eosc | FAIR-EASE
Building Interoperable Earth Science & Environmental Services

Integrating a community

Earth sciences use cases

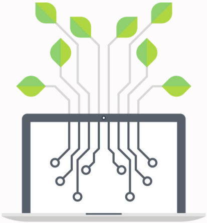
Samuel Keuchkerian - (CNRS)

Marie Jossé - Data Terra (CNRS)

FAIR-EASE has received funding from the European Union's Horizon Europe Framework Programme (HORIZON) - under grant agreement No. 101058785.

FAIR-EASE

Building an interdomain digital architecture for distributed and integrated use of environmental data



FAIR-EASE Data Discovery and Access Interdisciplinary Service

FAIR-EASE Virtual environments

FAIR-EASE Earth sciences use cases

5 pilots for an earth system model

Coastal Water Dynamics: focuses on the coastal marine environment near river estuaries, where important processes take place.

Earth Critical Zone: monitors land and soil degradation.

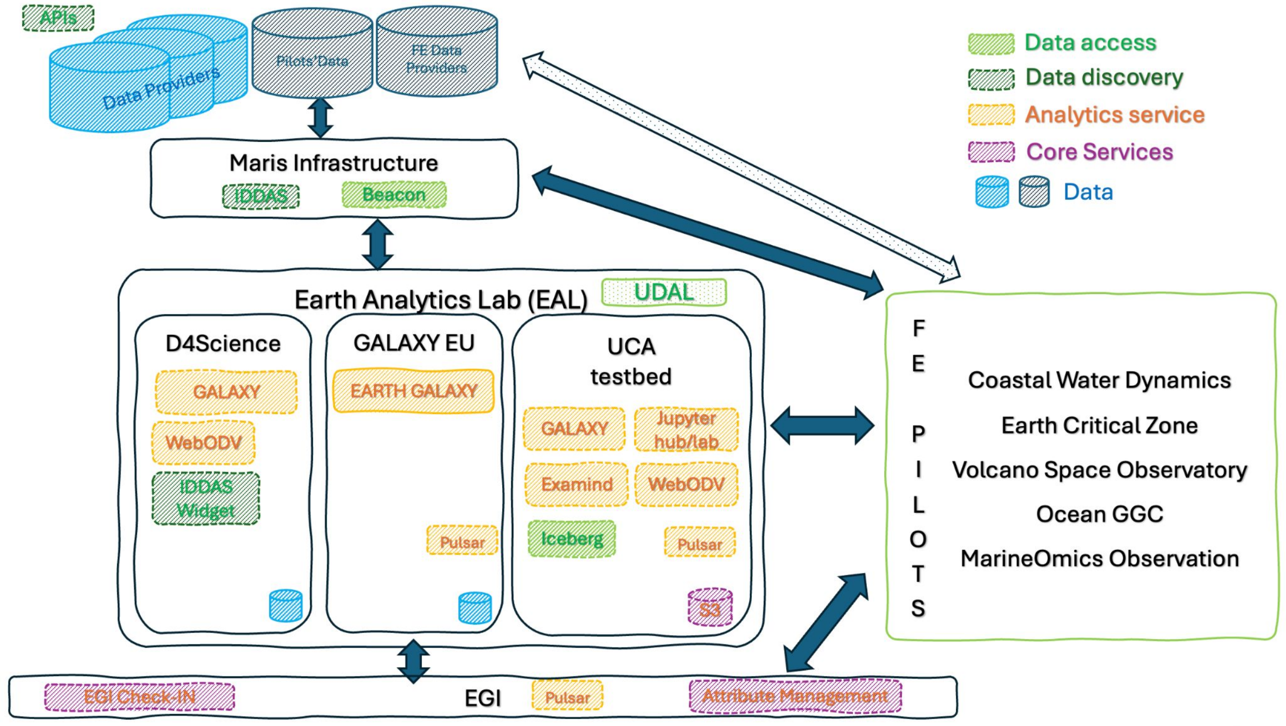
Volcano Space Observatory: monitors global volcanic activity, allowing the focus on any major volcanic eruption worldwide

Ocean Bio-Geochemical Observations: addresses fundamental scientific questions regarding the health of marine ecosystems (e.g. ocean acidification, ...) and needs for ocean resource management.

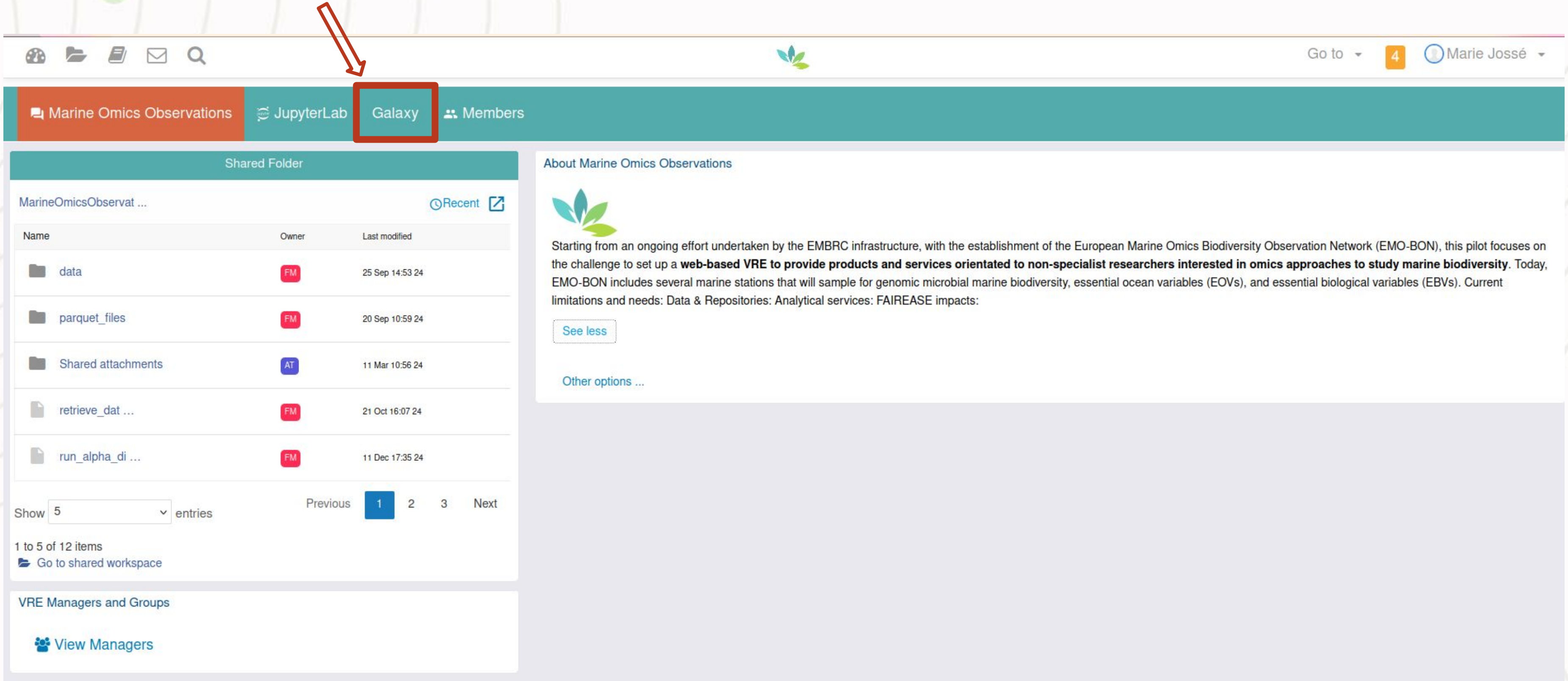
Marine Omics Observation: analyses of spatial- and time-comparable marine microbial metagenomics data sets for the exploration of biodiversity and its correlations with environmental quality

eosc | FAIR-EASE

FAIR-EASE datalake infrastructure



FAIR-EASE Pilots on D4science VLabs



Shared Folder

MarineOmicsObservat ... Recent

Name	Owner	Last modified
data	FM	25 Sep 14:53 24
parquet_files	FM	20 Sep 10:59 24
Shared attachments	AT	11 Mar 10:56 24
retrieve_dat ...	FM	21 Oct 16:07 24
run_alpha_di ...	FM	11 Dec 17:35 24

Show 5 entries Previous 1 2 3 Next

1 to 5 of 12 items

Go to shared workspace

VRE Managers and Groups

View Managers

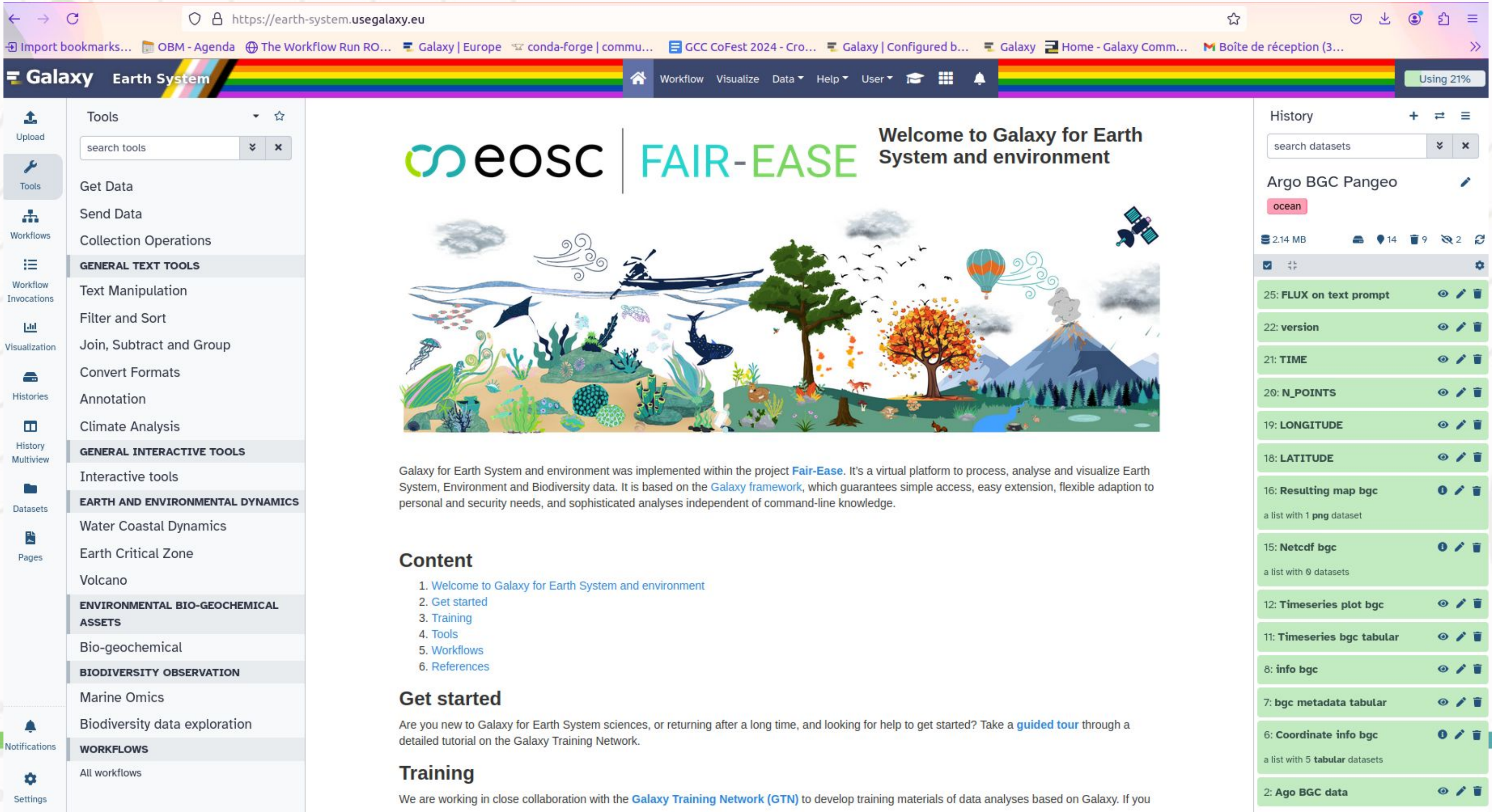
About Marine Omics Observations

Starting from an ongoing effort undertaken by the EMBRC infrastructure, with the establishment of the European Marine Omics Biodiversity Observation Network (EMO-BON), this pilot focuses on the challenge to set up a **web-based VRE to provide products and services orientated to non-specialist researchers interested in omics approaches to study marine biodiversity**. Today, EMO-BON includes several marine stations that will sample for genomic microbial marine biodiversity, essential ocean variables (EOVs), and essential biological variables (EBVs). Current limitations and needs: Data & Repositories: Analytical services: FAIREASE impacts:

See less

Other options ...

FAIR-EASE Pilots on Galaxy Europe



eosc | FAIR-EASE

Welcome to Galaxy for Earth System and environment

Galaxy for Earth System and environment was implemented within the project Fair-Ease. It's a virtual platform to process, analyse and visualize Earth System, Environment and Biodiversity data. It is based on the Galaxy framework, which guarantees simple access, easy extension, flexible adaption to personal and security needs, and sophisticated analyses independent of command-line knowledge.

Content

1. Welcome to Galaxy for Earth System and environment
2. Get started
3. Training
4. Tools
5. Workflows
6. References

Get started

Are you new to Galaxy for Earth System sciences, or returning after a long time, and looking for help to get started? Take a guided tour through a detailed tutorial on the Galaxy Training Network.

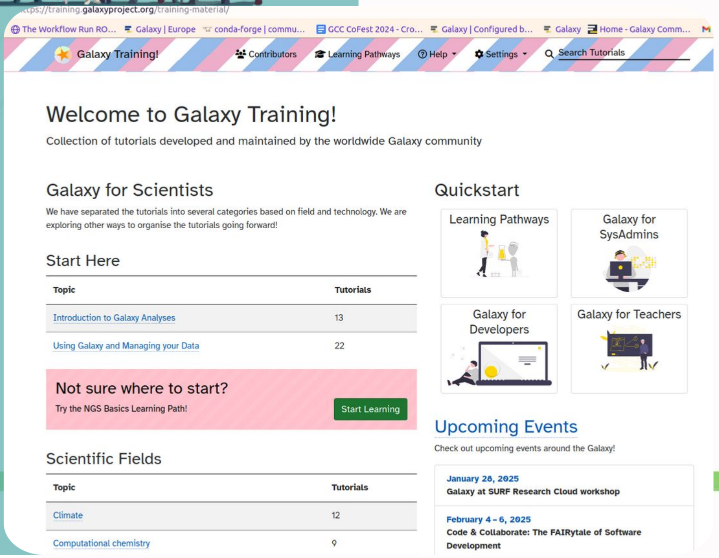
Training

We are working in close collaboration with the Galaxy Training Network (GTN) to develop training materials of data analyses based on Galaxy. If you

eosc | FAIR-EASE

Galaxy Training Network

An easy way to learn how to use Galaxy and improve your skills on various domains for instance a set of tutorials are available on FAIR management



Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Galaxy for Scientists

We have separated the tutorials into several categories based on field and technology. We are exploring other ways to organise the tutorials going forward!

Start Here

Topic	Tutorials
Introduction to Galaxy Analyses	13
Using Galaxy and Managing your Data	22

Not sure where to start?

Try the NGS Basics Learning Path!

Scientific Fields

Topic	Tutorials
Climate	12
Computational chemistry	9

Quickstart

Upcoming Events

Check out upcoming events around the Galaxy!

January 28, 2025
Galaxy at SURF Research Cloud workshop

February 4 - 6, 2025
Code & Collaborate: The FAIRytale of Software Development

- **A catalog of tutorials**
- **Pathways on a dedicated topic**
- **Classes, courses, webinars, and other interactive events**

A community for the community
https://training.galaxyproject.org/training-material/

Galaxy Training Network

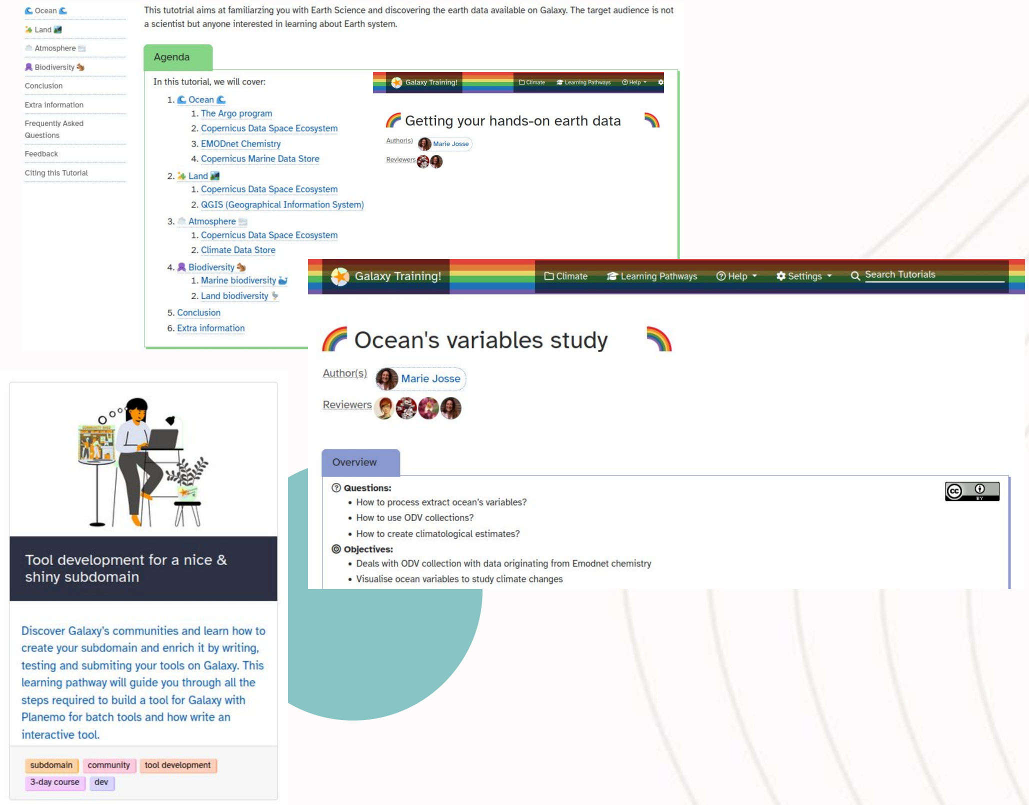
Galaxy Training Network

FAIR-EASE contribution (for earth sciences)

Earth sciences discovery tutorials

Thematic tutorials on ocean, land, atmosphere, and biosphere

Development tutorials end pathways to build a subdomain and a community with Galaxy



This tutotrial aims at familiarzing you with Earth Science and discovering the earth data available on Galaxy. The target audience is not a scientist but anyone interested in learning about Earth system.

Agenda

In this tutorial, we will cover:

1. Ocean
   1. The Argo program
   2. Copernicus Data Space Ecosystem
   3. EMODnet Chemistry
   4. Copernicus Marine Data Store
2. Land
   1. Copernicus Data Space Ecosystem
   2. QGIS (Geographical Information System)
3. Atmosphere
   1. Copernicus Data Space Ecosystem
   2. Climate Data Store
4. Biodiversity
   1. Marine biodiversity
   2. Land biodiversity
5. Conclusion
6. Extra information

Getting your hands-on earth data

Author(s) Marie Josse

Reviewers

Ocean's variables study

Author(s) Marie Josse

Reviewers

Overview

Questions:

- How to process extract ocean's variables?
- How to use ODV collections?
- How to create climatological estimates?

Objectives:

- Deals with ODV collection with data originating from Emodnet chemistry
- Visualise ocean variables to study climate changes

Tool development for a nice & shiny subdomain

Discover Galaxy's communities and learn how to create your subdomain and enrich it by writing, testing and submiting your tools on Galaxy. This learning pathway will guide you through all the steps required to build a tool for Galaxy with Planemo for batch tools and how write an interactive tool.

eosc | FAIR-EASE

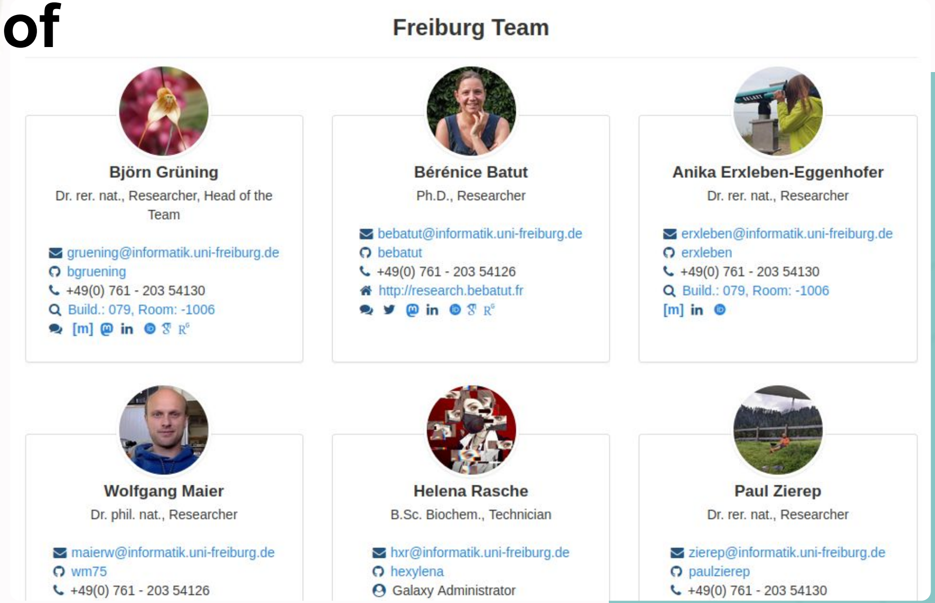
Importance of the team in the sustainability of Galaxy

A vibrant team of :

Administrators

Developpers

Trainers



Freiburg Team

Björn Grüning
Dr. rer. nat., Researcher, Head of the Team
gruening@informatik.uni-freiburg.de
bgruening
+49(0) 761 - 203 54130
Build.: 079, Room: -1006
[m] in

Bérénice Batut
Ph.D., Researcher
bebatut@informatik.uni-freiburg.de
bebatut
+49(0) 761 - 203 54126
http://research.bebatut.fr
in

Anika Erxleben-Eggenhofer
Dr. rer. nat., Researcher
erxleben@informatik.uni-freiburg.de
erxleben
+49(0) 761 - 203 54130
Build.: 079, Room: -1006
[m] in

Wolfgang Maier
Dr. phil. nat., Researcher
maierw@informatik.uni-freiburg.de
wm75
+49(0) 761 - 203 54126

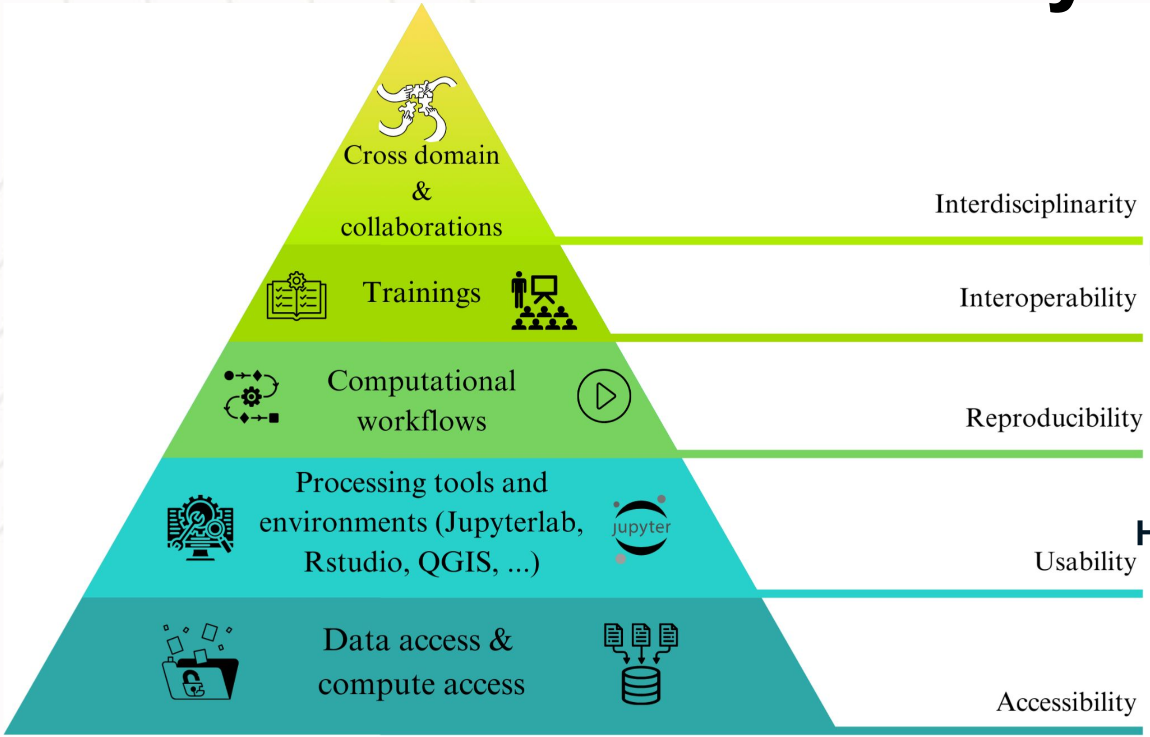
Helena Rasche
B.Sc. Biochem., Technician
hxr@informatik.uni-freiburg.de
hexylena
Galaxy Administrator

Paul Zierep
Dr. rer. nat., Researcher
zierep@informatik.uni-freiburg.de
paulzierep
+49(0) 761 - 203 54130

BUT with a strong focus on life sciences

Need the same team for earth sciences

Earth community



Cross domain & collaborations

Interdisciplinarity

Trainings

Interoperability

Computational workflows

Reproducibility

Processing tools and environments (Jupyterlab, Rstudio, QGIS, ...)

Usability

Data access & compute access

Accessibility

How to maintain it

How we built it

Galaxy for Earth System

Galaxy pyramid of needs

Services for EOSC in the proposal for a node

French national digital infrastructures

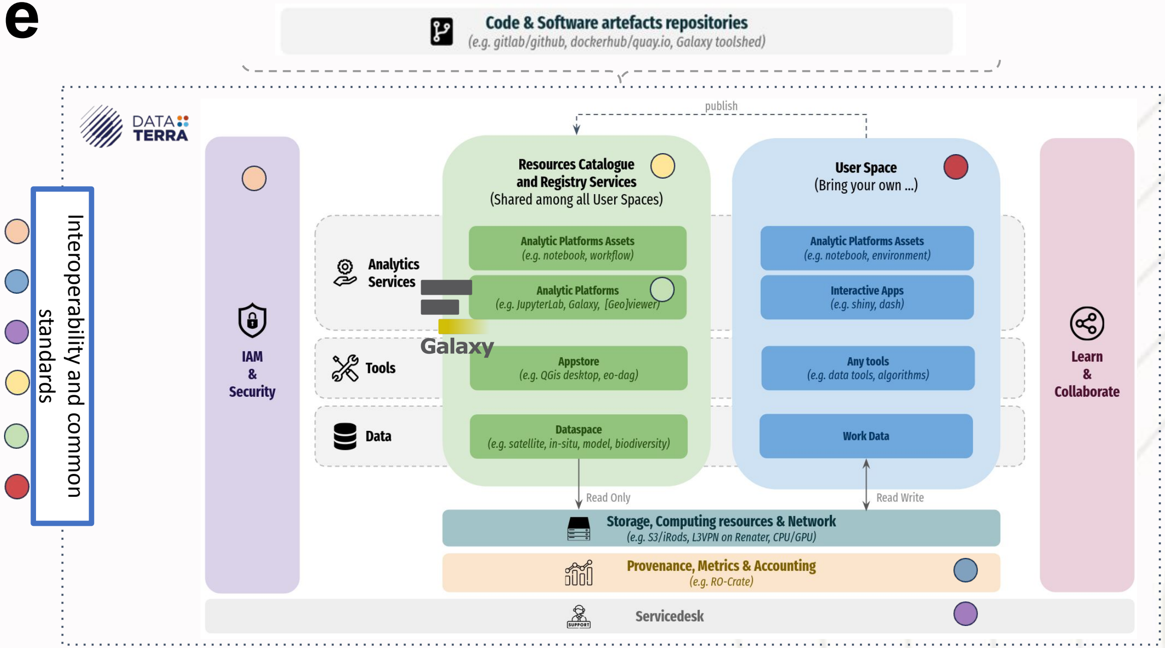
- Renater (Geant);
- France Grille (EGI), & Mesonet > GENCI / EuroHPC: IDRIS, CINES, TGCC;
- National & regional labelled data centres and meso-centres.

EOSC Federation

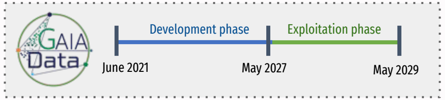
- D-T data and services accessible through the EOSC EU node;
- Services interoperability with thematic cluster nodes and related national nodes.

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)
Interoperability and common standards
DATA TERRA
publish
IAM & Security
Resources Catalogue and Registry Services
(Shared among all User Spaces)
User Space
(Bring your own ...)
Analytics Services
Analytic Platforms Assets
(e.g. notebook, workflow)
Analytic Platforms
(e.g. JupyterLab, Galaxy, [Geo]viewer)
Analytic Platforms Assets
(e.g. notebook, environment)
Interactive Apps
(e.g. shiny, dash)
Galaxy
Tools
Appstore
(e.g. QGis desktop, eo-dag)
Any tools
(e.g. data tools, algorithms)
Data
Dataspace
(e.g. satellite, in-situ, model, biodiversity)
Work Data
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
Learn & Collaborate



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

eosc | FAIR-EASE

Integrated Earth System Observation - Data Terra

e-Research Infrastructure to access, process and combine data

MAIN MISSION

Develop a global system for accessing and processing observation data (satellite, in situ), value-added products and services to observe, understand and predict in an integrated manner the functioning and evolution of the Earth system.

€42m (2020)

+1000 products & services

+15,000 users

100,000 TB (2022/2023)

Partners

French scientific organizations and universities

5 thematic hubs

Atmosphere	AERIS
Ocean	ODATIS
Solid earth	ForM@Ter
Land Surfaces	Theia
Biodiversity	PNDB Pôle National de Données de Biodiversité

Cross-cutting device

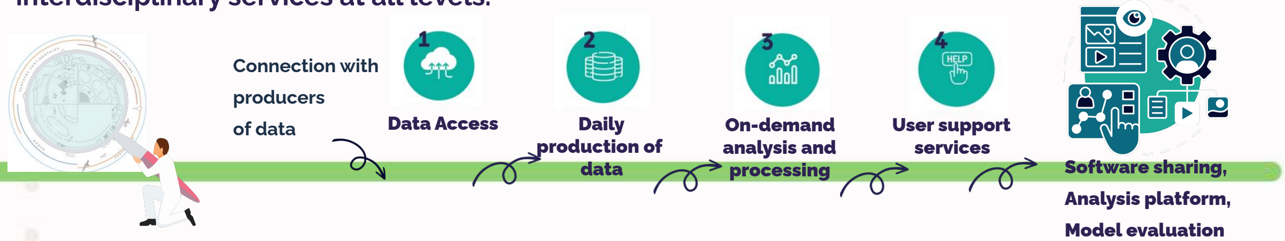
Satellite images — DATATERRA DINAMIS

- A multidisciplinary approach because it calls on work in several areas of Earth System sciences
- An inclusive project that goes beyond the scientific circle with an approach also oriented towards field actors and participatory data

OUR SERVICES

Facilitating the cross-referencing of observations and the modeling of Earth System data

The IR Data Terra offers services around Earth system observation data. The objective is to provide interoperable and interdisciplinary services at all levels.



Connection with producers of data

1 Data Access

2 Daily production of data

3 On-demand analysis and processing

4 User support services

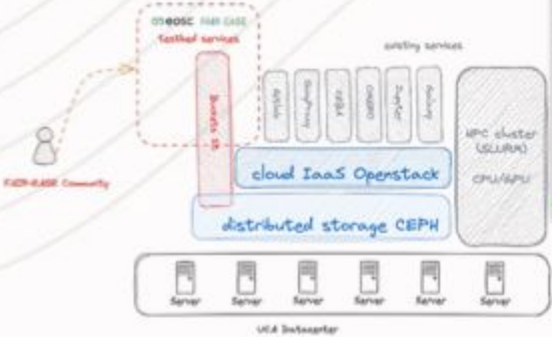
Software sharing, Analysis platform, Model evaluation

DATA TERRA

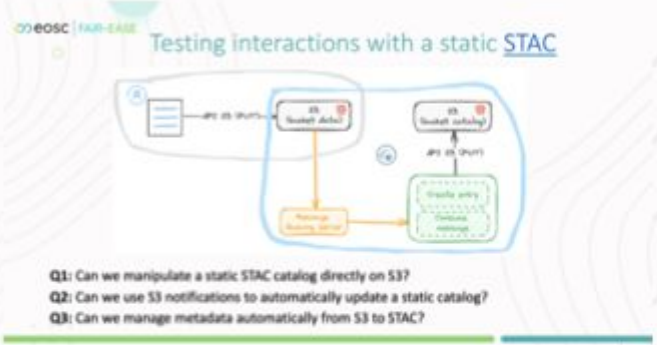
wednesday session

Integrating New Data Storage and Access Paradigms in pilot's practices

"S3 Testbed Tests" Antoine Mahul, David Sarramia, Damian Smyth



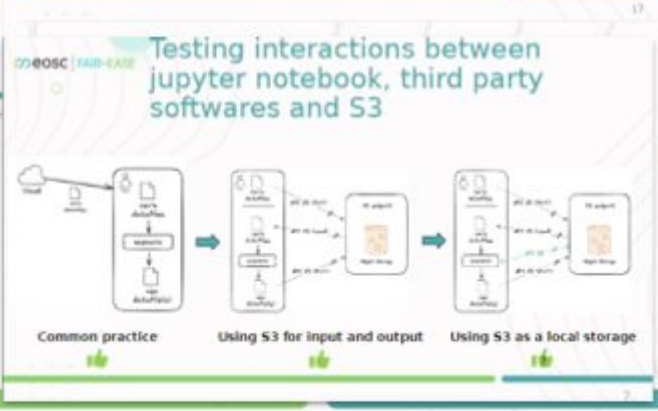
S3 concepts reminder

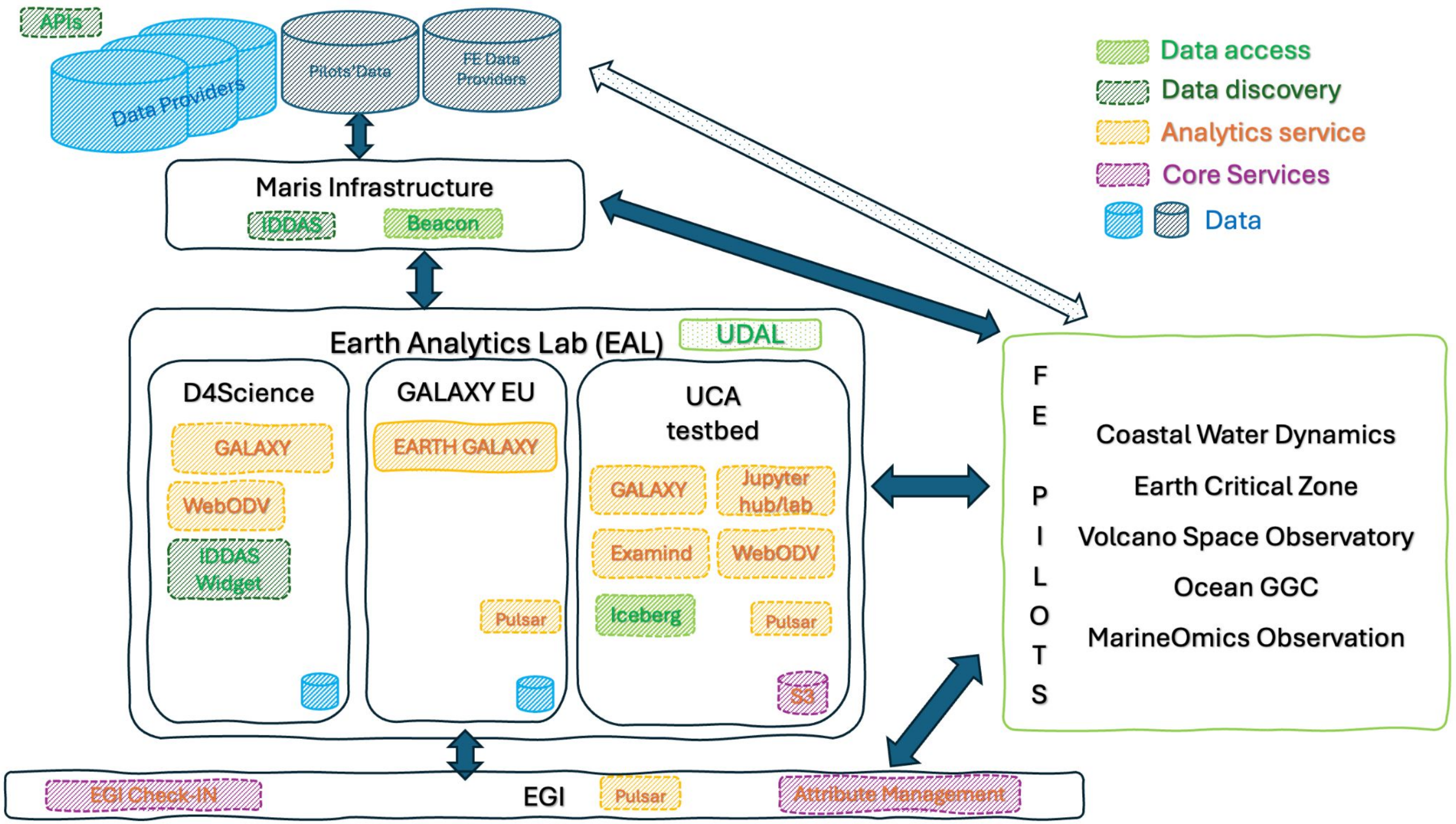

Testing interactions with a static STAC

Q1: Can we manipulate a static STAC catalog directly on S3?

Q2: Can we use S3 notifications to automatically update a static catalog?

Q3: Can we manage metadata automatically from S3 to STAC?


Testing interactions between jupyter notebook, third party softwares and S3



APIs

Data Providers

Pilots'Data

FE Data Providers

Maris Infrastructure

IDDAS

Beacon

Earth Analytics Lab (EAL)

UDAL

D4Science

GALAXY

WebODV

IDDAS Widget

GALAXY EU

EARTH GALAXY

Pulsar

UCA testbed

GALAXY

Jupyter hub/lab

Examind

WebODV

Iceberg

Pulsar

S3

EGI Check-IN

EGI

Pulsar

Attribute Management

FE PILOTS

Coastal Water Dynamics

Earth Critical Zone

Volcano Space Observatory

Ocean GGC

MarineOmics Observation

▨ Data access

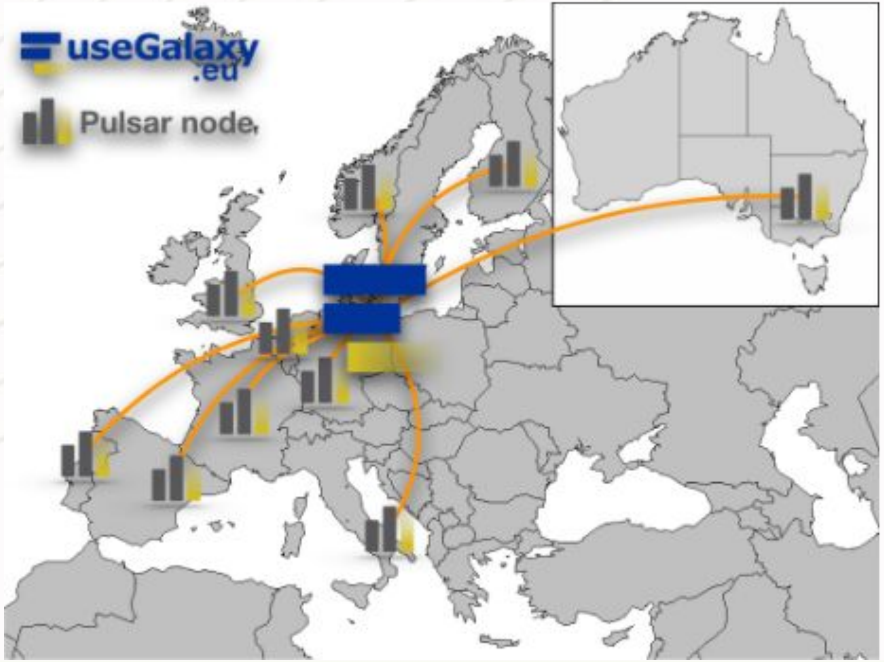
▨ Data discovery

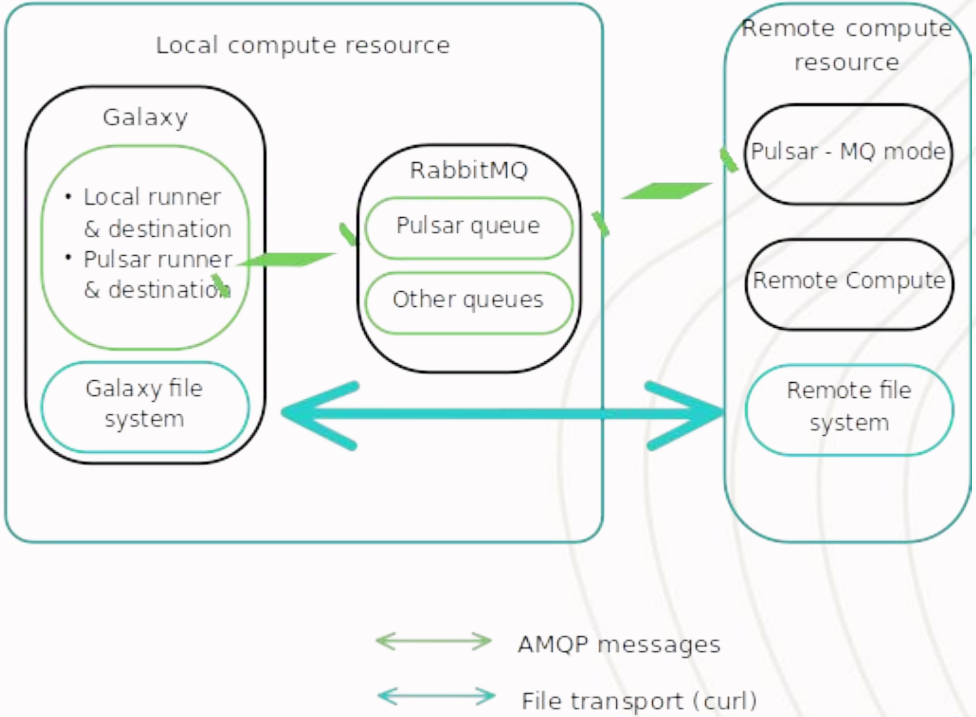
▨ Analytics service

▨ Core Services

▨ ▨ Data

Galaxy deals with remote compute resources and remote files system





Local compute resource

Galaxy

- Local runner & destination
- Pulsar runner & destination

Galaxy file system

RabbitMQ

Pulsar queue

Other queues

Remote compute resource

Pulsar - MQ mode

Remote Compute

Remote file system

AMQP messages

File transport (curl)

eosc | FAIR-EASE

UCA TESTBED INFRASTRUCTURE

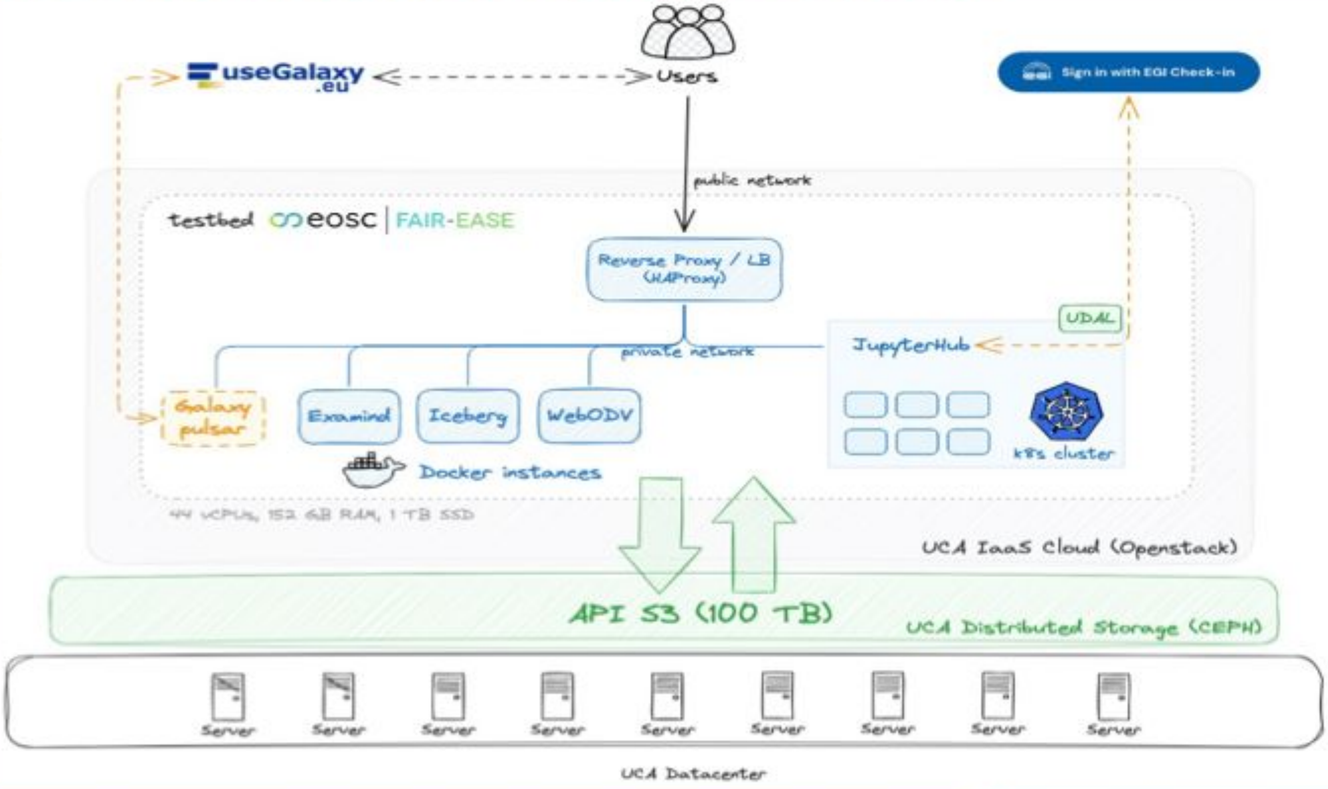
Examind / Iceberg/WebODV effective from this month end.

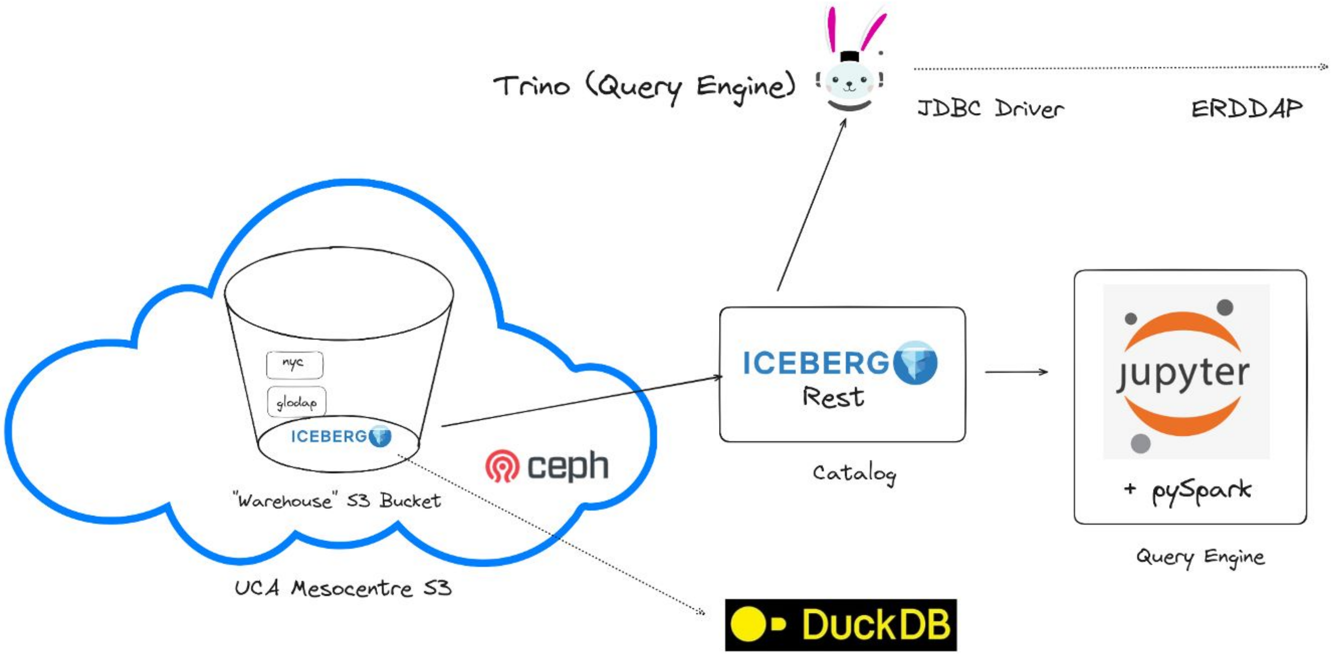
More precise elements:
Cloud IaaS : Openstack (community version) https://www.openstack.org/
total : 48 hypervisors, 800 CPU oores, RAM 7.6 TB,

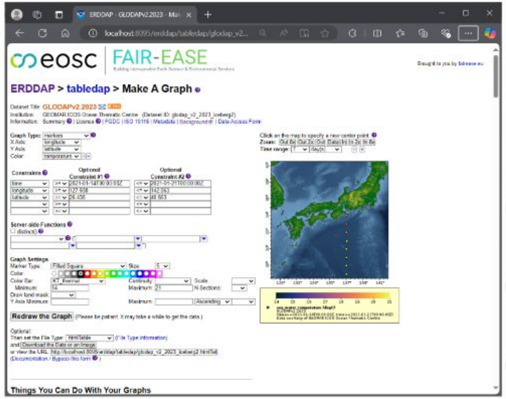
Distributed storage/ S3 : Ceph (community version) https://ceph.io/
total : 600 TB SSD (for vm openstack), 5.5 PB HDD (S3)
(replicated 3 times so /3)

Ressources allouée pour FAIR-EASE actuelles: 44 vCPUs, 152 GB RAM, 128 GB SSD, 100 TB via API S3

Cluster kubernetes FE : kubernetes v1.31, déployéed withTerraform et Talos Linux (https://www.talos.dev/)
3 vm for control plane
3 vm for workers qui vont executés les instances services client (JupyterHub/JupyterLab)






ERDDAP - GLODAPv2.2023 - Ma
localhost:8095/erddap/tabledap/glodap_v2...
eosc
FAIR-EASE
Brought to you by fairease.eu

ERDDAP > tabledap > Make A Graph

Dataset Title: GLODAPv2.2023
Institution: GEOMAR ICOS Ocean Thematic Centre (Dataset ID: glodap_v2_2023_iceberg2)
Information: Summary | License | FGDC | ISO 19115 | Metadata | Background | Data Access Form

Graph Type: markers
X Axis: longitude
Y Axis: latitude
Color: temperature

Constraints	Optional Constraint #1		Optional Constraint #2	
time	>=	2021-01-14T00:00:00Z	<=	2021-01-21T00:00:00Z
longitude	>=	127.938	<=	142.263
latitude	>=	26.436	<=	40.663
	>=		<=	
	>=		<=	

Server-side Functions
□ distinct()

Graph Settings
Marker Type: Filled Square Size: 5
Color Bar: KT_thermal Continuity: Scale:
Minimum: 14 Maximum: 21 N Sections:
Draw land mask:
Y Axis Minimum: Maximum: Ascending

Redraw the Graph (Please be patient. It may take a while to get the data.)

Optional:
Then set the File Type: htmlTable (File Type information)
and Download the Data or an Image
or view the URL: http://localhost:8095/erddap/tabledap/glodap_v2_2023_iceberg2.htmlTab
(Documentation / Bypass this form)

Click on the map to specify a new center point.
Zoom: Out 8x | Out 2x | Out | Data | In | In 2x | In 8x
Time range: 7 day(s)

Things You Can Do With Your Graphs

Apache Iceberg an open table format for huge analytic datasets (including metadata)

Iceberg adds tables to compute engines including Spark, Trino, PrestoDB, Flink, Hive and Impala using a high-performance table format that works just like a SQL table. https://iceberg.apache.org/docs/1.5.2/

Iceberg avoids unpleasant surprises. Schema evolution works and won't inadvertently un-delete data. **Users don't need to know about partitioning to get fast queries**.

- Schema evolution supports add, drop, update, or rename, and has no side-effects
- Hidden partitioning prevents user mistakes that cause silently incorrect results or extremely slow queries
- Partition layout evolution can update the layout of a table as data volume or query patterns change
- Time travel enables reproducible queries that use exactly the same table snapshot, or lets users easily examine changes
- Version rollback allows users to quickly correct problems by resetting tables to a good state

Reliability and performance🔗: Iceberg **was built for huge tables**. Iceberg is used in production where a single table can contain tens of petabytes of data and even these huge tables can be read without a distributed SQL engine.

- Scan planning is fast -- a distributed SQL engine isn't needed to read a table or find files
- Advanced filtering -- data files are pruned with partition and column-level stats, using table metadata

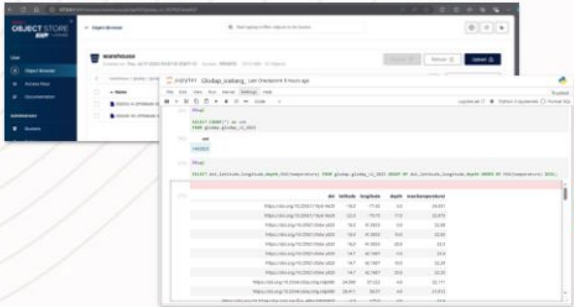
Iceberg was designed to solve correctness problems in eventually-consistent cloud object stores.

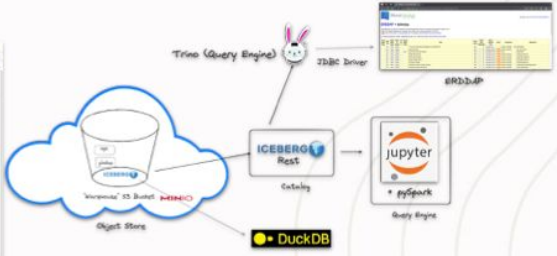
- Works with any cloud store and reduces NN congestion when in HDFS, by avoiding listing and renames
- Serializable isolation -- table changes are atomic and readers never see partial or uncommitted changes
- Multiple concurrent writers use optimistic concurrency and will retry to ensure that compatible updates succeed, even when writes conflict

Open standard🔗 designed and developed to be an open community standard with a specification to ensure compatibility across languages and implementations. Apache Iceberg is open source, and is developed at the Apache Software Foundation.

Integrating New Data Storage and Access Paradigms in pilot’s practices

“Open Data Lake House Format - Apache Iceberg - Quick start” Damian Smyth:





Joint meeting with D4T2 on New Data Storage and Access Paradigms on July 12th

https://fair-ease.atlassian.net/wiki/spaces/FAIREASE/pages/400818177/2024-07-12+Meeting

EXAMIND

The Examind software suite, developed by Geomatys, provides all the features you need t**o build a Geographic, Hydrographic, Oceanographic and Meteorologic geospatial data processing infrastructure**. Designed for interoperability, all the products in the suite comply with OGC standards, and integrate OGC Web services

Examind community is the open-source platform / server of the Examind ecosystem developed by Geomatys. This map server has a wide range of functions available, manages multiple data formats (Netcdf, geotiff, etc.), clouds native data formats (COGs, GIMI), multiple OGC (Open Geospatial Consortium) standards (WCS, WMS, WPS, OGC API, etc.), and has a number of ready-to-use processes for various uses in the geospatial world. Examind also offers several ways of managing and structuring data. The server can connect to an existing data source, hold the data locally, or generate new data via different processes (via WPS, OGC API Process, openEO); all via different paradigms, such as data cube structuring.

Examind supports several data storage options: local storage, FTP, S3 (AWS / Minio), via HTTP/HTTPS, or storage from a WMS / WMTS service, etc. The service can be deployed via a docker image and a docker-compose file on kubernetes infrastructures. It operates standalone and is based on the Apache SIS geospatial processing library.

Apache Iceberg is an open table format designed to manage large-scale analytic datasets reliably and efficiently, making it ideal for democratized data infrastructures like those used in research and Earth Sciences. It provides versioning, schema evolution, and ACID compliance, enabling robust and transparent data management across diverse platforms. For infrastructures such as the ones you've described, Iceberg can ensure data access, support for distributed query engines, and adaptability to regional and cloud-based resources, empowering collaborative and FAIR data-driven science.

Summary

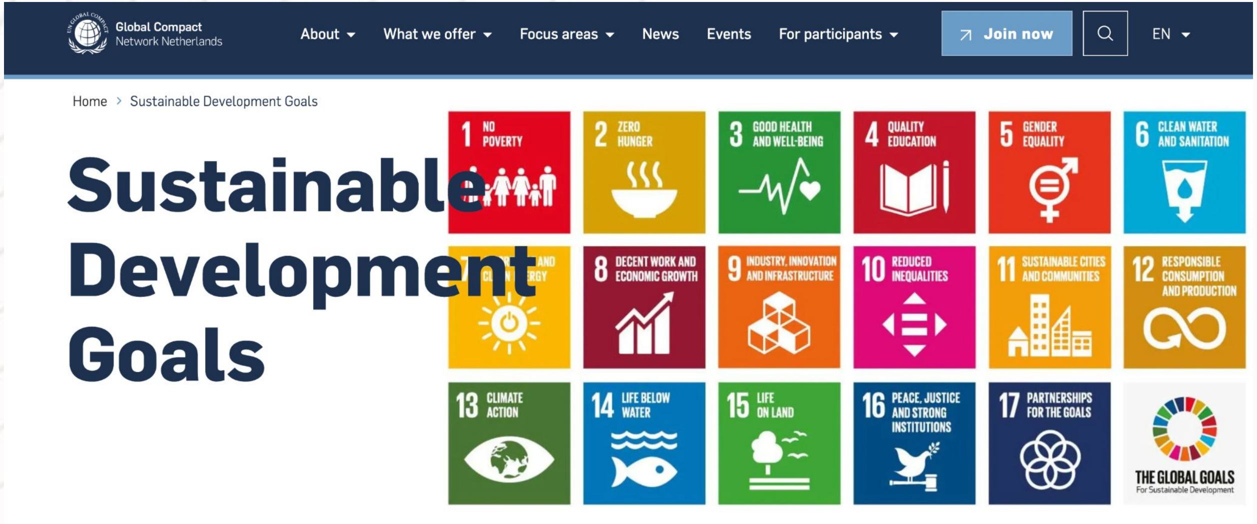
Iceberg excels in decoupling storage from compute, managing evolving schemas, and enabling scalable, FAIR-compliant data workflows, making it ideal for research infrastructures.

For FAIR, open data infrastructures like FAIR-EASE, Iceberg offers flexibility and adaptability while ensuring long-term usability and performance at scale.

Feature	Apache Iceberg
Primary Use Case	Data lake management with strong ACID compliance
***ACID Transactions**	Fully supported, scalable for large datasets
Schema Evolution	Rich support for evolving schemas without rewriting data
Partition Handling	Hidden partitioning; avoids explicit user-defined partitions
Integration	Compatible with Spark, Trino, Presto, Flink, Hive, etc.
Streaming Support	Experimental, gaining traction
Community	Open, strong Apache ecosystem
Adoption	Used by Netflix, Apple, LinkedIn

* In computer science, **ACID (atomicity, consistency, isolation, durability)** is a set of properties of database transactions intended to guarantee data validity despite errors, power failures, and other mishaps

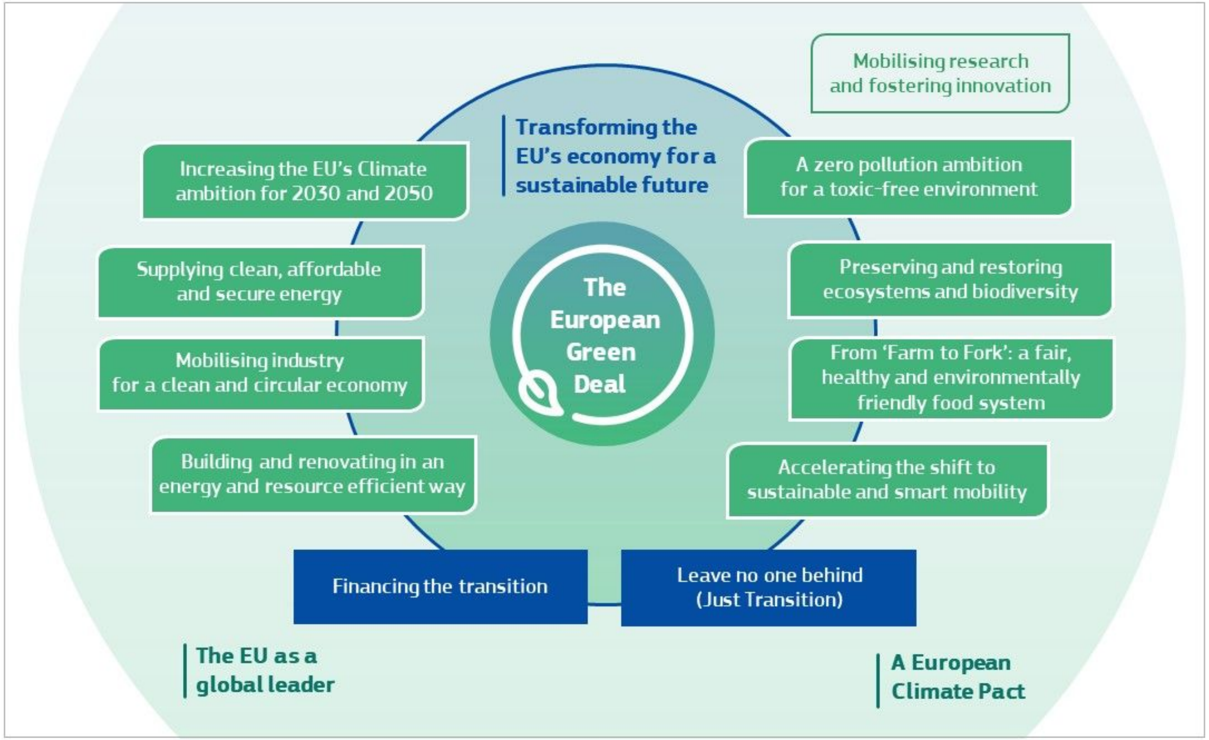
FAIR-EASE in the context of the European green deal


Global Compact
Network Netherlands
About
What we offer
Focus areas
News
Events
For participants
Join now
EN
1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
AND
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
THE GLOBAL GOALS
For Sustainable Development

Home > Sustainable Development Goals

Sustainable Development Goals

blablabla



The European Green Deal

Transforming the EU's economy for a sustainable future

Mobilising research and fostering innovation

Increasing the EU's Climate ambition for 2030 and 2050

Supplying clean, affordable and secure energy

Mobilising industry for a clean and circular economy

Building and renovating in an energy and resource efficient way

A zero pollution ambition for a toxic-free environment

Preserving and restoring ecosystems and biodiversity

From 'Farm to Fork': a fair, healthy and environmentally friendly food system

Accelerating the shift to sustainable and smart mobility

Financing the transition

Leave no one behind (Just Transition)

The EU as a global leader

A European Climate Pact

FAIREASE in the context of the european green deal

European green deal topics

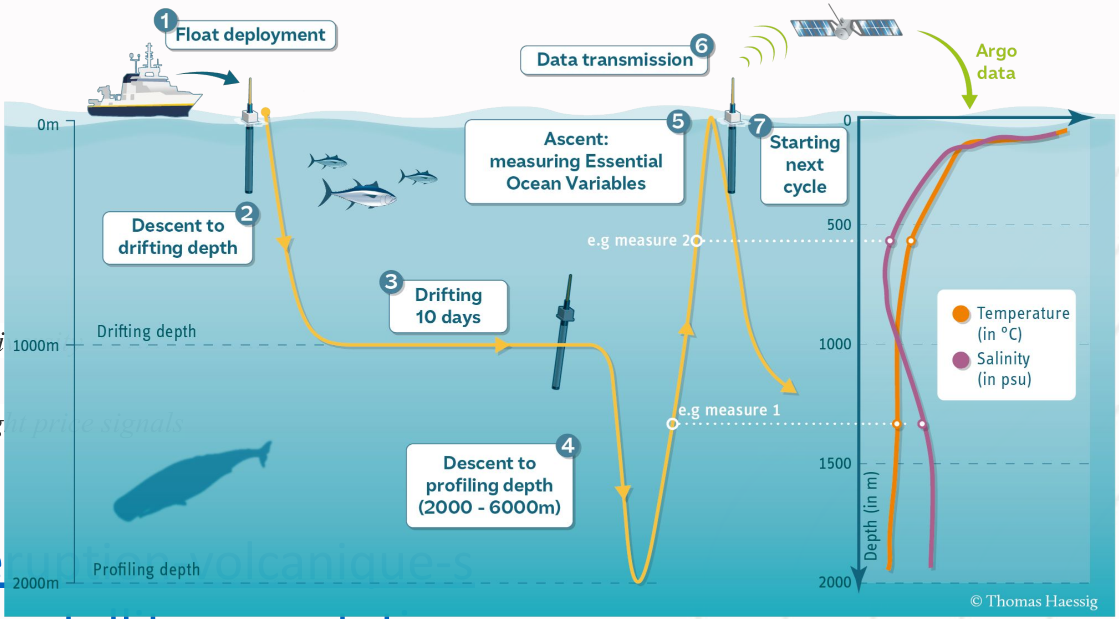
- **Designing a set of deeply transformative policies**
- **Mainstreaming sustainability in all EU policie**

- *2.1.7.*
- *Preserving and restoring ecosystems and biodiversity*
- *2.2.2.*
- *Greening national budgets and sending the right price signals*
- *2.2.3.*
- *Mobilising research and fostering innovation*

- *2.1.7.*
- *Preserving and restoring ecosystems and biodiversity*
- *2.2.2.*
- *Greening national budgets and sending the right price signals*
- *2.2.3.*
- *Mobilising research and fostering innovation*

https://www.aeris-data.fr/leruption-volcanique-sur-les-iles-tonga-vue-par-les-satellites-geostationnaires/

- *2.1.7.*
- *Preserving and restoring ecosystems and biodiversity*
- *2.2.2.*
- *Greening national budgets and sending the right price signals*
- *2.2.3.*
- *Mobilising research and fostering innovation*



https://www.aeris-data.fr/leruption-volcanique-sur-les-iles-tonga-vue-par-les-satellites-geostationnaires/