



EMSO ERIC Data Management Plan

Version 0.7

**EMSO ERIC
December 2023**



History of changes

Version	Date	Change	Author(s)
0.1	01/10/2023	First draft	Aljaz Maslo
0.2	17/10/2023	Format, Paragraph "Purpose", "Definitions", "Field of Application"	Valentina Tegas
0.3	01/11/2023	Filled up the paragraph of "Definitions". Added some clarifications to respond to comments from Valentina Tegas, Enoc Martinez and Marco Galeotti.	Aljaz Maslo
0.4	20/11/2023	Added Annexes A and B plus some clarifications to respond to Valentina Tegas comments.	Aljaz Maslo
0.5	27/11/2023	Document re-organization and updates throughout the document	Ivan Rodero
0.6	01/12/2023	Reorganization of the document and implementation of changes according to Ivan Rodero.	Aljaz Maslo
0.7	22/02/2024	DOIs added to ANNEX section	Aljaz Maslo



TABLE OF CONTENTS

DEFINITIONS	4
1. INTRODUCTION	5
2. FIELD OF APPLICATION	5
3. DATA COLLECTION	5
4. DOCUMENTATION AND METADATA	6
5. ETHICS AND LEGAL COMPLIANCE	6
6. STORAGE AND BACKUP	7
7. SELECTION AND PRESERVATION	8
8. DATA SHARING	8
9. RESPONSIBILITIES AND RESOURCES	9
ANNEX A: Regional Facility Data Management Plans	10
ANNEX B: EMSO ERIC Metadata Specification	10



DEFINITIONS

Term	Definition
Regional Facility (RF)	is an EMSO ERIC facility composed of the observation infrastructure, instrumentation and hardware, and other resources and services in a particular region.
EMSO ERIC Data	are quantitative and/or qualitative attributes of variables or sets of variables that have been gathered by EMSO ERIC regional facilities.
Metadata	describes information referring to data collected from instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions, electronic logbooks generated during the experiment and other logistical information.
Creative Commons (CC) Licenses	give everyone, from individual creators to large institutions, a standardized way to grant the public permission to use their creative work under copyright law. See https://creativecommons.org/
Third party	describes a party that is outside the main parties or not part of the principal group or organization involved (not part of EMSO ERIC and/or the RFs).
EMSO ERIC User	individual or institution that utilises the EMSO ERIC services to access data and data products and/or tools and software. Access includes discovery, download, execution, or any other use.
Data Owner	people or entity that possess the right and control over the acquisition, use and distribution of a data set.
Embargo Period	is a period during which access to data or publications is reserved exclusively to the data owners
FAIR data principles	means guiding principles to make data Findable, Accessible, Interoperable and Reusable.



1. INTRODUCTION

The purpose of this Data Management Plan (DMP) is to serve as a comprehensive guide to how the EMSO ERIC data are and will be managed within the EMSO ERIC consortium, from collection to long-term preservation and accessibility to the broader scientific community and stakeholders at large.

EMSO ERIC is a European Research Infrastructure that consists of a network of regional facilities (RFs) located in key sites in Europe, from the northeast Atlantic, through the Mediterranean, to the Black Sea. These RFs are connected by the common goal of providing high-quality, long-term time series data in deep water. These RFs have different historical backgrounds, research interests and technical capabilities and have developed their individual DMPs. The EMSO ERIC ecosystem aims to deliver high-quality and FAIR (Findable, Accessible, Interoperable, and Reusable) data from the aggregation of different RF data sources in a reliable and unified manner allowing the user to have a single point of access to all EMSO ERIC data.

EMSO ERIC DMP has a modular structure. The RF DMPs (Annex A) are linked together through this central DMP document. In this way, updating it and possibly adding new RFs in the future is greatly simplified. Please note that the DMP should be considered a living document that can be revised and will evolve based on progressing insights into the nature of the data collections while taking into account developing standards and data initiatives.

2. FIELD OF APPLICATION

This EMSO ERIC DMP applies to all EMSO Instruments and RF. Promoting an identical standardising approach to the management of scientific data will facilitate the research community using more than one facility and add to the overall transparency of the scientific process.

3. DATA COLLECTION

Observatories are platforms equipped with multiple sensors, placed along the water column and on the seafloor. They constantly measure a variety of parameters (oceanographic, biochemical, meteorographic, sound recordings, seismic data, images) with different temporal

and spatial coverage, that address natural hazards, climate change, and marine ecosystems. All the details about the EMSO ERIC collected data and their collection procedures are presented in the RF DMPs. RFs are encouraged to perform quality control (QC) procedures using a well-described convention/standard that is the most appropriate for their dataset and then report the used standard in the metadata. The QC procedures currently used by the RFs are described in their DMPs. EMSO ERIC is planning to set harmonized QC procedures for its data in the near future.

4. DOCUMENTATION AND METADATA

To ensure harmonization of the EMSO ERIC datasets RFs should follow the metadata standards described in EMSO ERIC Metadata Specification (Annex B) and via GitHub¹. 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 OceanSITES² Data Format Reference Manual v1.4 but adapted to the needs of EMSO ERIC and its federated data service based on ERDDAP. The EMSO ERIC Metadata Specification document is focused only on the specific metadata descriptions related to its datasets. For additional clarification on metadata specifications, including several examples, RFs may consult the OceanSITES manual.

With its commitment to data harmonization, EMSO ERIC ensures that its data sets, among others, comply with the "Interoperable" part of the FAIR data principles. Interoperable means that data should be readable for machines without the need for specialized ad hoc algorithms or translators and that metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

5. ETHICS AND LEGAL COMPLIANCE

RFs shall contribute to the EMSO ERIC dataset with the data cited in their Service Level Agreement (SLA). No sensitive data are provided by the RFs. Data users will have to accept the EMSO ERIC Terms and conditions to use the portal and download any data set. EMSO ERIC may also apply a registration process if seen as appropriate. Data produced by the RFs and by the financing agencies/companies are the owners of this data. Their terms and conditions to share the data (Distribution statements) are available in their DMPs and must be fulfilled in order to be shared through the EMSO ERIC platform. Most of the RFs use Creative Commons Attribution license (CC-BY) or Copernicus standards. Both licenses state that the data are public and free of charge if proper citation is displayed in any publication or product using

¹ <https://github.com/emso-eric/emso-metadata-specifications>

² <http://www.oceansites.org>

data. Moreover, both of these licenses ensure reusability which is in line with the FAIR data principles. When the RF data are shared through the EMSO ERIC platform, they should include at the end of the citation statement that the data are made freely available by EMSO ERIC (<http://emso.eu/>). RFs are encouraged to affix distribution statements and type of license, preferably CC-BY, to their metadata. Data shared by the RFs may be subjected to an embargo period. However even embargoed or restricted data should be discoverable through metadata. While it is understood that initial periods of exclusive data use are customary, the ultimate goal should be to make all data openly available as soon as possible and ideally within two years of acquisition.

When data produced (not shared by the RFs) by EMSO ERIC is shared with third parties EMSO ERIC shall retain all rights, interest, and title in such data. Access to data produced by EMSO ERIC shall, wherever possible be free and open to all members of scientific institutions and other stakeholders. To other users, EMSO ERIC may disseminate collected data for a fee. Hence, EMSO ERIC reserves the right to select, on a case-by-case basis, between two licensing options: the CC-BY license and, alternatively, the Creative Commons Attribution-NonCommercial license (CC-BY-NC). This latter license allows others to use, share, and build upon the data for non-commercial purposes while requiring proper attribution to the original creator. However, it does not permit commercial use without seeking additional permission from the creator. Under the CC-BY-NC license, you are free to use the data for academic or research purposes, but if you intend to use it for commercial purposes, you would typically need to contact the creator to negotiate a separate commercial use agreement or obtain permission.

Distribution statement for the data produced by EMSO ERIC, related to the CC-BY license:

These data follow the standards adopted by EMSO ERIC (<http://emso.eu/>); they are public and free of charge. The user assumes all risks for the use of the data. The user must display citations in any publication or product using data.

Distribution statement for the data produced by EMSO ERIC, related to the CC-BY-NC license:

These data follow the standards adopted by EMSO ERIC (<http://emso.eu/>); they are public and free of charge for any non-commercial use. The user assumes all risks for the use of the data. The user must display citations in any publication or product using data. The user must contact the lead PI prior to any commercial use of the data.

6. STORAGE AND BACKUP



The responsibility for storage and backup of the data lies within the RFs and is described in the RF DMPs. RF operators are recommended to create regular backups of the data. These backups should be stored at a different location than the actual data, on physically separated media. Both incremental (e.g. on weekdays; backups of all changed data files) and full backups (e.g. on weekends; backup of all data files) are suggested. For the central services and repository, the central IT support for EMSO ERIC will take up the role of backup and recovery.

EMSO ERIC will undertake all required efforts needed to protect the data, products, and services against unauthorized use. The primary responsibility to take necessary measures to ensure data security lies with the RF operators.

EMSO ERIC will undertake all efforts required to provide secure access to data. Where applicable, authentication systems will be used, requesting log-in before providing access to secured data and information. Furthermore, EMSO-ERIC will take measures to be compliant with the EU regulations regarding the protection of personal data.

7. SELECTION AND PRESERVATION

At the core of EMSO ERIC's mission is the collection, curation, and provision of high-quality oceanographic measurements to assess long-term trends. EMSO ERIC RFs collect various types of oceanographic data, including video and acoustic measurements. All data of long-term environmental value will be archived and preserved for sharing for the foreseeable future by the RFs. Data should be stored long-term unless it is stated otherwise or the data owner gives a specific reason to not preserve the data. RFs long-term preservation plans are described in their DMPs. EMSO-ERIC encourages the RFs to preserve the data in trustworthy accredited repositories where the necessary measures are in place to archive and preserve data over a long period. In order to achieve transparency about the conducted research, EMSO ERIC data linked to publications should always be made accessible and retrievable.

8. DATA SHARING

The primary source for sharing EMSO ERIC data is the ERDDAP platform, accessible at <http://erddap.emso.eu>. ERDDAP is a data server created by NOAA that provides a simple and consistent way to serve data on the web. ERDDAP is free and open source. It uses Apache-like licenses, so it can be adapted or enhanced to fit a user's requirements. Users can download subsets of gridded and tabular scientific datasets in common file formats and make graphs

and maps, which can be embedded in web pages and can be configured to update with the latest data available.

To avoid gathering all the data into one centralized dataset location EMSO ERIC selected a federated ERDDAP system to share its data. ERDDAP federation model consists of having one central ERDDAP server that references remote ERDDAP servers. ERDDAP federated model is well suited to the distributed data requirements of EMSO ERIC as data do not need to be transformed from the RF storage format of choice to one “master” format in order to be served. This means EMSO ERIC is not forcing data architecture decisions onto the RFs that may not fit with their existing architecture.

ERDDAP can be set up to serve data from the storage structures already in place at each RF and it automatically aggregates files within a dataset. ERDDAP also facilitates interoperability and data reuse since it can take a variety of formats as input and output them into user-preferred file formats. In the case that at the moment a RF cannot provide a local ERDDAP instance their data may be included in the central server.

Besides the ERDDAP platform data are also shared through the EMSO ERIC data portal (<https://data.emso.eu>). To enhance discoverability, data can also be made available through external data services like OceanSITES, EMODnet, and others. When possible, a persistent identifier (e.g., DOI) will be created for each shared dataset, enabling correct citation.

9. RESPONSIBILITIES AND RESOURCES

Implementation of the DMP is necessary at both the central and local levels. It requires a collaborative effort involving the RF data managers and their leader, the EMSO ERIC Data Officer, and the head of the EMSO ERIC IT team. The RFs will ensure that the DMP is implemented locally, covering aspects such as QC, data storage, backup, and metadata harmonization. This means that there shall be a dedicated DMP for every RF. Those responsible for implementing the DMP at the local level are indicated in the RF DMPs. EMSO ERIC will coordinate dataset harmonization between the RFs and provide support through the necessary training.



ANNEX A: Regional Facility Data Management Plans

The table below provides the pointers to the available RFs' data management plans (DMP) through their Digital Object Identifier (DOI).

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

ANNEX B: EMSO ERIC Metadata Specification

The EMSO ERIC Metadata Specification is under configuration management using a git-based code repository. The specification is summarized in a “living document” available at:

<https://doi.org/10.5281/zenodo.10669878>

