OBSERVING THE OCEAN TO SAVE THE EARTH

EMSO ERIC IT and Data Services

Ivan Rodero (ivan.rodero@emso-eu.org)

Geo-INQUIRE Seminar

11 May 2023 (Virtual)

EMSO ERIC – Introduction

EMSO ERIC was constituted on September 29, 2016 (EU Official Journal L268/59 October 1st, 2016)

A **CENTRAL HUB** AND INTERLINKED **FIXED POINT MUTI-SENSORS PLATFORMS**

EXCELLENCES

Delivering standardised: services, data process, scientific results.

High level of **INTEGRATION** among multi-sensor platforms and European institutions.

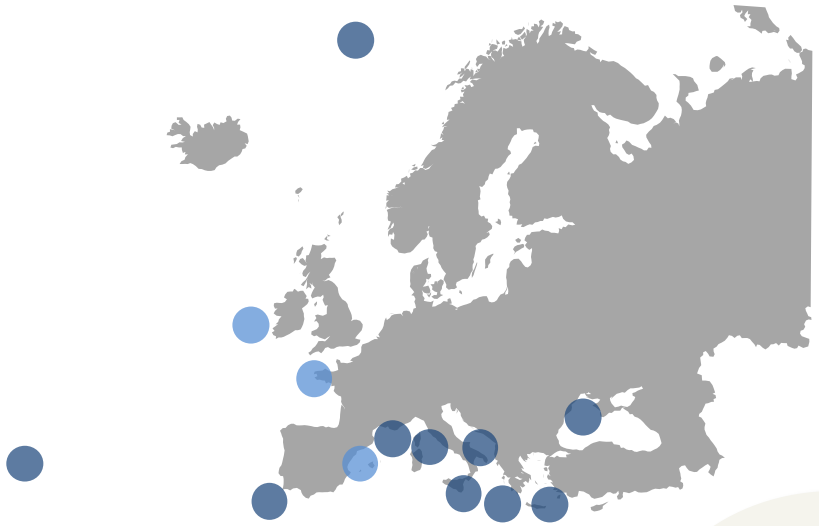
Common access policy and a single point of access for all users.

User programme designed to absorb capacity of the RI.

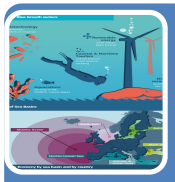
ADDED VALUE compared to the value of a single research cooperation network.

JOINT INVESTMENT STRATEGY to strengthening EMSO ERIC through its regional facilities/test sites and common and shared services.

DISTRIBUTED RESEARCH INFRASTRUCTURE



MISSION



To establish a comprehensive and smart sensor system in water column, seafloor, and sub-seafloor environments as part of the integrated and sustainable organization EMSO ERIC

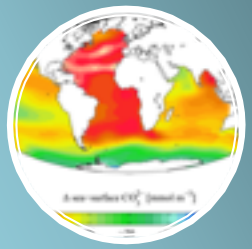
Science
Data Management
Engineering & logistics
Etc.

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Data & IT Supporting Scientific Challenges

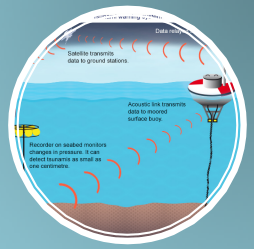
Research Infrastructure Challenges

To fulfil European societal scientific demands targeted in the EU´s H2020 Blue Growth Strategy

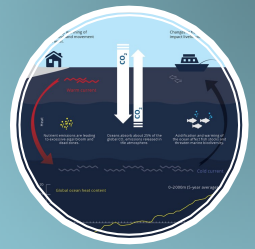


Global ocean warming and acidification

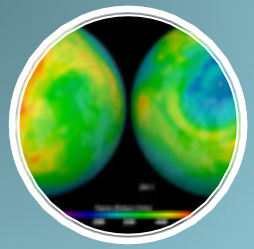
Impact and sustainability of Marine Resources exploitation



Real-time observations and early warning systems for earthquakes & tsunamis



Marine Ecosystems and Climate Change mitigation



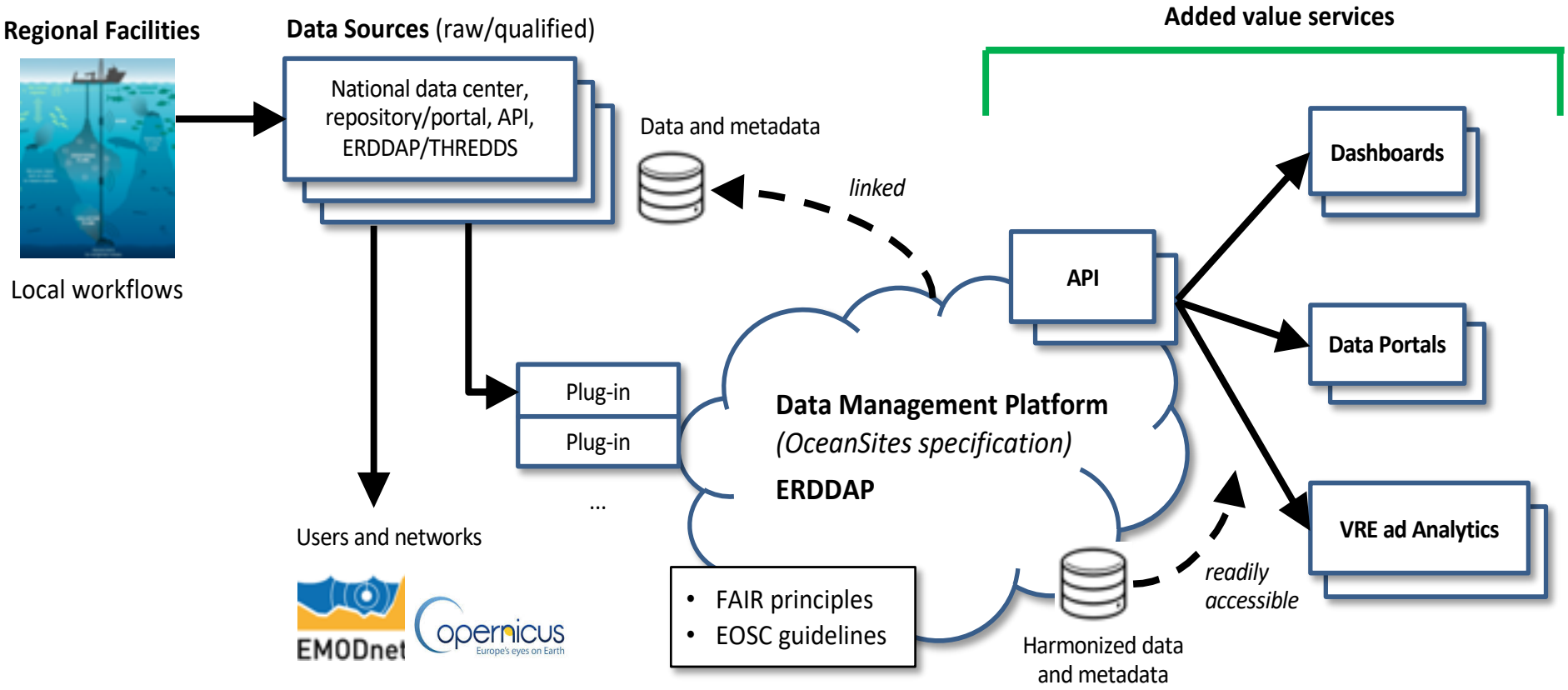
Earth interactions hydrosphere, biosphere,

lithosphere, atmosphere

Access HIGH QUALITY MARINE ENVIRONMENTAL DATA

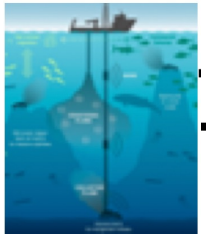
emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Data Management Architecture



Overarching Data Workflow

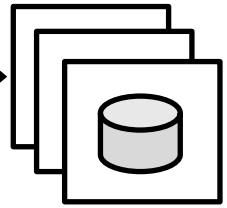
Regional Facilities



Local workflow

Data Sources

(raw/qualified)



National data center
Repository/portal
API
ERDDAP/THREDDS

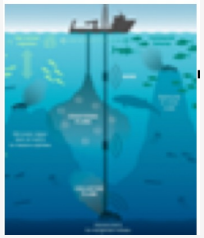
Data Networks
(e.g., EMODNet, Copernicus)

- Automated/on-demand **data products**
- Multi-node **data-based services**
- Harmonized/interoperable/standardized data/metadata
- Data portals/dashboards
- Interface with ENVRI/EOSC
- Contribute to EOOS
- Support for linked data/semantic queries
- Virtual Research Environment
- Community engagement/training
- Archival/backup, fail-over, config management, cyber-security

Data Workflow (What)

EMSO ERIC

Regional Facilities



Local workflow

Data Sources
(raw/qualified)

National data center
Repository/portal
API
ERDDAP/THREDDS

Value-added Products and Services

- Automated/on-demand data products
- Multi-node data-based services
- Data portals
- QC services
- Dashboards/visualization
- Virtual Research Environments
- Community engagement/training
- (…)

Networks/aggregators/repositories

Copernicus Europe's eyes on Earth

EMODnet

PANGAEA.

SeaDataNet

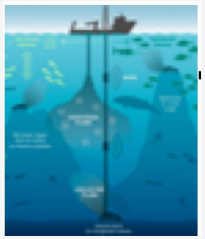
Metadata/data access
Data products/services

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Data Workflow (How)

EMSO ERIC

Regional Facilities



Local workflow

Data Sources
(raw/qualified)

National data center
Repository/portal
API
ERDDAP/THREDDS

Examples:
- Multi-node data visualization (e.g., EOVs)
- Multi-node data availability

Typical steps:
1. Download data from multiple sources
   - Different interfaces/mechanisms
2. Harmonize/clean data (format, units, etc.)
3. Generate the plots (e.g., Matlab, R, etc.)
4. Publish the plots/products
5. Repeat the process for new data

Value-added Products and Services

- Automated/on-demand data products
- Multi-node data-based services
- Data portals
- QC services
- Dashboards/visualization
- Virtual Research Environments
- Community engagement/training
- (…)

Networks/aggregators/repositories

Copernicus Europe's eyes on Earth

EMODnet

PANGAEA.

SeaDataNet

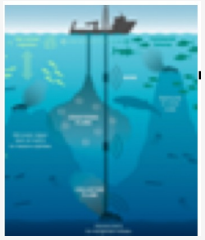
Metadata/data access
Data products/services

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Data Workflow (Implementation)

EMSO ERIC

Regional Facilities



Local workflow

Data Sources
(raw/qualified)

National data center
Repository/portal
API
ERDDAP/THREDDS

Data Management Platform + ERDDAP

- ✓ Tailored harmonization (EMSO)
- ✓ Rich APIs
- ✓ Robust and scalable
- ✓ QoS guarantees
- ✓ Full traceability
- ✓ EU-funded (e.g., EMSO legacy)
- ✓ Wide community

ALREADY AVAILABLE

API

Value-added Products and Services

- Automated/on-demand data products
- Multi-node data-based services
- Data portals
- QC services
- Dashboards/visualization
- Virtual Research Environments
- Community engagement/training
- (…)

Harmonized data/metadata

Networks/aggregators/repositories

- 3rd party "harmonization"
- 3rd party products/services

✓ Wide community
✘ Limited/no traceability
✘ Lack of control
✘ QoS guarantees?

Copernicus Europe's eyes on Earth
EMODnet
PANGAEA.
SeaDataNet

Metadata/data access
Data products/services

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Data Services (“External”)

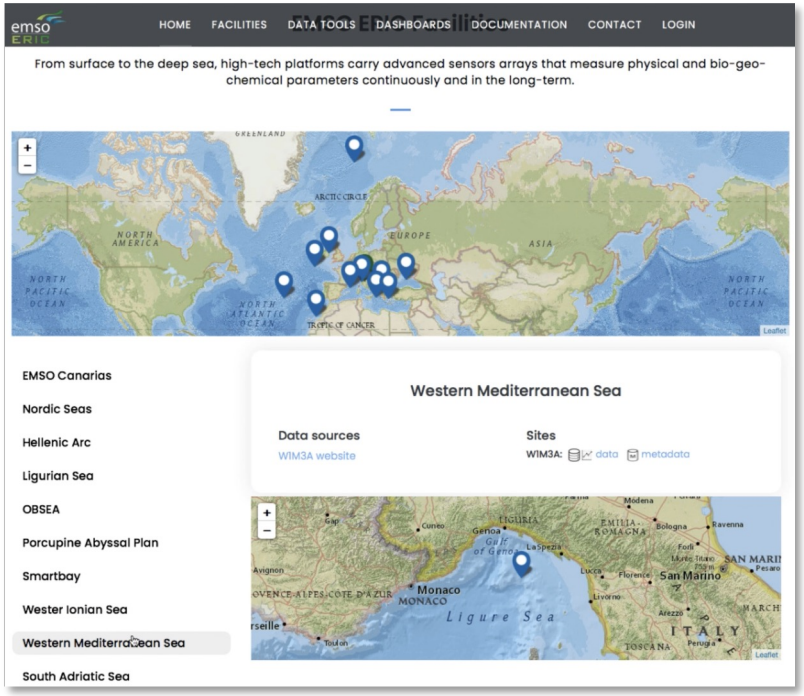
UPDATES

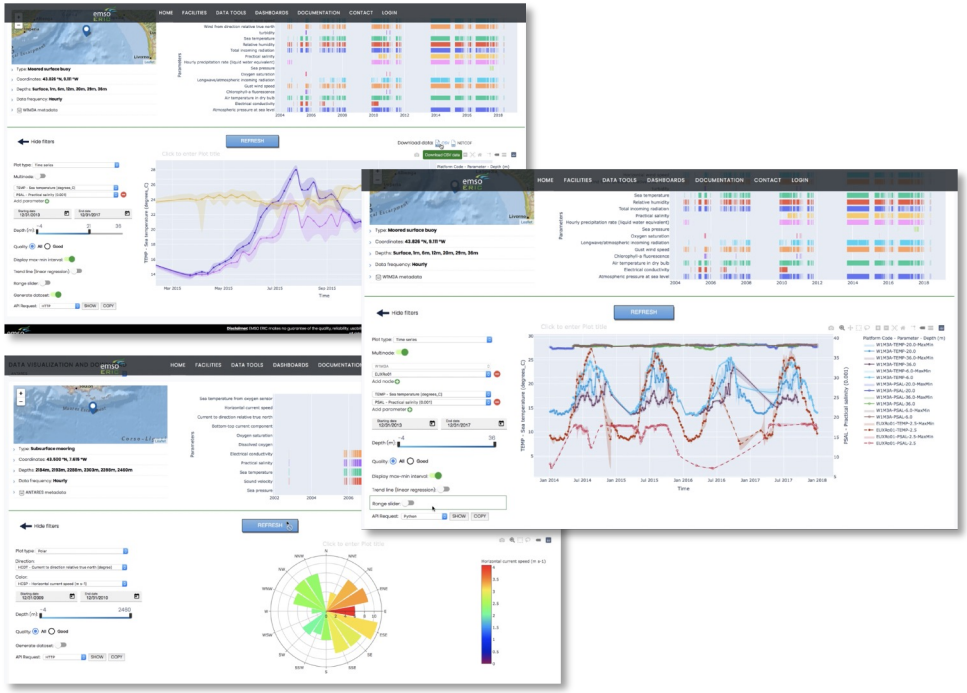
1. Data portal for data access and visualization
   - **New functionalities**
   - **Performance improvement**
2. Federated ERDDAP
   - **New deployments (e.g., ENVRI-FAIR demonstrator)**
   - **Increased data availability and access**
3. Authentication and Authorization Infrastructure (AAI)
4. Dashboards
   - **Increased data availability and access**
5. Application Programming Interface (API)
   - **Internal API for EMSO ERIC functionalities (additional security layer)**
6. Virtual Research Environment (VRE)
7. File explorer
8. Quality control toolset
9. Derived data product generation

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Data Portal and Services

Data access, visualization, download, reusability (PID/DOI)

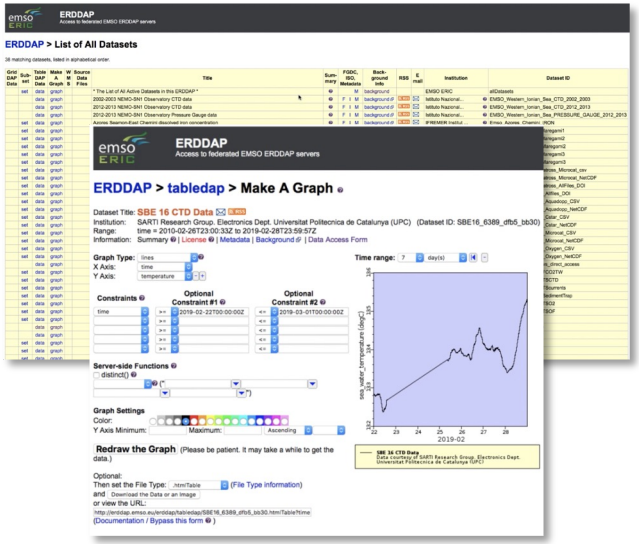




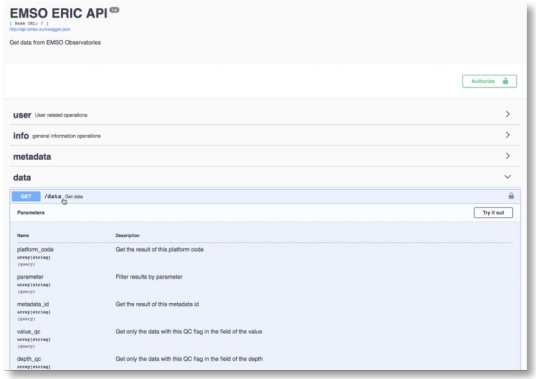
https://data.emso.eu

Data Portal and Services (2)

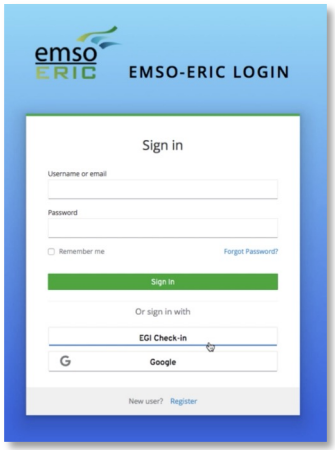
Federated ERDDAP (erddap.emso.eu)



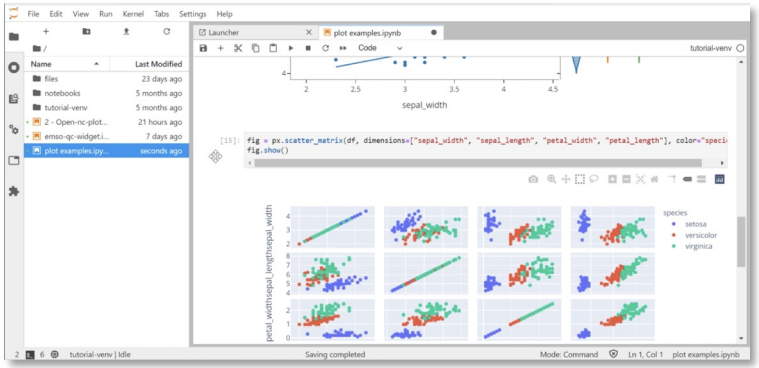
API (api.emso.eu)



AAI integration with EOSC (EGI Check-in)



Virtual Research Environment (jupyter.emso.eu)



- Dashboards
- File explorer
- Quality control toolset
- Derived data product generation
- PID management and more!

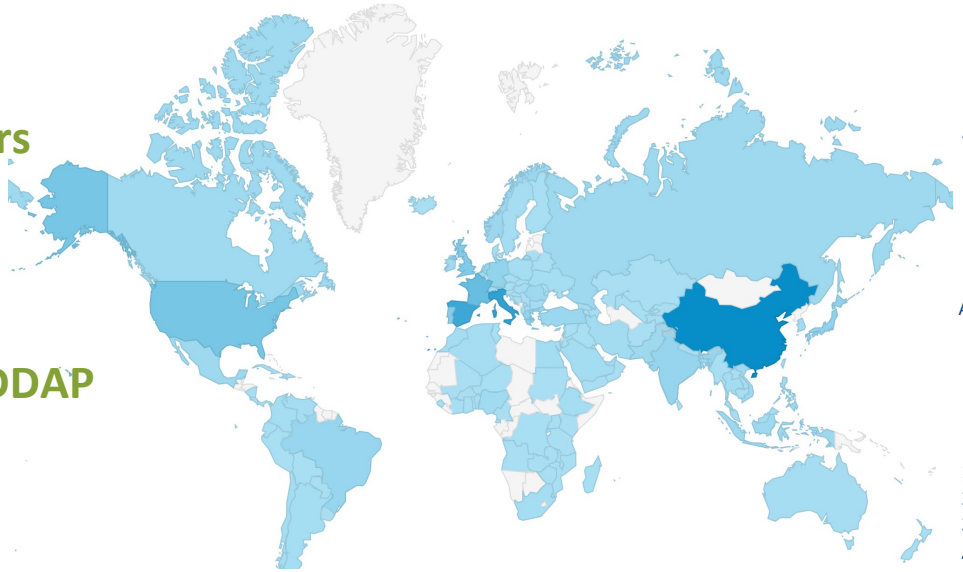
Operations in Pre-production (2020-Q3/2023-Q1)

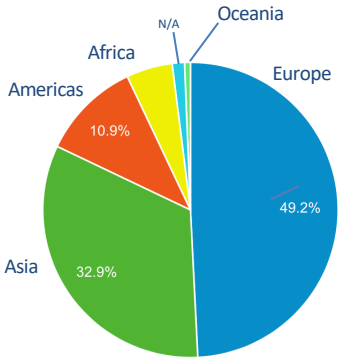
134 Countries

3.881 Distinct users

15.422 Page views

220.931 API and ERDDAP requests





1. China
2. Italy
3. Spain
4. France
5. USA
6. United Kingdom
7. Portugal
8. Japan
9. Greece
10. Others

KPI	2020-Q3 / 2023-Q1
Number of countries reached	**134**
Number of distinct users	**3.881**
Data portal page views	**15.422**
API and ERDDAP requests	**220.931**

Data Services (“Internal”)

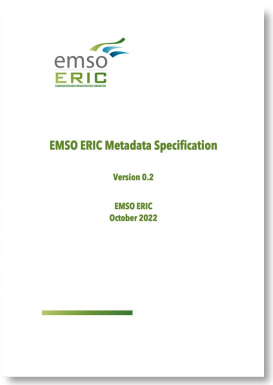
UPDATES

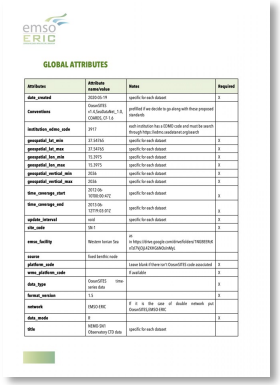
1. (Meta)data harmonization
   - **Initial specification available (living document)**
   - **Specification deployment undergoing (Regional Facilities)**
2. PID management
   - **Target: data queries/plots**
   - **Testing live (with warning notice)**
3. EMSO ERIC DOI service
   - **Internal testing**
4. NetCDF ingestion tool
   - **Internal testing**
5. Data management plan (consolidation process undergoing)
6. ERDDAP Checker
7. Training and capacity building / co-creation of code (DMSG)
8. Etc.

(Meta)data Harmonization

- Specification document draft (living document)

  (to be published in Zenodo)





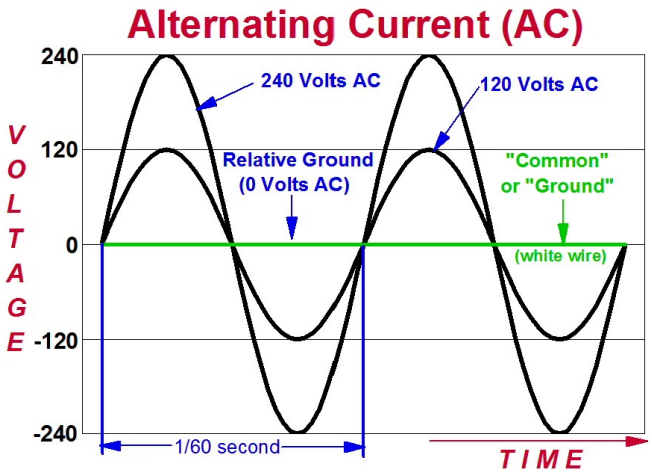
- Harmonization deployment progress (DMSG)
  - EMSO Metadata Specifications v0.1
    - Based on OceanSITES specifications v1.4
    - Focus on NetCDF files
    - E.g., CTD data
  - Leverages existing standards and formats
    - OceanSITES, SeaDataNet
    - Climate & Forecast (CF)
    - BODC Vocabularies

Name	Definition
TIME	Number of time steps
DEPTH	Number of depth levels
LATITUDE	Dimension of LATITUDE co-ordinate
LONGITUDE	Dimension of LONGITUDE co-ordinate

Name	Attributes	Attribute name/value
TEMP	long_name	temperature
	standard_name	sea_water_temperature
	units	degrees C
	comment	
	coordinates	TIME DEPTH LATITUDE LONGITUDE
	ancillary_variables	TEMP_QC
	_FillValue	-999
	sdn_parameter_name	Temperature of the water body
	sdn_parameter_urn	SDN:P01::TEMPPR01

Name	Attributes
TIME	long_name
	standard_name
	units
	axis
	ancillary_variables
	sdn_parameter_name
	sdn_parameter_urn
	sdn_uom_name

Specification



OceanSITES

Taking the pulse of the global ocean

Implementation

emso ERIC

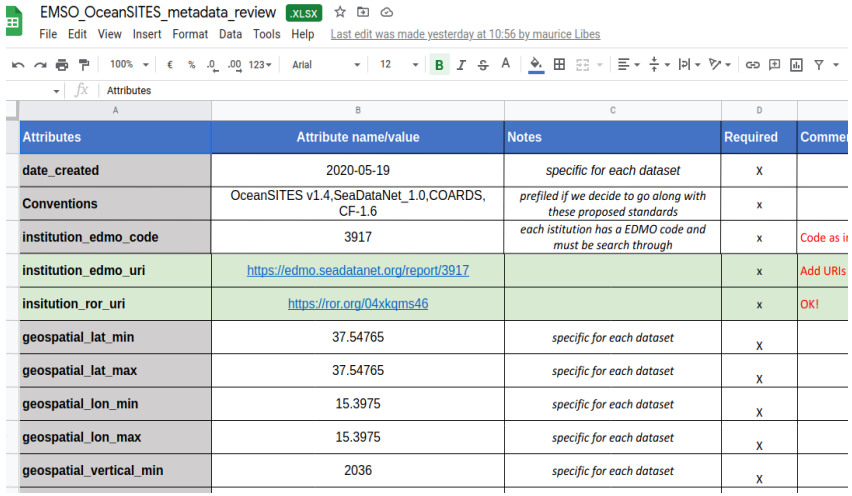
OceanSITES

Taking the pulse of the global ocean

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Metadata Specification

- New EMSO Metadata Specifications
- Based on v0.1 specifications
- Output of Harmonization working group:
  - Focus on NetCDF and **ERDDAP**
  - Generic for all Timeseries data



EMSO_OceanSITES_metadata_review .XLSX

File Edit View Insert Format Data Tools Help Last edit was made yesterday at 10:56 by maurice Libes

Attributes

Attributes	Attribute name/value	Notes	Required	Comme
date_created	2020-05-19	*specific for each dataset*	x	
Conventions	OceanSITES v1.4,SeaDataNet_1.0,COARDS, CF-1.6	*prefiled if we decide to go along with these proposed standards*	x	
institution_edmo_code	3917	*each istitution has a EDMO code and must be search through*	x	Code as i
institution_edmo_uri	https://edmo.seadatanet.org/report/3917		x	Add URIs
insitution_ror_uri	https://ror.org/04xkqms46		x	OK!
geospatial_lat_min	37.54765	*specific for each dataset*	x	
geospatial_lat_max	37.54765	*specific for each dataset*	x	
geospatial_lon_min	15.3975	*specific for each dataset*	x	
geospatial_lon_max	15.3975	*specific for each dataset*	x	
geospatial_vertical_min	2036	*specific for each dataset*	x	

ERDDAP Metadata Specification

This document includes a list of all the metadata terms required for a dataset to be compliant with the EMSO Metadata Specifications. The format is based on the 'OceanSITES Data Format reference Manual v1.4', but adapted to the needs of EMSO ERIC (European Seafloor and water-column Obseratory) and its federated data service based on ERDDAP.

Version: 0.1
Creation Date 2023-03-06
Last modification 2023-03-07

Genera

The followi

1. When

Global Attributes	Description	Compliance test	Required
date_created	Creation date	data_type#str	true
Conventions	conventions used in thd dataset	data_type#str	false
institution_edmo_code	EDMO code of the creator's organization	edmo_code	true
geospatial_lat_min	The southernmost latitude, a value between -90 and 90 degrees	coordinate#latitude	true

- Move specs to EMSO GitLab (on-going work)
  - Publicly available
  - Version control
  - Human readable (web)
  - Machine-actionable
  - Defines a set of compliance tests

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Metadata Specification (2)

- **EMSO Metadata Harmonizer tool**
  - Ensures compliance with EMSO Metadata specifications
  - Cross-platform (written in python3)
  - Preliminary version at GitLab

- **Metadata Harmonizer workflow:**
  - Accesses specifications from gitlab
  - Crawls through ERDDAP's datasets
  - Checks compliance with EMSO specs:
    - EMSO, OceanSITES, BODC, CF
  - Provides harmonization report

ERDDAP test report

variable	attribute	required	passed	message	value
global	date_created	True	False	not found	
global	Conventions	False	True		COARDS, CF-1.6, ACDD-1.3
global	institution_edmo_code	True	True		2150
global	geospatial_lat_min	True	True		41.18212
global	geospatial_lat_max	True	True		41.18212
global	geospatial_lon_min	True	True		1.75257
global	geospatial_lon_max	True	True		1.75257
global	geospatial_vertical_min	True	False	not found	
global	geospatial_vertical_max	True	False	not found	
global	time_coverage_start	True	True		2010-03-12T09:30:00Z
global	time_coverage_end	True	True		2023-03-23T09:30:00Z
global	site_code	True	True		OBSEA
global	emso_facility	False	False	not found	
global	source	False	False	not found	
global	title	True	True		CTD data from a SBE16 at OBSEA (30min average)
global	summary	True	True		CTD data collected by a SBE16 deployed at OBSEA, 30min average
global	principal_investigator	True	False	not found	
global	principal_investigator_email	True	False	not found	
global	license	True	True		Creative Commons Attribution 4.0 International (CC-BY-4.0)

```
Required tests passed: 39 of 120
Required tests passed: 1 of 15
   Total tests passed: 40 of 135
Required tests...  32%
Optional tests...  7%
Total tests...     30%
```

EOSC-Future Environment Dashboard

General information

- Implementation divided in two components:
  - Back-end, with a RESTful API

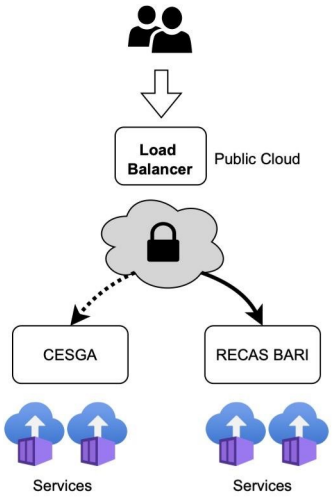
    https://env-dashboard.eoscfuture.eu:4000
  - Responsive front-end

    https://env-dashboard.eoscfuture.eu
- Source code on git-based repository -> available at:

  https://gitlab.emso.eu/eosc-future/
- Deployment and operations of the IT infrastructure using industry best practices and EOSC services:
  - High-availability
    (redundancy for failover and business continuity)
  - Deployment of independent services via containers
  - EGI Cloud Compute
    (https://marketplace.eosc-portal.eu/services/egi-cloud-compute)
  - AAI (currently EGI Check-in)
    (https://marketplace.eosc-portal.eu/services/egi-check-in)

IT Infrastructure

- Layer 4 load balancer sitting outside the perimeter
  - Failover is of a whole data center or individual services

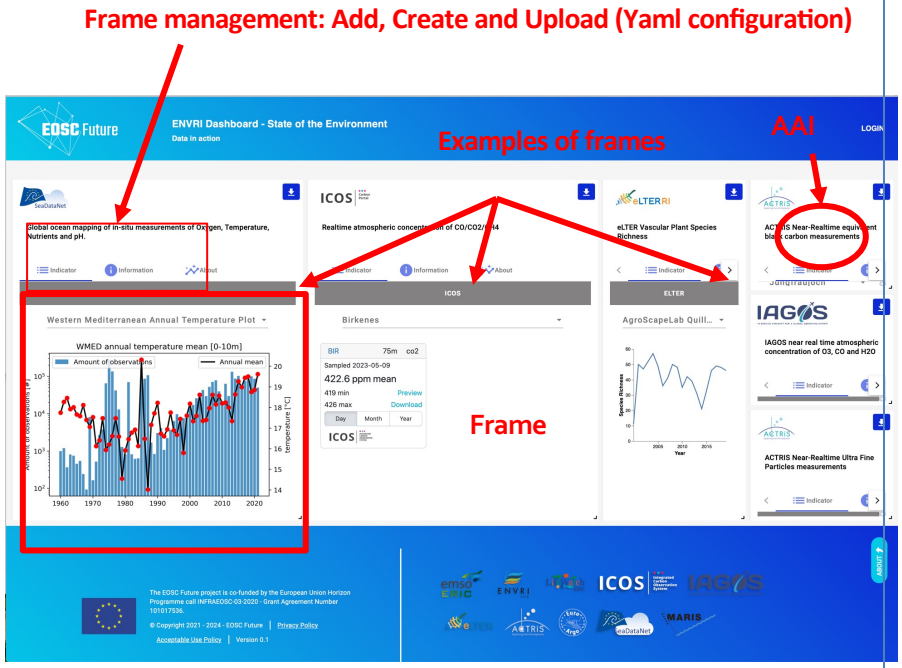


- Two data centers provided by EGI:
  - RECAS BARI (Italy)
  - CESGA (Spain)

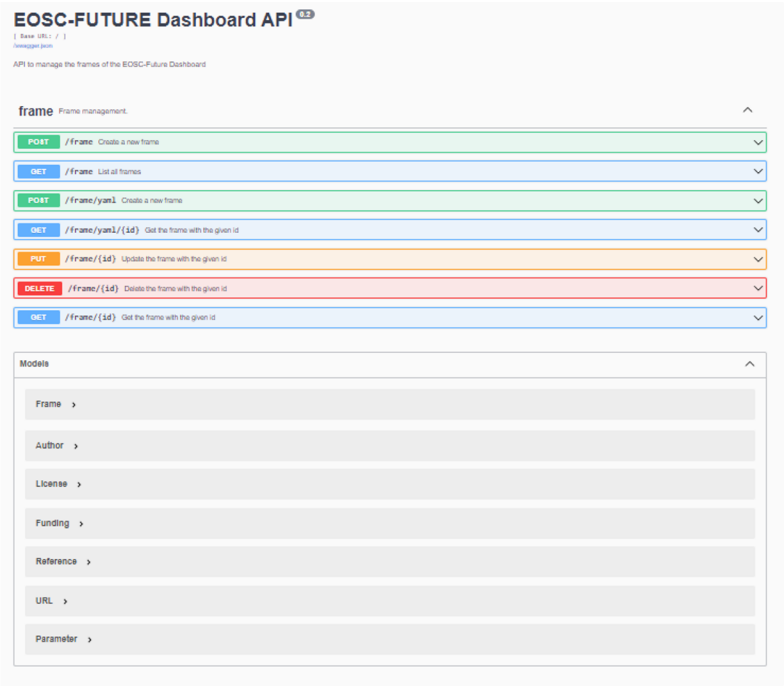
- EOSC core services integration undergoing:
  - EOSC Marketplace registration
  - Helpdesk
  - Monitoring

EOSC-Future Environment Dashboard (2)

Dashboard (Portal)



API with interactive documentation



https://env-dashboard.eoscfuture.eu/dashboard

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM

Thank you for your attention

www.emso.eu

Geo-iNQUIRE

ENVRI FAIR

EGI-ACE

EuroSea

MINKE

EOSC Future

emso ERIC
EUROPEAN RESEARCH INFRASTRUCTURE CONSORTIUM