



Blue-Cloud2026



# Data management and provision by EMSO ocean observing system

Raul Bardaji - EMSO & Blue-Cloud



Funded by  
the European Union

## Objectives of the Webinar

- ▶ Understanding EMSO ERIC
- ▶ Tools for Data Harmonization and Distribution
- ▶ Practical Application
- ▶ Strategies in Action
- ▶ Engaging with the Audience

**1 CENTRAL HUB and 14 INTERLINKED OBSERVATORIES AND MUTI-SENSORS PLATFORMS**

**27 European Institutions**

**Added value** compared to the value of a single research cooperation network  
**Joint investment strategy** to strengthen EMSO ERIC through its regional facilities sites and common and shared services



MARINE ECOSYSTEM



CLIMATE CHANGE



GEO HAZARD



TECHNOLOGICAL CHALLENGE

DISTRIBUTED RESEARCH INFRASTRUCTURE



## EMSO main activities

- ▶ Participation in research projects
- ▶ Providing physical access to regional facilities
- ▶ Providing access to data

# EMSO Calls for Access

## Primary Objective

Access to the EMSO infrastructure.

## Benefits for Participants

- ▶ Set up devices: sensors, instruments, systems, and emerging technologies.
- ▶ Conduct new procedures or experiments in a controlled and specialized environment.

## EMSO Data Access Strategies

- ▶ **Initial Approach:** Centralized data infrastructure complemented by a harmonization system for streamlined access.
- ▶ **Actual Approach:** Decentralized data infrastructure utilizing ERDDAP for enhanced data distribution and retrieval.

## EMSO Data Access Strategies

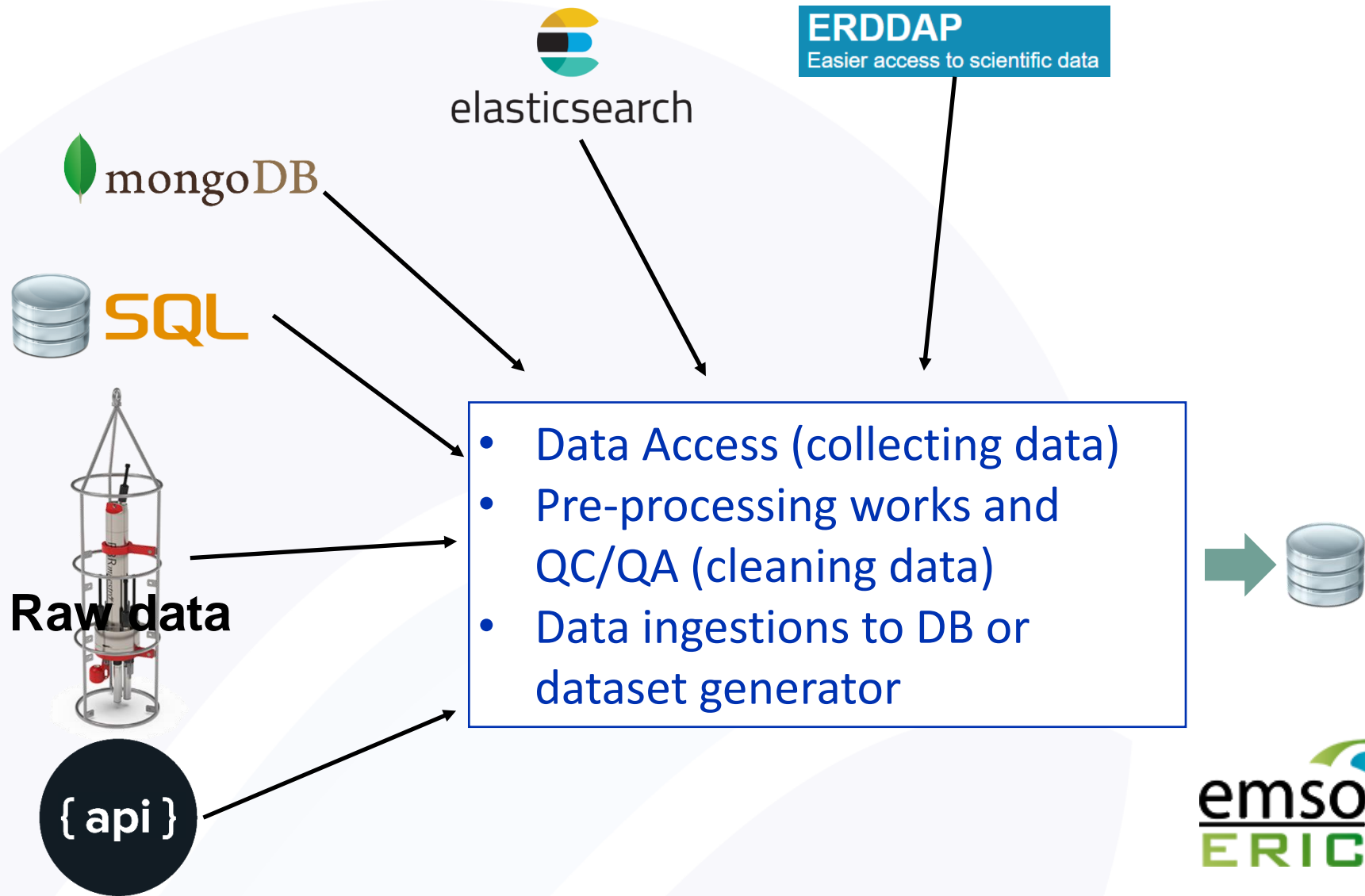
- ▶ **Initial Approach:** Centralized data infrastructure complemented by a harmonization system for streamlined access.
- ▶ **Actual Approach:** Decentralized data infrastructure utilizing ERDDAP for enhanced data distribution and retrieval.

# Data sources (DS) providing (FAIR) data





A technology for collecting, cleaning, and organizing data.





**WaterFrame**

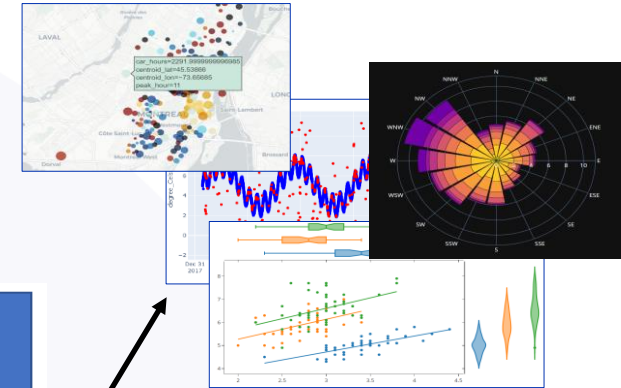
DEPTH	TIME	Variable1	Variable1_QC	...	VariableN	VariableN_QC
value	value	value	value		value	value
...	...	...	...	...	...	...
value	value	value	value		value	value

**metadata**

- Field 1: Info 1
- ...
- Field N: Info N

**vocabulary**

- Variable1
  - Field 1: Info 1
  - ...
  - Field N: Info N
- ...
- VariableN
  - Field 1: Info 1
  - ...
  - Field N: Info N



**Dataset generation and ingestion to DB**

**Pre-processing tasks and QA/QC**





<https://github.com/emso-eric/mooda>

\$ pip install mooda



# Why the Initial Strategy Faced Challenges

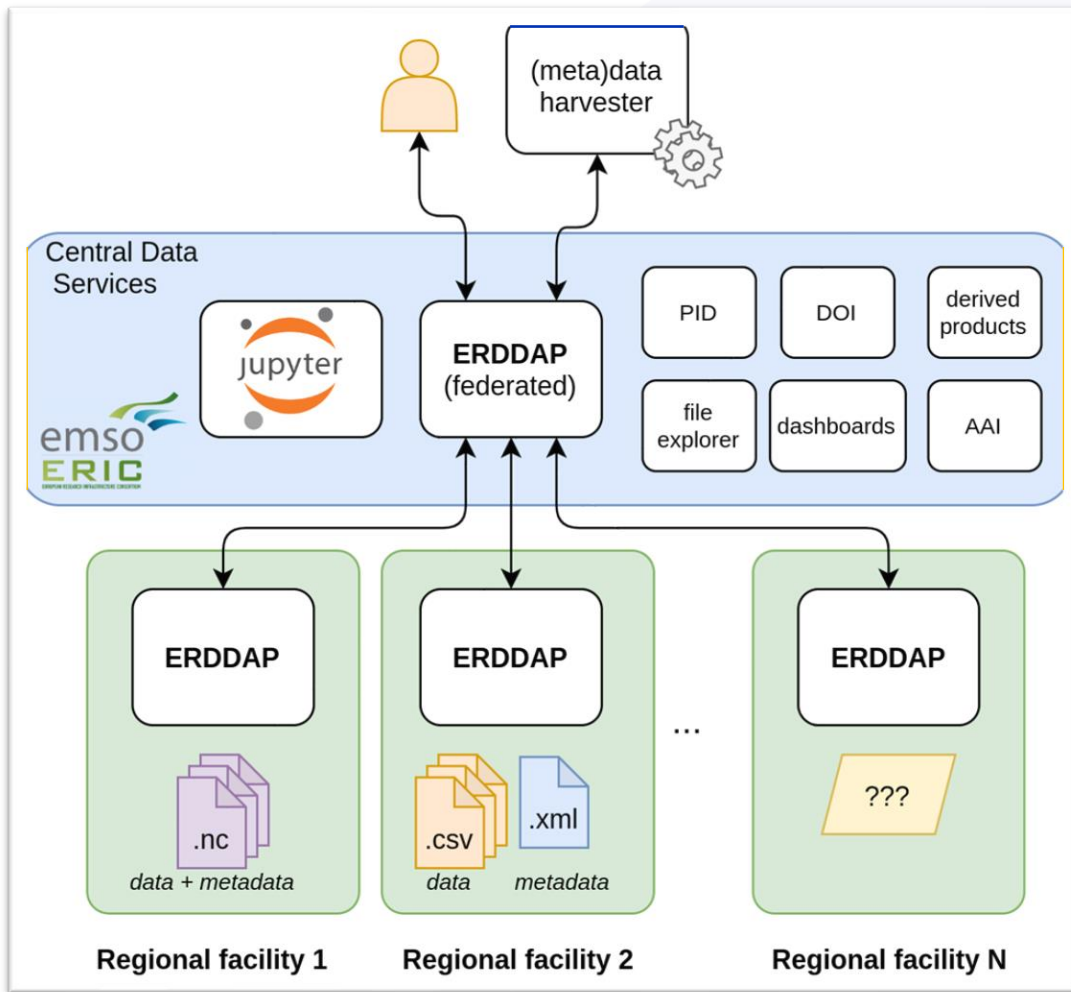
1. Diversity of Data Sources
2. Format and Quality Issues
3. Limited Human Resources
4. Viability and Sustainability



## EMSO Data Access Strategies

- ▶ **Initial Approach:** Centralized data infrastructure complemented by a harmonization system for streamlined access.
- ▶ **Actual Approach:** Decentralized data infrastructure utilizing ERDDAP for enhanced data distribution and retrieval.

# Data Discovery & Access mechanism

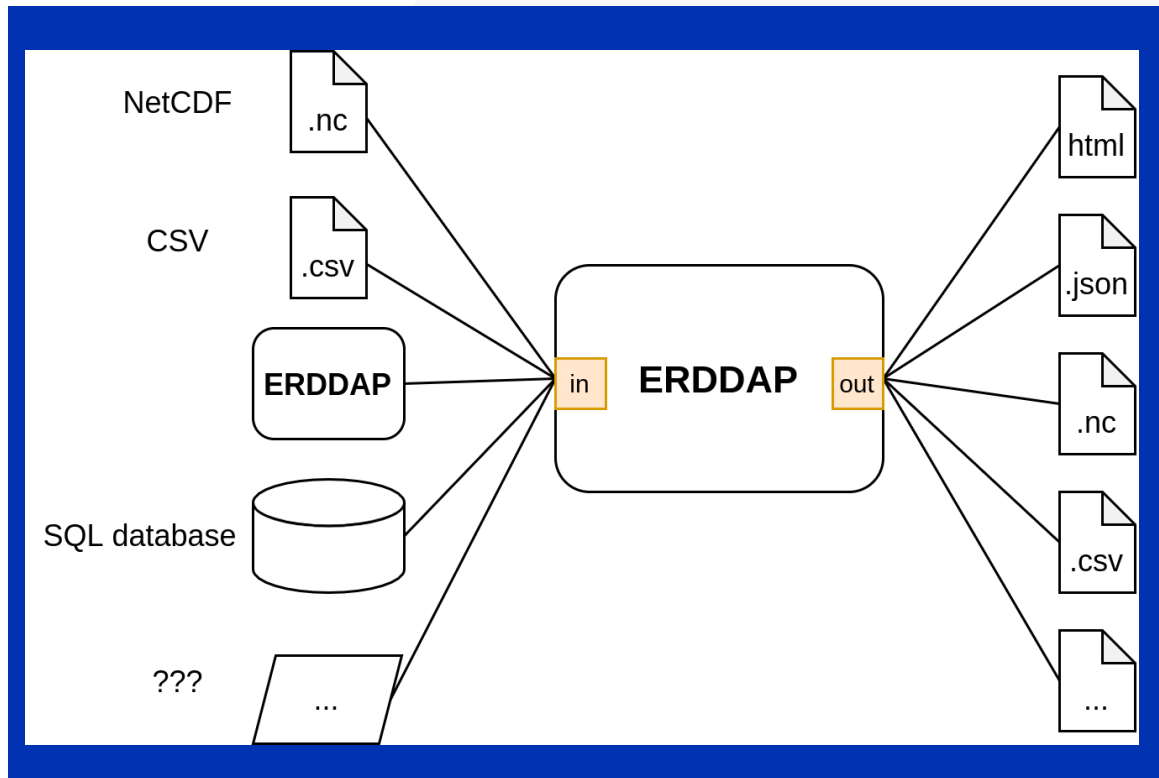


Distributed data  
 Federated ERDDAP  
 Common entry point



Not all sites have an ERDDAP yet

# Types of data and formats



## ERDDAP Service

- Gridded / Tabular data
  - Serves subsets of data
  - Multiple input formats
- Multiple input formats
- User-interface
- Simple to use REST API
- Allows federation

- ▶ **ERDDAP Playground:** Set up a simple ERDDAP service using docker.
  - ▶ <https://github.com/emso-eric/erddap-playground>
- ▶ **Metadata Harmonizer:** Tools to connect to an ERDDAP service and assess if the metadata is compliant with the EMSO Metadata Specifications.
  - ▶ <https://github.com/emso-eric/metadata-harmonizer>
- ▶ **EMSO Metadata Specifications:** This repository contains the EMSO Metadata Specifications, which define how datasets should be structured to be compliant with EMSO's data policy. The specifications apply to both NetCDF files and datasets served through ERDDAP.
  - ▶ <https://github.com/emso-eric/emso-metadata-specifications>



# Other web services

PID / DOI

File explorer

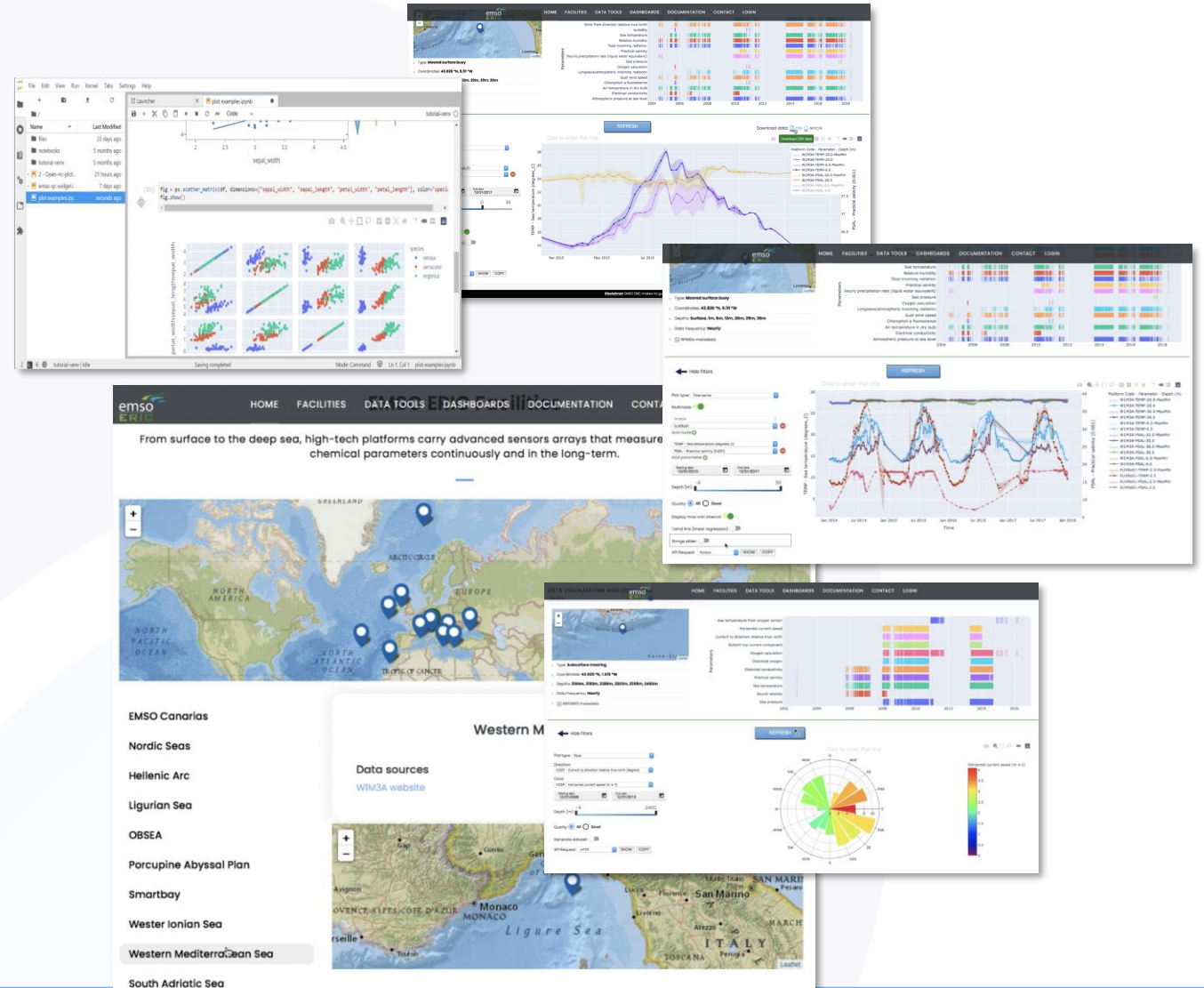
Research Env. – Jupyter Lab

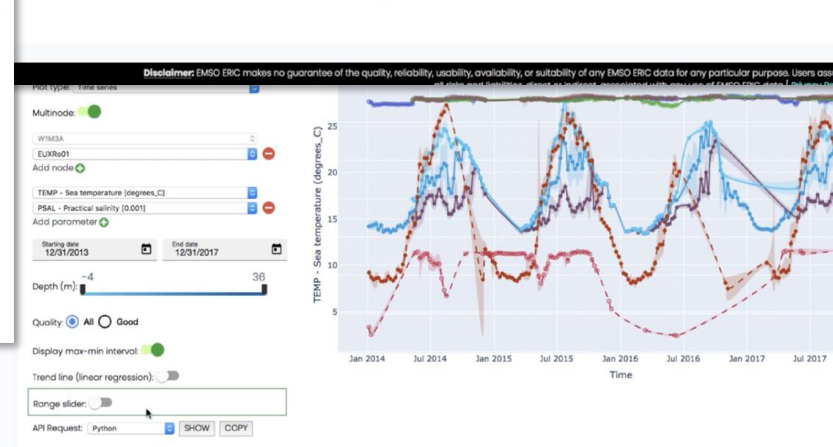
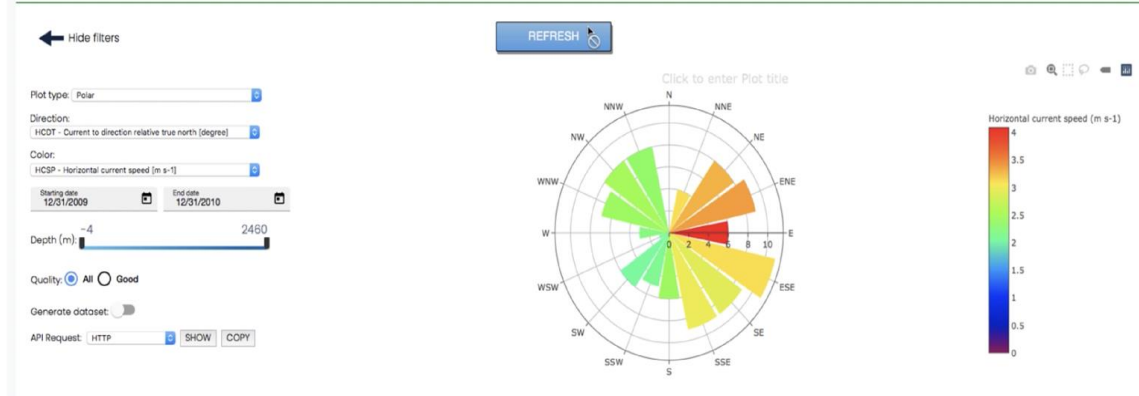
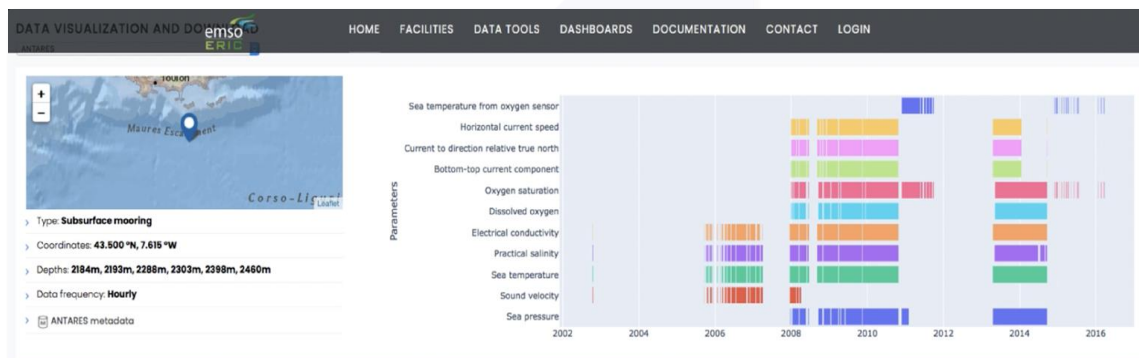
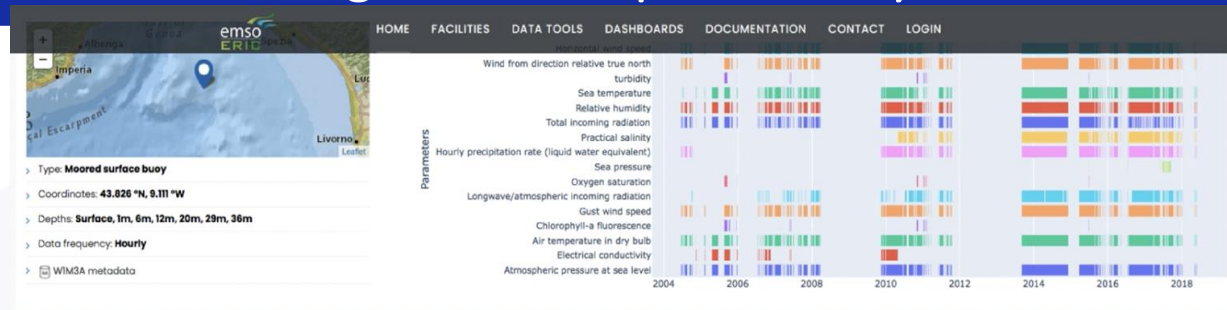
Derivate products

Dashboards

AAI

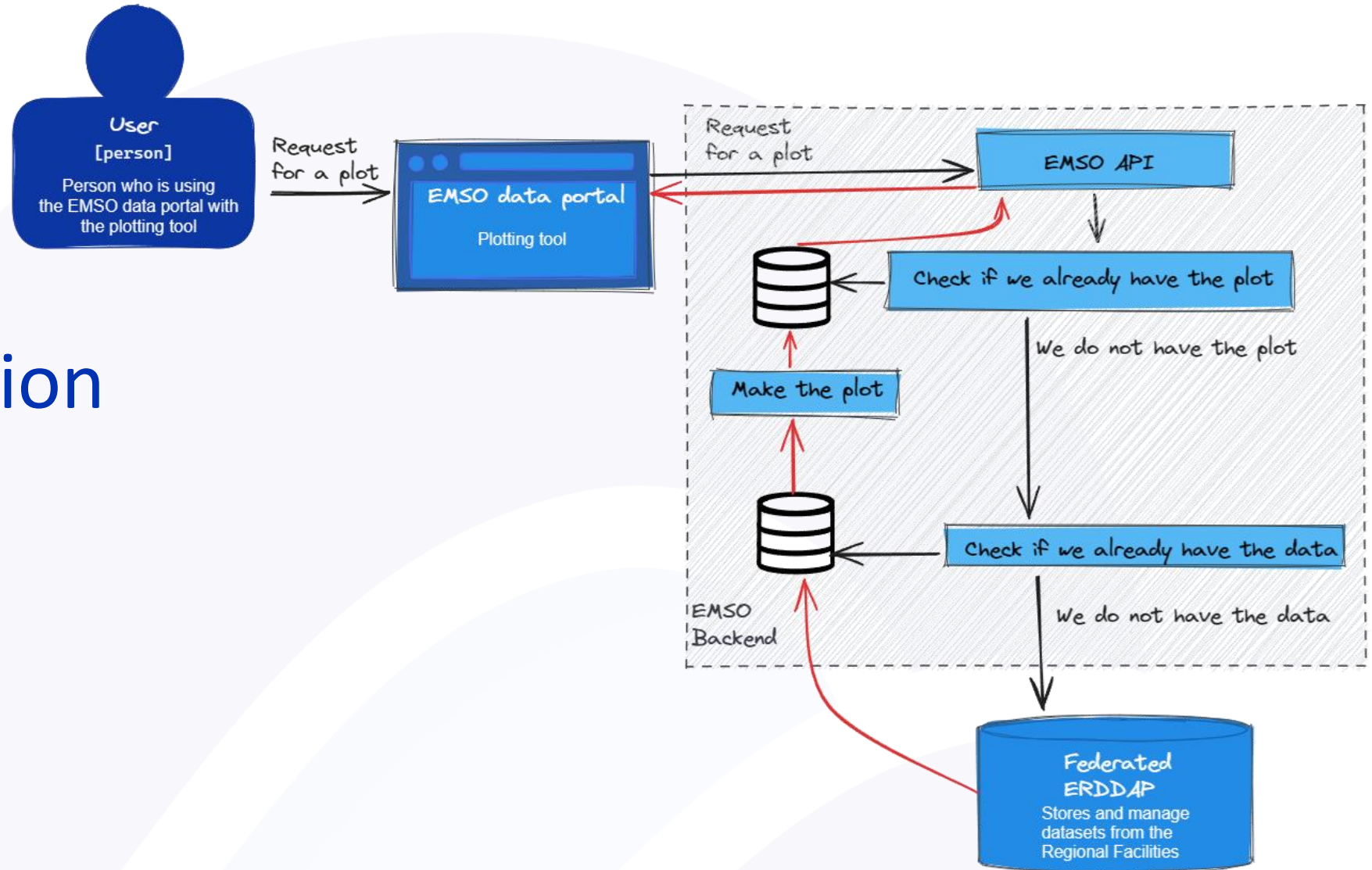
Internal API







# EMSO implementation



# Concluding Thoughts

- ▶ Effectiveness of the Initial Approach
- ▶ Challenges with Scalability
- ▶ Adopting a Conservative Actual Approach
- ▶ Utilizing ERDDAP for Federated Data Networks
- ▶ Limitations and Additional Layers

# eOSC | Blue-Cloud2026



[blue-cloud.org](https://blue-cloud.org)



[@bluecloudeu](https://twitter.com/bluecloudeu)



[blue-cloud org](https://www.linkedin.com/company/blue-cloud-org)



Funded by  
the European Union