



## Updated research infrastructure roadmap

2017-2020 Report annexed to survey

### DATA TERRA

Earth system data and services



### 2017-2020 Report

*Frédéric Huynh (IRD), Data Terra RI Director*

*Nicole Papineau (CNES), Data Terra Officer*

*Michel Diamant (IPGP), ForM@Ter Director*

*Emilie Deschamps-Ostanciaux (CNRS), ForM@Ter Deputy Director*

*Patrice Henry (CNES), AERIS Director*

*Sebastien Payan (Sorbonne Univ.), AERIS Deputy Science Lead*

*Gilbert Maudire (Ifremer), ODATIS Director*

*Sabine Schmidh (CNRS), ODATIS Deputy Director*

*Nicolas Bagdhadi (INRAE), Theia Director*

*Sylvie Galle (IRD), Theia/OZCAR In-Situ Data Officer*

*Richard Moreno (CNES), Data Terra Chief Technical Officer*

*Karim Ramage (CNRS), Data Terra Deputy Chief Technical Officer*

*Jean-François Faure (IRD), DINAMIS Executive Secretary*

*Caroline Blanke (CNRS), Data Terra Officer*

# Contents

1	Foreword	3
2	History and structure	3
3	Organization and coordinated activities	5
3.1	Governance	5
3.2	Science organization and data applications	6
3.2.1	Science working group activities	7
3.2.2	Coordination mission	7
3.3	Organization and technical activities	8
3.3.1	Interpole	8
3.3.2	Technical management team	9
3.3.3	GAIA Data	10
4	Links to Earth-observation research infrastructures	13
5	Data Terra and space	17
6	Active participation in European projects and international actions	18
6.1	European contributions	18
APPENDIX		20
Report of the data hubs making up Data Terra		20
1	AERIS data hub (Atmosphere, <a href="http://www.AERIS-data.fr">www.AERIS-data.fr</a> )	21
2	ForM@Ter data hub (solid Earth, <a href="http://www.poletterresolide.fr">www.poletterresolide.fr</a> )	26
3	ODATIS data hub (oceans, <a href="http://www.ODATIS-ocean.fr">www.ODATIS-ocean.fr</a> )	33
4	Theia data hub (land surfaces, <a href="http://www.theia-land.fr">www.theia-land.fr</a> )	39
5	DINAMIS national space data structure	43

## 1 FOREWORD

Data Terra is a research infrastructure (RI) dedicated to Earth system data, founded in 2016 and part of the 2019-2021 national roadmap. Its annual full mean cost is €40 million and it mobilizes more than 160 FTEs, spread across a cadre of more than 400 people from 30 partner organizations and universities.

Data Terra's main mission is to develop a structure for accessing and processing data, products and services geared towards observing, understanding and predicting in an integrated manner the history, workings and evolution of the Earth system in response to global changes. Its multi-source data, products and services are aimed at the scientific community and public and socio-economic stakeholders, and are accessible via a coherent, one-stop portal.

One of Data Terra's major ambitions is to federate and optimize existing institutions, structures and resources at national, European and international levels. Data Terra is built around four data hubs each focused on one of the key components of the Earth system: land surfaces, atmosphere, oceans and solid Earth, complemented by working groups and a cross-cutting service.

This document briefly describes activities pursued since 2014, when the four data hubs were created under the coordination of their main overseeing organizations. After outlining the context and history, it details progress status in a number of key areas, namely structure, governance, science strategy, technical developments, advances in European projects, links with Earth-observation RIs and space strategy. The activities of each data hub are briefly outlined in the document's appendix.

## 2 HISTORY AND STRUCTURE

The Data Terra RI has been built up in six main steps:

1. 2012-2014 focus group on Earth-observation data centres, coordinated by the French space agency CNES and the INSU national institute for universe sciences with contributions from numerous organizations; this led to the creation of the data centres and an Earth System RI to manage interactions between them
2. Inclusion as a proposed RI in national RIs/VLRIs\* – Ministry of Education, Research & Innovation roadmaps as part of the contribution to the Earth System and Environment domain in 2016
3. Establishment of open science and open data policies
4. Concerted effort to Europeanize RI activities through the European Strategy Forum on Research Infrastructures (ESFRI)
5. Services providing access to satellite and in-situ data (EOSC, Copernicus, etc.) coordinated since 2015
6. Establishment of international data and service sharing structures (GEO, GO FAIR, etc.)

\* Very Large Research Infrastructures

The inter-body focus group that drove the development of Data Terra in turn made six key recommendations:

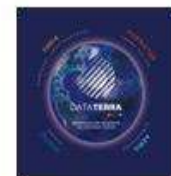
1. Create four closely interlinked national data hubs (Atmosphere, Ocean, Land Surfaces, Solid Earth) and work towards building a research infrastructure
2. Establish a strong and clear governance and executive
3. Make scientific expertise central to the data hubs
4. Leverage pooled technical resources
5. Develop access to data and computing and development processes
6. Embrace new technologies

## Key dates in the construction of Data Terra



## Structuration IR Data Terra- Dates clefs

- Nomination de chargés de mission - mise en place des pôles (2014/2015)
- Lancement du GT "Inter-pôles" (2015)
- Inscriptions sur la feuille de route nationale IR/TGIR du MESRI en 2016 et 2018
- Directeur de l'IR nommé en mai 2017
- Création de l'UMS CPST : CNRS en 2019, IRD en 2020, ...
- Lancement de DINAMIS (2019)
- Mise en place d'une Direction Technique (2019)
- Groupes de travail : GT TECH (2019), GT COM (2019), GT Science (2020), ...
- Projets nationaux (ANR, CPER), Européens (2018, 2020)
- Dépôt projet GAIA Data – AMI EQUIPEX+ (2020-2029)



**En parallèle, développement des pôles,  
composantes essentielles de l'IR**



### 3 ORGANIZATION AND COORDINATED ACTIVITIES

#### 3.1 GOVERNANCE

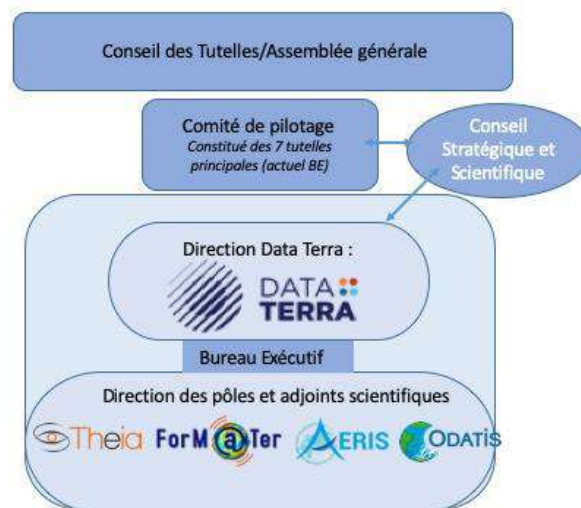
Data Terra is built around four data hubs each focused on one of the key components of the Earth system: land surfaces, atmosphere, oceans and solid Earth, complemented by a cross-cutting service (DINAMIS) supplying high-resolution satellite remote-sensing data to the hubs. This federation of competencies is illustrated in the figure opposite.



The Data Terra RI is composed of 26 organizations and universities.

Its governance revolves around:

- A General Assembly (GA) representing all overseeing organizations, signatories of RI and data centre agreements. The GA meets once a year.
- A Steering Committee with members from the main overseeing organizations (CNES, CNRS, Ifremer, IGN, INRAE, IRD, Ministry of Education, Research & Innovation and Meteo-France). This decision-making body meets two to three times a year.
- An Executive Bureau (EB) consisting of the Data Terra management team and data hub directors.
- A strategic and scientific oversight council. This external body provides its know-how and expertise in developing the RI and conceiving its strategy.



Data Terra's governance has to adapt to the specific needs of the different data hubs' communities and to the strategies individual to each Earth system component, while being coordinated and integrated around a shared overarching strategy.

This governance is geared towards addressing the multidisciplinary, multiscale science challenges that encompass several components of the Earth system and their interactions to enable a systemic approach to global change and its impacts.

This governance was established through:

- The signature of agreements covering the data centres.
- The creation of the joint service unit (UMS) coordinating the CPST Earth system data centres and services, complementing Data Terra, by the national scientific research centre CNRS in January 2019, after approval by CNRS's Scientific Committee (CS), and by the IRD development research institute in November 2019. The data hub directors are deputy directors of this joint service unit.
- The finalization of the joint service unit agreement and its signature planned end 2021, involving all of the overseeing members of the Data Terra RI steering committee.
- An agreement to form a consortium of 26 partner organizations and universities is in the process of being drafted for signing in 2021.

The RI's director leads activities with the support of a management team comprising:

- A Chief Technical Officer (see later in the document for how the technical management team is organized)
- An officer serving as mission secretary aiding definition and implementation of the cross-cutting science strategy
- A communications officer working with those at each data centre
- An institutional partnerships officer
- A Europe and international officer
- An administrator

The management team meets every fortnight and a management seminar is held every six months. This seminar brings together the members of the Executive Bureau, Data and Service Centre (DSC) directors and people involved in engineering and science activities.

Data Terra was first headed by Nicole Papineau (CNES/CNRS) on its launch in 2016, succeeded by Frédéric Huynh (IRD, 2017- ).

The data hubs are/were led by:

- AERIS: Nicole Papineau (CNES/CNRS), from its creation in 2014 to 2019, and by Patrice Henry (CNES) since 2020 with Sébastien Payan (Sorbonne Univ.) as Deputy Science Lead
- ForM@Ter: Michel Diamant (IPGP) since its creation in 2015, with Olivier Jamet (IGN) in its early precursor stages from 2012 to 2015
- Theia: Nicolas Baghdadi (INRAE) since its creation in 2012, with Sylvie Galle as In-Situ Data Officer since 2018
- ODATIS: Fabienne Gaillard (Ifremer) since its creation in 2015, with Gilbert Maudire (Ifremer) since 2018 and Sabine Schmidt (CNRS) as Deputy Science Lead

### 3.2 SCIENCE ORGANIZATION AND DATA APPLICATIONS

The Data Terra RI is predicated on making science central to everything it does. This is reflected by the fact that its data hubs are headed by scientists (a science lead and/or deputy science lead).



Scientific expertise is a key element of the data hubs:

- The hubs are focused on data and services, providing expertise, assuring value and quality of data and methods, leading and training communities, and conducting thematic foresight
- Scientific Expertise Centres (CES)—consortia of research laboratories or experts—mobilize/involve communities to meet their needs, promote research results (scaling up) and plan products and services
- A shift will be operated towards science data hubs encompassing several DSCs and working together as a network (see GAIA Data project), with science data hub activities co-led by data specialists and scientists

Planning of Data Terra’s activities includes those of the data centres while developing cross-cutting approaches to spur uptake of multidisciplinary data within an integrated Earth system approach. A Science Working Group was therefore set up in 2019-2020, with two people from each data hub and led by the Data Terra management team, to share approaches and lessons learned, and to cultivate and support cross-cutting themes and issues.

### 3.2.1 SCIENCE WORKING GROUP ACTIVITIES

On the basis of the scientific foresight of research organizations (INSU, CNES, IRD), the Science WG analysed the specific features of each data centre and what they have in common, and identified cross-cutting challenges.

This WG drafted a document in 2020 ([www.data-terra.org/groupe-de-travail-science/](http://www.data-terra.org/groupe-de-travail-science/)) setting out several priority challenges that cut across the data hubs, such as achieving an integrated approach to the land-sea continuum and the critical zone, urban dynamics, the climate and water cycle and food security, environmental hazards and impacts of society, biodiversity dynamics, agro-ecology and energy, etc. To meet these science challenges, it is vital to be able to extract and combine data from multiple sources. The new integrated services offered by Data Terra will have to be scaled to overcome the scientific and technological obstacles to accessing the full spectrum of data sources, extracting them and then combining them to develop high-quality synoptic data products.

### 3.2.2 COORDINATION MISSION

It is vital that work to coordinate activities and propose themes and cross-cutting projects be continued.

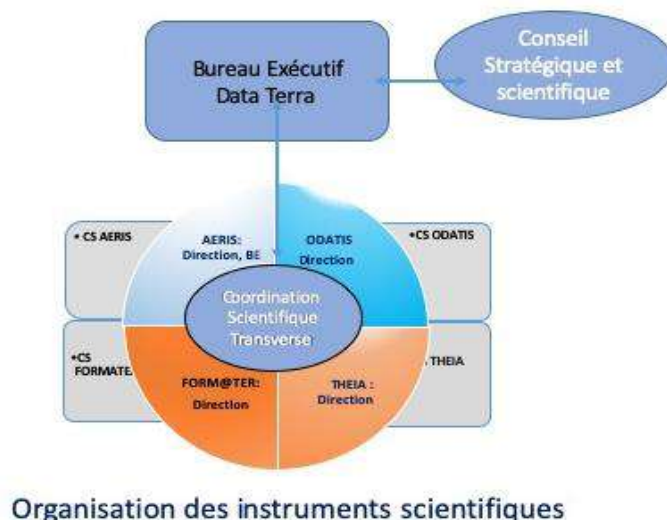
The Science WG is replaced by a coordination mission. A post for an officer is planned in 2021 to be the mission’s secretary. Their role will be to help the Data Terra management team shape and execute the RI’s science strategy. This cross-cutting, integrated strategy will be underpinned by and complement the science strategies of each data hub.

Data Terra has two instruments for coordinating its data hubs: a Strategic and Scientific Council and a Coordination Mission, interacting with data hubs’ scientific councils.

The Strategic and Scientific Council is an external body contributing its know-how and expertise in crafting the RI’s strategy and aiding its development.

The Coordination Mission is an emanation of Data Terra’s Executive Bureau comprising the RI’s management team and data hub management teams. This mission helps to execute Data Terra’s cross-cutting, integrated science strategy. It produces analyses and proposals for the Executive Bureau that may be submitted to the Strategic and Scientific Council to solicit its opinion and recommendations.

The figure below shows how these Data Terra instruments and data hubs' scientific bodies relate to each other:



### 3.3 ORGANIZATION AND TECHNICAL ACTIVITIES

Pooling and federating the data centres has been a goal since their creation in 2014. Prior to their creation, the overseeing organizations set up the Interpole group.

Initially led by Françoise Genova and Thierry Levoir from CNES, the first Interpole meeting was held in June 2015. In 2017, Richard Moreno, Data Terra's current Chief Technical Officer, co-led this group. There have so far been 11 Interpole workshops, held regularly every six months. For more details, see <https://www.data-terra.org/activites/dispositifs-transversaux/inter-poles/>.

#### 3.3.1 INTERPOLE

Interpole workshops were instituted because the data hubs had reached different stages of maturity. The aim of these workshops is to:

- Facilitate discussions between data hubs, taking advantage of the presence of scientific and technical leads and the participation of subject matter experts
- Establish exchanges on tools and techniques in domains of shared interest
- Start thinking about recommendations, standardization and interoperability

The National Biodiversity Data Centre (PNDB) joined the Interpole workshops in June 2017.

Workshops are usually held in face-to-face format and bring together 35 to 40 people, which is the capacity of meeting rooms. In December 2020, as a result of the COVID-19 crisis, the workshop was held virtually with 100 people.

Workshops run for two days. Morning sessions take the form of a plenary with presentations (progress reports on Data Terra and its data hubs, European projects, the Research Data Alliance, etc.) and time for discussion. Afternoon sessions take the form of a hackathon, with participants getting together in small groups to work on topics planned in advanced or decided on the spur of the moment.

Interpole workshops typically cover two or three topics (WG) and analyse one or two new potential topics, like for example archiving, SSO, thesaurus, formats, DOIs, catalogue, licences, computing,



DMPs/certification. Where deemed appropriate, the group may produce a best practice guide (e.g. for DOIs or SSO).

These workshops have enabled the data hubs' people to get to know one another, share their respective practices and work together. They have helped to nurture the Data Terra RI and continue to be a most useful forum for exchange.

### 3.3.2 TECHNICAL MANAGEMENT TEAM

Interpole proposes a bottom-up approach enabling all data centre stakeholders to share practices. Conversely, the conception and development of the Data Terra RI call for a top-down approach initially embodied by a small Technical Working Group (TECH WG) and since 2020 by a technical management team supported by this TECH WG.

The technical management team is made up of:

- A Chief Technical Officer (CTO): Richard Moreno (CNES)
- A Deputy CTO: Karim Ramage (CNRS/IPSL)
- Five people chosen for their expertise in crucial areas for Data Terra (system, in-situ data, catalogue)

To factor in the organization of the GAIA Data project, a representative of CLIMERI and a representative of PNDB have joined this team.

The figure below illustrates this technical organization:



The Data Terra technical management team meets fortnightly and is tasked with defining the RI's development plan and the Data Terra system's overarching technical principles. To this end, the team calls on the TECH WG comprising:

- The Data Terra technical management team
- Two representatives from each data hub (Chief Technical Officer and a subject matter expert)
- A representative of DINAMIS
- Two representatives from PNDB
- A representative of CLIMERI-France since mid-2020

The TECH WG meets every four weeks and breaks out into sub-groups—which may call on experts from outside the WG—to investigate technical topics in more depth or to lead identified activities. These sub-groups are:

- Architecture WG (focused on infrastructure, computing, networking and data grids)
- Catalogue and Terminology WG
- SSO Authentication WG
- Data Management Plan and DOI WG
- Data Terra warehouse activity, with support from BRGM
- Training activity

### 3.3.3 GAIA DATA

The infrastructure is spread around the country with widely varying IT resources (data storage and computing capacity), science disciplines (oceans, solid Earth, etc.) and types of data (satellite, in-situ, models, etc.). A key issue at stake for Data Terra is to facilitate cross-cutting applications, that is, involving heterogeneous IT resources and data. The aim is therefore to offer transparent and continuous access to all multisource Earth system data.

In this respect, the GAIA Data project selected in 2021 in response to the PIA3 Equipex+ call for proposals is crucial to Data Terra. This project will enable the RI's technical developments to be tightly structured around the distributed platform described in the roadmap survey.

The project is being led by three research infrastructures: Data Terra, CLIMERI and PNDB, working with 21 partners. Its full cost is €65m with €16.2m in funding from the government's PIA 3 future investment programme.

Full details of the project can be viewed at [www.data-terra.org/pia-3/](http://www.data-terra.org/pia-3/).

The goal is to develop and implement an integrated and distributed data and services infrastructure/platform for observing, modelling and understanding the Earth system, biodiversity and the environment, spanning the entire data cycle (observation and modelling) from acquisition (satellite, ground and in-situ data) through to multiple applications (qualification/validation, storage, processing/knowledge extraction, products and services) for the scientific community working to advance knowledge of these three areas.

This structure, built around interconnected, science-led data and service centres, will afford access to different sources of data and provide a continuum of distributed services for data storage (notably accommodating large volumes of satellite data and modelling), processing (including use of computing and artificial intelligence resources in partnership with the GENCI very large research infrastructure, CINES and IDRIS), post-processing (notably data cross-correlation), analysis and visualization. FAIR services tailored to the needs of scientific communities will be offered, covering the entire data cycle from acquisition through to multiple applications.

This project recently received a big vote of confidence, as these comments from the international judging panel show:

*“France already spends considerable efforts integrating data systems but GAIA DATA will, if funded, provide a major improvement to an already excellent set of infrastructures. It will substantially enhance French scientific leadership and will integrate the various data management systems in a way which has not been achieved elsewhere in the world.*

*It is now well established that the holistic view of the Earth system requires such databases and high-speed data exchange, so establishing this infrastructure will have a great impact on Earth science and all its disciplines. The proposed infrastructure will have a particularly significant impact on the French Earth science community.*

*The full socio-economic benefits of these activities can only be fully realized if the data obtained can be managed and used in a highly integrated manner. GAIA Data intends to provide this integration and hence to enable realization of the data potential for both the scientific and industrial sectors. Publications will certainly be made which would not have been possible without this level of integration and the return on investment will be very large in the short, medium and long term.*

*If policy is to be evidence-based, the very best use must be made of the data acquired and this is only possible if the data centres are integrated such as is proposed here. There is a huge community in France needing this infrastructure and the proposal links very well with existing or planned data infrastructures in Europe.*

*The consortium provides the required expertise to achieve the objectives. The impact of such a data infrastructure is huge and will lead to a new era in Earth science. Furthermore, the complementarity to many existing or planned observational infrastructures (including OBS4CLIM) is convincing.*

*[There is a massive advantage in developing a data management infrastructure of the magnitude proposed by GAIA Data.*

*It has the potential to generate major benefits for the scientific and technical community of France and with the excellent outreach ambitions expressed could exert very positive socio-economic benefits.*

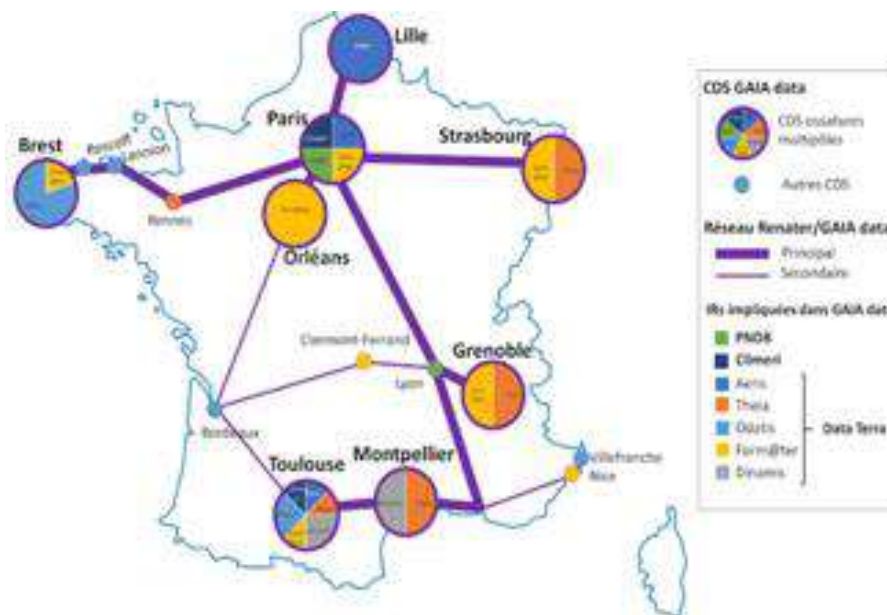
*There were several areas of concern expressed by the panel and these require attention, but the overall evaluation was very positive and the project definitively should be implemented.]*

*There is **massive advantage** in developing a **data management infrastructure** of the **magnitude** proposed by GAIA Data. It has the potential to **generate major benefits for the scientific and technical community of France** and with the **excellent outreach ambitions** expressed could exert **very positive socioeconomic benefits**. There were **several areas** of concern expressed by the panel and these require attention, but the **overall evaluation** was **very positive** and the project definitively should be implemented.*

*The Committee agrees that this project is of the **highest quality** and deserves to be **funded**”.*

The project is built around two main components:

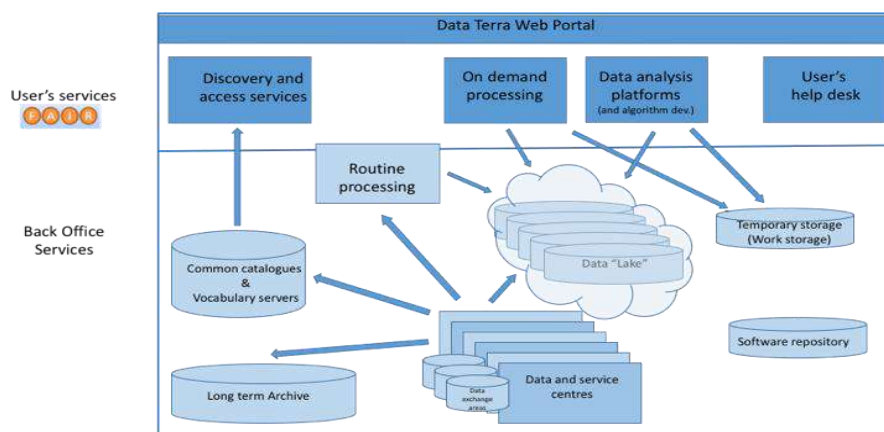
- A network of interconnected centres



- Services
  - Data discovery, access and management services
  - Cross-cutting services to enable cross-disciplinary research (data grid, interoperable cloud, knowledge portal, SSO)
  - On-demand analysis and processing services (virtual research environment, virtual analysis platform, interactive interface, sandpit, notebook, etc.)
  - Routine production and exploitation services (security, metrics, QoS, etc.)

Cross-cutting services leverage generic data services available in France, such as long-term archiving for example, and consolidate and adapt them where necessary.

These services complement one another: data discovery and access services implementing FAIR principles for data and products, development of algorithms spawning, for example, new on-demand processing services or even, following an ‘operationalization’ phase, routine production services.



[Schéma à corriger : User services]

#### 4 LINKS TO EARTH-OBSERVATION RESEARCH INFRASTRUCTURES

Data Terra is being put together in close collaboration with Earth-observation and experimentation infrastructures in the Atmosphere (ICOS, IAGOS, ACTRIS), Land Surfaces (OZCAR, RZA), Ocean (Euro-Argo, EMBRC-France, I-LICO), Solid Earth (RESIF-EPOS), Living Earth (Zones Ateliers (ZA) network, AnaEE Natura, Phenome-Emphasis, AnaEE Ecotron, ReColNat, RARE, etc.) and cross-cutting domains (SAFIRE, French Oceanographic Fleet, EMSO, I-LICO, REGEF project, etc.), with a view to serving their data needs.

A milestone has been reached in the last two years with a coordinated Data Terra approach. It will now be possible to pool methods more and strengthen interfaces with Earth-observation RIs. FAIR service developments and methods can be pooled. This collaboration is a longstanding feature for certain RIs and still in co-construction for others.

Some examples of cooperation are:

- IAGOS, for which Ether and then AERIS has been the European data centre since its inception, or Argo, for which the Coriolis DSC has been the European database from the outset. Another example of co-construction is ACTRIS, where AERIS is a key partner of this emerging European infrastructure's data centre. France is co-leading the data centre with Norway (NILU).
- Conversely, RESIF-EPOS has been built up with its own data centre, notably taking advantage of the construction of international seismology observatories, which had to structure their data at a very early stage.
- The in-situ land surfaces data information system is being built jointly by the Theia data hub and the OZCAR RI. OZCAR is one of the partners of the European eLTER RI (<https://www.lter-europe.net/elter-esfri>) currently in construction (eLTER-PLUS and PPP projects). The members of Theia/OZCAR (and RZA) are active in the eLTER WP tasked with structuring this RI's data and services.

The ODATIS ocean data hub and I-LICO RI interact closely and share the same commitment to support the elementary networks of I-LICO in proposing FAIR data and services, scaling up data management system requirements to match evolutions in coastal monitoring, and developing products for the littoral research community. These interactions are being structured through exchanges between ODATIS and I-LICO's coordination teams and governance structures.

The table below details the links and interactions between Data Terra and all the infrastructures in the environmental science field. As it clearly shows, certain RIs cover multiple disciplines and this is where Data Terra's intention to federate the different domains really comes into its own.

**Links between Data Terra and RIs working in the Environmental Science domain**

Scientific Field	French numerical RI	French RI/VLRI	European RI or partnership	PIA3 Project	Nature of collaboration with GAIA Data
Atmosphere	Data Terra/ AERIS	ICOS-FR	ICOS	OBS4CLIM	The data acquired by these three RIs (French components for ACTRIS and ICOS) are processed, formatted and stored in AERIS and disseminated by the Data Terra RI through the AERIS portal.
		IAGOS-FR	IAGOS		
		ACTRIS-FR	ACTRIS		
Land Surfaces and Interfaces	Data Terra/ Theia	OZCAR	eLTER	Terra Forma	The data produced by Terra Forma will be integrated in the Data Terra portal via Theia in-situ and in the PNDB portal.
Living Earth	PNDB	RZA			
		ReColNat	DISSCo	e-COL+	The e-COL+ project aims to acquire 3D-imaging to complete and extend the corpus of data and images produced by ReColNat, and to associate them with accelerated post-processing development thanks to artificial intelligence.
		RI AnaEE Ecotron AnaEE Natura	AnaEE		All the RIs in the "living Earth" domain plan to give access to and distribute their data (from in-situ or highly controlled experimental platforms) via the PNDB portal or also Theia for some physical environment data.  For Phenome-Emphasis, their internal solution relying on EGI services and on a Web 3.0 interface is a good use case to work the integration of new communities into Data Terra.
		RARe			
		Phenome-Emphasis	Emphasis		Data Terra will provide environmental data to Phenome for correlation between biota and environments.



Scientific Field	French numerical RI	French RI/VLRI	European RI or partnership	PIA3 Project, when submitted for same RFP	Nature of collaboration with GAIA Data
Ocean	Data Terra/ODATIS	Argo	Euro-Argo	Argo-2030	Argo manages the observation network of floats and produces observation data. Data Terra is responsible for data FAIRization. Thus, data produced by Euro-Argo will be managed by Data Terra/ODATIS.
Solid Earth	Data Terra/ForM@Ter	RESIF-EPOS	EPOS	MARMOR	Massive data produced by MARMOR will be integrated in the Data Terra portal via Data Terra/ForM@Ter.
Cross-cutting	Data Terra/ODATIS  ForM@Ter	EMSO-FR	EMSO		
	Data Terra/ODATIS  PNDB	EMBRC-France	EMBRC	Augmented biological observatory	Some data products (biogeography, species counts) produced by EMBRC-France will be integrated in the Data Terra portal via Data Terra/ODATIS. Environmental data will be provided by Data Terra for co-processing with genomic data.
	Data Terra/AERIS	SAFIRE	EUFAR	ANVOL E	All data acquired during airborne SAFIRE campaigns will be available through the Data Terra/AERIS Portal. Moreover, European aircraft data are available via EUFAR. ANVOL is dedicated to renewing the SAFIRE aircraft fleet.

Scientific Field	French numerical RI	French RI/VLRI	European RI or partnership	PIA3 Project, when submitted for same RFP	Nature of collaboration with GAIA data
Cross-cutting	Data Terra/ODATIS	French Oceanographic Fleet (FOF)	EuroFleets	Deep Sea innovation (submersibles)	Cruise Summary Reports and data from on-board sensors (echo sounders, seismic, CTD, submersibles data such as videos, etc.) will be managed and disseminated via Data Terra/ODATIS.
	ForM@Ter				Collaboration to offer production of satellite-based data for oceanographic cruises via Data Terra/ODATIS.
	Data Terra/ODATIS, Theia	I-LICO	-		Data produced by the elementary observation networks of I-LICO are processed, formatted and stored in Data Terra and disseminated by Data Terra via the web portal.
	AERIS				Creation of a joint product: a reference climatology of the French coastal environment.
	Data Terra: ForM@Ter	REGEF project			Data from the analytic platforms of the REGEF network will be accessible through Data Terra.
	ODATIS				
	Theia				

## 5 DATA TERRA AND SPACE



Data Terra has been structured around satellite Earth remote sensing data from the outset, making it a unique and original infrastructure in Europe offering one-stop access to ground and space data.

The French scientific community attaches a lot of importance to data from space systems to observe the key parameters of the Earth system and its evolving climate at regional, continental and global scales. Indeed, the combined use of space and ground data, survey data and aerial/balloon data has a significant multiplier effect that is advancing knowledge about the mechanisms at work in Earth's environment.

As Data Terra's partner, the French space agency CNES works with data and service centres to process, archive and disseminate science products derived from its Earth-observing missions.

AERIS is the Science Mission Centre for the POLDER, Parasol, Megha-Tropiques and Calipso missions, and soon for MicroCarb. CNES has also tasked AERIS with processing and distribution of gas concentration products from inversions derived from IASI data.

Theia is the mission centre for VEN $\mu$ S and operates value-added processing chains for Sentinel-1 and Sentinel-2 data archived at CNES. AVISO processes and distributes products from CNES's altimetry missions and products combined with other missions, and is a DSC of the ODATIS data hub. ForM@Ter develops interferometric processing services from Sentinel-1 data and ground deformation processing services from Pleiades very-high-resolution data. Form@ter's DSCs also specialize in processing and dissemination of space geodesy products derived from GNSS/GPS and DORIS data, and gravimetry products derived from GRACE/GRACE-FO data.

DINAMIS also provides one-stop access to high-resolution satellite imagery for the scientific and institutional communities. AERIS affords access to imagery from all satellites in the geostationary orbit arc via the SATMOS structure.

In addition to this strong national base, Data Terra has forged multiple ties with international space agencies, through partnership projects with CNES (NASA, ESA, Eumetsat, JAXA, ISRO, ISA, etc.) and directly to develop processing chains and gear them up for pre-operational production. Examples of the latter include IASI atmospheric chemistry products with Eumetsat, SMOS ocean salinity products with ESA, Calipso/CloudSat DARDAR products with NASA, altimetry products with Jason, Sentinel-3 and Sentinel-6 data for Eumetsat, snow cover products from Sentinel-2 and Hydroweb products for Copernicus.

Data Terra offers a range of services for users of space data and regularly adds to them. These include operationalization and implementation of new processing chains, on-demand data extraction from areas of interest, co-location of satellite and ground or survey data, visualization tools and support for users with real multi-disciplinary and multi-thematic scientific expertise.

## 6 ACTIVE PARTICIPATION IN EUROPEAN PROJECTS AND INTERNATIONAL ACTIONS

The data hubs have been involved in European and international projects from the outset. With the creation of Data Terra, a more coordinated policy was put in place.

Data Terra is a national research infrastructure (RI) strongly committed to European and international programmes and initiatives.

Data Terra is making a significant contribution through its data centres and scientific expertise to European and international initiatives like:

- European research infrastructures federated within the ENVRI cluster
- National space missions or bilateral (NASA, JAXA, etc.) and multilateral (ESA, Eumetsat, etc.) collaborations
- Close collaborations with Copernicus services, as in the marine (CMEMS), atmosphere (CAMS), climate (C3S) and land monitoring (CGLS) domains, contributing know-how in processing, data distribution systems and data standards
- EOSC initiatives
- The Destination Earth (DestinE) initiative to generate digital twins of the Earth
- International programmes like those led by IUGG, WMO, GEO and UNEP, through data from observation networks, international services, reference databases and data access and processing services for satellite, ground and in-situ data, including with South nations

Data Terra proposes a unique approach combining ground and space data and all components of the Earth system. This integrated approach squares with the aim of Destination Earth to obtain interconnected data on the Earth system and address the major environmental challenges that lie ahead.

Data Terra's European strategy will be accomplished in stages. Within the framework of ESFRI, Data Terra is a partner covering the entire data cycle through FAIR services with Earth-observation RIs. In the context of EOSC, it is helping to implement environmental data services through the EOSC-France initiative.

With respect to the European Space Policy, Data Terra is engaged at all levels of missions, from early concept phases right through to data utilization in partnership with space agencies.

### 6.1 EUROPEAN CONTRIBUTIONS

Close involvement in European projects and initiatives offers the means to strengthen Data Terra's services across the full data cycle and to consolidate its national, European and international position, working with other RIs and VLRI and exploiting synergies with the policies of organizations concerned.

The main projects in which Data Terra is playing an important role to structure European efforts are:

- ENVRI-FAIR (<https://envri.eu>): E-Infra environment cluster within the EOSC framework. Data Terra is involved in this project through the commitments of the European (ESFRI) ACTRIS, IAGOS, EPOS, Euro-Argo and ANAEE RIs and VLRI.
- EOSC-Pillar ([www.eosc-pillar.eu](http://www.eosc-pillar.eu), INFRA-EOSC, €12m, 30 partners): Data Terra, through CNRS and Ifremer, is tasked with coordinating the project's use cases and its FAIR services component for Earth observation.
- PHIDIAS ([www.phidias-hpc.eu](http://www.phidias-hpc.eu), CEF – DG Connect, €3m, CINES, CNRS, CSC, Ifremer, IRD, MARIS, University of Louvain): development of on-demand services from HPC centres for satellite Earth observation.

At national level, albeit with a strong European flavour, several projects have been selected through the French national research agency ANR's Flash open science call for projects: **Copilote** (<https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-copilote/>); **FairTOIS** (<https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-fairtois/>); and **CEDRE** (<https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-cedre/>) to contribute to furthering open science and FAIR data and service policies, and to certification of data and service centres.

Data Terra is actively contributing to GO FAIR with the start of the **GAIA Data Implementation Network** ([www.go-fair.org/2019/04/09/new-go-fair-implementation-network-gaia-data](http://www.go-fair.org/2019/04/09/new-go-fair-implementation-network-gaia-data)) approved in April 2019.

### Developing an international strategy

Data Terra and its data centres have partnered international projects and initiatives from the outset:

- Participation in international space missions
- Partnerships with international networks through national observation services (SNOs)
- Databases for the international community like ECADD, IUPAC and BGI

Data Terra has stepped up its international involvement through GEO, with support from the Ministry for Higher Education, Research and Innovation (MESRI), CNES and IRD. Data Terra's director sits on the Programme Board and numerous experts are taking part in Working Groups and initiatives in synergy with proposed products and services. In 2021, Data Terra will be assigning an expert under the aegis of MESRI, IRD and CNES to coordinate Earth-observation and SDG operations for GEO in Geneva ([https://earthobservations.org/geo\\_sdgs.php](https://earthobservations.org/geo_sdgs.php)).

It is worth noting increased international cooperation and strong overtures to internationalize and share Data Terra's ambitions, in particular with international agencies (UNEP) and nations in Africa (EU-AU cooperation, Lebanon, Morocco) and Latin America (Brazil).

## APPENDIX

### REPORT OF DATA TERRA'S DATA HUBS

Each data hub aims to ease access to satellite, aerial and in-situ data acquired and managed by research laboratories or federative structures (universe science observatories (OSUs), research federations (FRs), etc.), by national infrastructures like national observation services (SNOs) and environmental research observation and experimentation systems (SOERE), and by the French Oceanographic Fleet (FOF), aircraft, balloons and space missions.



## 1 AERIS DATA HUB (ATMOSPHERE, [www.aeris-data.fr](http://www.aeris-data.fr))



## 1.1 FOREWORD

The AERIS atmosphere data hub was created in 2014 but draws on over 20 years of experience in its domain. It encompasses four integrated and increasingly connected data and service centres (DSCs). The full mean cost of AERIS is €3.6m a year and it mobilizes 28.7 FTEs spread across 44 people from the 11 overseeing bodies, 28 of whom are permanent staff.

AERIS's main mission is to provide data, products, software, tools and services in the domain of atmospheric research. It works to describe, quantify and understand the atmosphere from a broad perspective through research areas like atmospheric dynamics, physics and chemistry. Its remit also covers work geared towards climate change research.

AERIS generates products from observations and offers a range of support services for data users, data survey campaigns and interfacing with models.



## 1.2 GOVERNANCE

From its inception in 2015, AERIS has established a governance structure to enable it to function efficiently under its overseeing organizations and with scientific supervision.

AERIS has held:

- Monthly meetings of its Executive Bureau, comprising the management team and DSC representatives
- Three to four meetings a year of its Scientific Council (renewed in 2020), tasked with crafting strategic proposals and providing a scientific perspective of projects undertaken
- Three meetings a year of its Steering Committee, comprising representatives from the 11 overseeing organizations and the Data Terra director



General Assemblies in the form of a two-day seminar bringing together 40 people working for AERIS were also held in 2017, 2018 and 2019. Among other things, they provided the opportunity to form inter-DSC working groups on topics like the catalogue, Web portal, infrastructure evolution and support for survey campaigns.



## 2019 - FIRST OPERATIONAL REVIEW

In the spring of 2019, AERIS conducted an Operational Review lasting several days, with visits to the different DSCs and a concluding presentation.

The review committee's members came from different horizons—scientists, data experts and IT system specialists—and all from outside the data hub itself. Its glowing report underlined that AERIS had accomplished most of the goals set when it was founded. This very positive assessment came with recommendations intended to help AERIS update its strategic plan. The review report is available on AERIS's website (<https://www.AERIS-data.fr/documentation/>).

### 1.3 DATA AND SERVICES

#### 1.3.1 ACCESS TO DATA AND SERVICES

One of AERIS's priority tasks was to develop and put in place a one-stop portal through which to access all 'atmosphere' data websites managed by the data centre (<https://www.AERIS-data.fr/>), as well as a broad catalogue offering users a centralized resource to search all existing data and products (<https://www.AERIS-data.fr/catalogue/>).

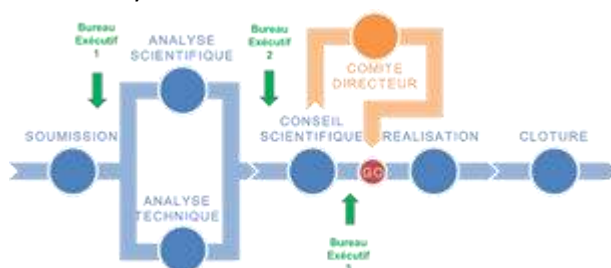
The AERIS website also keeps the atmospheric science community up to date with events, announcements, documentation and practical information.

A big effort has also been made to develop tools and services for users. In particular, AERIS proposes tools for extracting information from satellite imagery of points of interest (e.g. measurement points observed by ground platforms) and data pre-visualization tools. It has also put in place a service for assigning digital object identifiers (DOIs).

#### 1.3.2 OPEN CALL FOR PROJECTS

AERIS invites scientists to respond to an open Call for Projects via an online form on its website. Such projects are requests to AERIS to develop processing chains, format data, create services and databases, and provide access to data.

Since 2015, some 90 projects have been submitted and the vast majority have been accepted by AERIS's senior management team and validated by its Scientific Council. Projects are very varied in nature and duration (from a few days to several months), but they represent a fairly good balance between satellite data, ground measurements and survey data.



Flow diagram showing how projects submitted are processed

### 1.3.3 SPACE MISSION DATA MANAGEMENT AND PROCESSING

AERIS operates as a Mission Centre for CNES projects relating to atmospheric science. Continuing the work of the ICARE data centre, AERIS has completed processing of products from the Parasol mission and pursued operational processing of data from the Calipso and Megha-Tropiques missions, integrating new versions of processing chains (and associated reprocessing).

Level 2 inversion processing of IASI products has been accomplished with the IT resources of the AERIS/ESPRI data and service centre, and the same approach is now being applied to near-real-time processing and provision of products covering nine molecules and volcano dust.

Using data from all satellites in the geostationary orbit arc received by SATMOS, AERIS has compiled a unique global archive in NetCDF format accessible via the catalogue. These data are also used in automated processing chains to generate land aerosol products at the European Centre for Medium-range Weather Forecasts (ECMWF) for the Copernicus CAMS project.

More recently, AERIS has become involved in retrieving ESA Sentinel 5P products, in particular for the European PHIDIAS project.

AERIS is also partnering efforts in preparation for CNES's future MicroCarb and C3IEL missions, for which it will be performing part of science data processing and tasked with data distribution.

### 1.3.4 GROUND DATA MANAGEMENT AND PROCESSING / ACTRIS

AERIS plays an important part in managing and distributing data acquired from land sites or in laboratories. The volume of data is very small compared to satellite data, but this task is made more complex by their quantity and variety.

A big proportion of this activity is related to the ACTRIS research infrastructure.

AERIS had a stake as a data centre in constructing ACTRIS-France and the future European infrastructure. AERIS collects, processes, archives and disseminates measurements from this infrastructure's various observation platforms and laboratories.

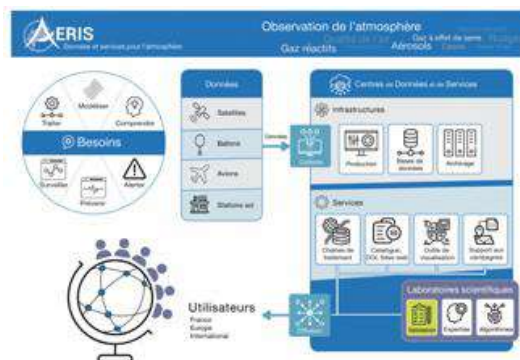
AERIS's director and deputy director sit on ACTRIS's governing bodies and members of AERIS are in the technical groups working on data.

A big proportion of these data is managed and made available to the Norwegian Institute for Air Research (NILU) for the future European ACTRIS ERIC research infrastructure.

### 1.3.5 AERIAL DATA MANAGEMENT

AERIS's catalogue includes the databases of the balloon-borne instruments previously managed at IPSL and OMP for the Ether data centre. It has continued this management role for new data. AERIS has joined the European HEMERA consortium as the centre for data acquired through this project from balloon platforms operated by CNES and SSC.

AERIS has two key activities concerning aerial data: the first is its historic role as a data centre for the IAGOS research infrastructure, and the second—more recent—involves





managing all data acquired from campaigns conducted by the SAFIRE unit operating French aircraft for the national weather service Meteo France, CNRS and CNES. This activity is set to be expanded to all disciplines covered by Data Terra.

### 1.3.6 CAMPAIGN SUPPORT

AERIS offers a survey campaign support service accommodating a wide range of data, from ground, in-situ or remote-sensing measurements to aerial measurements (by balloon and/or aircraft) and satellite acquisitions.

AERIS provides support:

- In preparing campaigns (with a portal for exchange and access to information)
- During campaigns (supplying external data and tools for visualizing balloon or aircraft measurements and trajectories)
- Post-campaign (with a website for access to data)

It is very often the case that measurements acquired during campaigns concern not only atmospheric parameters but also measurements to characterize sites of interest, both on land and at sea.

Major survey campaigns of note for which AERIS has been called upon in recent years include those for the AMMA and MISTRALS projects (ChArMeX, HyMeX) and more recently EUREC4A and MAGIC.

### 1.3.7 EUROPEAN AND INTERNATIONAL PROJECTS

A key recommendation of the CNES/INSU working group that led to the formation of the four French data hubs—and therefore AERIS—was the need to open up to Europe and combine national infrastructures within European consortiums.

AERIS has made excellent progress in this domain. Besides positioning itself as a data hub—or part thereof—for IAGOS, ACTRIS, EUROCHAMP and HEMERA, AERIS is closely involved in the ENVRI-FAIR project and a consortium member in the PHIDIAS, ANATOLIA and ATMO-ACCESS projects. It is also involved in the Copernicus CAMS and C3S projects.

AERIS also manages data from the AMMA and MISTRALS international survey campaigns and supplies data to ECMWF for the GAIA-CLIM project.

AERIS also serves as the French or European ground network database for international projects like AeroNet, GEWEX, GRUAN and NDACC.

It is also the data centre for the IUPAC (International Union of Pure and Applied Chemistry) database and the ECCAD (Emissions of atmospheric Compounds and Compilation of Ancillary Data) database.



## 2 FORM@TER DATA HUB (SOLID EARTH, [WWW.POLETERRESOLIDE.FR](http://WWW.POLETERRESOLIDE.FR))



## 2.1 FOREWORD

The ForM@Ter data hub federates existing expertise centres and data and service centres serving the solid Earth science community.

ForM@Ter's main mission is to ease access to space and in-situ solid Earth and geodesy data and to work towards creating new products and services that add value to available datasets.

ForM@Ter's ambition is to create value by helping existing data centres with their data and product and management practices and supporting processes for data discovery and availability (FAIR principles, interoperability). It will spark new services in fields of investigation that are currently poorly represented or not represented at all. These new services will be integrated into data centres and included natively in European and international structures.

The data hub will consolidate the solid Earth science community by giving it access to the data and products it needs to conduct research. ForM@Ter is working within Data Terra to develop cross-cutting research on the Earth system, covering major cycles like the water and carbon cycles, climate, the living Earth and natural hazards. It has recently developed services for measuring ground deformation with radar and optical satellite imagery and is involved in the European Plate Observing System (EPOS) research infrastructure.



## 2.2 HISTORY: HOW FORM@TER CAME INTO BEING

From January 2012 to October 2013, CNES and INSU tasked a focus group with making recommendations for the data hubs' future evolution. The group submitted its report to CNES and INSU in 2014.

At the same time, in January 2012 Olivier Jamet accepted CNES and INSU's proposal to lead the study to set up a solid Earth data hub. The aim was to specify data of interest to the future solid Earth data hub and define how it would be implemented, with initial funding from CNES at the start of the study and support from CNRS/INSU and IGN. The study called on scientists from 20 French research laboratories and institutes.

End 2014, an initial governance structure was established for the project with a 12-member set-up committee and a 20-member scientific committee that convened for the first time on 9 January 2015. After consulting science communities, it was decided that ForM@Ter would start by focusing on the **figure, movements and deformation of the Earth**.

Michel Diamant accepted the role of project officer in January 2015 and an executive bureau was established early in 2019.

The agreement establishing the ForM@Ter solid Earth data hub was signed at the end of 2020 by 13 entities: BRGM, CNES, CNRS, IGN, IGP, IRD, Paris Observatory, OCA, SHOM, Clermont-Auvergne University, Grenoble-Alpes University, Strasbourg University and Toulouse University.

The data hub was officially created following the first meeting of its Steering Committee on 15 January 2021, with Michel Diamant as its first director.

### MAIN GOALS

ForM@Ter's main goal is to ease access to space and in-situ solid Earth data and to work towards creating new products and services that add value to available datasets. It meshes closely with the national and European landscapes, working closely with infrastructures already in place and in construction.

To this end, ForM@Ter's mission is to federate existing centres to serve the solid Earth science community.

ForM@Ter is built around a web portal providing access to satellite, in-situ and experiment solid Earth data. Its ambition is to add value, notably for data and services in fields where data centres currently do not exist or need to be developed, working in combination with European and international structures. It must neither disorganize nor replace identified centres already in place at solid-Earth domain RIs or national observation services (SNOs). It must exploit synergies with the other structures to consolidate the solid Earth science community by giving it access to the data and products it needs to conduct research.

To optimize teams' efforts and ensure the closest convergence with European infrastructures, ForM@Ter's technical developments are leveraging wherever possible developments conducted and/or operated by French players involved in these European infrastructures, in particular EPOS, the core of which will be co-hosted in France.

It is envisioned that the ForM@Ter data hub will ultimately fulfil four functions:

- A data and service portal function offering one-stop access and simplified use of satellite and in-situ data
- An expertise function supporting data users
- A technology transfer function for the solid Earth domain disseminating methodological innovations from research laboratories to the wider scientific community, and beyond to public stakeholders and businesses
- A support function for existing services, notably international services, provided by the national community

## 2.3 STRUCTURE AND GOVERNANCE

ForM@Ter's three components are:

- A coordination and distribution centre based at IPGP
- Data and services centres
- Scientific expertise consortia (CES)



ForM@ter is led by a **Steering Committee**, advised by a **Scientific Committee**. Decisions are executed by an **Executive Bureau**.

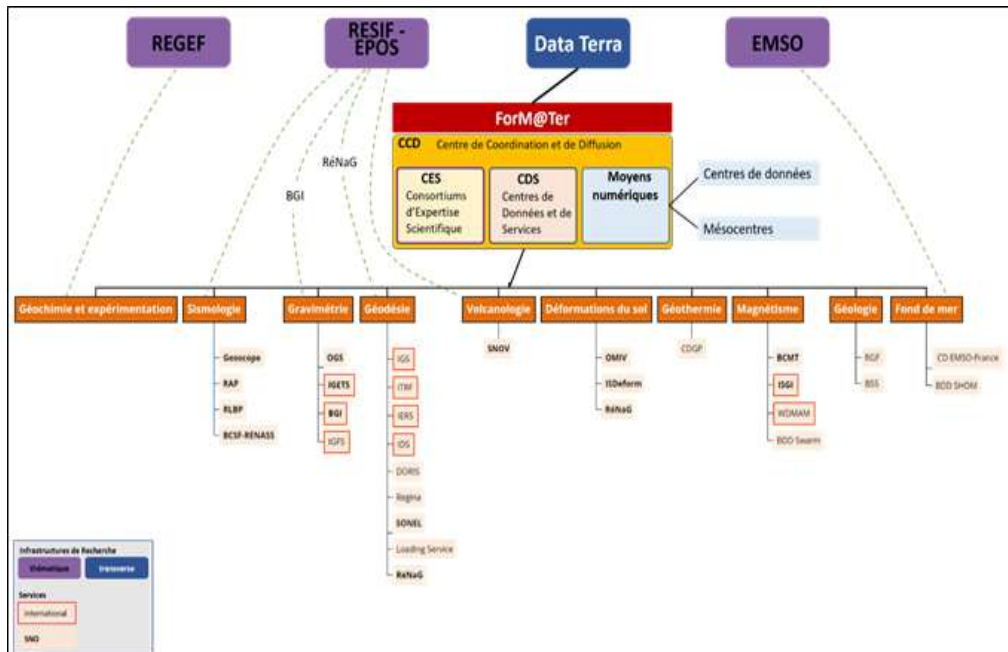
### 2.3.1 NATIONAL LANDSCAPE

ForM@Ter has found its place in a very rich national solid Earth science landscape, where certain disciplines are often already highly structured at both national and international level, either around an SNO or through international services like ISGI for magnetism and IERS for geodesy, or national research infrastructures like RESIF-EPOS for seismology and in part for gravimetry and geodesy, and REGEF for geochemistry.

In other solid Earth disciplines, the landscape is more fragmented, with widely distributed datasets and accompanying technical and scientific expertise.

In this context, a proposal is currently under study by the first Steering Committee to define DSCs for each of the major solid Earth disciplines. These DSCs would join existing services—SNOs, certified or otherwise—overseen by at least one party to the ForM@Ter agreement, contributing their own data and service distribution and/or archiving resources in the solid Earth domain.

The diagram below, in the process of being finalized, illustrates this proposal:

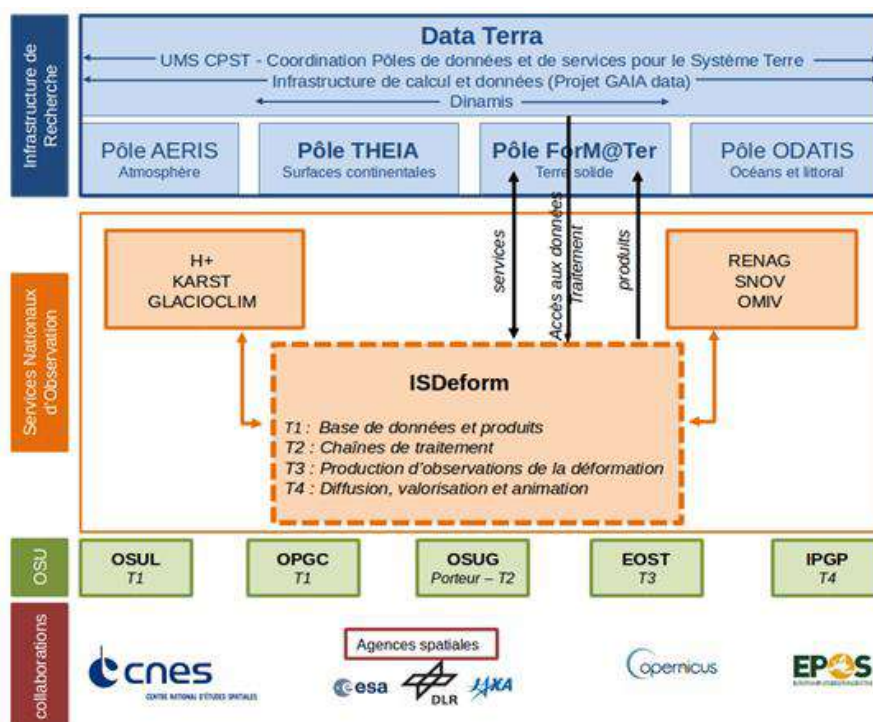


## SUPPORT FOR THE CREATION OF A NEW SNO: ISDEFORM

ForM@Ter has helped to put together and supported the proposal to create a new SNO in relation to the numerous satellite data processing services it has developed.

CNRS-INSU decided to certify ISDeform—for Imagerie Satellitaire pour comprendre les Déformations de la Terre\*—as a national observation service in February 2021.

\* Satellite Imagery to understand Earth's ground deformations



### 2.3.2 EUROPEAN LANDSCAPE

At European level, ForM@Ter is contributing alongside RESIF-EPOS and BRGM to the EPOS research infrastructure. It is notably a core player of the Satellite Data Thematic Core Service. ForM@Ter is also closely following Copernicus services and initiatives pursued by French teams in the solid Earth domain.

## 2.4 ACCESS TO DATA AND SERVICES

### 2.4.1 PORTAL

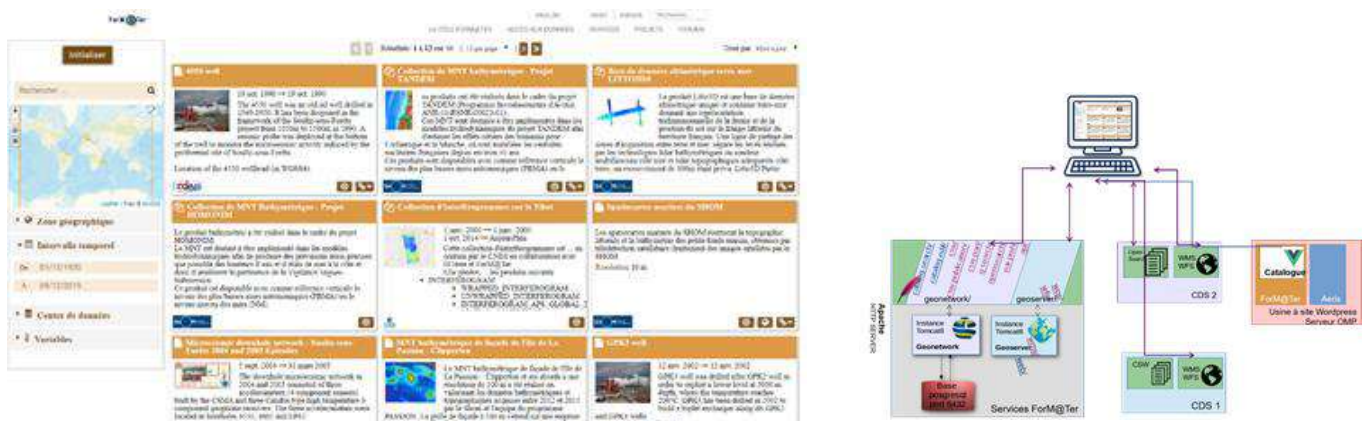
The first ForM@Ter website came on line end 2014. In the autumn of 2017, it was migrated to the Wordpress CMS administered by OMP, the Midi-Pyrenees Observatory. Tools, administration and the graphic charter are shared with AERIS. This charter will be adopted later by the other data centres.

ForM@Ter develops Web components to enhance certain functionalities and these components are also used by the other data centres. The portal is not only a communications tool but also a gateway to services developed by the data hub and to data via the meta-catalogue.

### 2.4.2 META-CATALOGUE

Development of the ForM@Ter meta-catalogue got underway in 2018 with:

- Choice of the development tool: Geonetwork
- Identification of existing thesauruses
- Initial integration/creation of DSC metadata files



### 2.4.3 UNIQUE AUTHENTICATION

In 2018, studies were launched to establish a unique authentication service. In 2020, a collaborative initiative pursued within Data Terra tested the Keycloak tool proposed by AERIS. The authentication service is now operational.



#### 2.4.4 VALUE-ADDED PRODUCTS

From end 2015 to 2020, ForM@Ter pursued a series of projects aimed at creating services for processing radar and optical satellite data. Most of these services will be operational in 2021. Flatsim and digital surface model (DSM) products will be available in 2021 via ForM@Ter's meta-catalogue. For the Flatsim service, a DOI was established at CNES in 2020 for distribution via the ForM@Ter portal.

##### Calculs à la demande

##### ■ GDM (Ground Deformation Monitoring)



##### GDM -SAR

Calcul d'interférogrammes; séries temp.

##### GDM -Optic

GDM-OPT-ETQ  
GDM-OPT-SLIDE  
GDM-OPT-ICE

Corrélation d'images; séries temp.



Principaux acteurs

##### ■ Production de MNS



Principaux acteurs

##### Calcul systématique

##### ■ Flatsim

(ForM@Ter large-scale multi-temporal Sentinel-1 interferometry processing chain in MUSCATE)

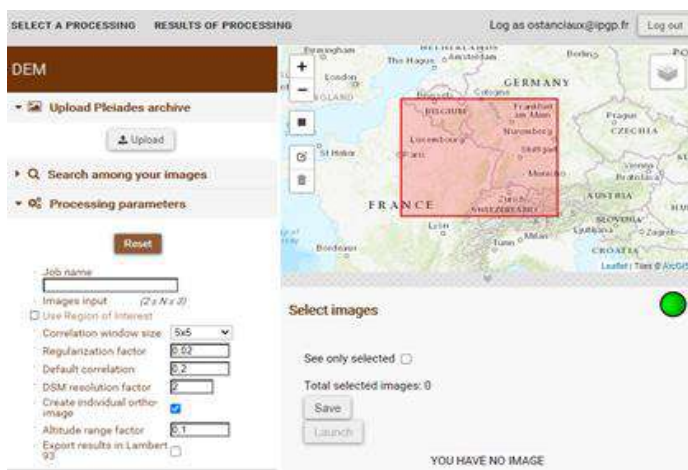


Calculs d'interférogrammes sur de grandes zones géographiques



Principaux acteurs

Radar and optical satellite data processing services developed by ForM@Ter



DSM service user interface (left) and GDM-SAR visualization interface (right)

### 2.4.5 CALL FOR IDEAS

A first call for ideas for the Flatsim service was issued in the spring of 2020. The plan is for such calls to be renewed annually:

- 13 responses were assessed technically and by a joint TOSCA TS-MDIS scientific committee
- 8 projects were selected (East and Northeast Tibet, Balkans, Turkey, Andes, Afar and Okavango Rifts, Mississippi River basin), covering a minimum surface area of 250,000 km<sup>2</sup>

### 2.4.6 DSC CERTIFICATION

**CEDRE** project led by Aude Chambodut for EPOS-France and ForM@Ter in response to ANR Flash open science call

*Towards Certification of solid Earth Data REpositories in France*

Cooperation between centres to build certification cases

WDS/Core Trust Seal certification

### 2.5 COMMUNICATION ACTIONS

- Participation at national and international congresses
- Organization of workshops (MDIS) every two years
- Webinars (2021)
- User reunions
- Newsletters
- Social media: Twitter account since end 2017; ResearchGate project
- Participation in Data Terra communications working group





### 3 ODATIS DATA HUB (OCEANS, WWW.ODATIS-OCEAN.FR)



#### 3.1 FOREWORD

The ODATIS ocean data hub was created in 2016 but draws on a heritage of more than 20 years acquired through the SALP altimetry and precise positioning department and the Coriolis in-situ operational oceanography centre. ODATIS encompasses nine data and service centres (two satellite centres and seven in-situ centres).

The full mean cost is €21.6m a year and the hub mobilizes 61.4 FTEs spread across 185 people from the 11 overseeing bodies, 167 of whom are permanent staff.

ODATIS's main mission is to provide data, products, software, tools and services in the domain of oceanographic research. It strives to provide and produce, with expert oversight, sets of qualified and described data designed for use in accordance with the very latest state of the art.

#### 3.2 HOW ODATIS CAME INTO BEING

The founding document of ODATIS was proposed by Fabienne Gaillard, Philippe Bertrand and Thierry Guinle in 2016. The current management team—Gilbert Maudire (Director), Sabine Schmidt (Chief Science Officer) and Gérald Dibarbourbe (Chief Technical Officer)—was then appointed in the autumn of 2017.

#### Les étapes réalisées

- Mars 2016: Présentation au comité de pilotage interpôles
- Mai 2016: Présentation au comité directeur du Pôle
- Juin 2016: Première réunion du conseil scientifique
- Juillet 2016: Ouverture du Web Odatis:

#### Automne 2016:

- Proposition de convention Odatis (basée sur la convention Aeris)
- Rédaction de la feuille de route 2017-2020 et plan de travail 2017.
- Présentation de la feuille de route au conseil scientifique (6 décembre)

#### Début 2017:

- Présentation de la feuille de route au Comité directeur

The priorities of ODATIS's roadmap are to:

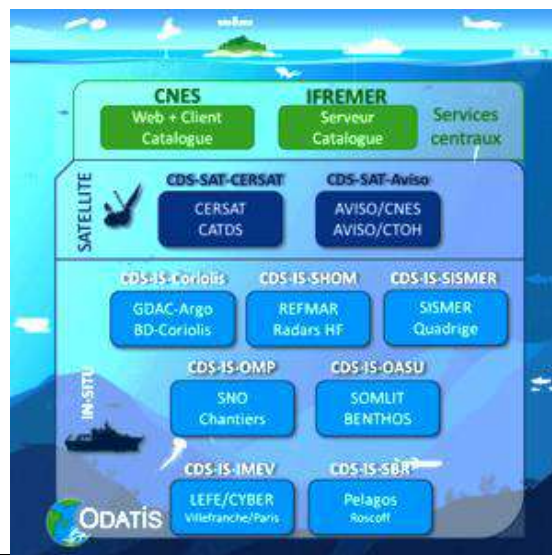
- Offer a synoptic picture of in-situ and satellite observations and related products
- Ease access, via a one-stop portal, to fully described and qualified data in accordance with current FAIR and TRUST standards, which implies long-term preservation of datasets and assuring that they are interoperable (spatially, temporally and in terms of disciplines)
- Promote combined use of data of different kinds (in-situ, satellite) and from different sources (operational networks, scientific experiments) in the field of oceanography research
- Provide IT tools and resources for searching, extracting and analysing data

ODATIS works to describe, quantify and understand the oceans from a broad perspective through the following research areas: ocean dynamics and thermodynamics, evolution of the oceans' physical and chemical properties, biogeochemical cycles, marine ecosystem mechanisms, ocean evolution and past links between ocean and climate (paleo-oceanography).

ODATIS also manages data covering topics specific to coastal regions, including estuaries and lagoons, with a particular focus on coastal morphodynamics, shorelines and sea level, pollution and eutrophication and littoral ecosystem changes.

### 3.3 DATA AND SERVICE CENTRE CHARTER

ODATIS encompasses nine data and service centres (two satellite centres and seven in-situ centres). These DSCs are tasked with managing data and routine generation of products (retention, long-term preservation, distribution, etc.), supported by the hub's founding bodies under a mandate (type of data, e.g. those of certified SNOs). They are interconnected through shared modules, i.e. a web portal (CNES) and data catalogue (Ifremer).



The heterogeneous nature of the DSCs, each with their own histories, prompted the need to describe all of the operational management and/or processing functions performed on their data. The definition reached is that a DSC assembles, harmonizes, maintains and makes available the dataset within its scope of activity.

ODATIS's Executive Bureau, comprising the hub's management team and other DSC directors, has worked to establish the Data Centre specification for Data Terra's ODATIS ocean hub. This specification covers functions, operating criteria, data management principles and interfaces. This document must therefore be viewed like the clauses of a Service Level Agreement (SLA) as set out in the ISO 20000 standard. The functions are operational functions to be performed for the agreed term of the partner organization(s), which undertake to provide the requisite human, financial and technical resources.

Not all Data Centres perform all functions. There are two types of centre: Assembling Centres, which chiefly manage the proximity interface with observation teams generating data, for example within OSUs; and Scientific Data Centres, which perform all functions, in particular organization of the long-term preservation of the dataset for which they have responsibility and which is considered the reference copy, and certified distribution and processing for users.



### 3.4 ACTIVITIES

#### 3.4.1 SERVICES FOR DATA PRODUCERS

The first version of the ODATIS website came on line in July 2016, followed by a second, more-complete version in December 2017. This website initially focused on data access. In May 2018, the Scientific Council recommended adding services for data producers and dedicated pages were subsequently included to support users with a range of specialized sections:

- Data management: this section includes an introduction to open science, FAIR and Core Trust Seal certification principles, guides on baselines and terminology, formats, attributes and template agreements to be used for oceanography data
- Data producer services: an interface guides and assists data producers, offering different possibilities for hosting and supplying DOIs, and support with formats and licences



### 3.4.2 TECHNICAL WORKSHOPS

Every year since 2017, ODATIS has organized two or three two-day technical workshops. These workshops bring together DSCs for presentations and panel sessions to define their technical direction and services, as well as hands-on practical cases. They meet the need for training and exchange to help DSCs fulfil their mission. Workshops are hosted by the different DSCs. The growing number of participants—from fewer than 20 in 2018 to more than 30 in 2020—shows the value of these workshops in helping to share best practices.

### 3.4.3 TOWARDS DSC CERTIFICATION (ANR COPILOTE)

Adoption of FAIR principles is very uneven across ODATIS's DSCs. Databases are very heterogeneous, mainly in-situ databases, and this remains a big obstacle to interoperability. To obtain support for bringing in-situ DSCs up to standard, the ODATIS management team submitted the ANR COPILOTE project (towards certification of the Data and Service Centres of the Ocean Data Cluster 2020-2022) in 2019 for the Flash open science call for research practices and open data. This effort is tapping into the experience of the partners already engaged in this process. The final aim is for all of ODATIS's DSCs to reach at least level 3—implementation phase—of the criteria required to achieve Core Trust Seal certification.

The objectives of this project started in April 2020 are to:

- Harmonize implementation of FAIR principles across all ODATIS Data Centres
- Improve the quality of data and services offered by the Data and Service Centres, in order to meet the requirements set out by the Research Data Alliance (RDA) for Core Trust Seal certification



Tasks are shown in the diagram below:



### 3.4.4 INVOLVEMENT IN EUROPEAN PROJECTS

The ODATIS ocean cluster is involved in several ongoing European projects:

- EOSC-Pillar, started in 2019, is based on the open science concept and on FAIR practices and data services to promote uptake across all oceanography data users communities.
- PHIDIAS, started in 2019, aims to develop and establish prototypes to exploit spatial and environmental Earth science data, leveraging high-performance computing (HPC) capabilities.
- Blue-Cloud Services is the result of a Horizon 2020 call for proposals and is the 'marine' strand of the European Open Science Cloud (EOSC). Launched in 2019, this project is pursuing a practical approach to exploit the potential of open science based on 'clouds' to implement a set of services geared towards gaining new insights into and managing the many aspects of ocean sustainability, through a series of five Blue-Cloud pilot demonstrators.
- ENVRI-FAIR is a Horizon 2020 project launched at the start of 2019 to connect the ENVRI cluster to EOSC. This cluster federates European research infrastructures focused on the environment and Earth system.

### 3.4.5 SCIENTIFIC EXPERTISE CONSORTIA (CES)

ODATIS is underpinned by longstanding data centres, in particular CDS-SAT-CERSAT and AVISO. The DSCs depend for the latest products and services on networks of scientists working in this field (e.g. altimetry, salinity, etc.).

However, to promote and exploit new processing methods and innovative products, the ODATIS ocean cluster has lent its support to the creation of Scientific Expertise Consortia (CES), covering in particular the areas of satellite, aerial or in-situ data on the ocean and its interfaces with the atmosphere, coasts and seafloor.

To meet the community's needs, three initial CESs have been established:

- Dissolved Oxygen (2019): this CES federates scientists working in the field of deoxygenation in the open ocean and coastal waters
- Flow Cytometry (2020): this CES federates scientists in France and from around the world working in the field of flow cytometry (FC)
- Ocean Colour: after a workshop in May 2019 bringing together this community, this CES was confirmed in December 2020

### 3.4.6 COMMUNITY OUTREACH

A pro-active approach is being pursued to raise the Ocean hub's profile through:

- Participation at national and international conferences, presentations of the cluster and its activities at programme committees and meetings
- A Tour de France of OSUs, with seminars and gatherings at OSUs hosting technical workshops
- Posting of news on the dedicated portal and other media
- Tutorials on YouTube, Twitter and through webinars





#### 4 THEIA DATA HUB (LAND SURFACES, WWW.THEIA-LAND.FR)



##### 4.1 FOREWORD

The Theia land surfaces data hub was created in 2012. It is structured around a Space Data Infrastructure (IDS) distributed across several stakeholders and a network of Scientific Expertise Centres (CES) and Theia regional mediation centres (ARTs) in mainland France and French overseas territories.

The full mean cost of Theia is €12.4m a year and it mobilizes 52.9 FTEs spread over 130 people from the 11 overseeing bodies, 112 of whom are permanent staff.

Theia's main mission is to foster uptake by the scientific community, and more broadly by public stakeholders and local authorities, of satellite, aerial and in-situ data on land surfaces. It provides national public bodies, the national and international scientific communities and private stakeholders with in-situ data and value-added products derived from satellite remote sensing, leveraging interoperable data and services infrastructures—currently four IDSs—and computing capacity (complementing that of the European Copernicus programme).

Theia is also seeking to structure the national scientific community, pool point data, images, processing and science expertise, and to raise the profile of national efforts on the international stage. It has formed Scientific Expertise Centres (CES)—currently 24—that share tools and processing methods with the community tailored to the various fields of land surfaces research. These CESs' teams are pursuing research and development work on innovative methods aimed at arriving at value-added land surfaces satellite data products. The third component of Theia are its regional mediation centres (ARTs)—currently nine—whose main mission is to federate and lead users—scientists and public and/or private stakeholders—at regional level, meshing with existing networks, particularly those of CRIGE regional geographic information resource centres.

In short, Theia is an innovation ecosystem supporting research, government action and economic development in the fields of the environment, hydro- and agro-systems and land planning, notably through the provision of data in France, Europe and South nations.

#### 4.2 SUMMARY OF ACTIVITIES SINCE 2017

- Some 24 CESs are now up and running, covering a broad spectrum of land surface domains including surface reflectance, biophysical vegetation parameters, land occupancy, snow cover, lake and river heights, quality of surface waters, soil moisture and land take.
- Production and distribution of Sentinel-2 data with the MAJA algorithm and from a new ESA ‘tile’ format, and of soil moisture data for numerous sites of interest in France and abroad, snow cover data, an annual land cover map of France, and lake and river height data (Hydroweb).



*Coverage of Sentinel-2 reflectance products from Theia*



*Coverage of snow cover extent products from Theia*



*Theia land cover map*

- Strong contribution from the Theia ART network (seven ARTs in mainland France and one in New Caledonia and South nations).

- Efforts to structure in-situ data got underway in 2017, with the initial aim of building a metadata portal to enable discovery of in-situ land surfaces datasets. The next step could be to offer data access and a data management service for producers who want it. Development of a Theia/OZCAR information system is supported by CNRS-INSU, IRD and ANR FairTOIS: Implementing FAIR principles in the Theia/OZCAR information system (2020-2022). The Theia/OZCAR Data Management Plan (ANR model) is available on line since December 2020.
- Development of the Theia/OZCAR information system: the volume of in-situ data is very low compared to that of satellite data, but the complexity of this task stems from the quality and variety of these data. The principles underlying the construction of this system, the terminology used and solutions proposed have been published in an article titled “Building the Information System of the French Critical Zone Observatories network: Theia/OZCAR-IS” in the “Data: opportunities and barriers” special issue of the Hydrological Sciences Journal (Braud et al., 2020). The system has been presented to the EGU and to European partners of the eLTER research infrastructure under construction.
- Beta version of the Theia/OZCAR data portal on line since January 2020 (7 out of 21 observatories accessible). A virtual workshop bringing together 50 participants was organized for the OZCAR 2021 RI Days, giving users the chance to provide critical feedback on existing requests and propose features to be added. This workshop also helped to define objects of interest that could be added to the portal’s query criteria.



*In-situ land surfaces data portal (<https://in-situ.Theia-land.fr/>)*

### 4.3 COMMUNICATION ACTIONS

- A yearly seminar in France with some 200 participants (scientists and public and private stakeholders). Over the least four years, two Theia seminars have been held abroad, one in Iran and one in India. The COVID-19 crisis has prevented other planned seminars—in Morocco and Uzbekistan—from going ahead.



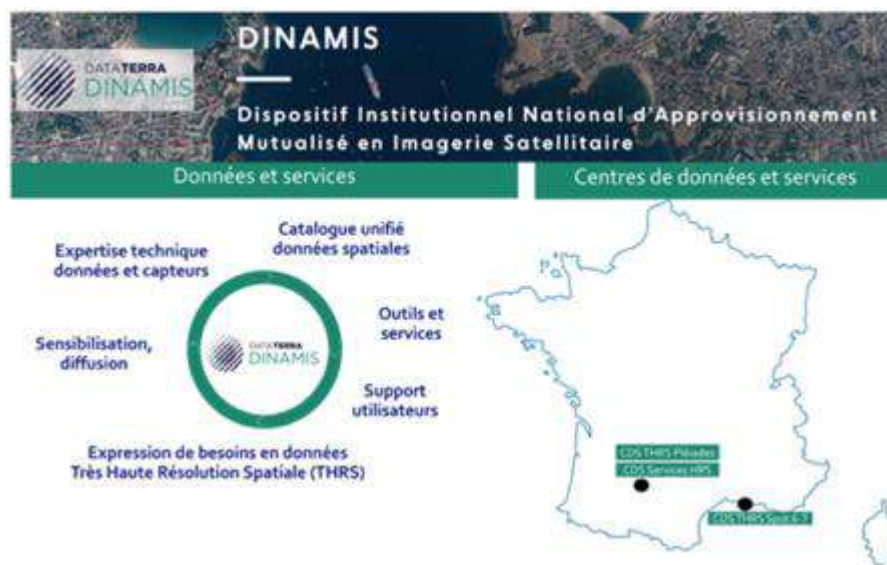
*Theia seminar in Iran (2018)*

- Two bulletins and four newsletters a year
- A website showcasing national expertise and the hub's accomplishments in different domains

The Theia scientific community is present in many European projects and certain Theia products—snow cover, lake and river heights—have even achieved a European destiny.



## 5 DINAMIS NATIONAL SPACE DATA STRUCTURE



### 5.1 FOREWORD

Various structures have been set up in France since 2000 to foster uptake of satellite imagery. These efforts to pool resources have enabled very significant advances in access to and use of such data. As of 2019, the GEOSUD/Theia structure had 530 subscribers (research laboratories, government departments, local authorities, associations and regional platforms). Its archives today hold more than 13,000 very-high-resolution images covering an area of more than 11 million sq.km.

To secure and further develop access to satellite imagery, it was decided to set up a one-stop cross-cutting structure called DINAMIS (for Dispositif Institutionnel National d'Approvisionnement Mutualisé en Imagerie Satellitaire), of which there is currently no equivalent.

Focused on satellite imagery access services, this structure will support users in France and, under certain conditions, from abroad in choosing and accessing archive imagery or tasking new acquisitions.

The structure offers access for free or at special rates made possible by pooling of resources to a bouquet of imagery at complementary resolutions, through provision of very-high-spatial-resolution (VHR) imagery as the core of DINAMIS and high-spatial-resolution imagery relays. Two modes of procuring commercial SPOT 6-7 and Pleiades VHR imagery are available to scientific stakeholders and local authorities: (1) commercial archive imagery and (2) new tasked imagery. All of this imagery is made available through the DINAMIS catalogue, allowing users to search and download images of shared interest in standardized formats. They can also express their requirements in fine detail, task acquisitions and add to the catalogue thanks to a dedicated application.

DINAMIS licences for Pleiades and SPOT 6-7 products distributed by the structure are granted to:

- All French public stakeholders
- All international scientific stakeholders
- Private stakeholders leading R&D projects or service providers for government procurement

DINAMIS has a two-pronged procurement policy for **archive and/or tasked very-high-spatial-resolution imagery**:

- Acquisition and pooling of reference SPOT 6-7 and Pleiades datasets for regions of shared interest, available to all
- Acquisition and pooling of ad-hoc datasets as they arrive and on demand through a permanent call for projects, available to all

There are procurement quotas to manage satellite capacity that is pre-funded yearly by the structure; above these quotas, pre-negotiated rates allow users to co-fund the imagery they need.

This policy governed by a contractual agreement with the supplier enables acquisition of imagery not only of French territory, but also all over the globe. Data procurement and all operations conducted by the structure are funded and led by a consortium of partners (CNES, IGN, CNRS, INRAE, IRD and CIRAD).

In addition to VHR imagery, the structure relays via its catalogue Sentinel-2 and Spot World Heritage (SPOT 1-5) imagery, as well as RapidEye data of French territories. Details of eligibility and access to products can be viewed on the DINAMIS website.



DINAMIS website, catalogue and dedicated application for expressing requirements



## 5.2 2018 - 2020 REVIEW

As of 2020, the GEOSUD/Theia structure had more than 560 subscribers—research laboratories, government departments, local authorities, associations and regional platforms—and more than 1,200 user accounts. Its archives today hold more than 90,000 very-high-resolution images. In 2020, more than 1 million sq.km. were added to the catalogue.

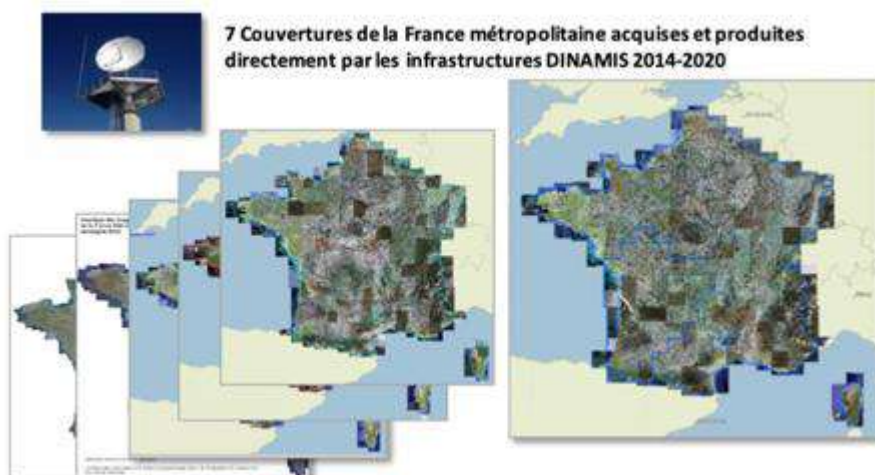
2018 was marked by DINAMIS coming on stream, with the launch of the dedicated application for expressing requirements. This first fundamental step enabled the procedures for procurement of Pleiades and SPOT 6-7 imagery to be harmonized. Provision of Pleiades imagery got underway as soon as the constellation was launched under the delegated public service agreement headed by CNES, while that of SPOT 6-7 imagery kicked off in 2014, organized by Equipex GEOSUD.

During the course of that year, more than 200 Pleiades requests were fulfilled, three-quarters of them from scientists and the remaining quarter from regional public stakeholders. For SPOT 6-7, 60 requests—90% from scientists—were handled by Equipex, on top of the yearly national coverage.

2019 was marked by the start of work on the catalogue, the purpose of which is to ensure smooth and easy access to imagery procured by DINAMIS. This tool is a meta-catalogue of data hosted by:

- CNES and IGN's Pleiades infrastructure
- Geosud's SPOT 6-7 infrastructure
- Airbus Defence & Space's infrastructure
- SPOT World Heritage (SWH) infrastructure (Theia and shortly CNES)
- PEPS Sentinel Product Exploitation Platform

During the course of that year, some 300 requests for Pleiades imagery and 80 requests for SPOT 6-7 imagery were received, generating more than 2,000 downloads corresponding to around 15 million sq.km.



In 2020, the catalogue came on line and Pleiades and SPOT 6-7 imagery already acquired was added to it; the website was laid out and put on line alongside the Data Terra architecture. The quantity of imagery acquired and delivered was comparable to 2019, generating more than 2,500 downloads corresponding to around 8 million sq.km.

### 5.3 EFFECT OF POOLING RESOURCES AND ECONOMIES OF SCALE

Pooling of resources by DINAMIS is fulfilling a key role: of all products downloaded, nearly half are archive data acquired in previous years for tracking and monitoring purposes. The same VHR imagery product may be downloaded up to ten times by different users.

Economies of scale achieved by DINAMIS are also significant: given the commercial value of the imagery it acquires, users could not have gained access to data for their research work or for land planning studies without it. Surveys conducted for the SPOT 6-7 component show that more than €100 million would have been needed for regular acquisition, since 2014, of coverages of mainland France, ad-hoc acquisitions and individual reuse of each image acquired for licence-holding communities. For specific applications requiring large volumes of imagery, like mapping of clear-cuts by INRAE, the survey shows that each euro invested in DINAMIS saved €60 in induced direct and indirect costs.

These savings are made possible by the discounts on Pleiades products granted under the delegated public service agreement, and by the fact that SPOT 6-7 imagery is procured by purchasing telemetry with direct reception of the satellite signal by the DINAMIS reception and image production infrastructure.

#### DINAMIS at a glance

