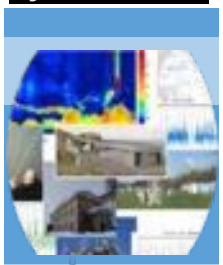
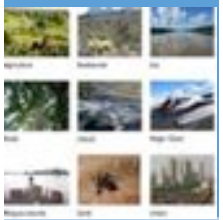
Infrastructure de Recherche Système Terre



AERIS

ForM@Ter

ODATIS



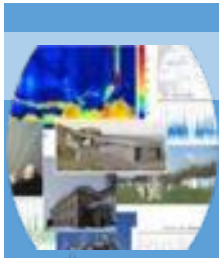
Theia

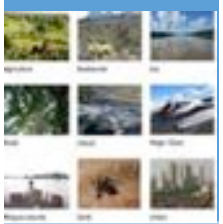
French Research Infrastructure Data and Services for the Earth System

Frédéric Huynh, directeur IR système Terre
Richard Moreno, directeur technique IR Système Terre

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

Infrastructure de Recherche Système Terre

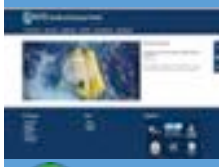


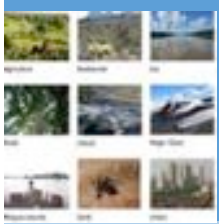


Presentation plan

- French EO Data & Services Hubs
- “Earth System” Research Infrastructure
- “Earth System” RI context
- Technical solution
- Technical strategy
- Conclusion

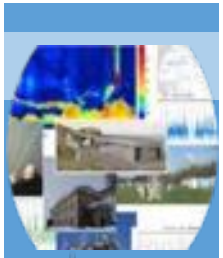
Infrastructure de Recherche Système Terre

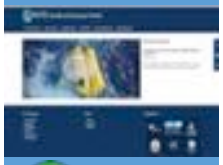


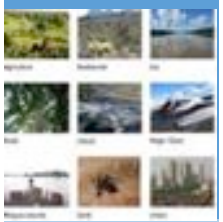


French EO Data & Services hubs

Infrastructure de Recherche Système Terre







Missions of the Data & Services Hubs

- Produce data series that are qualified and described according to accepted standards (Level 2 and +)
- To overcome spatial, temporal and disciplinary limits
- Promote the combined use of different data (satellites, in situ, campaigns)
- Develop and make available products combining different datasets
- Facilitate the exploitation of the information contained in the databases (visualization, interoperability, extraction, analysis exploration tool).
- Ensure long term preservation and facilitate the collection of heritage data
- Participate in the scientific, methodological and technical training of communities
- Contribute to the European &international promotion of French competences
- Provide support and expertise to users.

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

Infrastructure de Recherche Système Terre

Data & Services Hubs

Data from space missions funded by CNES are made available to the scientific community through data &services hubs specialized in one or more themes :

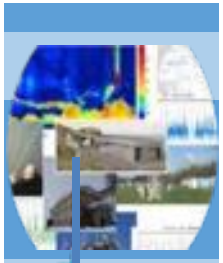
Earth Observation

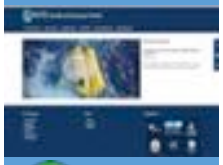
- **AERIS** gathers 4 atmospheric data centers (ESPRI, ICARE, SEDOO, SATMOS).
- **FORM@TER** created in 2012 is dedicated to Solid Earth domain.
- **ODATIS** is a portal devoted to oceanography from satellite data via AVISO + for altimetry but also from coastal or offshore data collected in-situ.
- **THEIA** created in 2012 is specialized in land surfaces.

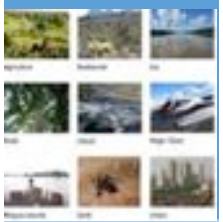
- **CDPP** specialized in the natural plasmas of the solar system was created in 1998.
- **CDS** is the oldest data center and has been collecting astronomical data since 1972.
- **MEDOC** created in 1995 is specialized in solar physics.

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

Infrastructure de Recherche Système Terre







AERIS

Data and services for the atmosphere
http://en.aeris-data.fr/

- AERIS « **Atmosphere and Service Data Pole** » has for objective to facilitate and enhance the use of atmospheric data
  - whether from satellites, ground, airplanes or balloons.
- For this, AERIS generates products from observations, but also provide many support services for the use of data, help to conduct synergies, campaigns or interface with models. AERIS aims to strengthen the existing systems consisting in four Data and Services centers that are defined as having the capabilities to manage data collectively. In addition, laboratories, laboratory networks or centers of expertise are essential actors for algorithmic development and prototyping. A significant advancement over existing data centers or bases is to develop an effective governance and value-added products for the community.

Infrastructure de Recherche Système Terre

AERIS

Data and services for the atmosphere
http://en.aeris-data.fr/

- Based on existing facilities:
  4 Services and Data centers

- Our strength

- More than 40 people working for AERIS
- A scientific community active, structured
- Observing systems
- Skills in Modelling and observing systems
- Involved in international network
- Databases for big campaigns

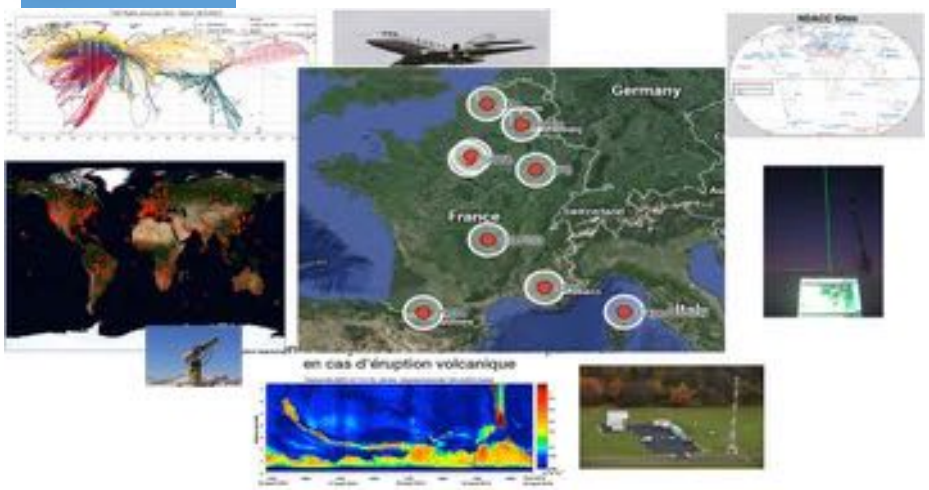
Infrastructure de Recherche Système Terre

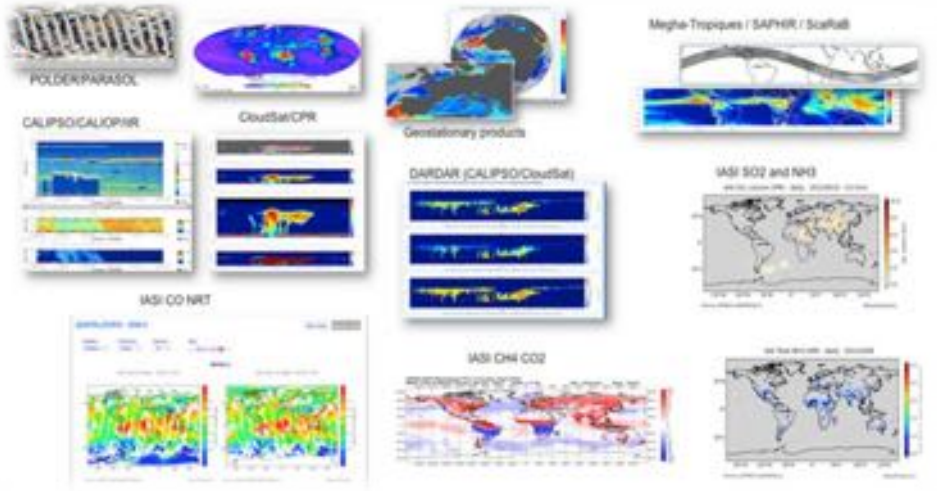
Data and services for the atmosphere
http://en.aeris-data.fr/

Data management, production, distribution

Remote sensing and in situ observations

Satellite Data Catalogue





Infrastructure de Recherche Système Terre

AERIS

Data and services for the atmosphere
http://en.aeris-data.fr/

Position in the european landscape

EOSCpilot

EOSC Open Science Cloud

ENVRI FAIR / INFRA EOSC

ENVRI plus

Complementary of ESA, EUMETSAT for satellite data

Copernicus
Europe's eyes on Earth

- CAMS, C3S and DIAS

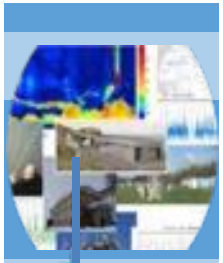
ACTRiS

- Strategic projects
  - ACTRIS, EUROCHAMP
  - HEMERA
  - EUFAR

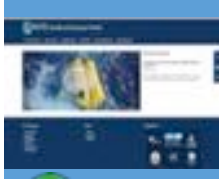
International databases

- Campaign MISTRALS
- Network NDACC
- Emissions ECCAD
- Chemistry IUGC
- Spectroscopic databases GEISA IASI

Infrastructure de Recherche Système Terre








Data and services for the solid Earth
https://en.poleterresolide.fr/

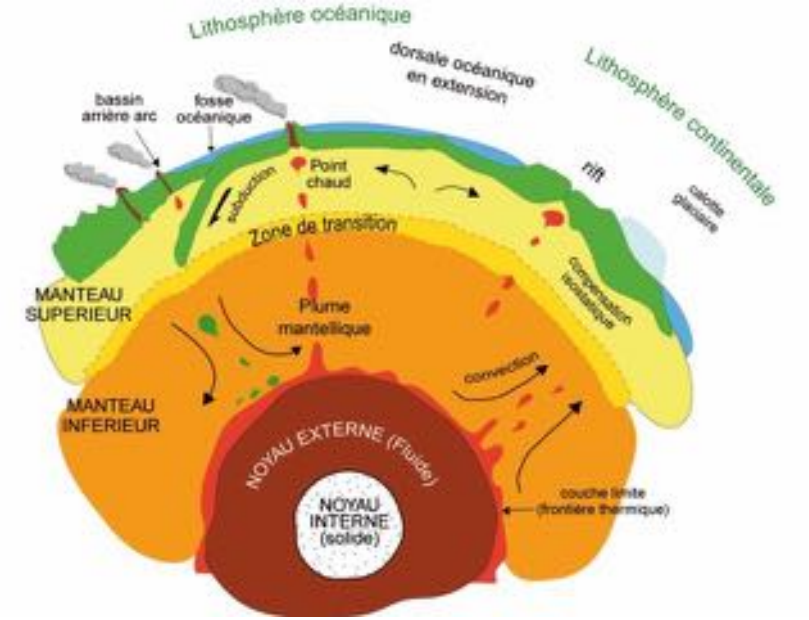
The purpose of the Solid Earth Centre is to facilitate access to data and contribute to the creation of new products and services by adding value to the available spatial and in-situ data. It is part of the national and European landscapes in close coordination with the infrastructure in place and under construction.

For this, Form@ter has the task to federate the existing centres in the service of the Solid Earth community.

Solid Earth…
Some specificities

An inaccessible environment directly

Very diverse data, digital but also physical (samples, drill cores, etc.)



Space, essential but not unique data

Variety of scales for Solid Earth

From the dimension of the atom to that of the planets.

From the fraction of a second to billions of years.

Data & Services for the Solid Earth

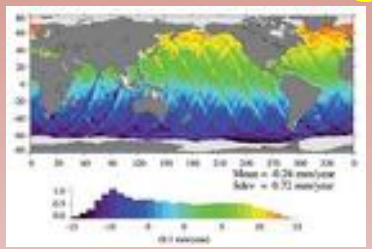
Solid Earth: scientific & societal issues

The formation of the Earth, its structure

The telluric risks

Resources

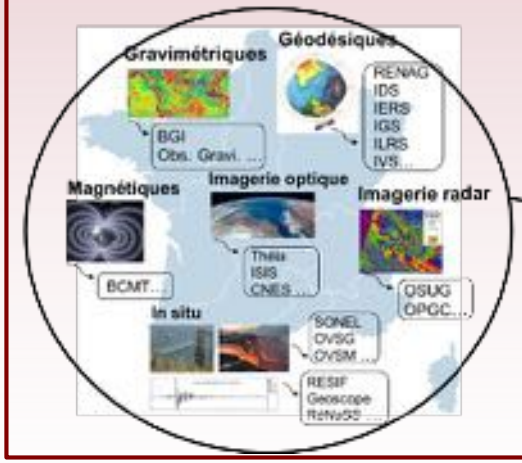
& crosscutting



Example: sea level

Objectives of Form@ter

Federate existing data centers and expertise



Positioning in the infrastructure landscape in Europe

Community and non-specialist services

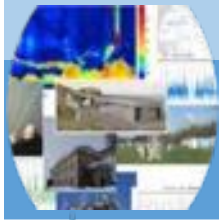
Access to data

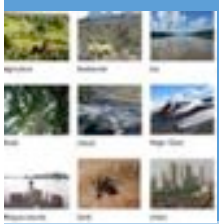
Tools & software

Community animation

Support to laboratories

Infrastructure de Recherche Système Terre





Theia Pôle Thématique Surfaces Continentales

Land Surface Data and services Center
http://www.theia-land.fr/en

- French Organisation born in 2012

➢ Objectives

➢ Promote and ease the use of space data, for science and public actors
  - imagery, altimetry,.. all wavelength
  - In-situ data
  - added value products,

➢ Develop added value products and services for the science communities and national public actors

➢ Develop networks of competences /expertise

➢ Support French achievements at European and international level

Infrastructure de Recherche Système Terre

Land Surface Data and services Center
http://www.theia-land.fr/en

Human networks

Science Expertise Consortium (CES)

- To Federate works of several labaratories
- Groups of scientists targeted to develop innovating algorithms for value-added products
- Examples of outcomes : Land cover classification; soil moisture, vegetal biophysical variables ; Imperviousness ; …

Regional Animation networks (ART)

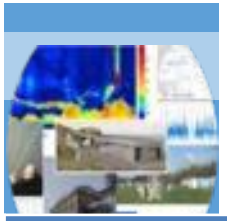
- To provide facilitation at regional level involving scientists, public and private users in order to :
  - Exchange information
  - Collect needs
  - Organize training

Networks spread at domestic and overseas level

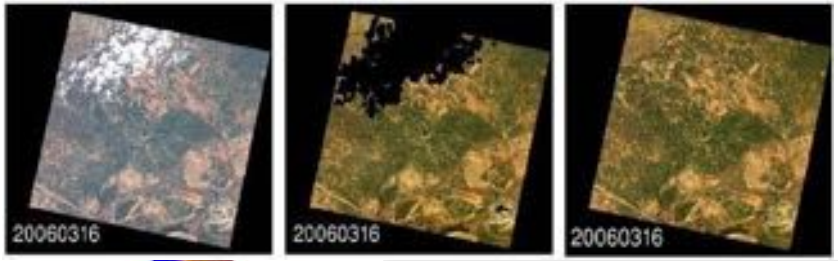
Infrastructure de Recherche Système Terre

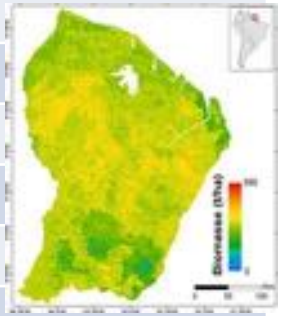
Product portfolio

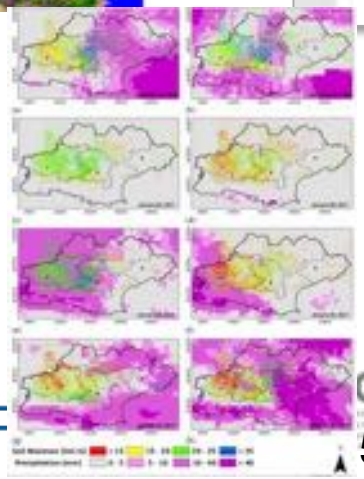
Added Value Products from Science Expertise Centres (~ 25 CES)

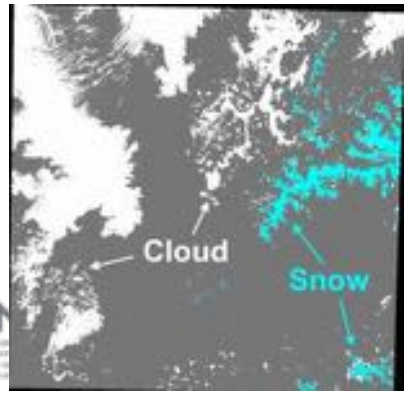


Added Value Products	Status
Surface reflectance	In production
Land Cover & Land Use	In production
Snow Coverage	In production
Soil moisture	In production
Lakes & Rivers height	In production
Decametric vegetal variables	Prototype being developed
Continental water quality	Prototype being developed
Epidemiology	
Irrigated surfaces	
Evapotranspiration	
Soil sealing-urban sprawl	
Forest biomass	
Soil mapping	
Albedo	
High frequency changes	
…	

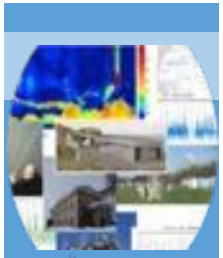


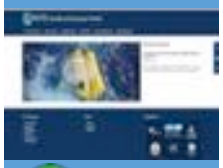


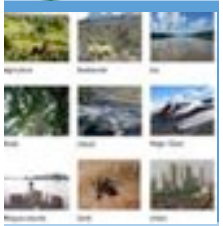




Infrastructure de Recherche Système Terre







Product portfolio

Added Value Products from Science Expertise Centres (~ 25 CES)

Very High resolution images free of charge for non commercial use

Product	Area	Access
Pléiades	Worldwide	Low fees for French public actors and their international partners
Spot 6/7	Worldwide	« Free » for French public actors and their international partners + French private Cies for R&D

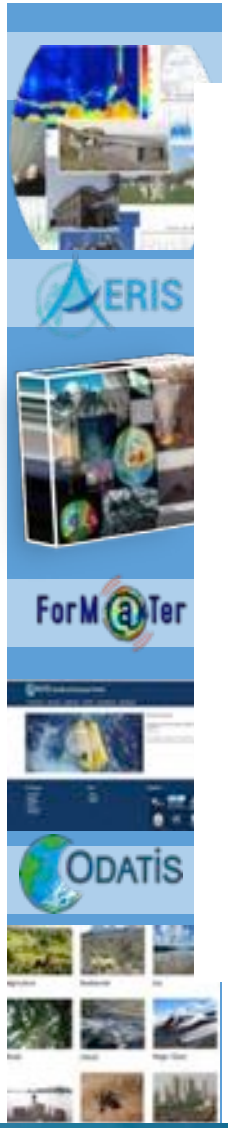
High resolution images & products

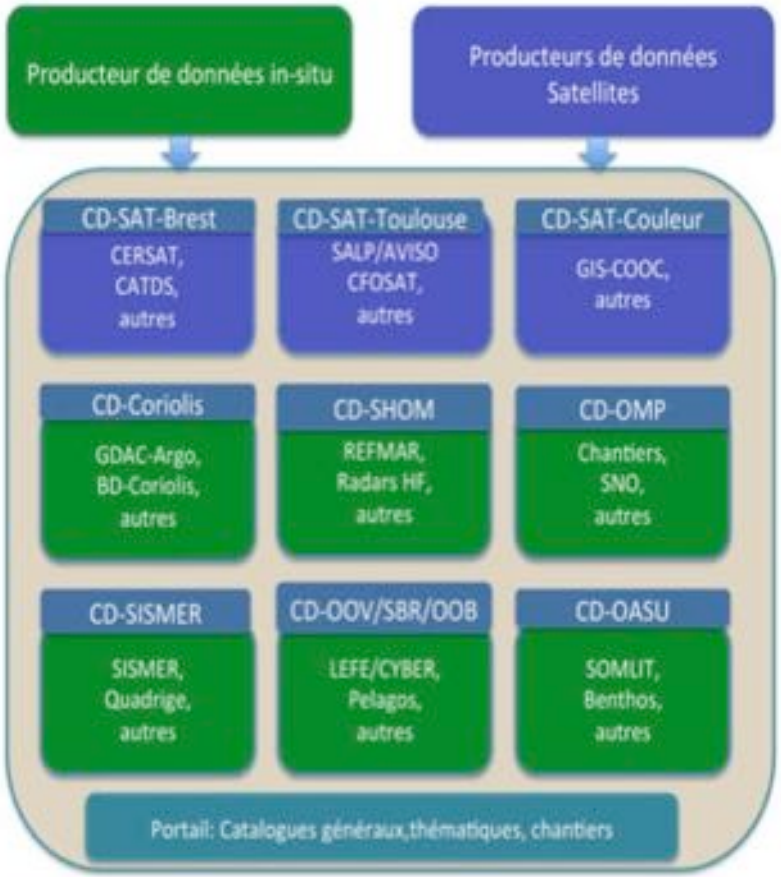
Product	Area	Period	Access	Availability
Sentinel-2 reflectances	Western Europe and other regions worldwide	2015 -	Free and open	Available
Landsat reflectances	Time series France	2005 – 2011 2013 -	Free and open	Available
Spot 4 (Take 5)	45 sites worldwide	Feb – June 2013	Free and open	Available
Spot 5 (Take 5)	120 sites worldwide	April – August 2015	Free and open	Available
Spot World Heritage (SWH)	400 000 images worldwide	1986 - 2011	Free and open (non commercial usage)	Available

Infrastructure de Recherche Système Terre

ODATIS

Data and services for the Ocean
http://www.odatis-ocean.fr/en/





http://www.odatis-ocean.fr/en/

Thematic actions

Global ocean and climate change

Ocean circulation, water cycle :
Global interoperability of offshore in situ and satellite databases (Altimetry, SSS, ARGO, ADCP, ...)

Bio-geochemical cycles :
Consideration of offshore data of biogeochemical interest

Coastal areas in global change

Interoperability of coastal databases: Quadriga, Benthos, Pelagos, SOMLIT, MOOSE (Low and High Frequency)

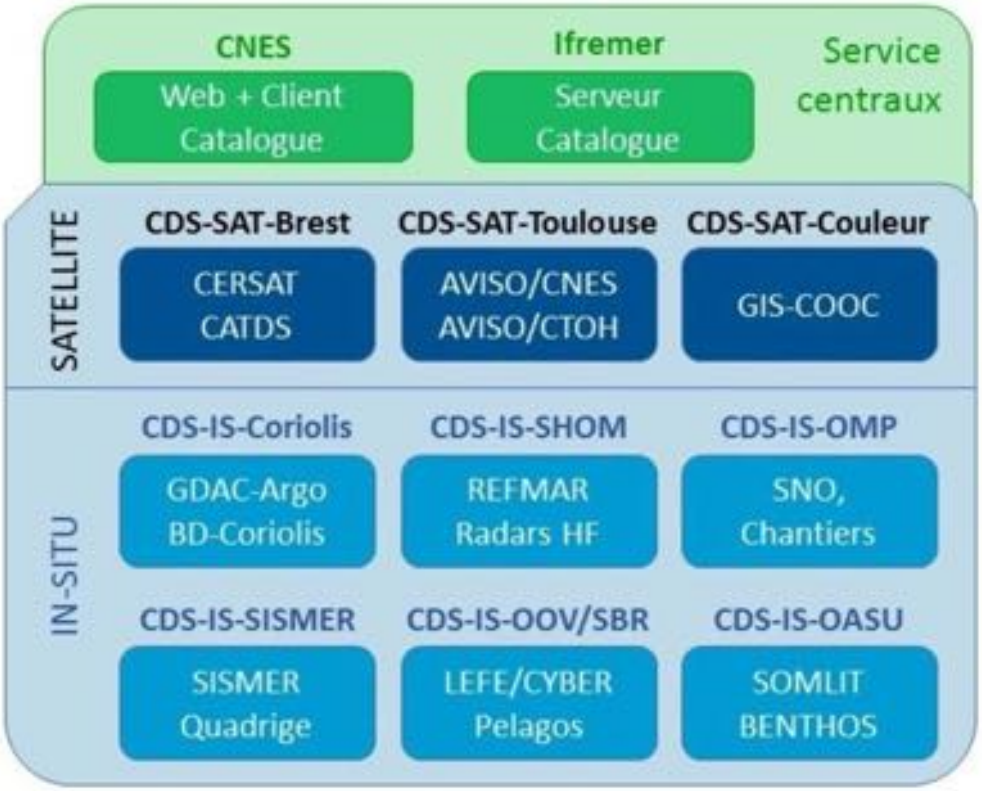
Water color and coastal issues

European integration

Seadatacloud, ENVRI PLUS/FAIR, Copernicus

cnes cnrs IGN Ifremer IRD METEO FRANCE

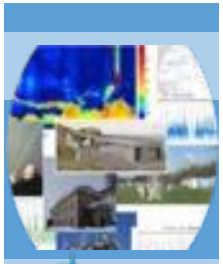
Data & Services hubs

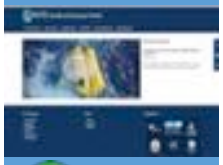


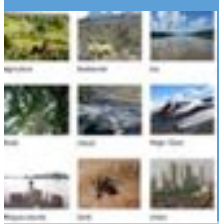
9 data centers (2018) :
- 3 for satellite data &
- 6 for in-situ data.
=> 12 data centers,(2019)

ODATIS is interfaced with several french research infrastructure:
- **TGIR** : French Oceanographic Fleet
- **IR EMSO** : Seabed observatories (ERIC),
- **TGIR ARGO** : Argo floats (ERIC, Euro-Argo)
- **IR I-LICO** : Coastal and nearshore observations, link with Theia
- **IR O-HIS** (under construction) : offshore observations outside oceanographic vessels.

Infrastructure de Recherche Système Terre





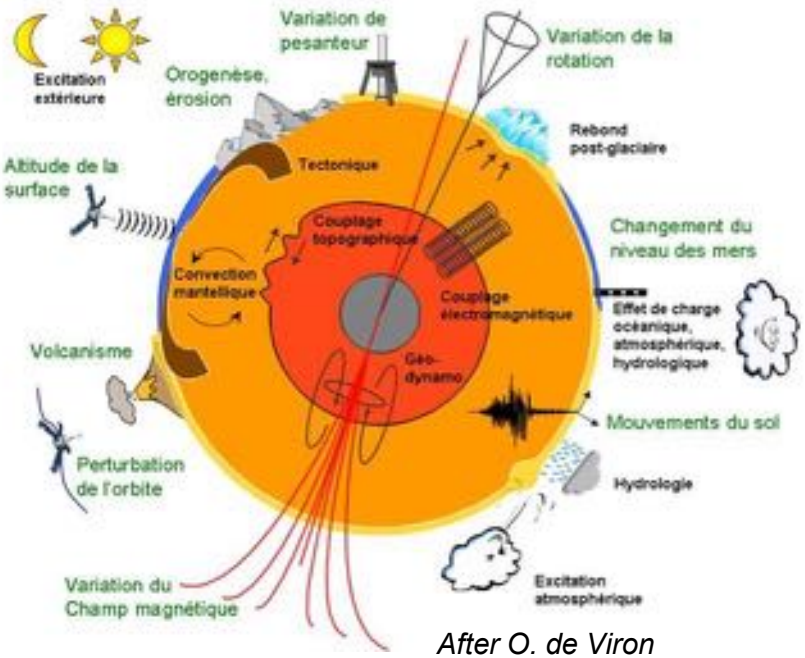


„Earth System“ Research Infrastructure

The Earth : a fascinating but complex system…

The Earth, a fascinating but complex system:

- numerous geophysical & geodynamic processes,
- with variable spatial and temporal scales,
- with many interactions, within and between its various compartments: *inner Earth, land surfaces, ocean, atmosphere, (not forgetting interactions with anthroposphere and also Universe...)*

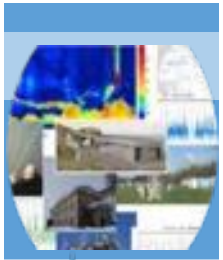


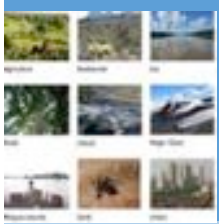
After O. de Viron

To understand these geophysical ,geodynamic and geoenvironement processes,
need to analyze numerous and very large datasets
(*satellite, in situ, campaigns,* ***long term observations*** *but also experimentation results, model outputs, AI, …*).

Scientists and decision makers need to have an easy access to all these data and associated products!

Infrastructure de Recherche Système Terre





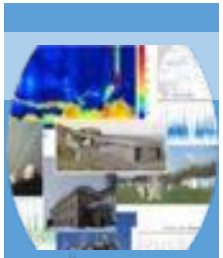
“Earth System” RI objectives

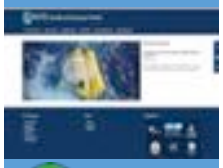
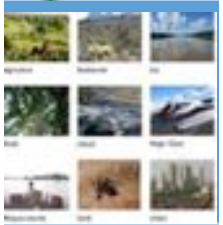
Develop a global **data, products and services** hub allowing to **observe, understand and predict** in an integrated manner the **history, functioning and evolution of the Earth System subject to global changes**.

- **implement**, at national, European and international level, **integrated approaches** for using **Earth Observation data** and derived information,
- **facilitate access** to high quality **data and products** across all compartments of the Earth system and their interactions,
- **fostering mutualisation, interoperability,** emergence of **multi- and interdisciplinary approaches and innovation** for scientific breakthroughs and the emergence of new services,
- **Serve** the **scientific** communities, the **actors of public action and innovation**

Infrastructure de Recherche Système Terre

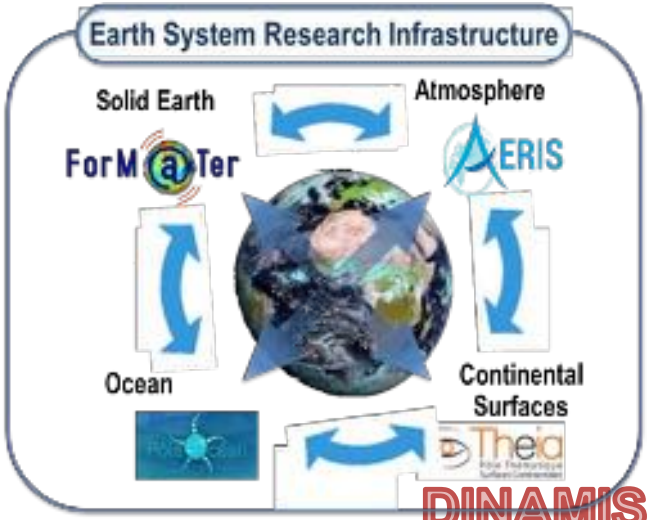
Missions of the "Earth System" RI



Missions of the "Earth System" RI

- **Federate** the Data & Services Hubs
- Develop **portals**
- Develop **tools to access & analyse satellite and in-situ** (ground, sea, airborne…) **data**
- **Ease and foster integrated & interdisciplinary research** to understand the processes associated with the **Earth System and Global Changes,**
- **develop European & international partnerships.**



DINAMIS
Inter-Pôles
...

*"Earth System" RI is positioned on the **whole data cycle** (in-situ and satellites), from their **production** (in synergy with other RIs and observatories) up to their **delivery** to **users** and to national, European and national **databases** and to national, European and international organizations (Copernicus, GEOSS, EOSC ...).*

Infrastructure de Recherche Système Terre

Structure & Organisation of the Earth System RI

- ✓ a RI on the national research infrastructure road map
- ✓ 4 Data & Service Hubs
  - ✓ THEIA : land surfaces
  - ✓ AERIS : atmosphere
  - ✓ ForM@Ter : Solid Earth
  - ✓ ODATIS : ocean

- ✓ Transverse activities
  - ✓ DINAMIS : Mutualised satellite imagery distribution
  - ✓ INTER-POLES committee : technical coordination
  - ✓ Working group Europe

- ✓ Governance
  - ✓ Steering board : 34 partners
  - ✓ Executive board : CNRS, CNES, IFREMER, IGN, IRD, IRSTEA, Météo France, MESRI
  - ✓ Directing Board: 4 Centre Directors & Transverse Actions managers
  - ✓ A Director & a team
  - ✓ Operational structure : UMS CPST "Coordination Pôles de données et de services pour le Système Terre"

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

Infrastructure de Recherche Système Terre

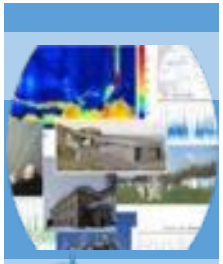
Priorities

- Reinforce cross-cutting activities

(DINAMIS, Inter-pôles WG, Europe/Int. WG, …)

- Active participation in European initiatives (H2020-EOSC and FP9, ESFRI, Copernicus,…)
- Active participation in international initiatives : GEO/GEOSS, ONU-Env, GO FAIR, ENVRIfair...
- develop an efficient integrated information system (storage, archiving, processing, AI, cloud services, portals)
- Synergies and joint activities with with Space Climate Observatory (SCO)
- Development services based on artificial intelligence

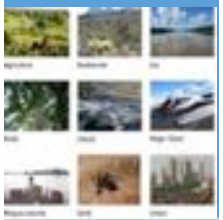
Infrastructure de Recherche Système Terre



AERIS

ForM@Ter

ODATIS



Theia

„Earth System“ RI context

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

Infrastructure de Recherche

Système Terre

Status of the Data and Service Hubs

- 4/5 Data and Services Hubs (AERIS, FORM@TER ODATIS, THEIA & FRB)
  - **Very different**
  - Not the same level of FAIRisation
  - Data from French / European satellites, but also from other countries (NASA, JAXA, USGS, NOAA, …)
  - **In-situ data, models, …**
  - Each Hub is distributed among several data & services centers
  - The current state of the Data & Services Hubs is to be taken into account
  - The volume of data is increasing
  - A mandate to open up to the downstream sector
- A few figures
  - 15 Data & services infrastructures
  - 30 CES : scientific expertise consortium
  - 7 000 To (2017) - 50 0000 To (2022)
  - 150 ETP full time equivalent – 250 scientists, data scientist, engineers, technicians
- In progress & in discussion:
  - 5th data & Services hub on biodiversity

FRB FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

‘Technical harmonization’

- Interpole working group
  - Created in 2014 to promote technical exchange between the Data & Service Hubs
  - A two-day workshop every 6 months
  - 2 in depth topics handled in parallel
  - Example of topics
    - Long term preservation, Authentication & Autorisation, catalogues, formats, DOI, Licences, processing, …
- RDA – Research Data alliance
- H2020 : ENVRI+ / ENVRI FAIR in an ESFRI context
- GO FAIR initiative
- Space context
  - CEOS/WGISS – Working Group for Information Systems and Services
  - ESA DCB (Data Coordination Body) and their working groups

Infrastructures – environment and earth system

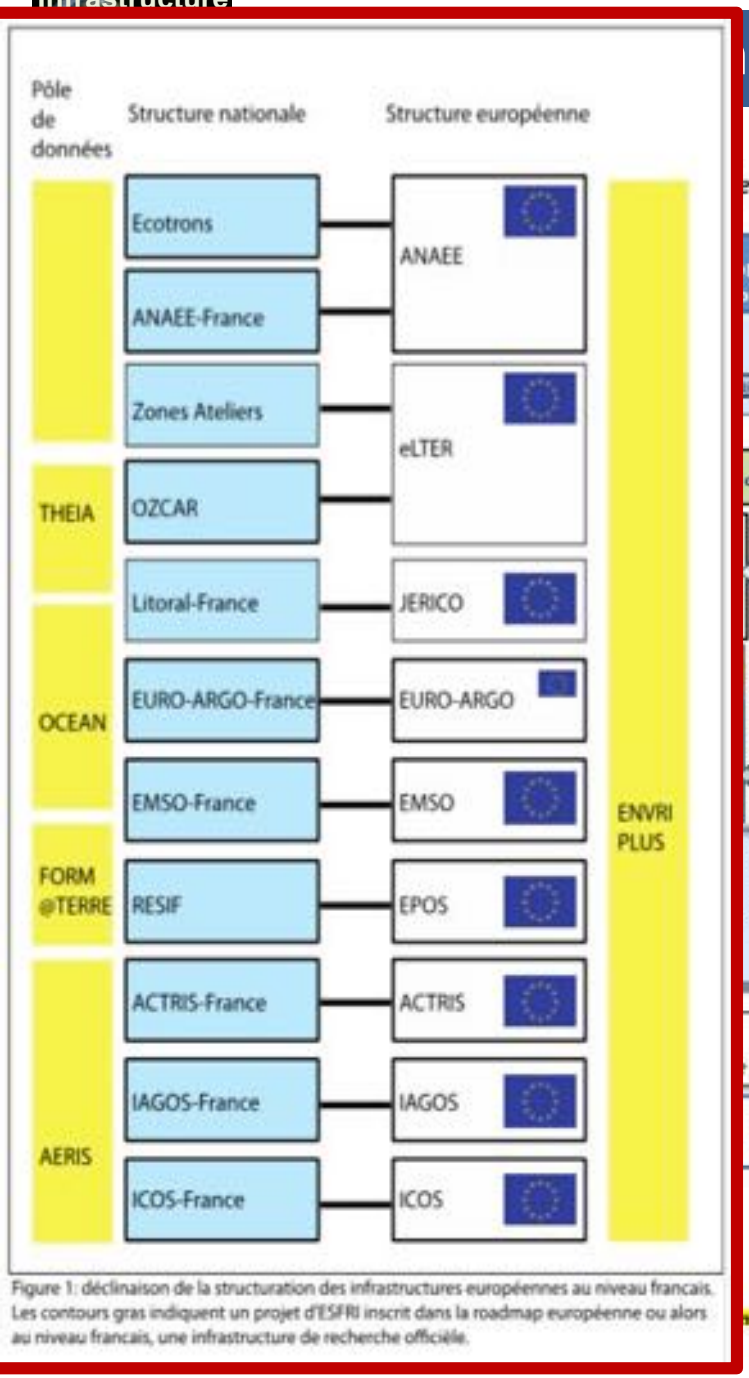
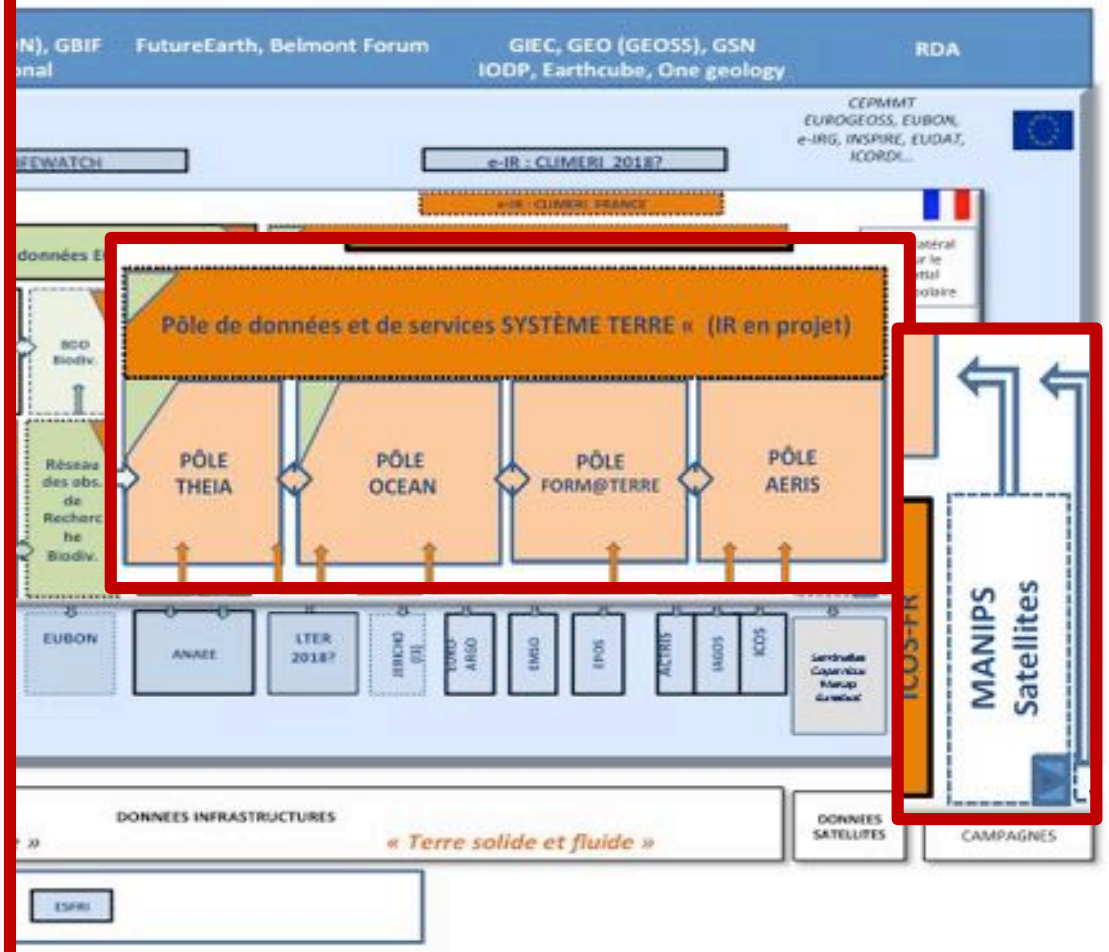


Figure 1: déclinaison de la structuration des infrastructures européennes au niveau francais. Les contours gras indiquent un projet d'ESFRI inscrit dans la roadmap européenne ou alors au niveau francais, une infrastructure de recherche officiéle.

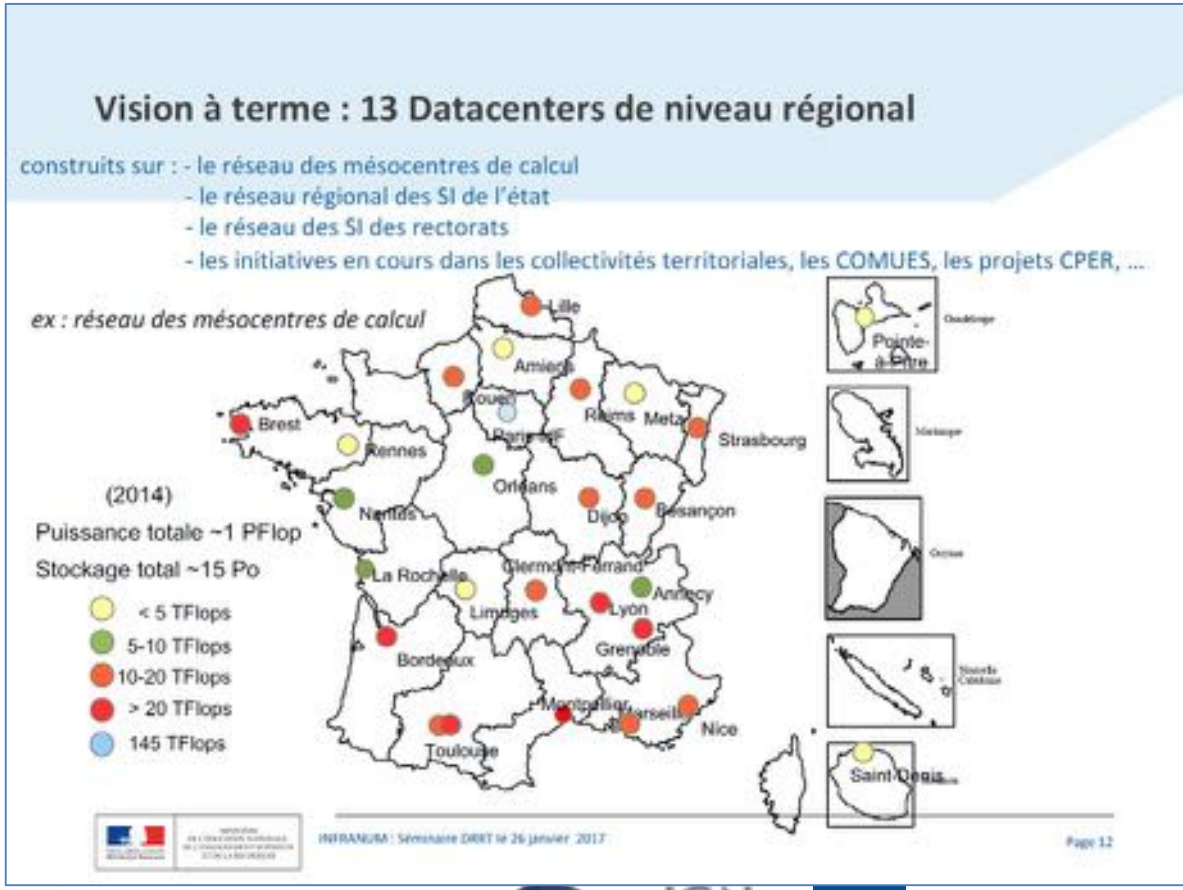


Infrastructure de Recherche

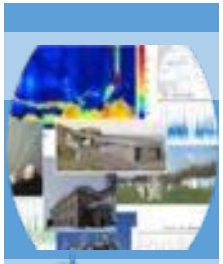
Système Terre

INFRANUM

- Project from the MESRI ministry to rationalize the computing infrastructure of the research and of the higher education
- Data and processing should converge toward big regional clusters



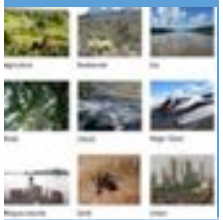
Infrastructure de Recherche Système Terre



AERIS

ForM@Ter

ODATIS

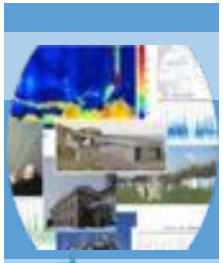


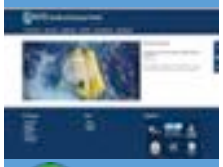
Theia

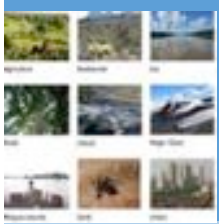
Technical Solution

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

Infrastructure de Recherche Système Terre







Technical work to be done

- Develop a WWW portal for communication
- Develop a portal allowing a state of the art access to data and services
  - Allowing to discover, access and process the datasets and the data (i.e. granules)
  - MMI & web services with rich criteria at the level of the data
- Elaborate a technical harmonization for the Data & Services Hubs
- Take into account the future constraints and opportunities
  - **INFRANUM**, EOSC, DIAS
  - EOSC => BlueCloud (IFREMER) / FoodCloud (INRA)
- Take into account the European and International contexts
  - ESFRI & ENVRI* ⇔Les infrastructures de recherche européennes
  - ESA 'EO innovation Europe'
  - Evolution of GEOSS & EuroGEOSS
- Increase the FAIRness of the IR ST data
- Propose interoperable services
- Provide means (software and hardware) allowing to combine data from all the Earth System compartment

Infrastructure de Recherche Système Terre

Technical work to be done

- Inspiring models:
  - NASA EOSDIS hub Opensource
    - Common Metadata Repository / Unified Metadata Model
    - Combined with their progressive migration toward cloud computing
  - Hub Pangeo (~datacube) Opensource
    - Already used for atmosphere, ocean and climate data
  - GeoDAB: GEOSS data hub
    - And the H2020NextGEOSS initiatice Opensource
      - Data hub combined with cloud processing capabilities
  - ESA Initiative : 'Network of Exploitation platforms'
    - And their cloud initiatives:DIAS, TEP/MEP/MAP
  - ENVRIfair H2020 project
    - Naturally linked to EOSC
    - French Data&Services Hubs are part of the consortium
      - IAGOS, ACTRIS, EURO-ARGO, EPOS, EMSO, ANAEE, …

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

High level functions

Exploitation teams

operate

Users

Products Tools Services

Users feedback

Training Incubation Algorithms Chains

Eo-lab & InSpace

Support

Data & Services Infrastructure

Data & Services Centers — CDS

Repository / Publication — DPO

Advanced services for dissemination — HDSK

Quality of service — QoS

IDS

User feedback

Users support

Help Desk — HDSK

Support to data producers — SPRD

Collaboration

Support to developpers — SDEV

Animation centers (regional and national) — CART

SUPP

Expertise

Scientists

Products definition ATBD & proto

Operational Processing chains

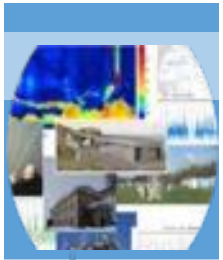
Dev teams

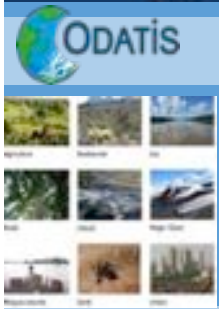
Support to developpers

Synergies

AI initiatives eg Datapole

Infrastructure de Recherche Système Terre





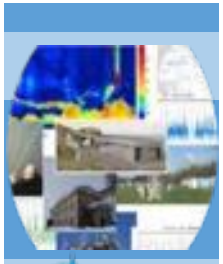
Data: IR ST data catalog

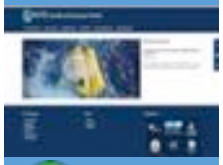
- A unique, operational catalog
  - Containing all data metadata and pole services
    - Like the NASA EOSDIS CMR (https://earthdata.nasa.gov/about/science-system-description/eosdis-components/common-metadata-repository)
  - Requires a unified data model (~ UMM NASA & HMA ESA)
    - Taking into account spatial data, in-situ, models, ...
    - Work in progress as part of the GT Interpole catalog
  - May also contain metadata for products that are not IR ST
- With interfaces
  - interoperable
    - **INSPIRE**, CEOS opensearch, Linked data / RDF, GeodCAT, WIGOS / WMO standards, ...
  - Adaptable (nothing is fixed)
  - Two step search: collections then granules
    - CEOS connected data assets, GEOS, NASA EOSDIS, ESA FedEO, ...
  - Enabling community portals for sophisticated data and services research
    - Example: SCO or poles

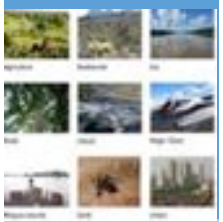
Processing

- Solutions to the needs
  - Have a centralized / common way for multi-source treatments
    - => Cf INFRANUM
  - Offer cloud compatible technology solutions
  - Ability to launch projects on DIAS and EOSC
    - To answer calls for projects CE and ESA
    - Compatibility of IR interfaces with DIAS and EOSC
  - Propose (after user needs analysis)
    - An innovative solution for data recovery like:
      - datacube + Jupyter
      - Pangeo
      - Knowledge graph
      - SeaScope (Ocean Datalab)
    - Strong Artificial intelligence capacities

Infrastructure de Recherche Système Terre







Technical Strategy

Computing Infrastructures

- Target = INFRANUM

- Progressive but determined approach
  - Migration when data is used in a transverse IR ST framework (SCO example)
  - Migration when IT means become obsolete
  - Some datasets may not be migrated
    - Old and little used
    - Necessary proximity of the producer while being of a reasonable volumetry: case of certain **in-situ data** => concept of technical cache

- Ability to be distributed across multiple sites
  - At the beginning the IR ST will still be spread over several CDS themselves distributed over several sites
  - On opportunities to be able to switch to external means like DIAS or EOSC
    - Eg for European projects

Infrastructures

- INFRANUM not operational within 3 years

- **Intermediate step** on relatively limited means
  - Temporary infrastructure to choose
    - CINES
    - CNES
    - DIAS
    - EOSC ⇔ EUDAT, EGI, …
  - A way to learn and validate the consideration of user needs
  - Software for the intermediate stage
    - Limit specific developments (at least for this step)
      - Catalog and data hub (CKAN, NASA CMR, CNES datalake, ...)
      - 'Modern' means of processing data: Datacube + Jupyter / Pangeo / ...
  - Rapid Implementation of Development / Migration Support service

Data Management Plan

– All data and services will not be migrated simultaneously
  - Need to prioritize
    => Inventory of data and services

TO BE FINDABLE:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource
F4. metadata specify the data identifier.

TO BE ACCESSIBLE:

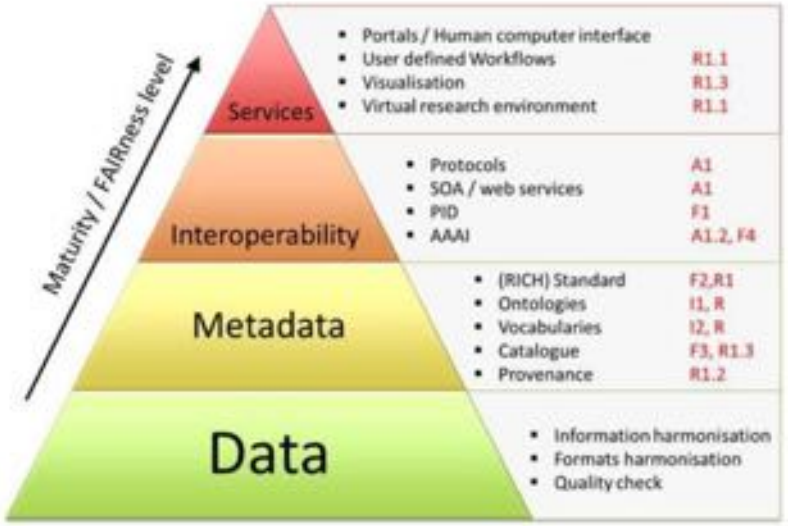
A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

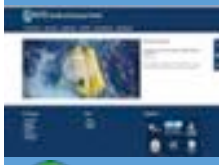


Inventory via a FAIRness maturity matrix for data, metadata and services

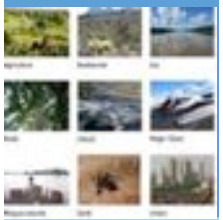
Infrastructure de Recherche Système Terre

AERIS

ForM@Ter



ODATIS

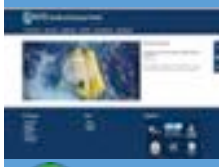


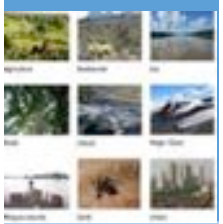
Theia

Conclusion

irstea Ifremer IRD cnes cnrs IGN METEO FRANCE

Infrastructure de Recherche Système Terre





Conclusion

- "Earth System" Research Infrastructure
  - Will allow to combine data from all Earth System compartment
    - Ocean
    - Land surface
    - Solid Earth
    - Atmosphere
    - Biodiversity
  - **=>** removal of the existing silos between communities
- Data and processing chains
  - Will be clusters in big computing infrastructure (EOSC)
  - Allowing to efficiently combine the data
- This wealth of data will
  - Be exploited by "modern" processing tools
    - eg Pangeo / Datacube
  - will be perfectly suited for techniques like Artificial Intelligence