

EMSO ERIC Data Federation and Management via ERDDAP

Blue-Cloud GA, Federation workshop
6 November 2024, Lisbon

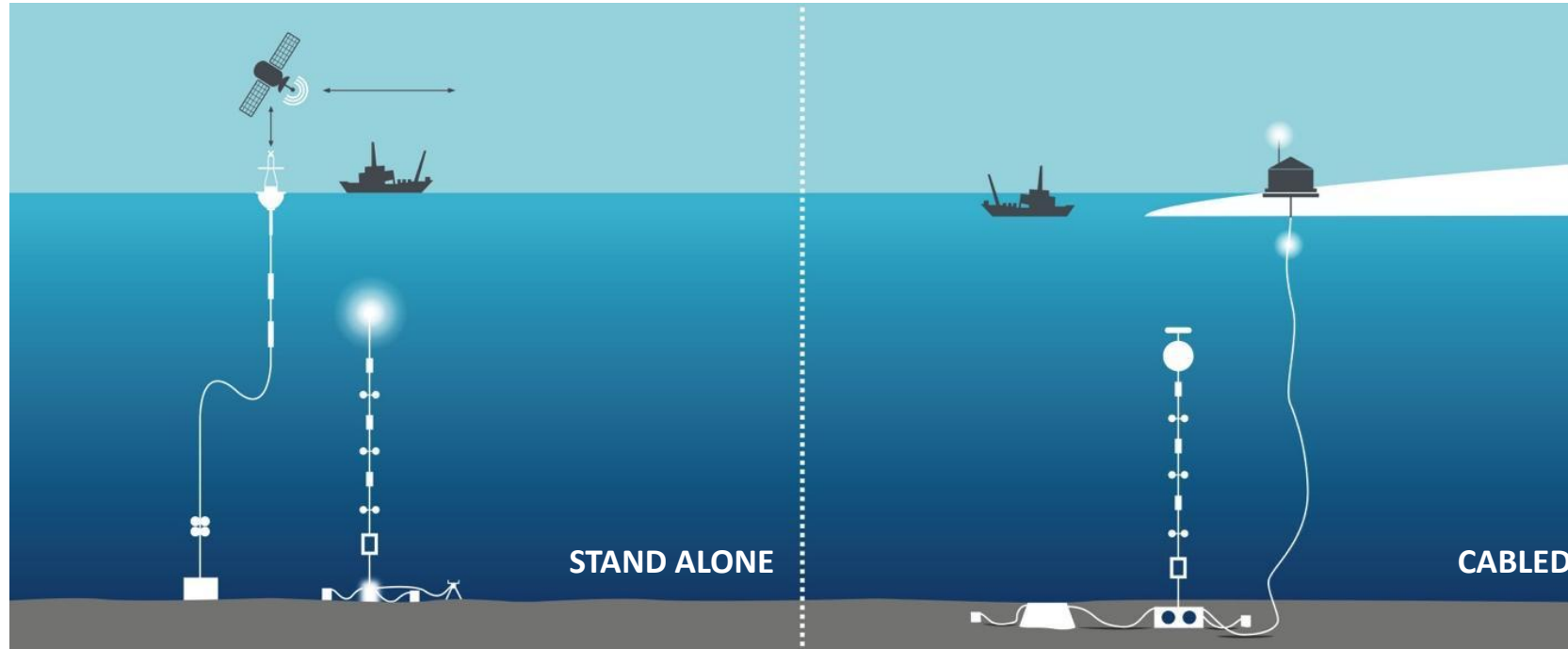
Ingrid Puillat

EMSO ERIC Director General

Ingrid.puillat@emso-eu.org

EMSO ERIC – Introduction

EMSO ERIC: a legal entity since 2016



EMSO Specificities:

- Open ocean (deep part vs Shallow),
- Long time series,
- In fixed regional area
- Oceanology and geophysics questions

EMSO ERIC – Introduction

EMSO Mission

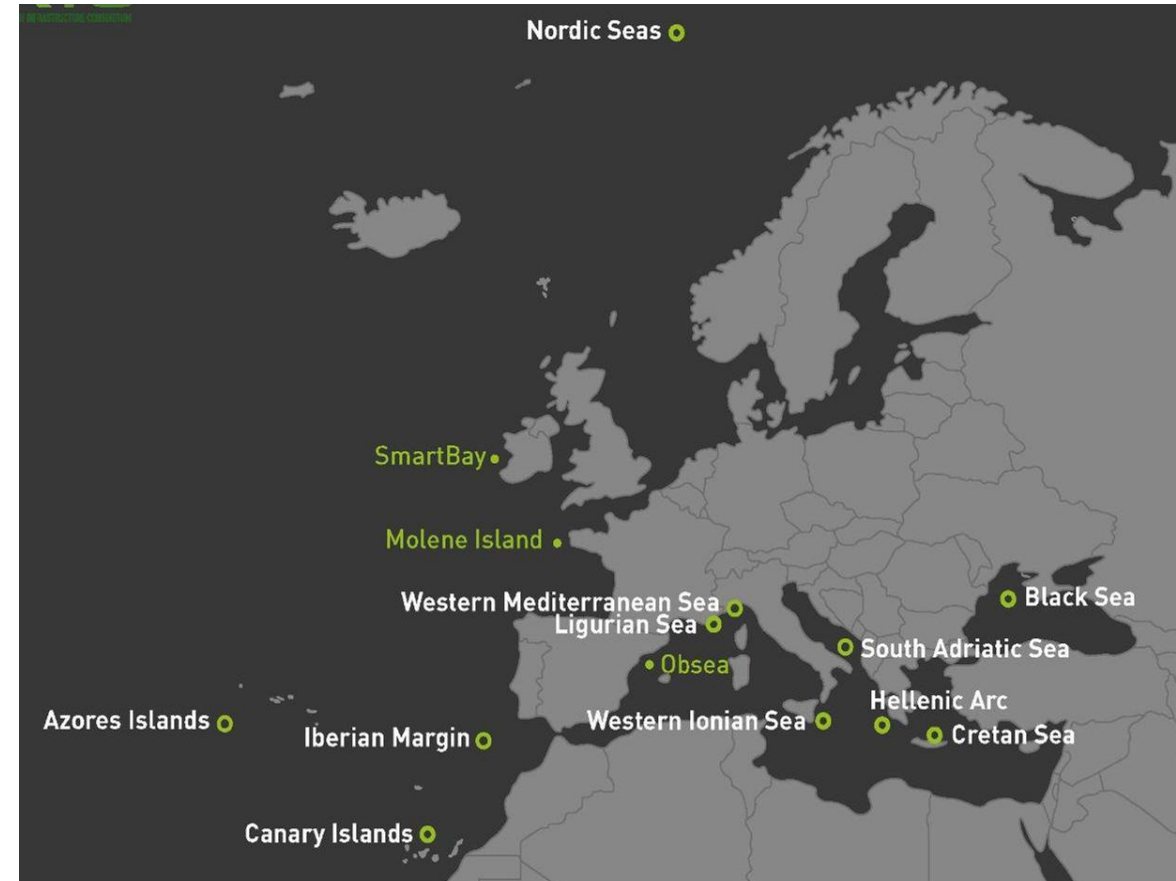
*As a European marine research infrastructure, EMSO ERIC aims at the advance in the knowledge of deep ocean and water column processes in key oceanic **regions** in the context of **global change**.*

*The **operational** scope of EMSO is the provision of **services** for both the **long-term** repeated **observations** and **analysis** of Essential Ocean Variables as well as other physical and environmental variables retrieved by deep-sea observatories.*

EMSO delivers Findable, Accessible, Interoperable and Reusable data.

EMSO ERIC has to render them Visible, Sustainable and Inspiring thanks to dedicated tools and products for the research community and the society.

14 Observatories, 27 European Institutions



EMSO ERIC – Introduction

2035 Vision of EMSO ERIC

A formulation to be validated

“The sustainability of our ocean stands in our **integrative** capacity to observe, study, understand and advocate it from its deepest part to the upper water column on long terms.

In that context, by 2035, EMSO will be recognised as the European reference for the **long-term observation and analysis** of the marine environment **variability** and marine geohazards from the seabed to the water column, in key marine **regions.**”

EMSO Data Infrastructure

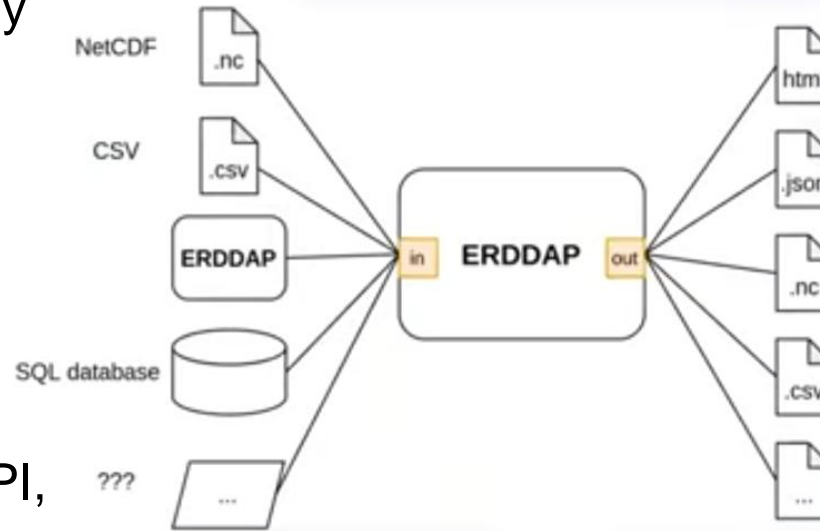
- **Initial Approach:** Centralized data infrastructure complemented by a harmonization system.

Problem: Heterogeneous data would require significant investment of time and effort from RF to achieve harmonization.

- **Actual Approach:** Federated data infrastructure utilizing **ERDDAP** for enhanced data distribution and retrieval.

EMSO Data Infrastructure: ERDDAP overview

- Open-Source Software developed by **NOAA**
- Input data: NetCDF, CSV, acoustic, databases...
- Output formats: NetCDF, CSV, JSON, images...
- User Graphical Interface & REST API, data download, interactive plotting...
- Supports the **creation of federated networks** of ERDDAP servers

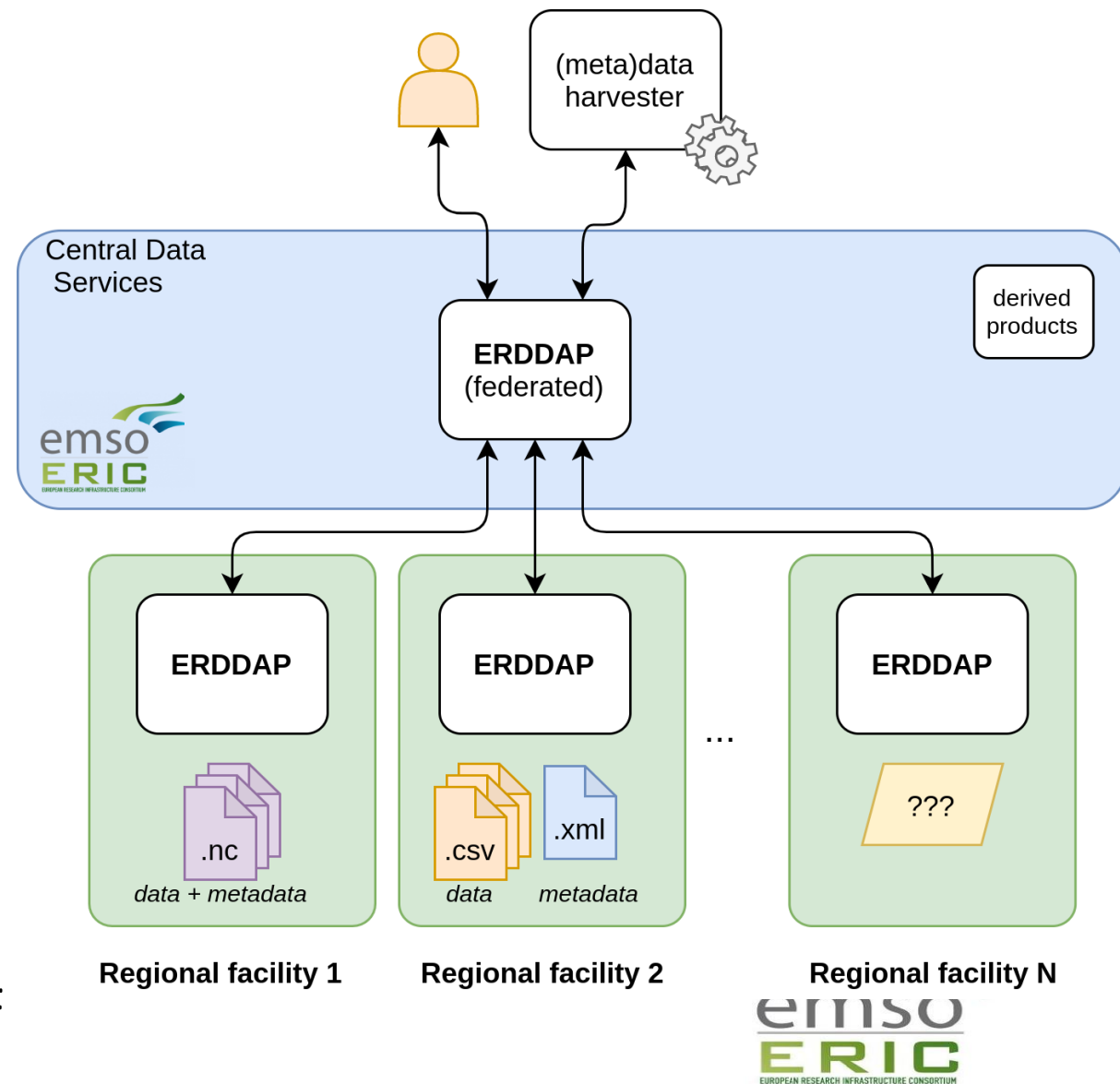


ERDDAP Service

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

EMSO Data Infrastructure: ERDDAP

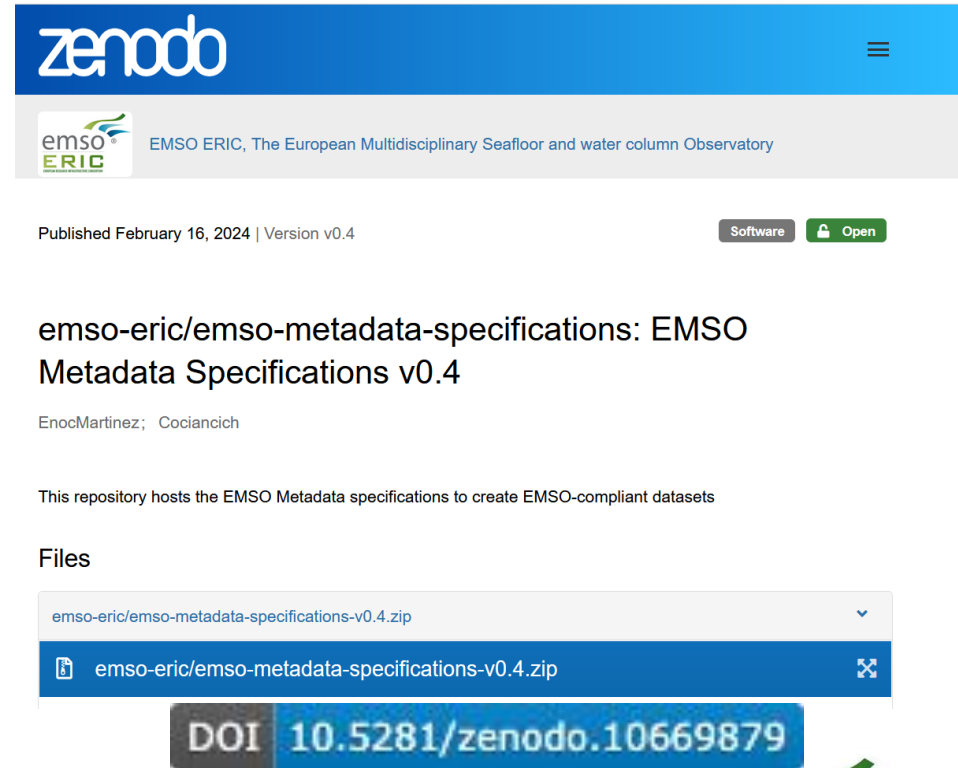
- **ERDDAP**: EMSO's data backbone
- ERDDAP servers deployed at regional facilities
- **Central ERDDAP** links regional facilities
- Data Managed at RF level
- EMSO provides instructions on **GitHub** how to deploy ERDDAP <https://github.com/emso-eric/erddap>
- Data needs to be fully **harmonized** at every RF



EMSO Metadata Specifications Document

- Based on the **OceanSITES Data Format** (NetCDF, CF, ACDD) but adapted to the needs of EMSO.
- Available in **GitHub** and **Zenodo**
- In March 2023: **~50** datasets, **~ 30 % data harmonized**
- In November 2024: **~170** datasets, **86 % data harmonized**

About 130 variables harvested
based on EoV



The screenshot shows the Zenodo repository page for the project 'emso-eric/emso-metadata-specifications: EMSO Metadata Specifications v0.4'. The page features the Zenodo logo at the top left, the project name, and the version number. It also displays the publication date (February 16, 2024) and the version (v0.4). The page includes a 'Software' label and an 'Open' button. The description states: 'This repository hosts the EMSO Metadata specifications to create EMSO-compliant datasets'. The 'Files' section shows a single file: 'emso-eric/emso-metadata-specifications-v0.4.zip'. Below the file list, a DOI is displayed: 'DOI 10.5281/zenodo.10669879'.

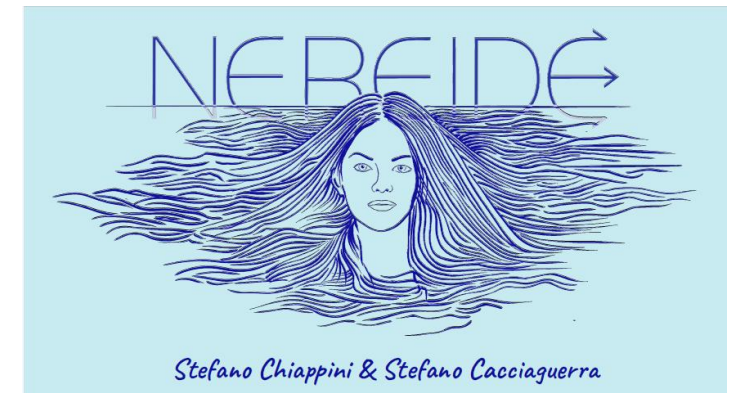
Upcoming Developments in EMSO ERDDAP

- Not all Regional Facilities are currently federated.
- We **plan to federate** the **Black Sea, PLOCAN, and Iberian Margin** facilities shortly. Some of these RFs have employed **new personnel** specifically for this purpose.
- **Integrating additional data sources** (images, DAS)

NEREIDE:

EMSO Data Management - recent rearrangements

- **NEw REsearch Infrastructure Datacenter for EMSO**
- Stefano **Chiappini** and Stefano **Cacciaguerra** INGV
- **Funded** by the Italian National Operational Programme (PON) InSEA,
- Located in **Sicily**
- Number of cpu cores: **2600**
- Memory capabilities: **3.6 PB = 3600 TB** raw disk
- Virtual Research Environment

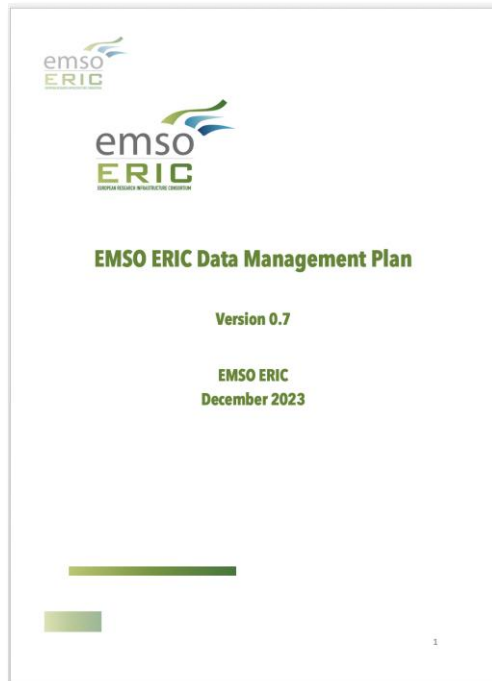


[Guide document](#) for NEW users

**Thank you for your
time and attention!**

Federated: EMSO Data Management Plan

ESFRI Monitoring Report: “EMSO ERIC has a very **well elaborated** DMP that is in line with the FAIR principles and could serve as a **template** for other RI.”



DOI 10.5281/zenodo.10697877

Regional Facility DMP	DOI that points to the latest version of the document.
Canary Islands	https://doi.org/10.5281/zenodo.10589053
SmartBay	https://doi.org/10.5281/zenodo.10523017
Hellenic Arc and Cretan Sea	https://doi.org/10.5281/zenodo.10473373
South Adriatic Sea	https://doi.org/10.5281/zenodo.10401727 https://doi.org/10.5281/zenodo.10412750
Western Mediterranean Sea	https://doi.org/10.5281/zenodo.10391619
OBSEA	https://doi.org/10.5281/zenodo.10417096
Ligurian Sea	https://doi.org/10.5281/zenodo.10478953
Iberian Margin	https://doi.org/10.5281/zenodo.10672121
Nordic Seas	https://doi.org/10.5281/zenodo.10454442