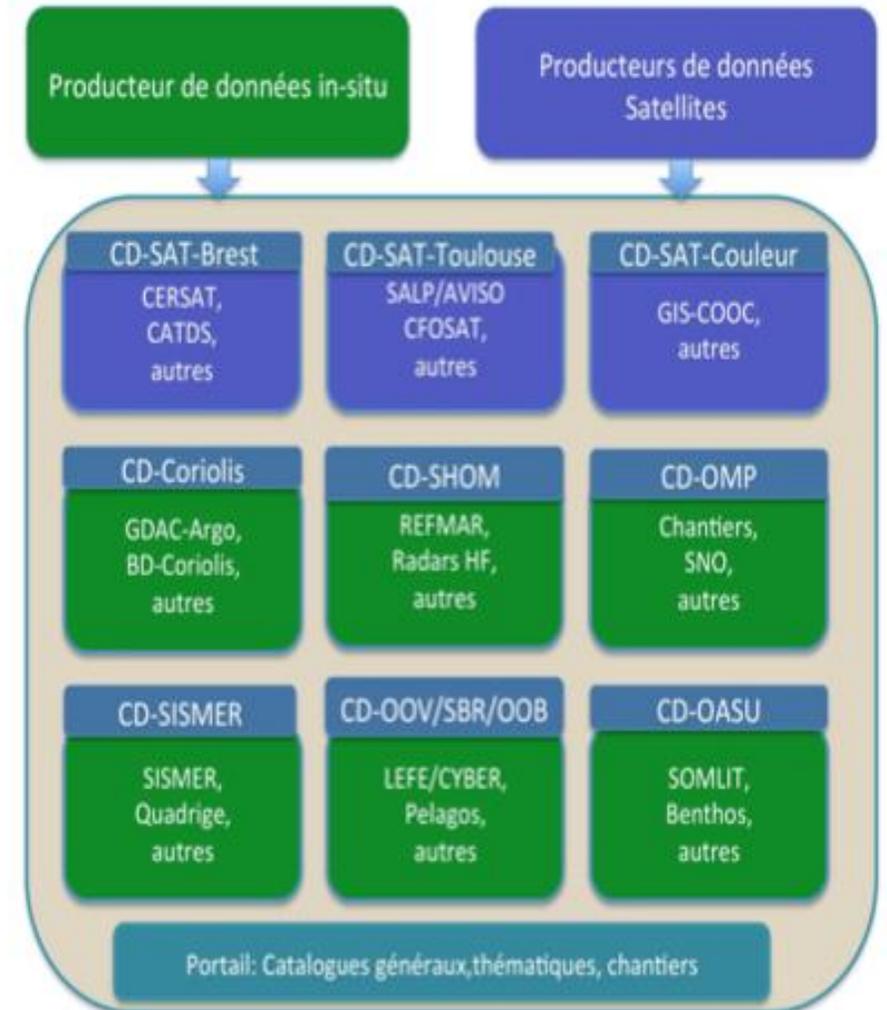
Président:

Sabine Schmidt
22 membres



ODATIS - CDS

- 12 Centres de Données
 - Données satellites
 - Données in situ
- ODATIS interfacé principalement dans le domaine des données in situ, avec plusieurs IR / TGIR
 - TGIR : Flotte Océanographique Française
 - IR EMSO : Observatoires fond de mer (ERIC)
 - TGIR ARGO : Flotteurs Argo (ERIC, Euro-Argo)
 - IR I-LICO : Observation côtière et littorale, lien avec Theia
 - IR O-HIS (en projet) : observation hauturière hors Argo et hors navires océanographiques

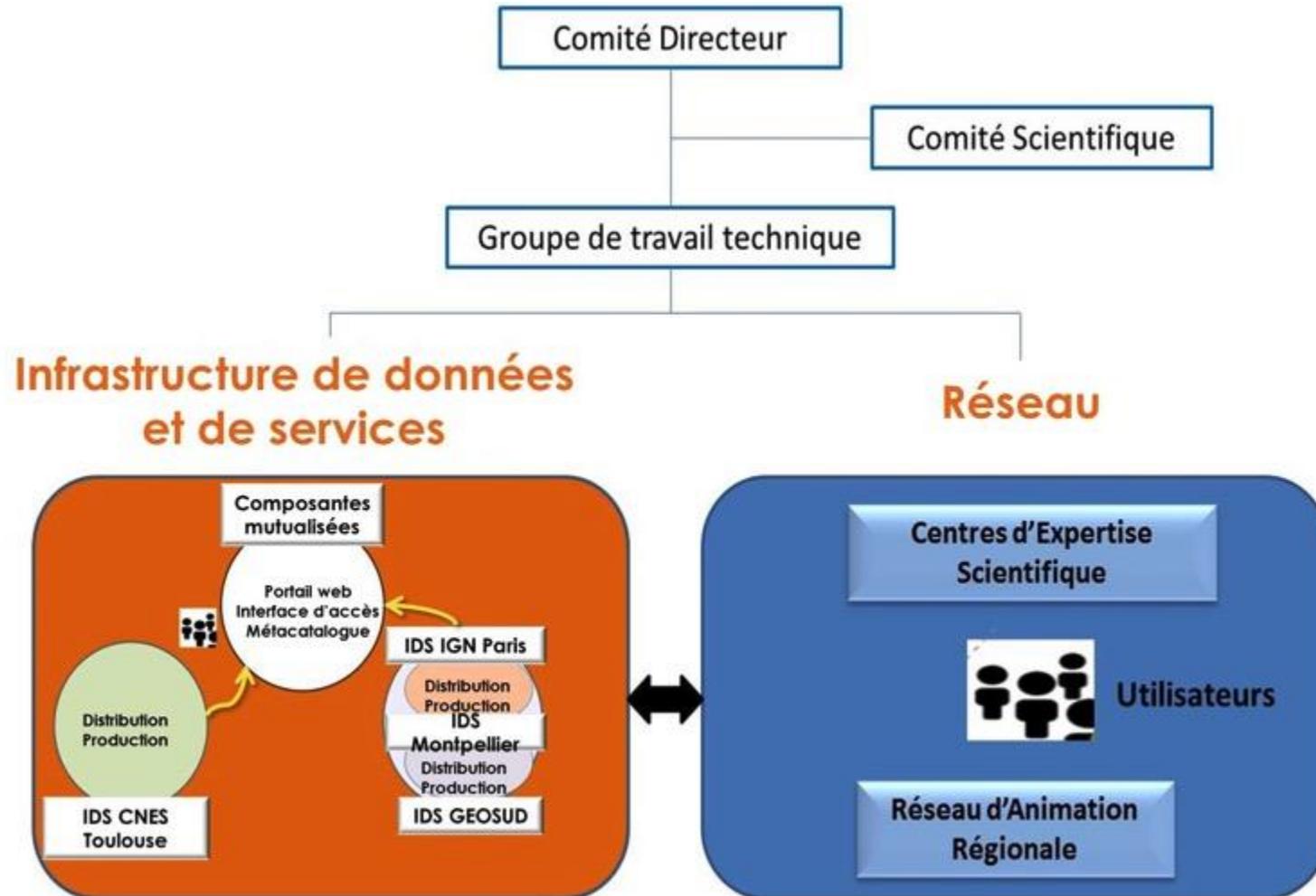


THEIA

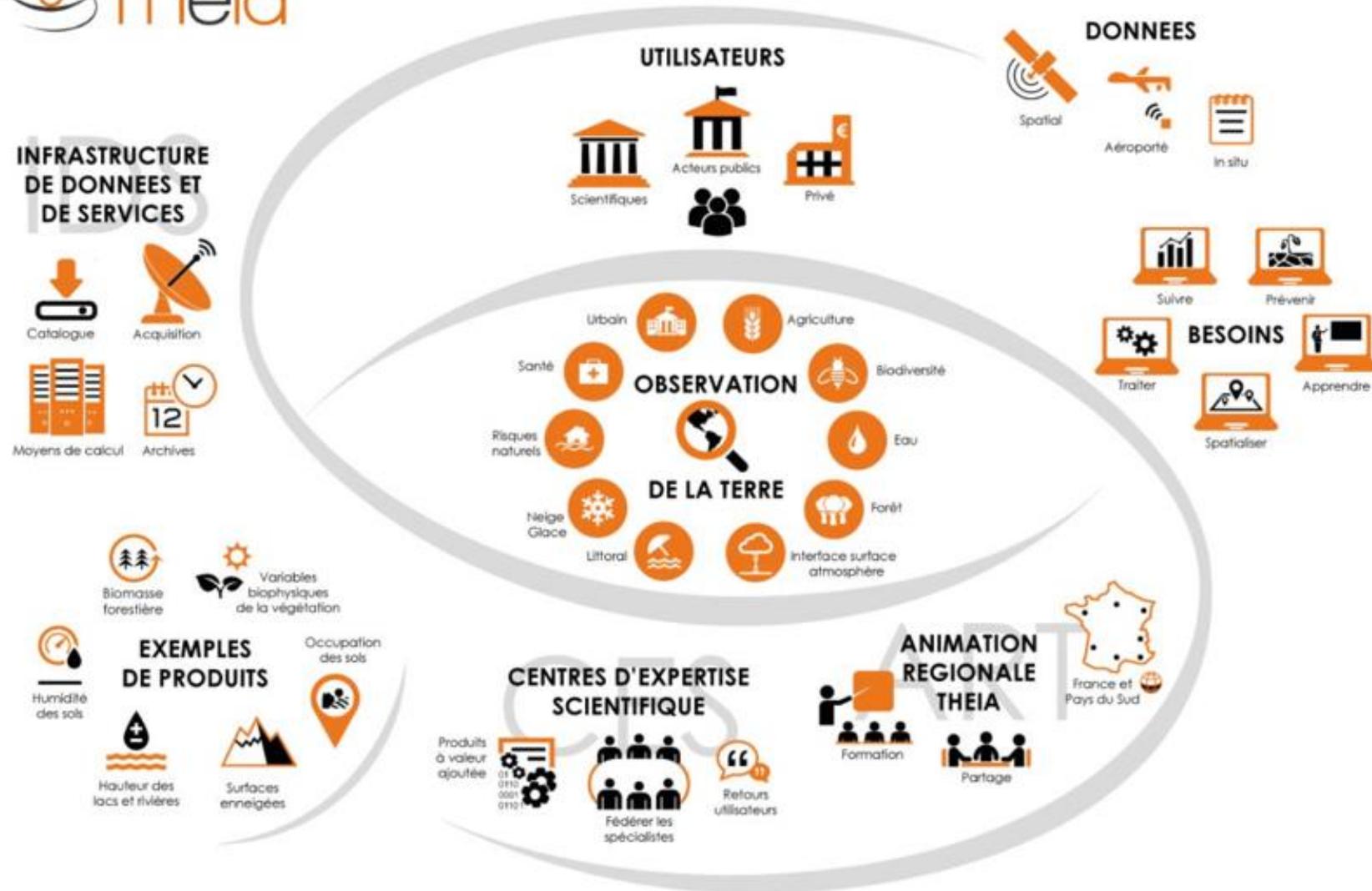
- Pôle de données et de services surfaces continentales
- <https://www.theia-land.fr>
- Objectifs
 - Faciliter l'usage des images issues de l'observation des surfaces continentales depuis l'espace
 - Construire un dispositif commun capable de produire des données spatiales (échelles locale à globale) à valeur ajoutée pour la communauté surfaces continentales et fournir des services en lien avec les besoins des utilisateurs
 - Favoriser le partage d'expérience et la capitalisation des méthodes
 - Rendre visible les réalisations nationales à l'échelle européenne et internationale



THEIA - Organisation



THEIA - Services





DATA
TERRA

Dispositifs et activités transverses

Dispositifs transversaux

- Ateliers et groupes de travail Inter-Pôles
- Groupe Technique Data Terra
- Dispositif transversal d'accès aux données spatiales :
DINAMIS
- Projets transversaux en cours d'identifications (MISTRALS, SAFIRE+, etc.)
- Projets européens

Inter-Pôles

<https://www7.obs-mip.fr/interpole/>

- Plénière tous les 6 mois
- Bilan des Pôles
- Point Europe (ENVRI, RDA, etc.)
- Mise en place de Groupes de Travail
- Terminés
 - Recommandations DOI
 - Recommandations Authentification
 - Recommandations Formats de données
- En cours
 - GT Catalogue → Data Terra
 - GT Licences
 - GT Infrastructure / Calcul → Data Terra

- Produire des recommandations pour la communauté Data Terra
- Espace d'échange

- Nouveaux GT
 - GT DMP
 - GT Certification

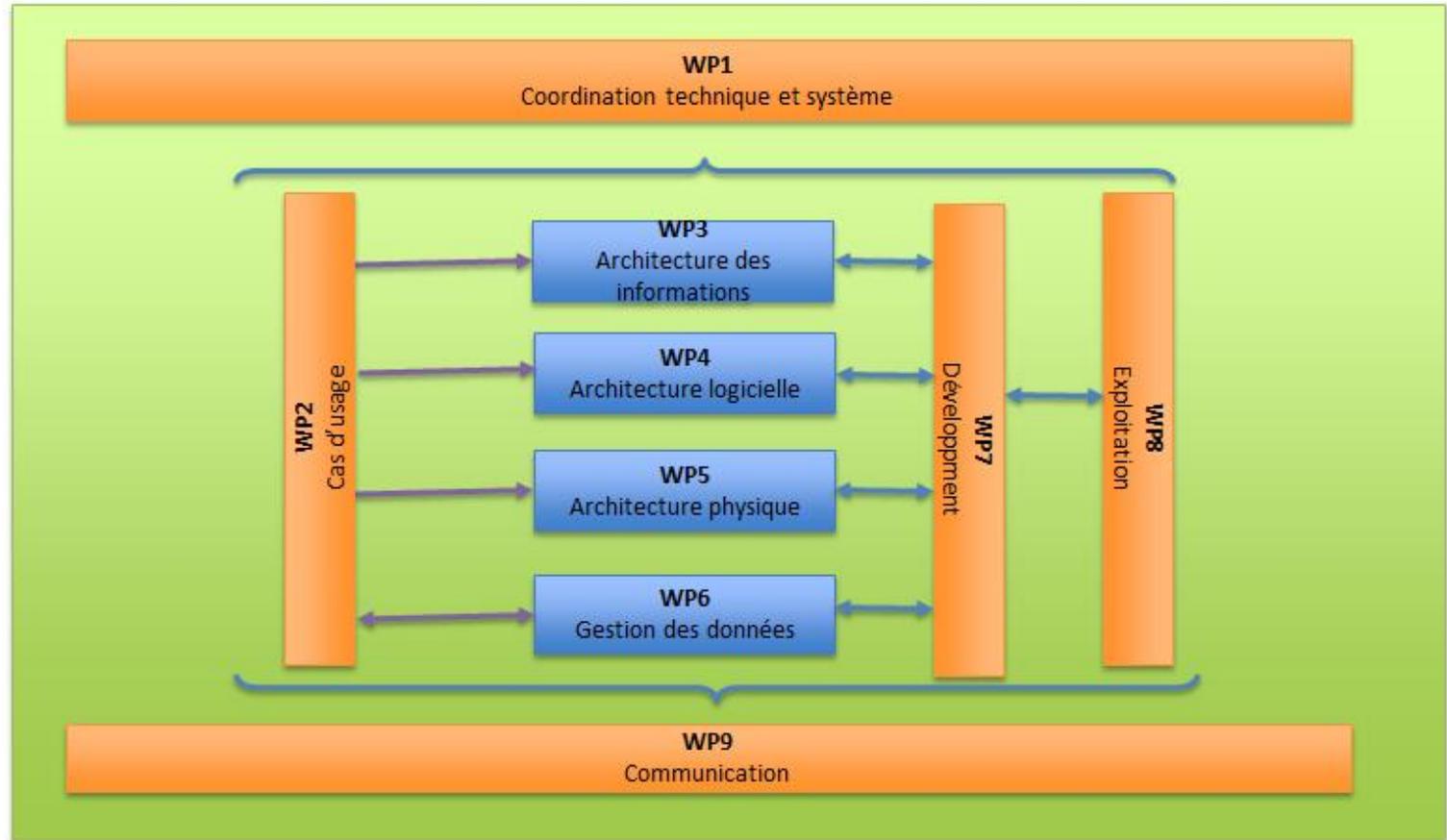
RDA

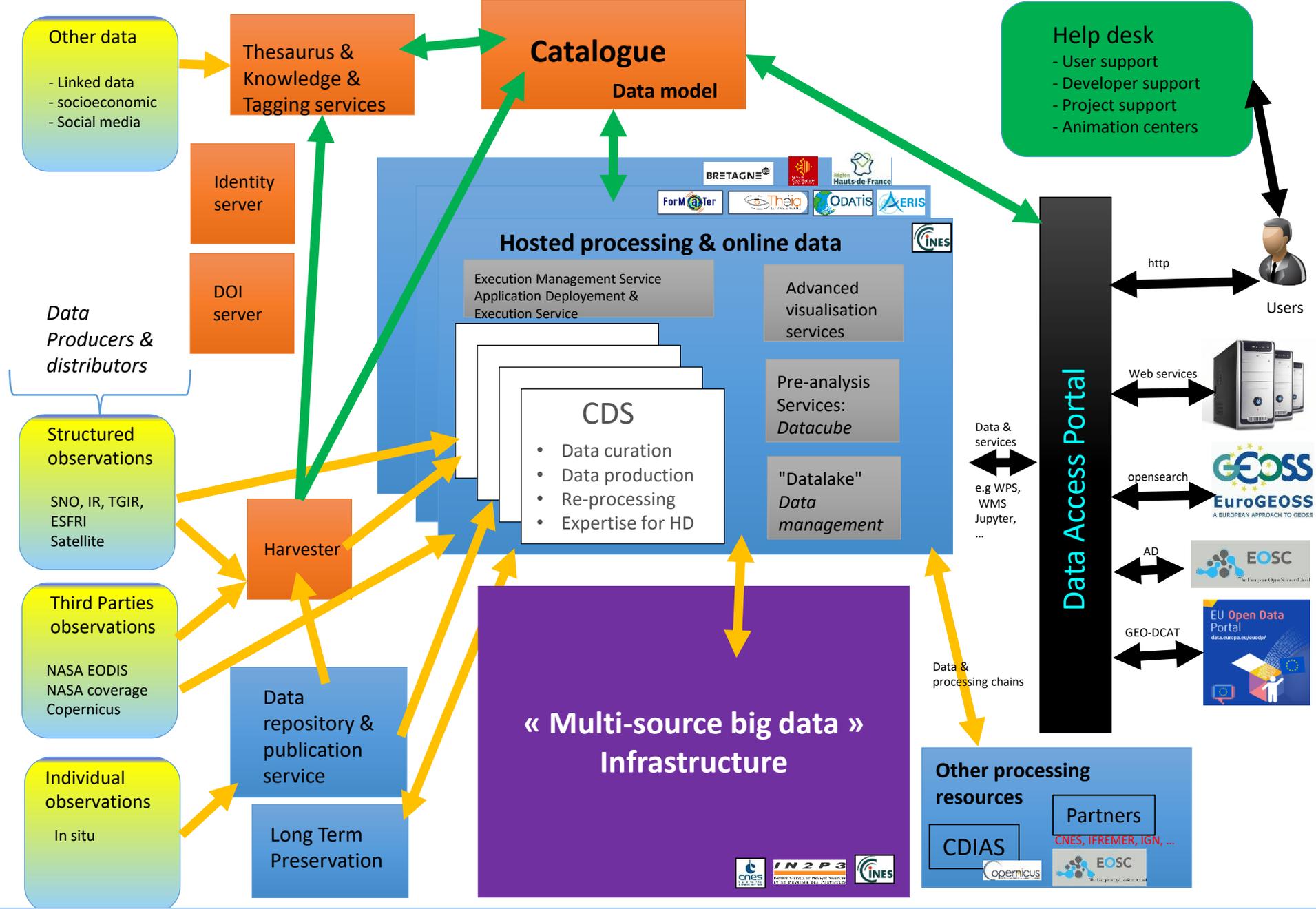


- Recensement de toutes les recommandations RDA utiles
- Participation aux activités RDA (Working et Interest Groups, plénières)
- Débriefing aux ateliers Inter-pôles
- Lien fort avec ENVRI FAIR
- Beaucoup de recommandations
 - PID/Métadonnées
 - Citation <https://www.rd-alliance.org/groups/rda-france>
 - Terminologies/Vocabulaires
 - Sémantique
 - Provenance/Versionnement des données
 - Certification (CoreTrustSeal)
 - Evaluation de la « FAIRness »

Groupe de travail technique Data Terra

- Les 5 Pôles
- Réunions mensuelles
- Point Europe
- Actuellement focus sur les infrastructures et le catalogue de métadonnées Data Terra





Other data

- Linked data
- socioeconomic
- Social media

Thesaurus & Knowledge & Tagging services

Catalogue
Data model

Help desk

- User support
- Developer support
- Project support
- Animation centers

Identity server

DOI server

Hosted processing & online data

Execution Management Service
Application Deployment & Execution Service

Advanced visualisation services

Pre-analysis Services: *Datacube*

"Datalake" *Data management*

CDS

- Data curation
- Data production
- Re-processing
- Expertise for HD

Data Producers & distributors

Structured observations

SNO, IR, TGIR, ESFRI, Satellite

Third Parties observations

NASA EODIS
NASA coverage
Copernicus

Individual observations

In situ

Harvester

Data repository & publication service

Long Term Preservation

« Multi-source big data » Infrastructure

Data & services
e.g WPS, WMS, Jupyter, ...

Data Access Portal

http

Users

Web services

opensearch

GeoSS
EuroGeoSS
A EUROPEAN APPROACH TO GEOSS

AD

EOSC
The European Open Science Cloud

GEO-DCAT

EU Open Data Portal
data.europa.eu/euodp/

Data & processing chains

Other processing resources

CDIAS

Partners
CNES, IFREMER, IGN, ...

opernicus

EOSC
The European Open Science Cloud

Transversal services

- SSO & AAI & DOI
- QoS surveillance
- Usage analytics

Projets européens

- ENVRI-FAIR : Démarrage janvier 2019 ; IAGOS, ACTRIS, EPSO, EURO-ARGO, ANAEE ; 37 partenaires, 19M€ - 16% Fr (CNRS, Ifremer, BRGM, INRA)
- PHIDIAS : Démarré, Coordination CINES, resp. 4 WP, budget cumulé 1 M€ (CNRS, IRD, IFREMER, sociétés partenaires Geomatys, Spascia) sur 2,6 M€

Développer et mettre en place des prototypes pour l'exploitation de données spatiales et environnementales de Sciences de la Terre en s'appuyant sur des capacités de calculs intensifs

- EOSC-Pillar : Accepté, appui à la gouvernance EOSC, coordination WP6, (CNRS, IFREMER, INRA, INSERM, INRIA, etc.)

Pratiques et les services de données FAIR pour les promouvoir dans l'ensemble des communautés

- Blue Cloud : EOSC océan : accepté, coordination Ifremer/ODATIS, CNRS, Sorbonne Université

Réaliser un ensemble de services permettant de mieux comprendre et gérer les nombreux aspects de la durabilité des océans, à travers une série de cinq démonstrateurs pilotes

The EOSC will allow for universal access to data and a new level playing field for EU researchers

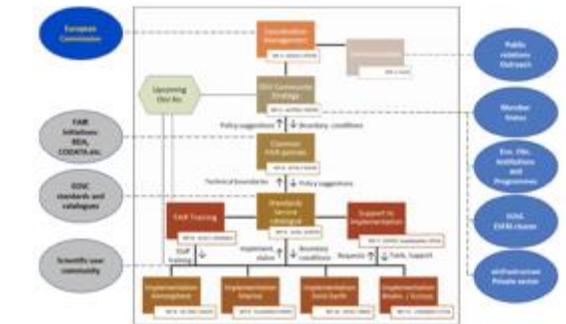
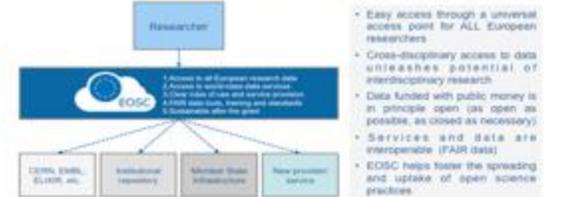


Figure 5. Work package structure and work flow of ENVRI-FAIR.

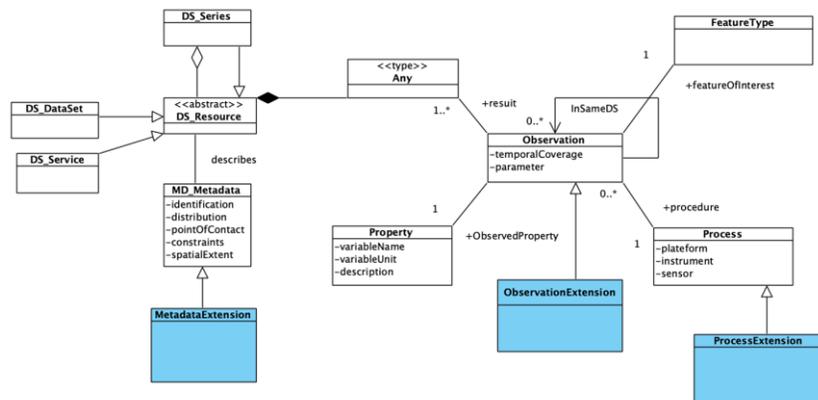
Part. N°	Participant organization name	Country
1	Consorzio GARR (GARR)	Italy
2	Consiglio Nazionale delle Ricerche (CNR)	Italy
3	CONIECA Consorzio Interuniversitario (CONIECA)	Italy
4	Istituto Nazionale di Fisica Nucleare (INFN)	Italy
5	Fondazione Centro Euro-Mediterraneo sul Cambiamento Climatico (CMCC)	Italy
6	Universität Wien (UNIVIE)	Austria
7	Centre Informatique National de l'Enseignement Supérieur (CINES)	France
8	Centre National de la Recherche Scientifique (CNRS)	France
9	Institut National de la Recherche Agronomique (INRA)	France
10	Institut National de la Recherche en Informatique et Automatique (INRIA)	France
11	Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)	France
12	Institut National de la Santé et de la Recherche Médicale (INSERM)	France
13	Karlsruher Institut für Technologie (KIT)	Germany
14	Deutsches Klimarechenzentrum GmbH (DKRZ)	Germany
15	Forschungszentrum Geophysik (FGZ - Forschung für Angewandte Forschung e.V./Fraunhofer)	Germany
16	Helmholtz Zentrum Potsdam, Deutsches GeoForschungszentrum (GFZ)	Germany
17	Geert University (Ghent)	Belgium
18	Triest IT Srl (TRUST-IT)	Italy

GT Catalogue de métadonnées

Objectifs

- Elaborer un socle de recommandations et spécifications afin de bâtir un catalogue des données et des services associés pour l'ensemble des données de Data Terra
- Sur la base des pratiques et catalogues élaborés par les pôles
- Eux-mêmes souvent élaborés par moissonnage des Centres de Données et Services (ou d'initiatives conjointes)

- Chaque pôle a développé un catalogue
 - Bloc commun de métadonnées
 - Services plus ou moins avancés (Recherche, téléchargement, visualisation, calcul)
- GT Interpole → GT Tech Data Terra
- Garantir l'interopérabilité
- Actions en cours
 - Modèle pivot pour les profils de métadonnées
 - Identifier un bloc commun de métadonnées
 - Granularité des données (granule, collection)
 - Observation & Measurement pour la découverte
 - Vocabulaires
 - Variables, plateformes, instruments (NetCDF-CF, ECV, EOVS, GCMD, etc.)
 - Publication
 - Services standards (OGC, OpenDAP, Thredds, etc.)





DATA
TERRA

Solutions implémentées dans les Pôles

Offre web Data Terra

Objectifs

Fonctionnels

- Mise en place rapide de sites dédiés à des projets spécifiques
- Fonctionnalités classiques : news, articles, pages, etc.
- Fonctionnalités spécifiques : gestion d'équipes, lettre d'information, etc.
- Contenu géré par les responsables scientifiques

Techniques

- Homogénéité ergonomique → Identification rapide
- Hébergement mutualisé → Limitation des coûts de maintenance et sécurité
- Services web pour accéder aux données
- Réutilisation des développements

Services

Usine à sites basée sur Wordpress

- Popularité
- Forte expérience au SEDOO

Fonctionnalités

- Gestionnaire de contenu (CMS)
- Gestion des rôles pour l'édition
- Multilinguisme

Thème commun AERIS

- Thème personnalisable
- Utilisé par les 3 autres Pôles

Hébergement centralisé

- Facilite la maintenance
- Fonctionnalités partagées

Statistiques : Piwik

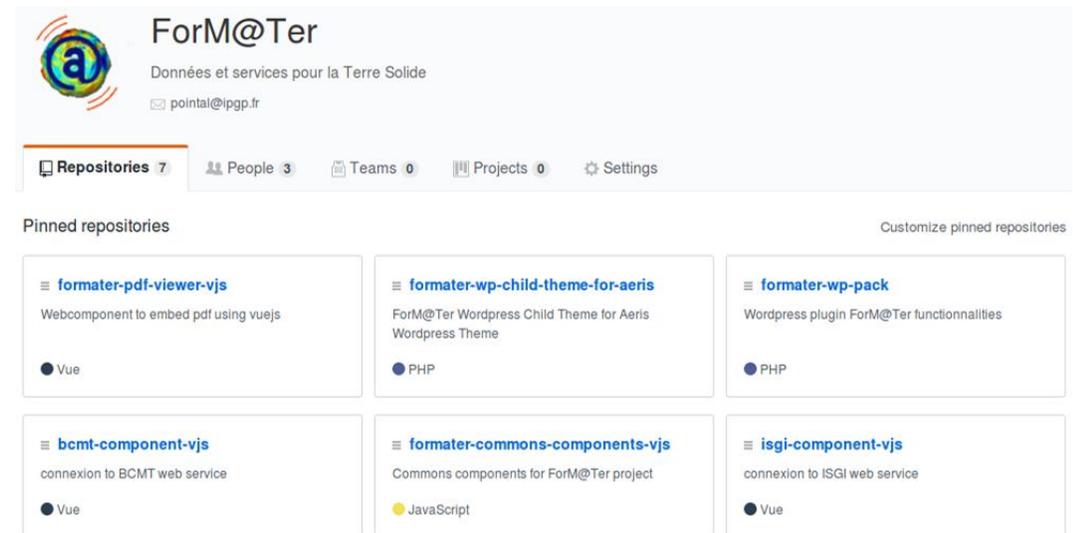
Offre web Data Terra

- 3 portails de Pôles hébergés
- 2 thèmes disponibles et maintenus
 - En constante évolution
 - Thème Data Terra en cours de développement
- Utilisée pour tous les nouveaux projets AERIS
- **Disponible pour tous les projets relevant de Data Terra**



Composants web

- Composants riches
- Réutilisables
- Développement mutualisé
- Technologie VueJS
- Services web RESTful
- <https://composants.aeris-data.fr/>



The screenshot shows the GitHub profile for 'ForM@Ter'. The profile header includes the repository count (7), people (3), teams (0), and projects (0). Below the header, there are six pinned repositories:

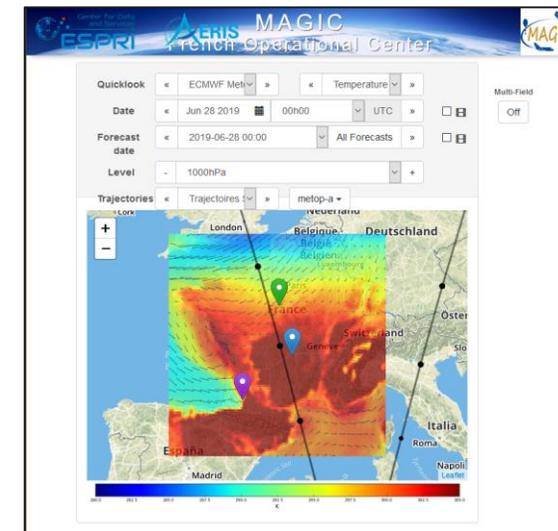
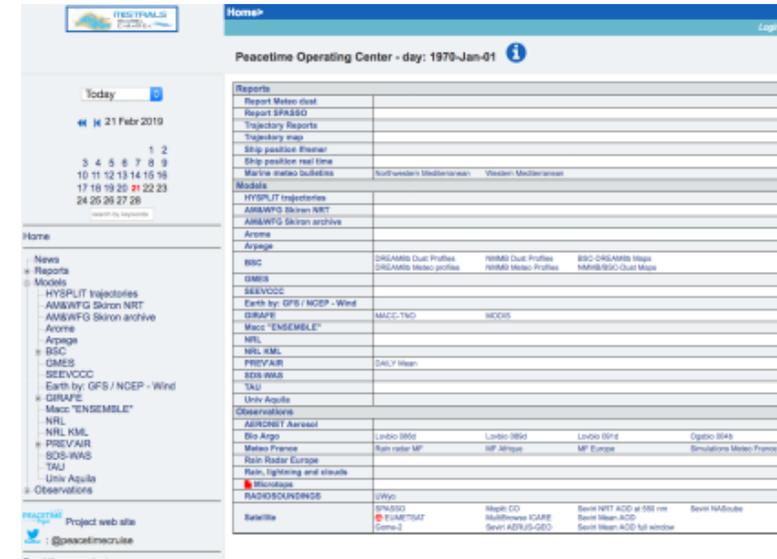
- formater-pdf-viewer-vjs**: Webcomponent to embed pdf using vuejs. Language: Vue.
- formater-wp-child-theme-for-aeris**: ForM@Ter Wordpress Child Theme for Aeris Wordpress Theme. Language: PHP.
- formater-wp-pack**: Wordpress plugin ForM@Ter fonctionnalités. Language: PHP.
- bcmt-component-vjs**: connexion to BCMT web service. Language: Vue.
- formater-commons-components-vjs**: Commons components for ForM@Ter project. Language: JavaScript.
- isgi-component-vjs**: connexion to ISGI web service. Language: Vue.



Offre outils de campagne AERIS

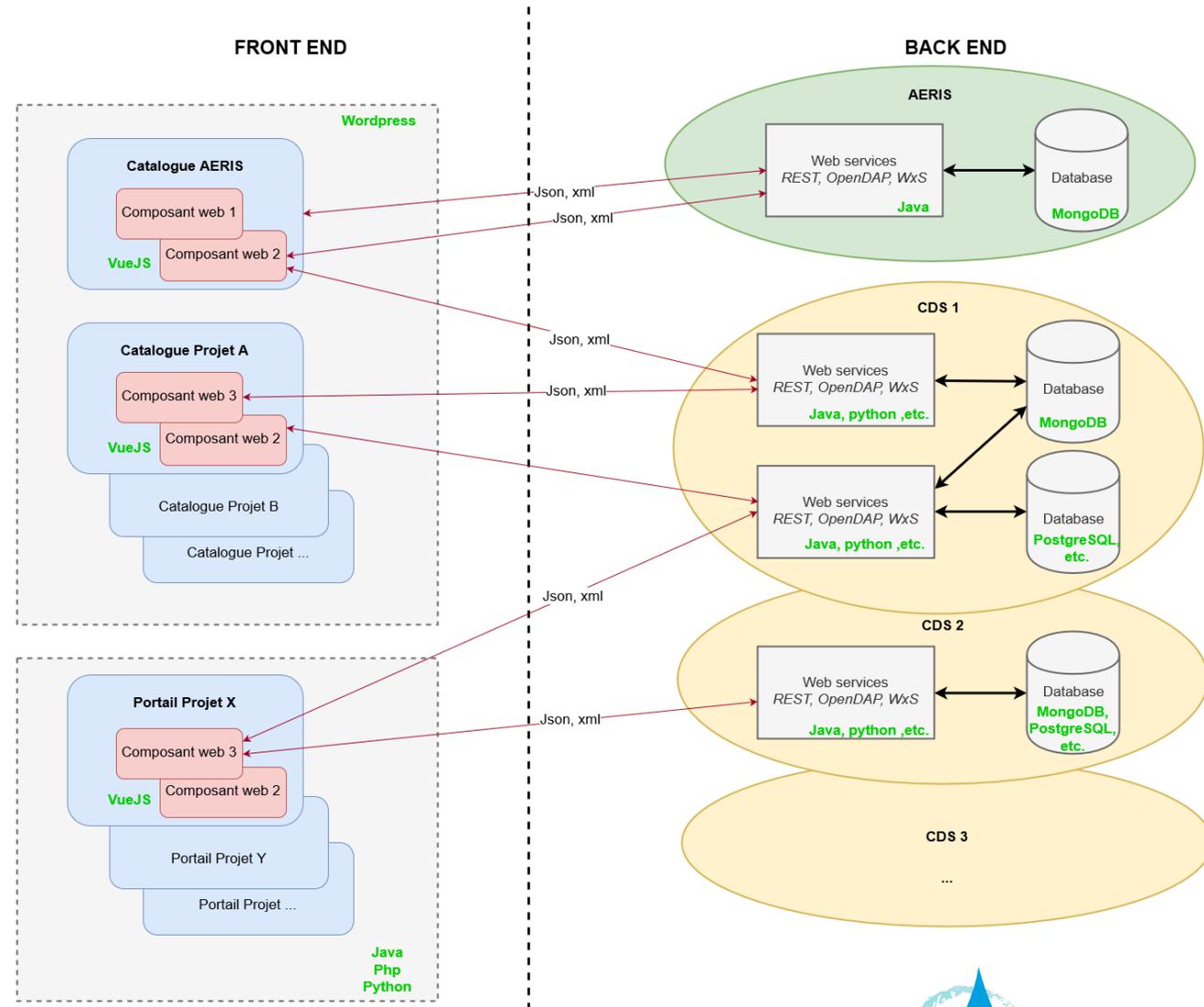
Site de support pendant les campagnes

- Servir les besoins opérationnels, temps réel des observations pendant la campagne : planification des vols, fourniture de données, visualisation, etc.
 - Fournir a posteriori une archive complète de la campagne
 - Archiver toutes sortes d'informations : rapports, plans de vol de avions, compte rendu de briefings, etc.
- ➔ Fournir un solution commune AERIS
- ➔ Etendre aux autres pôles?



Authentification et autorisation centralisée AERIS

- Mettre en place un système pour tous les utilisateurs AERIS
 - Système d'autorisation pour gérer de manière fine les droits des utilisateurs par projet
 - Centralisé et commun à tous les projets (portails de données et services web)
 - Authentification avec des systèmes existants (ORCID, Renater, etc.)
- Indispensable pour le téléchargement de beaucoup de données AERIS et pour les statistiques d'utilisation
- En accord avec la politique européenne (embargo, etc.)
- Extensible à Data Terra et aux autres Pôles



Certification

- Prévues dans tous les Pôles
- Pour tous les CDS
- Reconnaissance locale et Internationale
 - Profil renforcé et exposition mondiale
 - Visibilité accrue
 - Amélioration des perspectives de financement
- Engagement manifeste à l'égard de la science ouverte
 - Adhésion au principe Open Science & FAIR
 - Vers la qualité et la gestion à long terme
- Performances et souplesse accrues
 - Interactions et échange de données facilités (interopérabilité)
 - Découverte et citation améliorées (réputation et reconnaissance)
 - Pratiques et processus améliorés

Certification - CoreTrustSeal (CTS)

<https://www.coretrustseal.org/>

- Certification atteignable au niveau Data Terra
- Toutes les recommandations ne sont pas obligatoires pour avoir la certification



Certification



- Appel Flash “Science Ouverte : Pratiques de Recherche et Données Ouvertes”
 - ODATIS : COPILOtE – “CertificatiOn PoLe OcEan”
 - FORMATER : CEDRE – “towards Certification of solid Earth Data REpositories in France”
- Grande hétérogénéité structurelle et culturelle
- Projets de même philosophie visant à promouvoir les bonnes pratiques de Data Management d'une manière durable dans une démarche «FAIR» auprès des communautés scientifiques et techniques

Certification - COPiLOtE

- 9 Centres de Données et de Services (CDS)
- Harmonisation nécessaire pour faciliter la vie d'utilisateurs accédant à des jeux de données multiples, selon deux axes convergents :
 - « FAIRisation » des données
 - Certifications des CDS

« Auto-évaluation » initiale puis soumission d'un dossier à *CoreTrustSeal* :

- Critères de *CoreTrustSeal*
- Principes « FAIR »
- Décliné dans le cahier des charges des CDS

Production de guides de « bonnes pratiques »

(Ateliers préexistants : InterPôles et Odatis + experts)

- Métadonnées et vocabulaires
- DOI
- Formats supportés
- Services
- Outils
- Pérennité et sécurité

Implémentation progressive

- 3 Centres Pilotes
- 3 cours par an / ensemble des CDS
 - Coordinateurs
 - *Data managers*
 - Informaticiens
- Support en continu

Certification - CEDRE

- Consortium de 13 centres de données et de services
- Approche coopérative → transformer le processus individuel de préparation à la certification en un effort collectif,
- Répondre aux questions communes comme individuelles
- Travail conjoint et croisé de révision → transfert de compétences et de savoir-faire
- Chaque CDS engage des ressources pour mettre en œuvre les améliorations requises en vue de la certification CoreTrustSeal
 - Minimum de 2/3 des CDS impliqués finalisant leur dossier de certification,
 - 100% des CDS impliqués améliorant de manière significative leurs pratiques en gestion de données

Certification – CertifyMyRepo



- Outil permettant de suivre l'évaluation d'une certification Core Trust Seal d'un dépôt de données
 - Suivi des versions
 - Collaboratif
- Offre de service gratuite et ouverte à tous
- IAGOS et ECCAD

The screenshot shows the 'My repositories' page on the CertifyMyRepo website. The page has a dark blue header with the 'CERTIFYMYREPO' logo and a 'CREATE A NEW REPOSITORY' button. A sidebar on the left contains navigation links for 'My repositories', 'My Profile', and 'Login'. The main content area displays three repository cards, each with a unique color and a 'Requirements Radar Chart'.

IAGOS (Green Card):
Keywords: AERIS, In-service Aircraft for a Global Observing System, MOZAIC
Contact: info@iagos.org
Requirements Radar Chart: A radar chart with 17 axes (R0-R16) and a scale from 0 to 4. The chart shows a blue line representing the repository's performance across various requirements.

SSS (Red Card):
Keywords: LEGOS, Sea Surface Salinity, oceanography
Contact: sss-contact@legos.obs-mip.fr
Requirements Radar Chart: A radar chart with 17 axes (R0-R16) and a scale from 0 to 3. The chart shows a blue line representing the repository's performance across various requirements.

ECCAD (Orange Card):
Keywords: Emissions of atmospheric Compounds and Compilation of Ancillary Data, AERIS, GEIA
Contact: contact@eccad.fr
Requirements Radar Chart: A radar chart with 17 axes (R0-R16) and a scale from 0 to 4. The chart shows a blue line representing the repository's performance across various requirements.

At the bottom right of the page, there is a pagination control showing 'Items per page: 10' and '1-3 of 3'.

Citation dynamique

- Mise en place d'un Querystore : enregistrement des requêtes utilisateurs dans une base
 - PID sur chaque requête
 - Utilisation de fragment pour l'identifiant de la requête dans le cas d'AERIS
- ➔ Landing page spécifique à une requête (redirection du DOI + fragment)

