French National Tripartite Event
Paris, 11-12/09/2025

DATA TERRA

EOSC France 2025

Événement National Tripartite

11-12 sept. 2025 Paris (France)

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

Direction générale
de la recherche
et de l'innovation

CNrs

EOSC Federation candidate Data Terra node experience

Frédéric Huynh, Director, Data-Terra Research Infrastructure
Jean-Pierre Vilotte, CNRS/INSU Scientific Deputy of High Performance Computing and Data
Allessandro Rizzo, EOSC Node Data Terra coordinator

september, 11th, 2025

01 National **E-infrastructure** for the **Earth system science**, Climate, Environment and Biodiversity of the French national **RIs road-map** receiving long-term support from the French Ministry of Research and **34 Research organisations (CNRS, CNES, INRAE, IFREMER, IRD, Météo-France, MNHN, BRGM, IGN, CEA, Sorbonne Univ., ...)**

02 D-T provides **access and treatments to multi-source and multi-component FAIR data and services.** It is organised into **thematic Data Hubs and transversal services,** also deploying national instances serving the implementation of **RIs at the European level (ESFRI).**

03 D-T is **reference Data and Services E-Infrastructure** for major national research projects, and its activities are supported by nationally labelled **observation services** (SNO)
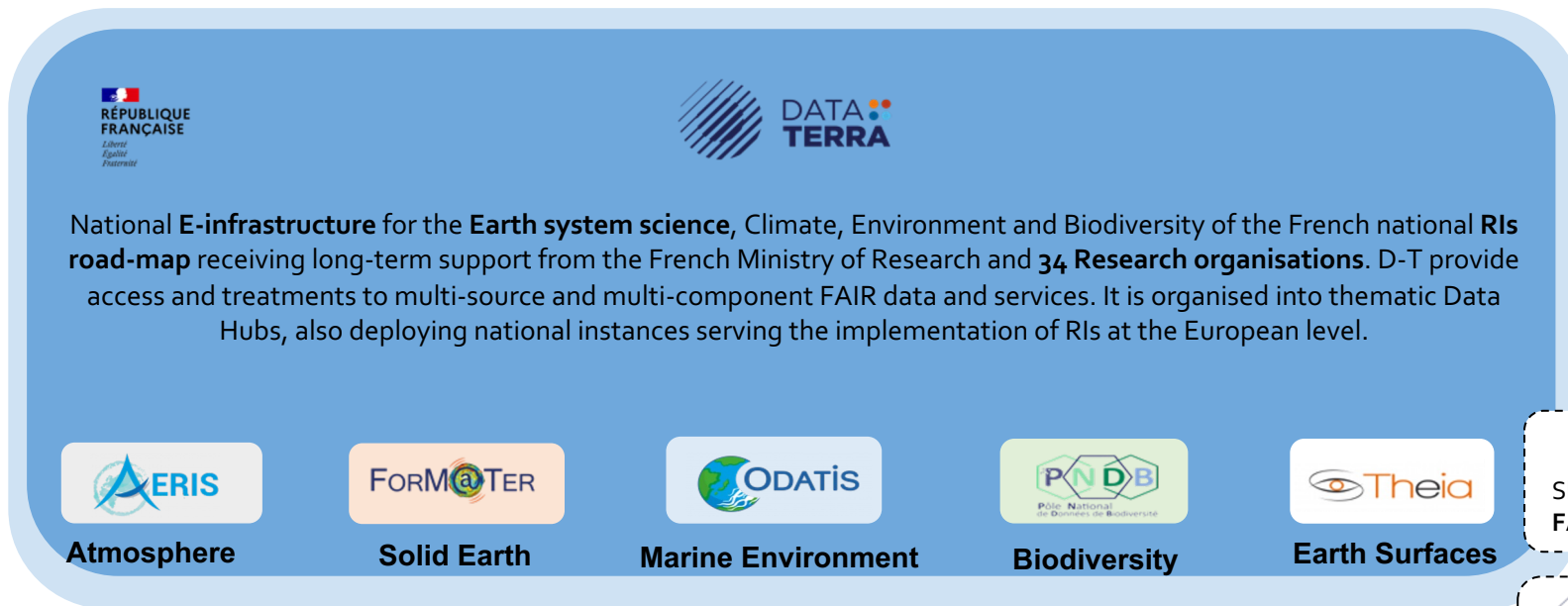
04 The French **CNRS is the legal entity for the Data-Terra EOSC node**, but is supported by all the members of the D-T steering committee.
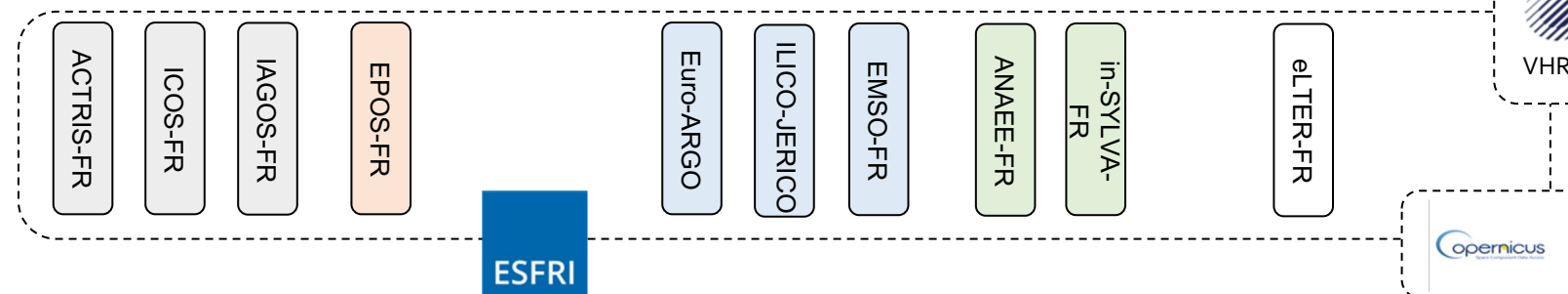
Data-Terra: Organisational Capacities

- **50 M€/year**
- 34 National Research Organisations and Universities
- 30 Observation Data and Services Centres (CDOS)
- 32 Scientific expertise consortium (CES)
- 500 scientists, engineers and technicians

National **E-infrastructure** for the **Earth system science**, Climate, Environment and Biodiversity of the French national **RIs road-map** receiving long-term support from the French Ministry of Research and **34 Research organisations**. D-T provide access and treatments to multi-source and multi-component FAIR data and services. It is organised into thematic Data Hubs, also deploying national instances serving the implementation of RIs at the European level.

AERIS — Atmosphere
ForM@Ter — Solid Earth
Odatis — Marine Environment
PNDB — Biodiversity
Theia — Earth Surfaces

DATATERRA INTER-PÔLES — Sharing expertise for **FAIR data** and services

DATATERRA DINAMIS — VHR **satellite imagery**

✓ Reference Data and services E-Infrastructure for major national research projects
✓ Federation layer for French contributions to Environment ESFRIs and ERICs that would otherwise be weakly coordinated
✓ Involved in several EU EOSC-related projects (coordination, partner)

✓ Supported by nationally labelled observation services (SNOs)
✓ Articulated nationally around the Observatoires des Science de l'Univers (OSUs) which include the Universities

ACTRIS-FR | ICOS-FR | IAGOS-FR | EPOS-FR | ESFRI | Euro-ARGO | ILICO-JERICO | EMSO-FR | ANAEE-FR | in-SYLVA-FR | eLTER-FR | Copernicus

3

Technical architecture blueprints aligned for offering services to EOSC

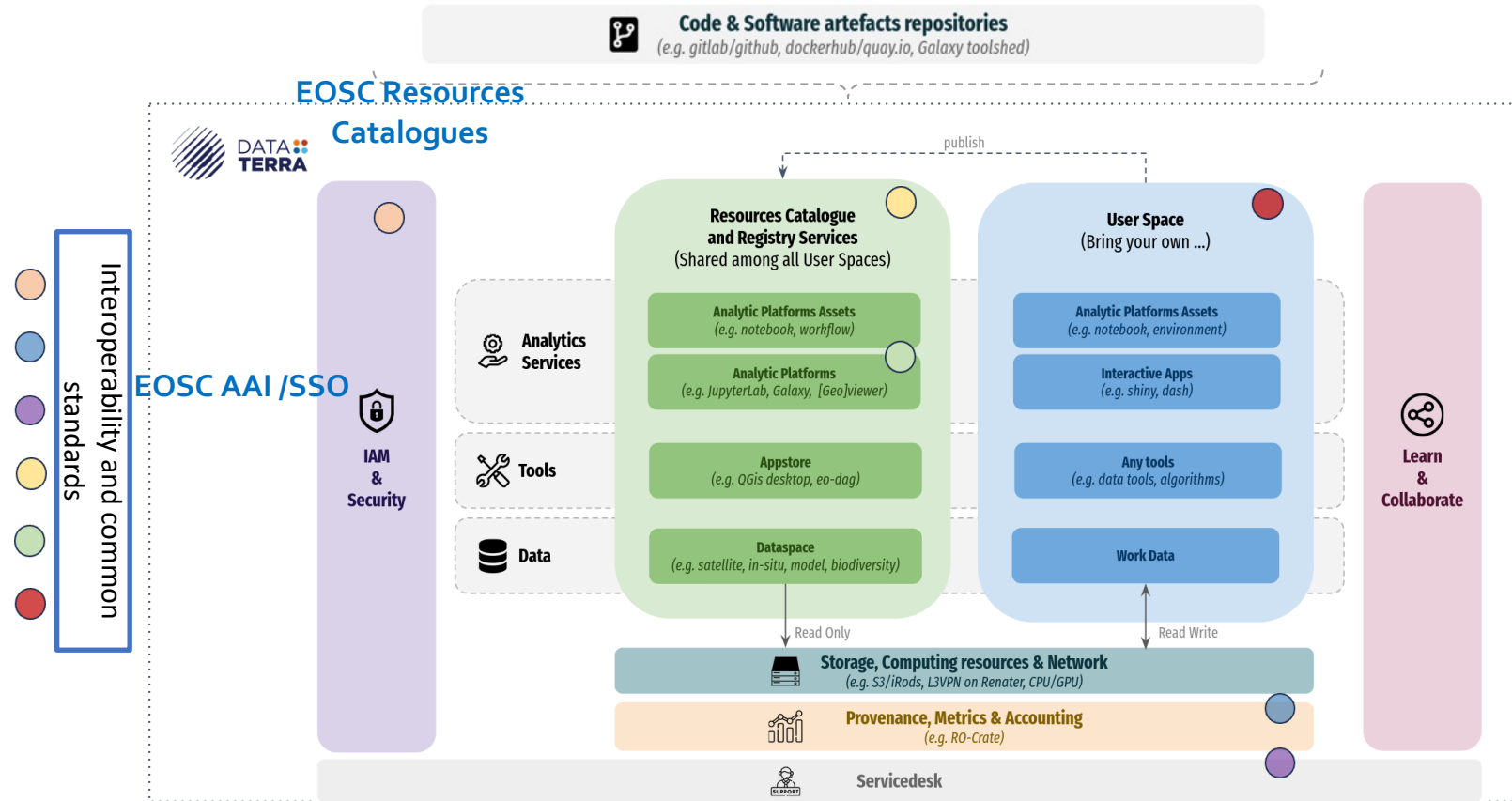**French national digital infrastructures**
- Renater (Geant);
- France Grille (EGI), & Mesonet > GENCI / EuroHPC: IDRIS, CINES, TGCC;
- National & regional labelled data centres and meso-centres.

EOSC Federation
- D-T data and services accessible through the EOSC EU node;
- Services interoperability with thematic cluster nodes and related national nodes.

Core services
- Distributed data storage and management;
- Large data transfer (files, objects);
- User spaces (interactive notebooks, virtual machine, container images);
- HPC/Cloud computing services
- Federated AAI.

Code & Software artefacts repositories
(e.g. gitlab/github, dockerhub/quay.io, Galaxy toolshed)

EOSC Resources Catalogues

DATA TERRA

publish

EOSC AAI /SSO

Interoperability and common standards

IAM & Security

Analytics Services

Tools

Data

Resources Catalogue and Registry Services
(Shared among all User Spaces)

- Analytic Platforms Assets *(e.g. notebook, workflow)*
- Analytic Platforms *(e.g. JupyterLab, Galaxy, [Geo]viewer)*
- Appstore *(e.g. QGis desktop, eo-dag)*
- Dataspace *(e.g. satellite, in-situ, model, biodiversity)*

User Space *(Bring your own ...)*

- Analytic Platforms Assets *(e.g. notebook, environment)*
- Interactive Apps *(e.g. shiny, dash)*
- Any tools *(e.g. data tools, algorithms)*
- Work Data

Learn & Collaborate

Read Only — Read Write

Storage, Computing resources & Network
(e.g. S3/iRods, L3VPN on Renater, CPU/GPU)

Provenance, Metrics & Accounting
(e.g. RO-Crate)

Servicedesk

Development phase — June 2021 — May 2027 — Exploitation phase — May 2029

GAIA Data

eosc | FAIR-EASE
Sept 2022 — Aug.2025

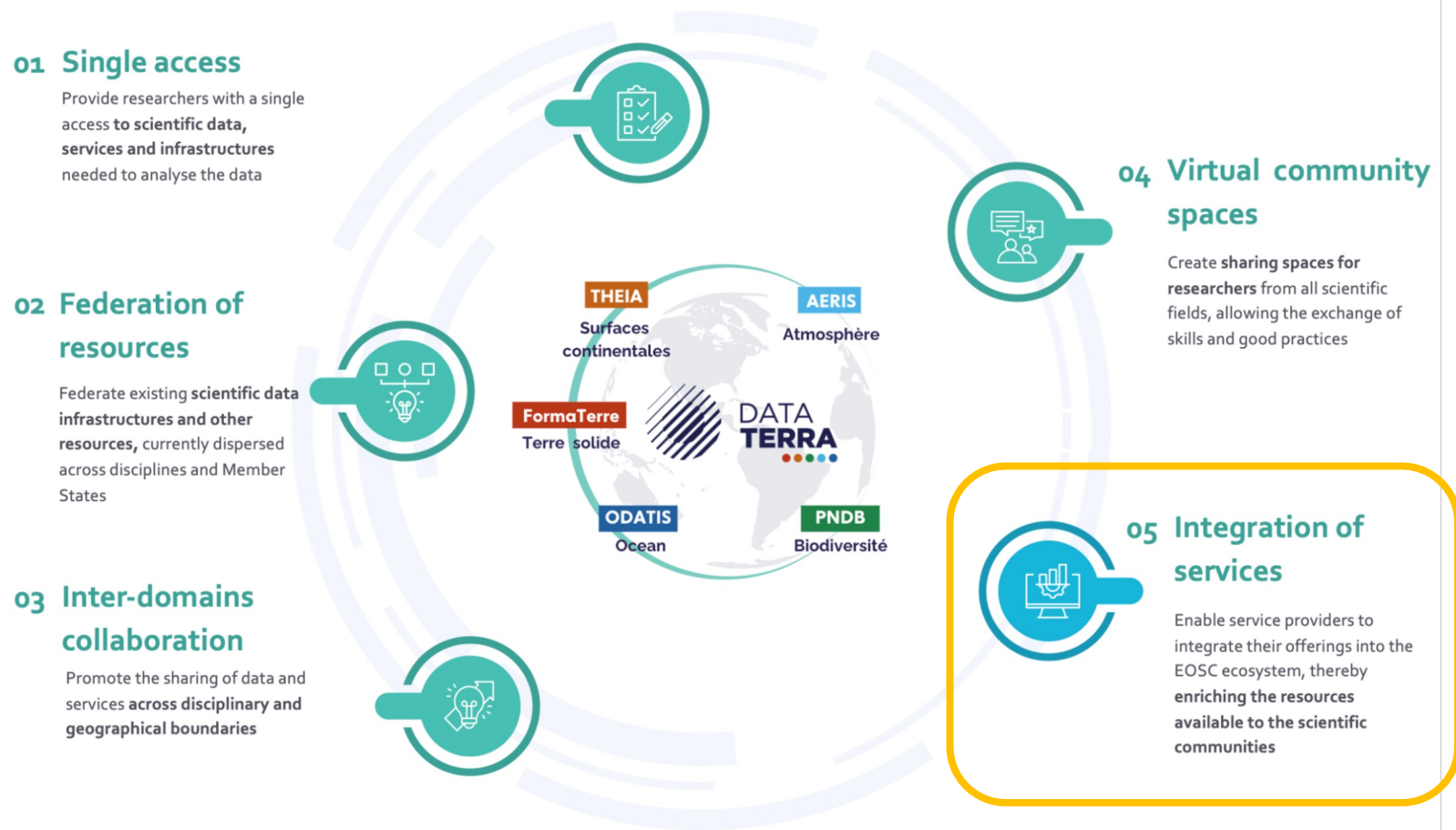Gaia data timeline allows to co-develop with EOSC EU node and third party (EOSC Federation)

Data Terra services designed to support EOSC EU Node and third party core services

EOSC Federation (EOSC EU node and other nodes)

- Identity Management
- Monitoring and Accounting
- Service Management System
- Resources Catalogues and Registry Services
- Application Workflow Management
- User Space

CNRS   DATA TERRA

Data-Terra: a federated earth system sciences "e-infrastructure EOSC node"

01 Single access

Provide researchers with a single access **to scientific data, services and infrastructures** needed to analyse the data

02 Federation of resources

Federate existing **scientific data infrastructures and other resources,** currently dispersed across disciplines and Member States

03 Inter-domains collaboration

Promote the sharing of data and services **across disciplinary and geographical boundaries**

04 Virtual community spaces

Create **sharing spaces for researchers** from all scientific fields, allowing the exchange of skills and good practices

05 Integration of services

Enable service providers to integrate their offerings into the EOSC ecosystem, thereby **enriching the resources available to the scientific communities**

THEIA — Surfaces continentales

AERIS — Atmosphère

FormaTerre — Terre solide

DATA TERRA

ODATIS — Ocean

PNDB — Biodiversité

Data-Terra key functions to the EOSC ecosystem

A **seamless access** to high-quality, trusted FAIR and AI-ready multi-domain and multi-modal (space, airborne, in-situ) data for the Earth, climate, environment and biodiversity system with rich metadata, semantic interoperability and provenance information

Cross-domain data analysis workflows addressing emerging urgent multidisciplinary research challenges in relation to global changes, adaptation, extreme events characterisation and societal impacts enhancing the linkages with other Data-Spaces and Data-hubs at the European and global scales

A **federated layer** towards the environment-oriented ESFRI/ERICs in coordination with other thematic and national nodes of the EOSC Federation, as well as with European organisations and programs such as Copernicus, DestinE, ESA, ECMWF, and EUMETSAT

DATA TERRA

Use-Cases and external dependencies

Use-Cases aim to **demonstrate the value contribution to the EOSC Federation** through enrolling and onboarding capabilities towards the Federation, namely the EOSC EU Node and other nodes. The UCs have already reached a technical maturity (TRL) which makes them **suitable for deployment in an operational environment**.

UC1: Datalab for the Earth system sciences

Provide a robust and scalable computational framework to support Earth System sciences within the EOSC Federation

eosc | FAIR-EASE

August 2025

- Test proof the use of Galaxy for Earth System sciences
- Create of a dedicated subdomain on Galaxy Europe
- Populate the subdomain with tools related to the pilots (e.g. QGis desktop, ODV, STAC client)
- Build-up analytic workflows according to the best practices and share on Workflowhub
- Write fitting tutorials to guide researchers and users
- Improve Galaxy itself (e.g. RO-Crate, interactives tools, OGC:WPS interface)

GAIA Data

May 2027
Operational phase

- Bring entirely new earth use-cases virtual lab (e.g. Jupyter, Galaxy)
- Set up a Earth system dedicated subdomain on Galaxy France in collaboration with IFB
- Plug Galaxy France to new distributed resources

eosc

March 2026

- Develop new tools following the users demand
- Improve interoperability layer to interface galaxy with all relevant earth sciences services (Earth portal and new services…)
- implement trainings, webinars, courses, dedicated support on how to use Galaxy

- Enhance linkages with other European Infrastructures
- Collaborate with other projects and nodes (Blue-Cloud, Elixir, NFDI) for improving the resources catalogue
- Connect the user environment with other EOSC services
- Implement tutorial and training sessions for users

DATA TERRA

UC2: Volcano Space Observatory

Multi-domains interactive web platform dedicated to near real time and retrospective tracking and analysis of volcanic events and impacts

eosc | FAIR-EASE → **August 2025**

- Integrate online tools in a unique interactive platform
- Scale up computing infrastructure for EoSAR processes
- Use Copernicus DataSpace S3 bucket (subsetting via HTTP Range)
- Open API access for machine to machine computation calls
- Evaluate integration of marine data (ARGO, etc) in the platform
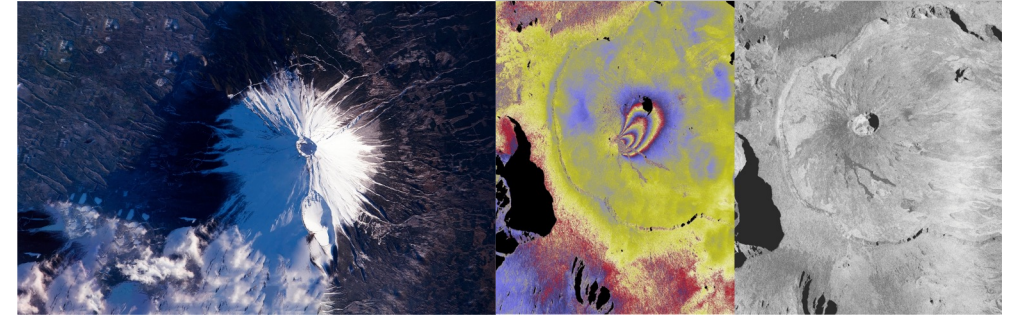
GAIA Data → **May 2027**
Operational phase

- Promote object storage (S3) and ARCO formats (e.g. Zarr, COG, GeoParquet)
- Promote new OGC services (e.g. STAC, OpenEO, API:processes)
- Add ability to run conterairized processes on French national grid
- Evaluate data access to CNES GEODES for inSAR data
- Test support of Argo data in the platform

eosc → **March 2026** - - ->
┌──┐
│ - Full support of marine data (Argo) in the platform │
│ - Set up daily alert mechanisms   │
└──┘

- Faster dedicated access to Copernicus Datasets
- Direct access to Argo datasets trough Blue-Cloud node
- Add ability to run conterairized processes on the EOSC Federation
- Evaluate and use european digital infrastructures such as Copernicus, Eumetsat

DATA TERRA

Added value and external dependencies

Use cases	Contributions & Value to Users	External Dependencies

Galaxy, a datalab for the Earth system sciences	• **Multi-node scientific workflows**: connect the user environment across different EOSC nodes.
	• **Data Access & Management**: connect the user environment with EOSC services, providing better access to data via APIs, S3 buckets, and direct links.
	• **Collaboration & Sharing**: facilitate interdisciplinary collaboration by improving Galaxy's interoperability and connecting Galaxy users across different European research infrastructures.
	• **Training & User Support**: tutorials, webinars, and training sessions.
		• Successful enrolling of EOSC EU node services and standards (i.e. federating capabilities).
		• Successful onboarding and enrolling with EOSC nodes like Blue-Cloud, NFDI, Elixir:
		  • Access to additional data,
		  • Integrate resources from other EOSC Nodes services,
		  • Run processes on EOSC Nodes resources.

Volcano Space Observatory	• **Data Access & Management**: evaluate faster access to Copernicus datasets; evaluate data processing on other European Federated infrastructures; and evaluate access and processing on european digital infrastructures such as Copernicus, Eumetsat.
	• **Multi-node scientific workflows**: access to marine data to be integrated into the platform.
		• Successful enrolling of EOSC EU node services and standards (i.e. federating capabilities).
		• Successful onboarding and enrolling with EOSC nodes like Blue-Cloud:
		  • Access to additional data,
		  • Integrate resources from other EOSC Nodes services,
		  • Run processes on EOSC Nodes resources.

DATA TERRA

Data-Terra

Expectations:

Data Terra a national and thematic node integrated to the EOSC Federation in order to:

- ✓ provide access to FAIR high-quality, trusted and AI-ready multidomain and multitype (space, airborne, in-situ) data for Earth, climate, environment and biodiversity system;
- ✓ enable cross-domain data analysis workflow to address emerging urgent multidisciplinary research challenges in relation to global changes, adaptation, extreme events characterisation and societal impacts;
- ✓ provide national resources and services to support HPC/HPDA/AI workflows across a continuum of HPC/Cloud/Data infrastructures through the EU node;

Users perspective:

- ✓ increase the impact and sustainability of the results of INFRA-EOSC research and innovation projects (i.e. FAIR-EASE, FAIR-IMPACT, …), through onboarding and integration in the Data Terra national node (tools, services, standards, guidelines);
- ✓ Improve multi-domain data discovery, interoperability and cross analysis;
- ✓ Accelerate open science data products workflows.

Coordination Efforts:

- ✓ boost collaboration and coordination with other national thematic initiatives in EU member states (i.e. Italy-CNR, Germany-NFDI4Earth, Germany-NFDI4Biodiversity) and in other associated countries (i.e. UK);
- ✓ contribute to the coordination among thematic ESFRI Eric and to the emergence of EU thematic cluster for the Earth, climate, environment and biodiversity system enabling interdisciplinary research;
- ✓ leverage national digital infrastructures capacities (France Grilles/EGI, MesoNet, National Data Centres, GENCI/EuroHPC);
- ✓ contribute to deploy and support the EOSC EU Node and third parties services;
- ✓ implement services compliant with technical interfaces defined in the EOSC Interoperability Framework (IF) to connect with other nodes.

Challenges:

- ✓ adapt the business model for providing services in a sustainable way (a specific activity is already ongoing and supported by the French Ministry of Higher Education and Research);
- ✓ size properly the operational means on the medium and long run (after 2025-2026).

Some insights from our experience

Challenges

Onboarding resources from third parties;

Taking over results from INFRAEOSC projects and stakeholders engagement;

Alignment with common shared standards;

Multi-node experiences for enabling the Federation;

Making the Federation as normative as enough (policy harmonisation); and more generally… time & Resources in the short term!

Lessons learned

1-to-1 meetings with the EOSC EU Node to progress on technical federation;

Targeted meetings with other nodes (multi-node or even single-node) in order to put in evidence potential synergies and common Use-cases;

An internal governance structure and organisation to be adapted to face new challenges.