

EMSO ERIC Physical and biogeochemical data curation and use

DOORS RI and data training

April 2025

Aljaz Maslo

EMSO ERIC Data Officer

aljaz.maslo@emso-eu.org

EMSO ERIC – Introduction

Distributed European RI

Network of **seafloor** and **water-column**
observatories

8 Member States, **14** RF, **27** Eu Institutions

~ **130** variables

Goal: offering **FAIR** data & services to scientific
community



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 ERIC Data Infrastructure: ERDDAP

ERDDAP: EMSO's data backbone

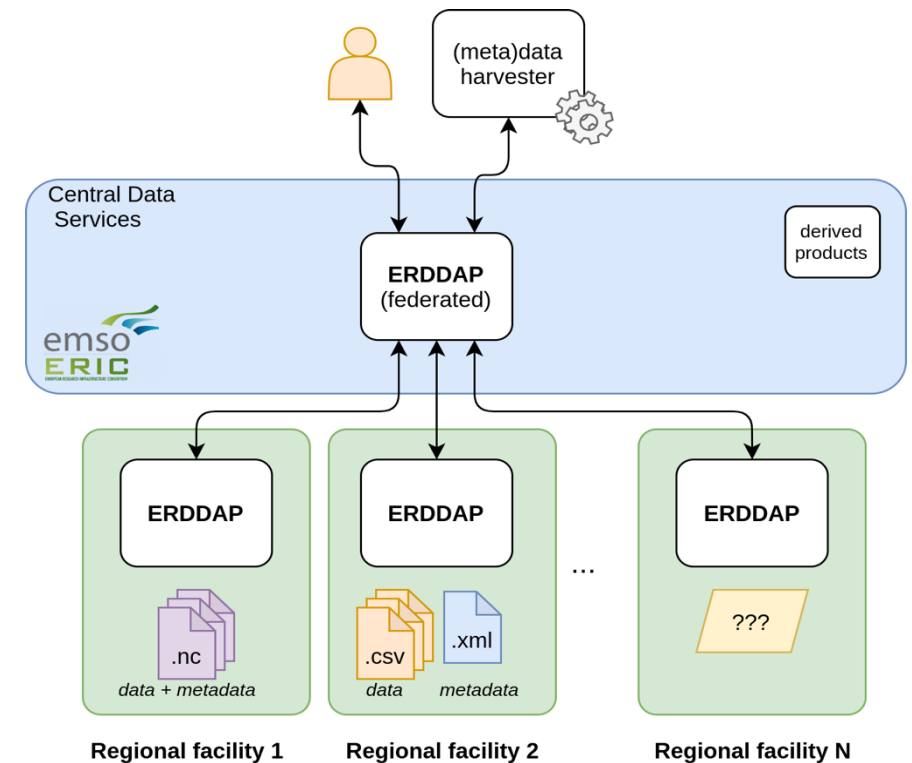
Open-Source Software developed by **NOAA**

ERDDAP servers deployed at regional facilities

Central ERDDAP links regional facilities

Data Managed at RF level (collected, stored)

Data needs to be fully **harmonized** at every RF



EMSO Metadata Standard

Custom EMSO Metadata Specification Document

Based on the global OceanSITES Data Format but adapted to the needs of EMSO. Version control, human readable, machine-actionable, defines a set of compliance tests

Available in **GitHub** and **Zenodo**



EMSO ERIC choices on Vocabularies

**EMSO metadata standard
makes use of several
controlled vocabularies from
the NERC Vocabulary Service**

hub.com/emso-eric/emso-metadata-specifications/ 100% ☆

emso-metadata-specifications / EMSO_metadata.md

Preview Code Blame 198 lines (163 loc) · 20.1 KB Raw ↗ ↕

Controlled Vocabularies

This metadata standard makes use of several controlled vocabularies from the [NERC Vocabulary Service](#)

vocab ID	description	URL
EDMO	European Database of Marine Organizations	EDMO
ROR	Research Organization Registry	ROR
P01	parameter vocabulary	P01
P02	parameter codes	P02
P06	units	P06
L05	sensor types	L05
L06	platform types	L06
L22	sensor models	L22
L35	sensor manufacturers	L22
SPDX	software licenses	github

EMSO Metadata Standard

EMSO Quality Control (QC) same as OceanSITES

QC: Temperature 68 %, Salinity 64 %, Dissolved Oxygen 7 %

Quality Control Attributes

QC Attributes	Description	Compliance test	Required
long_name	human-readable label, it is suggested to use parameter long_label and add "quality control flags" at the end	data_type#str	true
conventions	OceanSITES QC Flags	data_type#str	true
flag_values	0,1,2,3,4,5,6,7,8,9	data_type#str	true
flag_meanings	unknown good_data probably_good_data potentially_correctable_bad_data bad_data nominal_value interpolated_value missing_value	data_type#str	true

EMSO FAIR Data

FAIR:

Findable: 53 % DOI,

Accessible: ERDDAP public URL

Interoperable: Use standard data formats and vocabularies

Reusable: license CC-BY 76 %, CC-BY-NC 20 %, CLIVAR 4 %

EMSO ERIC Metadata Service

EMSO Metadata Harmonizer tool:

Checks compliance with EMSO specs.

Crawls through ERDDAP's datasets

Tips to improve harmonization

Provides harmonization **score**

Available in GitHub

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%

EMSO ERIC Metadata Service

EMSO Metadata Harmonizer tool:

General report

dataset_id	emso_facility	institution	total	required	optional
EMSO_OBSEA_Buoy_Ai	OBSEA	Polytechnic University of C	85	91	68
EMSO_OBSEA_CTD_30i	OBSEA	Polytechnic University of C	91	99	68
EMSO_OBSEA_CTD_full	OBSEA	Polytechnic University of C	91	99	68
EMSO_OBSEA_CTD_SBI	OBSEA	Polytechnic University of C	81	84	71
EMSO_OBSEA_AWAC_c	OBSEA	Polytechnic University of C	83	90	67
EMSO_OBSEA_AWAC_c	OBSEA	Polytechnic University of C	83	90	67
EMSO_OBSEA_Besos_B	OBSEA	Polytechnic University of C	85	91	68
EMSO_OBSEA_Besos_B	OBSEA	Polytechnic University of C	85	91	68
E2M3A_METEO	South Adriatic Sea	National Institute of Ocea	84	100	41
E2M3A_MRDT	South Adriatic Sea	National Institute of Ocea	69	84	33
E2M3A_PCO2PROA	South Adriatic Sea	National Institute of Ocea	77	99	33
E2M3A_PCO2PROW	South Adriatic Sea	National Institute of Ocea	77	99	33
E2M3A_SAMI	South Adriatic Sea	National Institute of Ocea	80	100	35
E2M3A_SBE16PLS	South Adriatic Sea	National Institute of Ocea	83	100	39
E2M3A_SBE37O	South Adriatic Sea	National Institute of Ocea	82	100	38

EMSO ERIC Metadata Service

EMSO Metadata Harmonizer tool:

Specific report
for dataset

attribute	variable	required	passed	message	value
date_created	global	TRUE	TRUE	ok	2023-11-16T11:52:26Z
Conventions	global	FALSE	TRUE	ok	CF-1.10, Copernicus-InSituTAC-ParametersList-3.2.1,
institution_edmo_code	global	TRUE	TRUE	ok	164
institution_edmo_uri	global	TRUE	TRUE	ok	https://edmo.seadatanet.org/report/276
geospatial_lat_min	global	TRUE	TRUE	ok	35.7382
geospatial_lat_max	global	TRUE	TRUE	ok	35.7382
geospatial_lon_min	global	TRUE	TRUE	ok	25.1425
geospatial_lon_max	global	TRUE	TRUE	ok	25.1425
geospatial_vertical_min	global	TRUE	TRUE	ok	-3
geospatial_vertical_max	global	TRUE	TRUE	ok	1000
time_coverage_start	global	TRUE	TRUE	ok	2008-02-27T06:00:00Z
time_coverage_end	global	FALSE	TRUE	ok	2009-07-25T21:00:00Z
update_interval	global	TRUE	TRUE	ok	P1M
site_code	global	TRUE	TRUE	ok	E1M3A
emso_facility	global	FALSE	TRUE	ok	Cretan Sea
source	global	FALSE	TRUE	ok	moored surface buoy
platform_code	global	FALSE	TRUE	ok	E1M3A
wmo_platform_code	global	FALSE	FALSE	not an inte	6101007
data_type	global	FALSE	TRUE	ok	OceanSITES time-series data
format_version	global	FALSE	TRUE	ok	1.4

Harmonization Status

RF harmonization status					
	Regional Facility	Nov 2023	Nov 2024	Jan 2025	Feb 2025
1	SmartBay	25%	78%	81%	75%
2	OBSEA	28%	89%	90%	90%
3	Cretan Sea	71%	87%	87%	87%
4	Ligurian Sea	73%	78%	78%	78%
5	Hellenic Arc	74%	84%	84%	84%
6	Azores	87%	87%	88%	88%
7	Nordic Seas	In progress	99%	99%	98%
8	Western Ionian Sea	96%	100%	100%	100%
9	South Adriatic Sea	100%	89%	89%	89%
10	Western Mediterranean Sea	100%	100%	100%	100%
11	Iberian Margin	96%	In progress	96%	96%
12	Black Sea	In progress	In progress	94%	94%
13	Canary Islands	In progress	In progress	In progress	In progress
	total average	75%	89%	91%	90%

Accessing EMSO data through ERDDAP

Find data

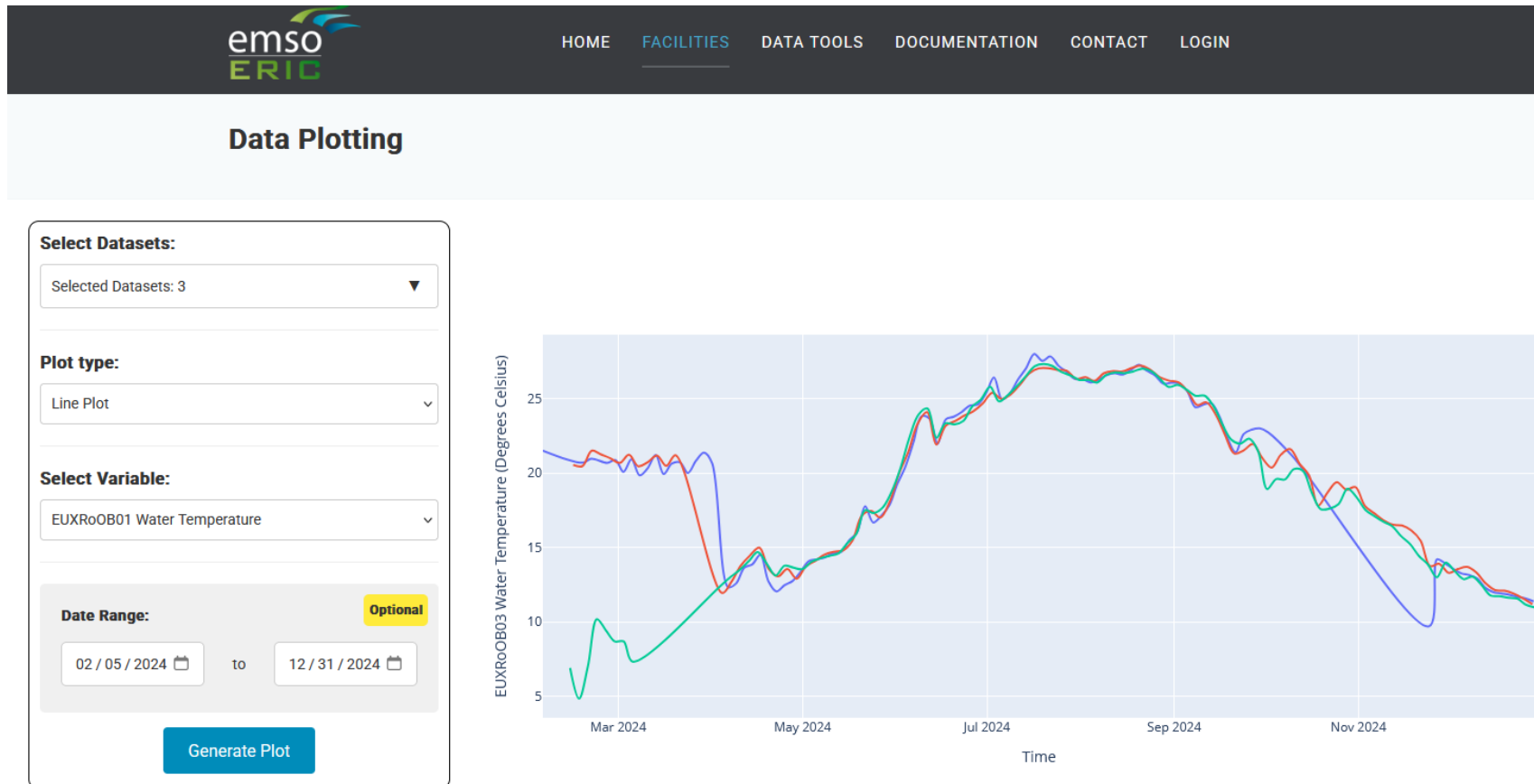
Plot data

Download data/figure

Machine to machine URL

Plotting EMSO data through DataPortal

Simple plots



Why Harmonised FAIR Data in one place, Is Important

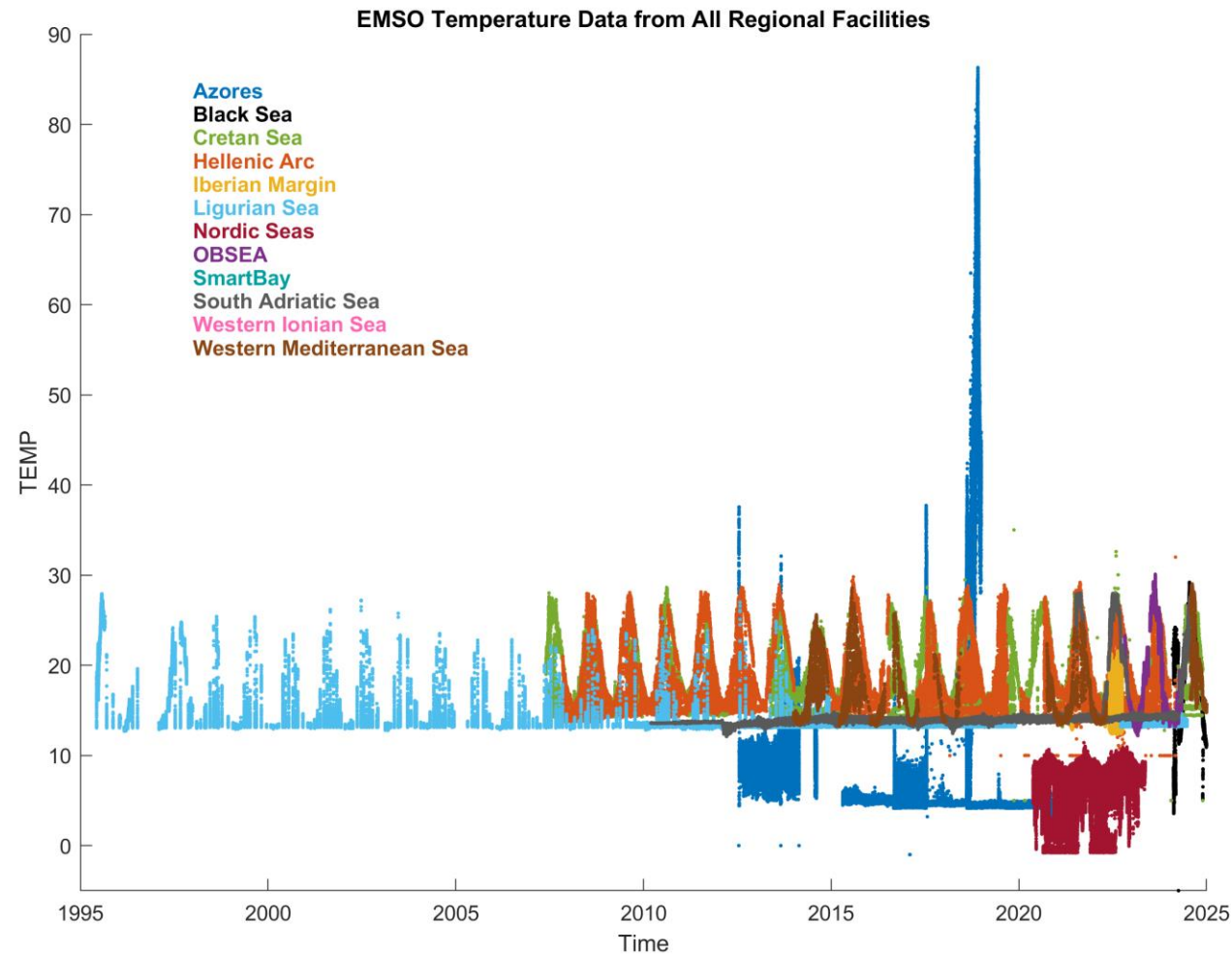
Scientists should spend **time on research**

Less time searching, more time **collaborating**

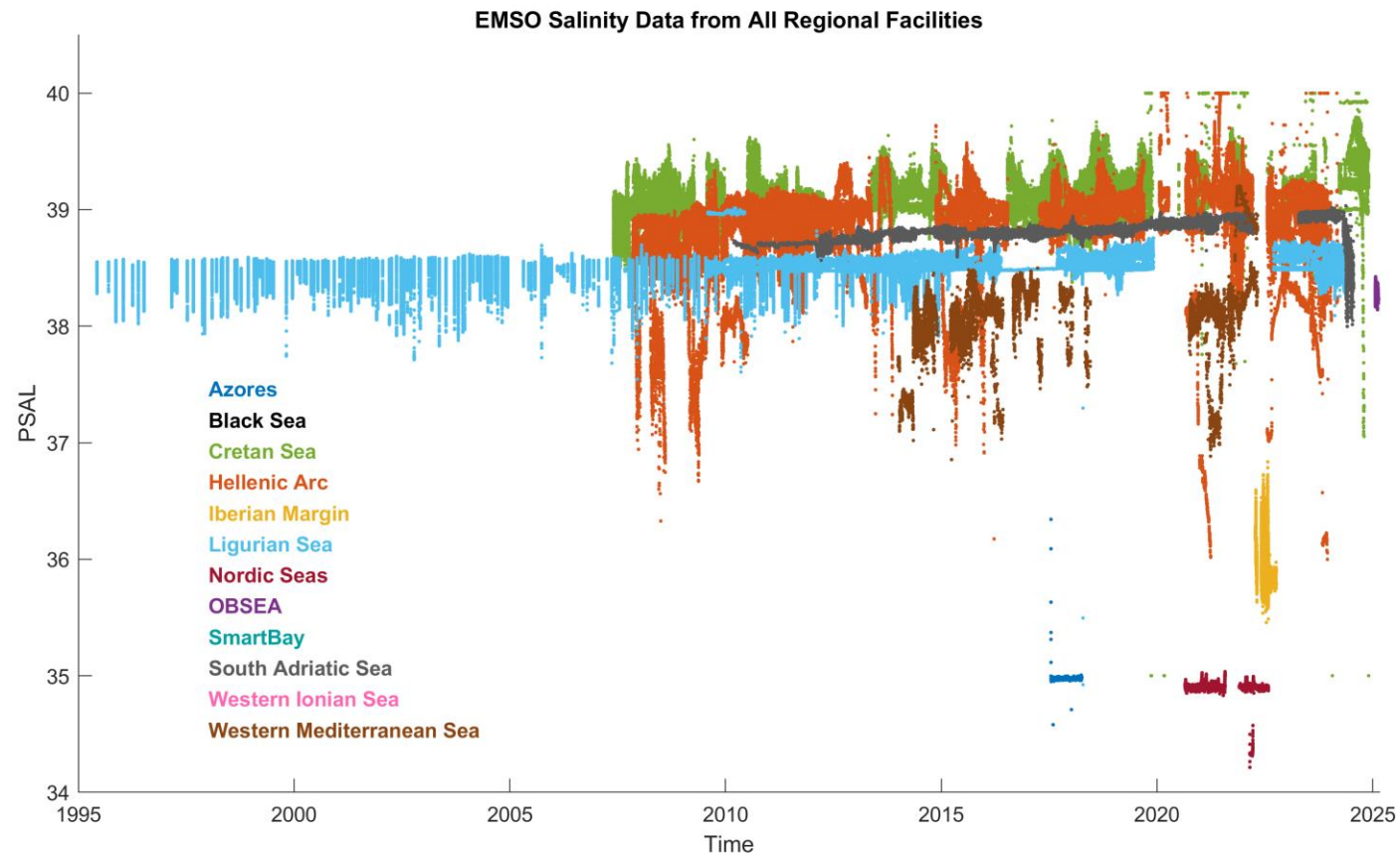
Standardised data improves **reproducibility**, and trust

Accelerates scientific progress by removing technical barriers

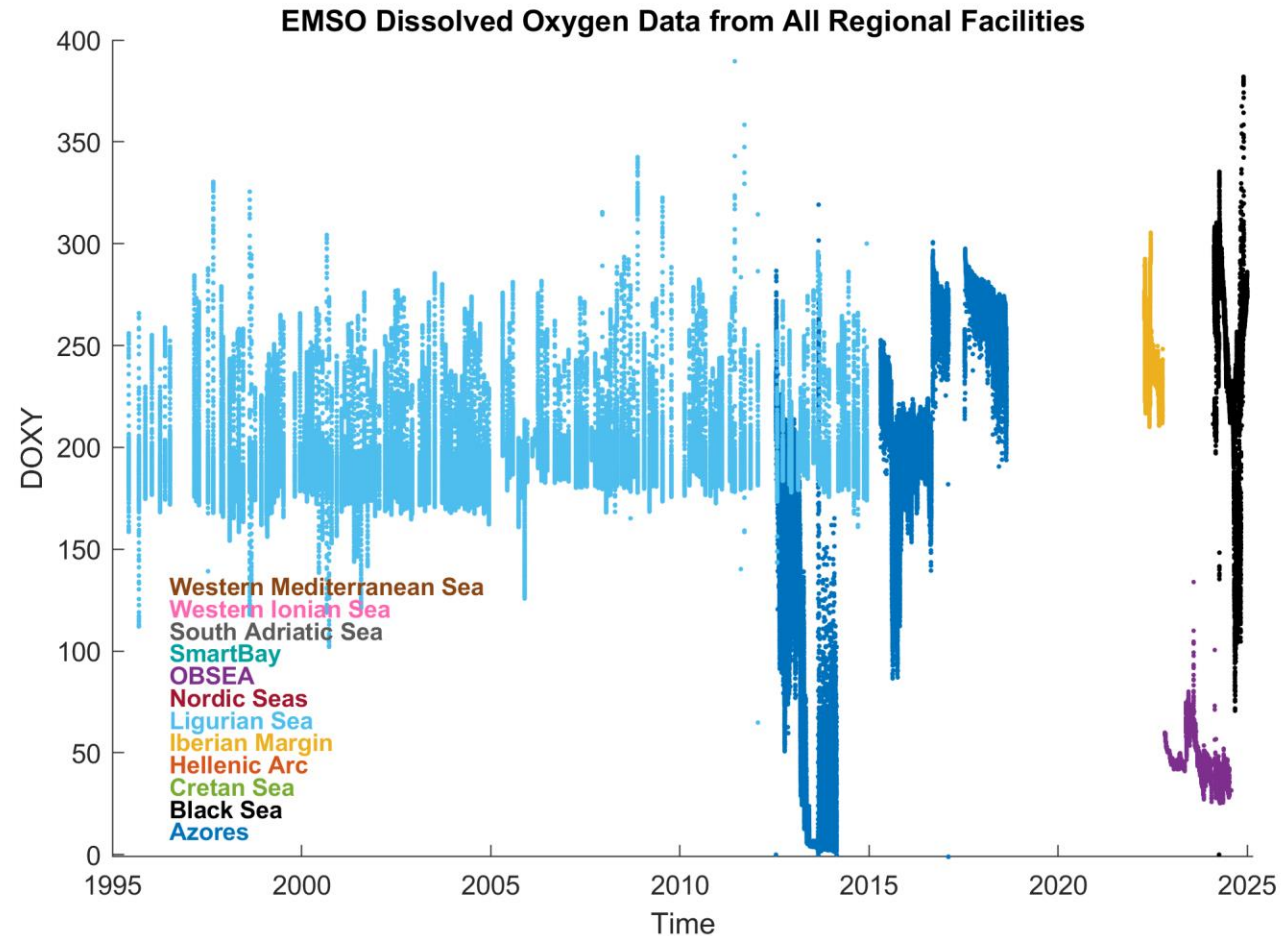
EMSO Temperature Data from all RF



EMSO Salinity Data from all RF



EMSO Dissolved Oxygen Data from all RF



Summary

ERDDAP is the backbone

Harmonization is **improving (90 %)** due to EMSO

Metadata Specification Document & Metadata

Harmonizer tool.

ERDDAP, DataPortal

Examples of EMSO Data

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

Aljaz Maslo

aljaz.maslo@emso-eu.org