



# CS - MACH1

## D1.5 Data Management Plan (DMP)

Antonio Novellino (ETT), Peter Thijssse (MARIS), Viviana  
Piermattei (CMCC), Camilla Campanati (CMCC)

November 2025



**Funded by  
the European Union**

This project funded by the European Union (GA 101214613). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the granting authority (REA). Neither the European Union nor the granting authority can be held responsible for them.

## Document details

Project Acronym / Name	CS-MACH1 / Marine Citizen Science Data Horizon
Project URL	<a href="https://cs-mach1.eu/">https://cs-mach1.eu/</a>
Project Type	HORIZON Coordination and Support Action
EU Call	HORIZON-MISS-2024-OCEAN-01
Grant Agreement No.	101214613
Project Start Date	1st June 2025
Project end date	30th November 2027
<b>Work Package</b>	
Work Package	1: Project Coordination and Management
<b>Deliverable</b>	
Deliverable	1.5
Due date of Deliverable	30th November 2025 (M6)
Actual Submission date	28.11.2025
Lead Beneficiary for this deliverable	CMCC
Reviewed by	Camilla Campanati, Viviana Piermattei, Antonio Novellino, Peter Thijssse
Revision	
Type of Document	DMP
Dissemination level	PU - Public
Number of pages	24



## Document history

Version	Date	Comment	Modification made by
1.0	10/10/25	Draft document shared with partners	
2.0		Review by Peter Thijssse MARIS and Antonio Novellino ETT	
3.0	28/11/25	Minor edits	

### To cite this document:

Novellino A., Thijssse, P., Piermattei, V., Campanati, C. (2025). D1.5 Data Management Plan (DMP). Deliverable Report of the Project Horizon Europe CS-MACH1 (Grant Agreement No. 101214613).

### Document Disclaimers:

This document is based on the Horizon Europe Data Management Plan Template [V1.1 April 2022] and adapted to provide the CS-MACH1 partners and CS-MACH1 stakeholders a living guide for the project Data Management Plan. The community is not responsible for any use that might be made of the content of this publication.



## Table of Contents

<b>Table of Contents</b> .....	<b>4</b>
<b>List of referenced acronyms</b> .....	<b>5</b>
<b>1. Executive Summary</b> .....	<b>6</b>
<b>2. Scope of the document</b> .....	<b>7</b>
<b>3. Project output</b> .....	<b>8</b>
3.1 Project output summary.....	8
3.1.1 Observation datasets.....	9
3.1.2 Data products.....	10
3.1.3 Other materials.....	10
<b>4. FAIR Data</b> .....	<b>10</b>
4.1 Making data findable.....	11
4.1.1 Observation data.....	11
4.1.2 Data products and other research output.....	12
4.2 Making data accessible.....	12
4.2.1 Observation data.....	12
4.2.2 Data products and other research output.....	13
4.3 Making data interoperable.....	14
4.4 Increase data re-use.....	15
<b>5. Allocation of resources</b> .....	<b>17</b>
<b>6. Data security</b> .....	<b>18</b>
7.1. Ethics and Code of Conduct.....	18
<b>8. Other issues</b> .....	<b>20</b>
<b>9. Intellectual Property and Licensing</b> .....	<b>20</b>
<b>10. Data Sharing and Long-Term Preservation</b> .....	<b>21</b>
10.1 Further recommendations and measures on long term preservation of informed consent.....	21
<b>11. References</b> .....	<b>22</b>



## List of referenced acronyms

CARE	Collective benefit, Authority to Control, Responsibility, and Ethics
CDI	Common Data Index
CMCC	Centro Euro-Mediterraneo sui Cambiamenti Climatici
CS	Citizen Science
CS-MACH1	MARine Citizen science data Horizon
DMP	Data Management Plan
DOI	Digital Object Identifier
DTO	Digital Twin Ocean
ECSA	European Citizen Science Association
EDITO	European Digital Twin Ocean
EMODnet	European Marine Observation and Data Network
EOSC	European Open Science Cloud
ETT	ETT SPA
EU	European Union
EurOBIS	European node of the Ocean Biodiversity Information System
FAIR	Findable, Accessible, Interoperable and Re-usable
GBIF	Global Biodiversity Information Facility
GDPR	General Data Protection Regulation
GEOSS	Global Earth Observation System of Systems
IPR	Intellectual Property Right
MARIS	MARIS B.V.
MCS	Marine Citizen Science
MCSI	Marine Citizen Science Initiatives
MCSDN	Marine Citizen Science Data Network
NOC-BODC	National Oceanographic Centre
NVS	NERC Vocabulary Server
OBP	Ocean Best Practices
PM	Person Months
PMO	Project Management Office
SeaDataNet	Pan-European infrastructure for ocean & marine data management
SCOOP	Solutions for Cost-effective Ocean Observation Platform
SMHI	Swedish and Meteorological Institute
SO	Specific Objective
VLIZ	Flanders Marine Institute
WP	Work Package



# 1. Executive Summary

This document encompasses the first version of the CS-MACH1 Data Management Plan (DMP) and outlines the project's approach to achieve FAIR data and project output sharing policy.

The outputs of CS-MACH1, managed as much as possible in accordance with the FAIR principles, include:

- (1) Observation data, collected from CS during the WP5 use-cases;
- (2) Scientific publications;
- (3) Training and educational materials;
- (4) Best Practice documents;
- (5) Policy briefs;
- (6) Metadata for all datasets, including provenance, collection protocols, and sensor specifications.

For these outputs, this first version of the DMP defines the approach for data sharing per FAIR metric, differentiating between observation data generated and data products created, vs. other project outputs like documents, publications and policy briefs. Organising a FAIR data flow for the MCS observation data is an activity at the core of the project. The aim is to prepare the data flow from observers to the EU aggregators with EMODnet as the main target. In this document the basic data management principles will be described, while in the WPs (WP2, WP3, WP4, WP5), the detailed best practices will be co-designed and co-created towards FAIR data sharing, especially the I(nteroperability) and R(eusability), which will pose the biggest challenge for citizen data. For the other project outputs the main emphasis will be on F(indability) and A(ccessibility), because I and R metrics are not relevant nor achievable for such outputs.

Designed at the project's start (M6), the DMP is a living document and will be reviewed and updated regularly, in subsequent versions to submit by M18 (D1.6) and M28 (D1.7) respectively.

## 2. Scope of the document

This document presents the initial version of the MARine Citizen science data Horizon (CS-MACH1) Data Management Plan (DMP) and it is based on the Horizon Europe Data Management Plan (DMP) template.

Major challenges for the success of CS in the marine field include the lack of FAIR (Findability, Accessibility, Interoperability, Reusability) data, particularly in terms of Interoperability and Reusability blocking integration and assessing trust in the data, and thus uptake/re-use of the data. CS-MACH1 aims at overcoming existing CS empowerment barriers, such as accessibility, recognition, trust and sharing, of clustered and localised MCS data. A key objective (SO3) is to **enable and demonstrate enhanced FAIRness of the CS metadata and data**, ensuring it can feed into international data infrastructures such as EMODnet, SeaDataNet, EurOBIS and GEOSS, and ultimately contributing to the Digital Twin Ocean (DTO).

To achieve this, CS-MACH1 will mobilise the capacities within the consortium required for FAIR data management, aligned with EMODnet standards.

CS-MACH1 will map and mobilize the MCSI and develop a network for actors involved in MCS data collection in and beyond Europe (WP2). The focus is on several actors. Once these have reached the right level of readiness, the data is increasingly FAIR and apt to be integrated in the EU main data infrastructures, and therefore ready to be used by their end-users. The partners will also mobilise their own respective communities of experts to become part of the network.

The DMP outlines the Data Sharing Policy and strategies for handling research data with FAIR principles. Designed at the project's start (M6), the DMP is a living document and will be reviewed and updated regularly, in M18 and M28 respectively. It firstly defines which project output is in focus for the DMP, distinguishing between research outputs (e.g. Documents, reports, software, etc) and data (e.g. Newly acquired observation data, data held temporarily from MCSI, and processed data products). For both categories a pathway will be defined aimed to be as FAIR as possible, and taking into account Confidentiality, Open Access, Intellectual Property Rights (IPR) and knowledge management protection. The DMP ensures compliance with the Horizon Europe Open Science and FAIR (Findable, Accessible, Interoperable, Reusable) principles while respecting IPR, confidentiality, and GDPR (Reg. EU 2016/679).

For research outputs such as reports and best practices, it is important to define a clear publication strategy that ensures these materials effectively reach end-users both within and beyond the project. This strategy should also support their continued accessibility and reuse after the project's completion. For observation data, it is equally important to define how the collected data will be processed, stored, shared, and preserved throughout CS-MACH1. Particular attention should be given to how the MCSI could develop the capacity to carry out these processes effectively, ensuring that CS data can be **re-used** as much as possible by research and beyond.

To support the re-use of MCS data by integration into EU platforms like EurOBIS, SeaDataNet, and EMODnet, CS-MACH1 will work in close cooperation with the MCSI on applying standards, defining and/or pointing to best practices for sensor deployment and data management,



observation protocols for using cost-effective devices, and developing training materials to guide MCSI coordinators in managing their data flow. Appropriate devices, protocols, and practices will be recommended to ensure trustworthy data, aligning with CARE principles. Details of this approach will be documented e.g. as output of Task 3.1 [M1-M12], which will develop an overview of what is needed to create FAIR dataflows and from Task 3.2 [M9-M27], which will provide support to MCSI representatives and technology developers to streamline FAIR data flows as mentioned above.

Each subsequent DMP version (D1.6 and D1.7) will take into consideration results from these tasks in WP3. Case-specific pathways examples will further guide the data management

*The document recalls the scope of the CS-MACH1 project and describes the CS-MACH1 framework for the research data management according to the requirements in the Annotated Grant Agreement, article 16 and 15.*

## 3. Project output

### 3.1 Project output summary

The main objective of CS-MACH1 is to develop a MCS data network (MCSDN) and its web-based hub to empower MCSI towards improved data uptake and integration by users. This will be achieved by fostering continuous interaction among MCSI, technology providers, data management experts and researchers. In addition, the project will provide the necessary cost-efficient technology, data management best practices and standards, training and support to enable the production of FAIR data flows.

Different project WPs will generate and reuse different types of data:

- WP2 will activate and manage the MCSDN, including communication with all key actors through a network secretariat. It will determine the most accessible way to appropriately catalogue MCSI data, optimise data sharing and management.
- WP3 will develop comprehensive tools, protocols and best practices for collecting and managing MCS data in line with FAIR principles, enabling CS and cost-efficient sensor data submission to existing EU aggregators
- WP4 will deliver workshops and provide related e-learning materials focusing on using the MCSDN to provide FAIR data and prepare the data for EMODnet, supporting environmental monitoring, policy compliance, and ocean literacy.
- WP5 will collect stakeholder feedback to refine technology and data workflows. The Literature Tracking Service offered by GBIF will enable the retrieval and recognition of participant contributions listed in the literature tracking list (new extended FAIR/CARE service).

- WP6 (PEDCR, D6.1) will outline the plan for the exploitation, dissemination and communication of project results. The project website will serve as a central web-based hub, providing access to data ingestion portals, training modules, protocols and resources from WP2-WP5 activities and feedback mechanisms.

Summarising, the outputs of CS-MACH1 will include:

- (1) Observation data: Collected from MCSI and low-cost sensors during the WP5 use-cases;
- (2) Data products: Processed datasets from various sources as outcome from WP5 integration activities.
- (3) Other materials
  - a. Scientific publications;
  - b. Training and educational materials;
  - c. Best Practice documents;
  - d. Policy briefs

All outputs will contribute to the overarching goal of producing **interoperable, open, and FAIR-compliant data** that enhances the integration of MCS with traditional marine data infrastructures.

The data and related products will be valuable to environmental agencies, marine researchers, policy makers, and citizen science initiatives seeking interoperable marine data e.g. filling current gaps in monitoring and decision-making.

### 3.1.1 Observation datasets

The CS-MACH1 project will generate and manage multiple datasets related to MCS, technology and environmental monitoring. These include observational data from low-cost sensors and CS contributions (including physical data, biodiversity data and litter characterisation related to plastic pollution). Existing in-situ data from data infrastructures such as EMODnet, SeaDataNet, and EurOBIS, and, where relevant, local monitoring datasets, will be reused for comparative analysis and validation (WP5).

Observation and sensors' data will comply with EurOBIS, EMODnet, SeaDataNet standards for data structure and metadata. Cooperation with the SCOOP platform will allow accessing and redirecting users towards sensor ID catalogue and manuals, defining best practices for the appropriate use, deployment and calibration of sensors (i.e. which sensors are best suited for specific environmental parameters and observation conditions).

The management of biodiversity data and litter characterisation related to plastic pollution will be evaluated with the project progressing, to ensure at least collected datasets include metadata defining methodologies and acknowledgement of MCSI contributors. As the project progresses involving the various actors of the MCSDN, the formats of the collected data will be further detailed in the following DMP versions. Nevertheless the aim is to homogenise as much as possible towards data being interoperable, open formats (e.g. CSV,

NetCDF, JSON, XML, PDF). Likewise data volumes ranges (GB scale expected) across sites and time-series, will be characterised in the following DMP versions, depending on sensor deployment density and temporal coverage.

The generation and re-use of “traditional observational data (in WP5) directly supports the project’s goal to : demonstrate how CS can complement established marine monitoring systems; Enable FAIR and interoperable data exchange between CS and EU infrastructures; Support policy compliance, environmental monitoring and ocean literacy; Provide training, tools and standards that enhance data quality and integration. This data will only be stored temporarily within the project, used in the demonstrators and not be re-shared (only as part of a derived product, see 3.1.2).

### 3.1.2 Data products

The CS-MACH1 project will also generate data products, including **derived datasets, aggregated and processed data collections**, integrating multiple data sources, in the form of for e.g. maps, graphs. CS-MACH1 will explore publication through EMODnet catalogues to support marine data integration into EU infrastructures and ensure long-term visibility and accessibility.

### 3.1.3 Other materials

Other materials will include:

- a. **Scientific publications** reporting project results and methodological advances;
- b. **Training and educational materials**, including e-learning modules and workshop outputs
- c. **Best Practice documents**, summarising protocols and FAIR data management guidelines;
- d. **Policy briefs**, translating project findings into actionable recommendations.

All materials will be available through the project website and Zenodo CS-MACH1 community repository (<https://zenodo.org/communities/cs-mach1/>), to ensure transparency, accessibility and lasting impact beyond the project’s lifetime.

## 4. FAIR Data

CS-MACH1 will create a MCSDN to facilitate FAIR data flows, for MCSI, cost-efficient technology providers, and data management experts. It will also deliver training material to empower citizens to monitor and contribute to marine science through platforms such as SeaDataNet, EurOBIS and EMODnet. CS-MACH1 will facilitate the ingestion of standardised in-situ marine CS data into EU aggregators, The approach to achieve FAIRness for the specific data types in CS-MACH1 is detailed in the following sections.

### 4.1 Making data findable

The FAIR metrics considered for *Findability*[1]are:



- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F2. Data are described with rich metadata.
- F3. (Meta)data are registered or indexed in a searchable resource.
- F4. Metadata specify the data identifier.

#### 4.1.1 Observation data

Data are findable when described with sufficiently rich metadata and registered or indexed in a searchable resource that is known and accessible to potential users. Additionally, a unique and persistent identifier should be assigned such that the data can be unequivocally referenced and cited in research communications. This ensures persistent linkages between data, metadata, and related materials, such as code, models and research literature that support data discovery, interpretation and reuse.

CS observation data will be made Findable in two possible and complementary ways:

- Initial publication as-is by the MCSI (where available)
- Publication via EMODnet Data Ingestion, which provides the first step towards Findability and Accessibility of data as-is, promoting CS data, for a wider EU audience. In a next stage some datasets could be selected to ultimately be integrated into EurOBIS, SeaDataNet and used in EMODnet data products. Each dataset, registered through EMODnet Data Ingestion, will receive a persistent identifier (DOI or Handle). High-quality, complete metadata will enable experts to assess data quality and potential for re-use. CS-MACH1 WPs 3 and 2 have the aim to strengthen this pathway.

Cost-efficient sensor data constitutes a specific case. CS-MACH1 will actively promote the use of cost-efficient sensor data for physical and chemical data collection within CS activities. By cooperating with cost-efficient sensor developers and documenting best practices for the data management and deployment, the project will ensure data are **as FAIR as possible at the source** and of constant quality. Initial storage of data and publication of their metadata and service endpoints in accessible catalogues are key steps for Findability, enabling automatic **indexing and harvesting**. Operational (real-time) data may flow directly into **EMODnet Physics**, while offline datasets (locally analysed data) should be deposited in **EMODnet Data Ingestion** with essential metadata and a DOI for Findability and referencing.

Metadata will comply with the disciplinary standards of the respective aggregators (e.g., SeaDataNet CDI, OBIS Darwin Core, and EMODnet metadata formats) to ensure interoperability and reusability. Where such standards do not exist for specific citizen science data types, project-defined templates will capture essential contextual information, including data origin, collection methods, sensor characteristics, geographic coverage, and temporal resolution.

To enhance discovery, metadata records will include standardized **keywords** from controlled vocabularies such as the **SeaDataNet Common Vocabularies** and **NERC Vocabulary Service**,

and will be made available in catalogues that support **automatic harvesting and indexing** by external systems, ensuring visibility across repositories and aggregators.

These cases and pathways will be further detailed during the project (e.g. in WP3).

### 4.1.2 Data products and other research output

Data products and other research output (e.g. scientific publications, educational materials, best practice documents and policy briefs and guidance) will be digital open-access and publicly available on **Zenodo repository CS-MACH1 created community section** (<https://zenodo.org/communities/cs-mach1/>), on the CS-MACH1 project website and where applicable published on the **Ocean Best Practices website**. The dedicated Zenodo community will enhance Findability of documents, by assigning **persistent identifiers (DOIs)** and enabling **metadata tagging** for efficient discovery.

All digital outputs such as software will be documented and, where possible, released as open-source via GitHub CC0/CC-BY (or equivalent) licenses.

## 4.2 Making data accessible

Accessible datasets can be obtained by humans and machines through well-defined and universally implementable protocols, with appropriate authorisation where necessary. Anyone should be able to access at least the metadata. It is important to emphasise here that Accessible in FAIR does not mean “open without constraint”. *Accessibility* ensures that both humans and machines are provided - via metadata - with clear information on how and under what conditions data can be accessed. The underlying protocols and mechanisms must allow scalable access across the web while ensuring security and compliance.

The FAIR metrics considered for *Accessibility* [1] are:

A1. (Meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 . The protocol is open, free, and universally implementable.

A1.2. The protocol allows for an authentication and authorization procedure, where necessary.

A2. Metadata are accessible, even when the data are no longer available.

### 4.2.1 Observation data

Accessibility of CS data is essential for uptake and trust by potential users. Following the Findability process (Section 4.1.1), datasets made available through international aggregators such as **EMODnet**, (and some via **EurOBIS**, and **SeaDataNet**) will be **retrievable via persistent**

**identifiers** (DOI or Handle) using open, free, and standardized protocols and technology (e.g., HTTPS, OPeNDAP, or ERDDAP).

Metadata are always accessible without authentication/authorisation needed. .

Where applicable, **restricted access** may apply to:

- i. Datasets containing personal information (e.g., citizen participants' identifiers);
- ii. Proprietary technology details from sensor developers; or
- iii. Data under temporary embargo for scientific publication or intellectual property protection.

Such datasets will be shared **under controlled access**, with requests managed through the repository's AAI system. The identity of users will be verified through institutional credentials or ORCID authentication.

At present, **no dedicated data access committee** is foreseen, as access control is managed directly through the repositories' established procedures.

If embargoes are applied (e.g., to secure publication or patent filing), the aim is that this will be **limited to a maximum of 12 months** and fully justified in the project's data inventory.

#### 4.2.2 Data products and other research output

All gathered research output and generated data products (as mentioned under *Findability*) will be retrievable by their identifier using a standardised communications protocol in the CS-MACH1 community on Zenodo (<https://zenodo.org/communities/cs-mach1/>), which is a **trusted EOSC-federated repository**.

Data and outputs will be open-access under **CC-BY** or **CC0 licences**, following the principle "as open as possible, as closed as necessary."

Metadata will always remain open under **CC0**, including authorship, licensing, and funding information. Metadata will include sufficient information to enable data access and interoperability and will be **harvestable** by external catalogues.

Restricted datasets, where justified (e.g., those containing personal data or proprietary sensor details), will be managed under controlled access. Access will be granted upon request and subject to GDPR compliance.

All related **software, scripts, or documentation** required to access or reproduce results will be openly shared via **GitHub** (under CC-BY, MIT or Apache 2.0 licences) or referenced within metadata records, ensuring transparency and reusability.

Data and metadata stored in Zenodo, EMODnet, and related infrastructures will remain **accessible and findable for at least 10 years after the project's end**, in line with repository retention policies.

## 4.3 Making data interoperable

Regarding Interoperability and Reusability CS-MACH1 will focus primarily on observation data as community standards and vocabularies for other research outputs (e.g., educational materials, policy briefs) are not yet established. For these non-standard outputs, the project will concentrate on ensuring **Findability and Accessibility (F and A)** compliance.

The FAIR metrics considered for *Interoperability* [1] are:

11. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
12. (Meta)data use vocabularies that follow FAIR principles.
13. (Meta)data include qualified references to other (meta)data.

CS-MACH1 aims at establishing a MCSDN that supports MCSI in producing interoperable observation data, to reach as many end-users as possible (and thereby strengthening their chances of sustained funding).

By aligning with EU-wide marine data standards and metadata formats, as well as best practices for observation, the project will ensure that CS data can be **seamlessly integrated** with institutional and professional datasets, enhancing trust, quality, and reusability across domains.

These solutions do not exist yet, or are not known to the CS initiatives. WP3 and WP2 will develop and describe the pathways for CS datatypes, including best practices for deployment as well as metadata and data formats, and will also provide training material. Specific use cases will support and validate this work. The first concepts to achieve interoperability will be described in this chapter (and section 4.4). As the project progresses, the actual results will follow and the DMP will be expanded accordingly.

*Interoperability* between different data sources depends on the use of shared semantics and domain-specific vocabularies within metadata to consistently describe attributes such as parameters, units, and platform type. To enable seamless integration across EU marine data systems, CS-MACH1 will follow or contribute to **community-endorsed interoperability best practices established** by **EMODnet**, **SeaDataNet**, and **OBP** to ensure harmonisation with EU infrastructures. All CS-MACH1 datasets and metadata records will include qualified references to related metadata and follow community standards. Established community standards, metadata, and vocabularies include:

- **Metadata standards:** ISO 19115, SeaDataNet CDI profile;
- **Data and metadata vocabularies:** NERC Vocabulary Server (NVS), World Register of Marine Species (WoRMS), European Directory of Marine Organisations (EDMO),

**Controlled vocabularies and thesauri:** BODC Parameter Usage Vocabulary (P01), SeaDataNet Platform Vocabulary (L06), Device Vocabulary (L22);

- **Repositories and frameworks:** EMODnet, EurOBIS, SeaDataNet, and Ocean Best Practices (OBP).

The approach during the project: CS-MACH1 will document best practices in formatting data and metadata, and create training material on this topic. Practically CS-MACH1 will analyse selected existing CSI dataflows, metadata, and services to recommend best practices and suggest solutions for generating metadata aligned with vocabularies.

CS-MACH1 aims to prepare the observation data from cost-efficient sensors for integration into the relevant EU data aggregators. This puts automatically requirements on the metadata, that should follow the ISO19115 metadata model as basis, as more specifically the SeaDataNet CDI profile in which specific vocabularies are used for e.g. parameters, instruments, organisations, and important for accessibility the usage licence and access restrictions. Currently, data from cost-efficient sensors do not follow the required metadata schema and use of vocabularies. CS-MACH1 will work with the sensor developers to improve this, and increase and extend the metadata as close to the source as possible. The cost-efficient sensor developers will receive training (material) to update their metadata and data formats for better interoperability.

Additionally, EU metadata models and vocabularies, particularly those managed by NOC-BODC, will be expanded to include cost-efficient sensor and CS data.

## 4.4 Increase data re-use

As mentioned in Section 4.3, CS-MACH1 strives for CS data interoperability and reuse by compliance (as far as possible) with EMODnet metadata and data standards. **The focus regarding Interoperability is on the semantics**, which (depending on the theme) are vocabularies from SeaDataNet, EurOBIS, and some dedicated to EMODnet: e.g. NERC Vocabulary Service NVS, WORMS, EDMO, EDMERP, etc).

Regarding Reusability the focus is on specific elements of the metadata that are needed to determine if the data is re-usable: How reliable is the data, what is the provenance (how was it collected, how was it processed), and if it is allowed to be re-used. And to support reuse and integration the metadata and especially the data needs to follow community standards (formats).

The FAIR metrics considered for *Reuse [1]* are:

- R1. (Meta)data have a plurality of accurate and relevant attributes.
- R1.1. (Meta)data are released with a clear and accessible data usage license.
- R1.2. (Meta)data are associated with their provenance.
- R1.3. (Meta)data meet domain-relevant community standards.



Similar to Interoperability CS-MACH1 will (in WP3 and WP2) develop and describe the pathways for the CS datatypes to become better Reuseable, including e.g. best practices for metadata and data formats, how to include provenance information and licensing information, and will also provide training material. The first concepts to achieve reusability will be described in this chapter, but during the project more results will follow and the DMP will be expanded accordingly.

The main goal is that collected CS data (during the project and in the MCSI) will be thoroughly described with a plurality of accurate and relevant attributes, fit to at least the minimum requirements of EMODnet to optimise their reuse. All deposited datasets will remain multi-year available and reusable through trusted platforms (e.g. EMODnet, EurOBIS) as defined by these aggregators, in line with their retention and preservation policies.

CS-MACH1 will publish detailed guidance on the metadata and data formats and how to semantically populate these.

Metadata of datasets should include methodological descriptions, quality/processing information, and provenance information. Each dataset will be accompanied by detailed documentation (e.g., README files, data dictionaries, and processing notes) describing methodology, variable definitions, units, and data quality checks, ensuring users can validate and correctly reuse the data.

To overcome trust issues in CS data, an important component of provenance and quality information is the reference to a sensor identifier and applied collection/deployment protocols. For cost-efficient sensors, referencing the sensor-ID in a catalogue (SCOOP) allows users to assess data quality and resolve doubts about data reusability. For other data, such as biodiversity, metadata completeness, details on the protocols followed, and the user's expertise are crucial for ensuring data reliability.

Clear licensing and acknowledgment guidance will accompany all reusable data. Open licenses (CC-BY or CCO) will be promoted and applied as much as possible.

+ making CS data Acknowledgeable: A key focus of the project is on acknowledging contributors, particularly critical in the case of CS, throughout the data cycle, where knowing who reported, validated, and curated the data is vital. This information will be included in metadata to enhance data provenance, ensure sustainability of volunteer networks, and support data quality assessment. CS-MACH1 will develop a good/best practice on how this can be achieved.

+ Develop best practices and training around use cases: Some observations (e.g. Sea Level from CMCC low-cost sensors), already integrated in EMODnet, are proof of usability, as anyone can download the data. For other kinds of data, the integration in other platforms may need to be assessed and metadata vocabularies cross-mapped to support reusability. Pathways will be provided (e.g. through a decision tree) as the data types become available, linking MSC1 with the type of data they collect and/or sensors used, and guiding them on how to organise and deposit the data into nodes or platforms. This procedure will be structured

before the WP5 on-site training and will be used as guidelines or protocols for the training workshop.

Other project outputs (e.g. Educational materials, policy briefs) will follow the same FAIR-aligned documentation and licensing principles. Ethical and data security aspects, including GDPR compliance for any personal data, are addressed in Section 7 of the DMP.

## 5. Allocation of resources

Costs for FAIR data management (storage, curation, metadata creation, and repository fees) are included in all WPs budgets, with main PM allocated to WP1, WP2 and WP3. Besides having a Data Management team, each partner has some allocated budget to implement FAIRness according to the principles and recommendations described in this document.

Data management is included in a dedicated project work package (WP1) led by CMCC, which will oversee the coordination, ensuring the compliance with FAIR, GDPR.

ETT will co-lead the technical implementation of DMP and will manage the CS-MACH1 dedicated community in the Zenodo repository, supported by CMCC.

Responsible for the data management is Antonio Novellino (ETT), who has long experience in data management and ocean data management and sharing for both Copernicus Marine Service (he was the CMEMS Dissemination Unit deputy), EMODnet (he is the Coordinator of EMODnet Physics, and leader in EMODnet Ingestion operational data ingestion flow), and several H2020 and HEurope marine data and services projects.

More specifically he will be supported by:

The PMO (CMCC) and in particular, Viviana Piermattei, CS-MACH1 Project Coordinator, and Camilla Campanati, CS-MACH1 Project Manager.

MARIS, co-leader of WP3, in particular Peter Thijsse, data management expert, will actively support the data management for data standardization and interoperability and metadata models.

WP3 partners, NOC-BODC, MARIS, VLIZ, ETT and SMHI partners will analyse the observation protocols and data outputs from the cost-efficient instruments, as well as data from ongoing CSI. They will then develop and document best practices, training material and actual training to engage and support selected CS groups and cost-efficient technology concepts (WP3).

WP2 and WP6 will engage all stakeholders through two-way communication and knowledge exchange, complying to ethics and GDPR defined in section 7.

All partners contributors provide metadata and follow DMP guidelines.

It is worth highlighting, core players in data management practices for EMODnet (which are based on SeaDataNet and EurOBIS) and GEOSS are project partners, with specific expertise on FAIR data flows, standards and best practices as in use in target international data infrastructures such as EMODnet, SeaDataNet, EurOBIS, GBIF and GEOSS, and onwards to the European Open Science Cloud (EOSC) and DTO related developments.

Long-term preservation will, depending on the output, be ensured through target infrastructures (EMODnet, EurOBIS etc) or, whenever data do not fit the infrastructures, by the project Zenodo repository, which itself follows a long-term data preservation and guarantees retention for the “lifetime of the repository “ (i.e. committed to maintain the repository for the next 20 years)

## 6. Data security

CS-MACH1 project outputs and data are initially stored in the internal project drive (on secure institutional servers managed by CMCC) and with restricted access to the consortium partners and daily backups. Finalised project outputs (e.g. revised deliverables) will be uploaded in Zenodo-dedicated community repository, managed by ETT, which provides trusted long-term storage and curation. As mentioned in Section 5, Zenodo repository guarantees long-term data preservation and retention. Likewise, ingested data in EMODnet platform provides trusted long-term preservation.

As a project involving a range of different actors and stakeholders and their views, personal data will be collected. The data will be kept as basic as possible, and not include sensitive information such as health status, religion or sensitive accurate locations.

Data collected about vulnerable and endangered marine species will be anonymized and lowered in geographical accuracy when stored or shared, and the consortium will follow strict ethical review when designing the methods for data collection and analysis. All personal data will be handled in line with the GDPR/General Protection Regulation (2016/679/EU).(See Section 7).

Pre-existing data such as in public repositories and previous studies will be used and will not include personal information. Appropriate consent for secondary use of these data will be ensured.

## 7. Ethics and GDPR

Ethical compliance is ensured under Horizon Europe guidelines and the CS-MACH1 Grant Agreement.

Personal and citizen-contributed data will follow informed consent procedures. Anonymization will be applied to remove identifying elements and sensitive information (health, religion, precise locations). No vulnerable groups will be targeted. As already stated in Section 6, data collected about vulnerable and endangered marine species will be anonymized and lowered in geographical accuracy when stored or shared

### 7.1. Ethics and Code of Conduct



Partners must carry out their actions in accordance with ethical principles that include the highest standards for research integrity and the EU Charter of Fundamental Rights and the European Convention for the Protection of Human Rights and Fundamental Freedoms and its Supplementary Protocols. Partners must pay particular attention to the principle of proportionality, the right to privacy, the right to the protection of personal data, the right to the physical and mental integrity of persons, the right to non-discrimination, the need to ensure protection of the environment and high levels of human health protection. In addition, the partners must respect the fundamental principle of research integrity as set out in the European Code of Conduct for Research Integrity. This means following the principles of the reliability of quality of research in design, methodology, analysis and use of resources. As well as honesty in developing, undertaking, reviewing, reporting and communicating research in a transparent way. Furthermore, the respect for colleagues, society, ecosystem, cultural heritage [2] and the accountability for the research from idea to publications, management, organisation, training, supervision, mentoring and for impacts.

This also means that the partners must ensure that persons, carrying out research tasks, follow the good research practices including ensuring, where possible, openness, reproducibility and traceability and refrain from the research integrity violations described in the Code of Conduct.

## 7.2. Personal Data Regulation

Some CS-MACH1 activities, such as workshops, meetings, training etc. may collect personal data that falls under GDPR. In these cases, CMCC develops a document specifying the reasons for personal data collection and its intended use. Personal data will not be used for any other purposes or shared outside the project and the specific task that collected the data. This document also informs participants of their GDPR rights, including the rights to access, rectify, erase, restrict processing, data portability, and object to data processing. Derived anonymized/grouped figures (e.g., the number of attendees) may be used for the project reporting purposes. A Consent Form will be presented to individuals for acceptance to participate in the data collection activities related to stakeholder engagement and co-creation as well as for any audio/video recordings during participation of project meetings and events. The document will strictly abide by provisions of Regulation (EU) 2016/679. All individuals will be aware of the aims, objectives and methods of the study, including the nature of questions, the time and tasks required and contact details of the researchers.

All personal data collected for the sake of event organization will be fairly and lawfully processed; processed for limited purposes; adequate, relevant and not excessive; accurate; not kept longer than necessary; processed in accordance with the data subject's rights; secure; not transferred to countries without adequate protection. Limited personal data (name, email) will be collected for the purpose of meeting/workshop/event registration.

Moreover, CS-MACH1 adopts the following approach to GDPR:



- Perform a periodic Data Protection Impact Assessment, to verify that CS-MACH1 partners are applying the recommendation on data protection
- Informed Consent, to ensure that participants provide explicit, informed consent for the collection and use of their personal data, and maintain records as evidence that proper consent was obtained.
- Collect Only Necessary Data, to ensure that the personal data collected is adequate, relevant, and limited to what is necessary for the purposes of the research
- Keep the documentation (DMP, consent forms, data protection policy, etc) updated

For the specific case of surveys and questionnaires, whenever applicable, CS-MACH1 uses the EU Survey tool [3]. Hosted by the European Commission's Directorate-General for Informatics (DG DIGIT), is available free of charge to all EU citizens, the EU Survey is the official survey management tool of the European Commission. It is primarily used for conducting public opinion surveys, as well as self-assessment forms and registrations and is native compliant to EU GDPR. CS-MACH1 DMP recommends not to use web forms that are not validated or compliant to EU GDPR.

## 8. Other issues

The DMP aligns with institutional and national data management policies of participating partners and the European Open Science Cloud (EOSC). Updates will be provided in Month 18 and Month 28 (Interim and Final DMP versions), guiding data management processes during the project and for five years after its completion.

## 9. Intellectual Property and Licensing

CS-MACH1 aims at enabling more CS data to be used for community benefit, hence any new data collected by the project in the demonstration cases will be open and free (Open Creative Commons licenses CC-BY/CC0 for non-confidential data) by default.

CS-MACH1 adheres to open science practice hence foreground IP will be made publicly available (in publicly accessible repositories and community tools e.g. Zenodo, CS-MACH1 website).

This general approach is described in detail in the DMP (D1.5, this document and updated its subsequent versions D1.6 and D1.7) which all partners will adhere to.

IPR is handled in accordance with Article 16 of the Grant Agreement. The management of intellectual property (IP) will balance Open Science, FAIR guidelines, and IP protection, aiming to enable result exploitation by other entities through result transfer. Proprietary or restricted data (e.g., pre-commercial sensors) will be justified and documented in DMP updates.

CS-MACH1 is going to develop scientific articles, educational materials, policy-making resources, etc. for which the IP will be the citation and credit to the project (and authors). Any

deviation or any specific case for which the IP rights cannot be considered under an open framework (e.g. CC-BY/CC0) will be described and reported in (one of) the DMP.

## 10. Data Sharing and Long-Term Preservation

### 10.1 Further recommendations and measures on long term preservation of informed consent

Whenever CS-MACH1 will implement surveys, questionnaires, or collect personal data for any reason (e.g., attendance to events), EU GDPR regulation will be used as reference and the user will be informed about the use of personal data. In general, CS-MACH1 will not transfer personal data (e.g., email addresses) to other entities and the only use will be setting up a distribution list to inform users about project progress. Users will be always able to change their consensus and ask for being removed from the distribution channel.

For stakeholder engagement activities, coordinators will collect only the necessary information, including name and surname, affiliation, and activity sector/field of expertise.

Information will be collected by using information sheets and consent forms. Any further contact by the coordinator the activity should be explicitly authorised by the data subject, and it will be strictly limited to the purposes stated in the consent form.

Contact information will be used to involve stakeholders who gave their consent in follow-up communications and efforts, in order to guarantee consistency with the CS-MACH1 multi-stakeholder approach. Workshops and dissemination events will be conducted according to two different approaches throughout the project timeframe, based on the purpose and scope of the initiative:

- i. Events without registration: for these events, coordinators will collect no data about participants, in line with the principle of data minimisation. This approach will be followed whenever it does not impair the fulfilment of CS-MACH1 objectives.
- ii. Events with registration: in order to take part in these events (e.g., workshops), the participants will need to submit a registration form, with a limited number of personal data fields (up to 7). In this case, basic personal information and contact data will be used to facilitate the management of the meeting, giving participants access to preparatory materials. Participants will have the possibility to give their consent to receive information on the organisation of further initiatives and related activities.

In accordance with the EU Charter of Fundamental Rights and with the GDPR, CS-MACH1 will safeguard the following rights throughout the project duration: (i) right to be informed about the personal information shared, (ii) right of access to such information, (iii) right to rectification (in case the information is incorrect or incomplete), (iv) right to erasure of the

data, (v) right to restrict processing of the data, (vi) right to data portability from a service to the other, (vii) right to object.

Concerning the right to be informed (art. 13 and art. 14 GDPR) about the personal data shared with the Consortium (or with one of the partners), when giving the consent to data processing, data subjects will be adequately informed, and will be able to access information on data sharing on the portal of the activity for which data is collected.

Concerning the right to access information shared (art. 15 GDPR), data subjects will have the possibility to request a copy of such data. Based on the internal procedures, the data stored by the organisation will be collected within 30 days. Any potential delay or obstacle (due to the anonymisation/pseudonymisation procedure or to the data storage techniques) will be adequately notified and justified to the data subject.

Similarly, concerning the right to rectify incorrect or incomplete information (art.16 GDPR), CS-MACH1 will ensure that data subjects have the possibility to submit a request of rectification, tp will carry out the necessary procedures within 30 days. Any potential delay will be adequately notified and justified to the data subject.

Concerning the right to be forgotten (art. 17 GDPR), CS-MACH1 will ensure the erasure of the data collected and of the files stored in local archives within 30 days from the official request. When sharing personal data, the subject will be informed of the procedures to obtain the erasure of such data, according to the internal policies and data management infrastructures of the partner leading a certain activity.

Concerning the right to portability of data and their limitations (art. 18 and art 20 of GDPR), the subject will be informed on the procedures to obtain the transmission of data to the physical or legal person designated, based on the internal policies of the responsible partner. The partner will be responsible for the assessment and management of the request and for the aggregation of data in an intelligible format in due time.

## 11. References

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016).

<https://doi.org/10.1038/sdata.2016.18>

[2] CARE principles: <https://www.gida-global.org/care>

[3] EU Survey tool: <https://ec.europa.eu/eusurvey/home/welcome>

