

MARHAB

Improving marine habitat status by considering ecosystem dynamics

Grant agreement number: 101135307

Topic: HORIZON-CL6-2023-BIODIV-01-4

Starting date: 1 January 2024

Duration: 48 months

D1.1

Title: Data Management Plan

Date of delivery: 2024-06-30

Organization name of lead participant for this deliverable: IMR

Deliverable number	D1.1
Deliverable title	Data Management Plan
Work Package	WP1
Due Month	M6
Published data	2024-06-30
Lead Beneficiary	Institute of Marine Research (IMR)
Dissemination level	PU - Public
Deliverable type	R – Document, report

Authors		
Name	Organization	E-mail
Rocío Castaño Primo*	IMR	rocio.castano.primo@hi.no
Camilla S. Landa*	IMR	camilla.landa@hi.no

*Corresponding authors

Keywords
Data Management Plan, Open Access, FAIR

Executive Summary
This document is the Data Management Plan (DMP) for the MARHAB project. It includes a description of the project and the types of data used and generated by the project. The plan includes sections describing how to ensure data FAIRness, as well as security and ethical aspects of the data. This DMP is considered a living document and will be updated throughout the project.

Document history		
Version	Date	Description
1.0	2024-06-30	The first version of the DMP, delivered to the EC as D1.1 at M6.

List of acronyms	
Acronym	Meaning
MARHAB	Improving marine habitat status by considering ecosystem dynamics
DMP	Data Management Plan
WP	Work Package

Table of Content

MARHAB	Error! Bookmark not defined.
1. Introduction.....	4
1.1 About the MARHAB project	4
1.2 The project's Open Science strategy.....	5
2. Data used and generated in MARHAB	6
3. Data Flow?	10
4. Ensuring FAIRness	10
4.1 Making data Findable	Error! Bookmark not defined.
4.2 Making data Accessible	12
4.3 Making data Interoperable	13
4.4 Making data Reusable	14
5. Data Security	14
5.1 During project	Error! Bookmark not defined.
5.2 After project	Error! Bookmark not defined.
6. Ethical aspects	14
7. List of available datasets	15
8. References	15

1. Introduction

This document is the first version of the Data Management Plan (DMP) for the MARHAB project and deliverable D1.1 with due month 6 (M6). The DMP will be updated during and at the end of the project, constituting the deliverables D1.2 at M36 and D1.3 at M48.

The DMP will guide the consortium to manage the data quality and protection issues that will arise throughout the project. The DMP and its deliverable will identify the type of data to be generated by the project team, as well as the standards that will be used to ensure the quality and scientific relevance and impact, conciliating with the Open Research requests of Horizon Europe, respecting IPRs and ensuring FAIR principles and coherence with the EU and national regulations, under the philosophy of making data “as open as possible, as closed as necessary”.

In this document, we give a short summary of the MARHAB project; describe the types of data used and generated across the WPs; and describe how these data will be managed internally during the project, and their availability after the project lifetime. Next, we describe how to ensure data FAIRness, as well as security and ethical aspects of the data. There is also a final section listing all datasets as they become available.

1.1 About the MARHAB project

The vast majority of marine habitats and species protected under the EU Habitats Directive still show an unfavorable conservation status. This situation is particularly pressing in Northern Europe, where the conservation status of marine habitats in the Kattegat-Skagerrak region is consistently bad. MARHAB’s objective is to improve the conservation status of marine ecosystems by demonstrating an ecosystem dynamics approach to restoration and maintenance of protected habitats (Figure 1). By employing cutting edge technology (genomics, tracking, in situ observations, machine learning) in state-of-art ecosystem research, the project will provide the scientific underpinnings needed to bring about the long overdue reconciliation of fisheries management with biodiversity conservation. Moreover, it will generate a leap in knowledge urgently needed to sustainably manage the productive but severely overexploited ecosystems harbored in the Kattegat-Skagerrak Seas and adjoining coastal waters.

MARHAB will focus on data-deficient habitats and data-poor/ data-limited species which are critical to address within the framework of restoring function. Likewise, we will address habitats and species in the worst status (IUCN red list, EU red list, national red lists), or with declining trends, or concerning levels of targeted- or incidental catch in fisheries and pinpoint the management tools with greatest utility in turning negative trends. We also address species for which recovery of populations has created tension with stakeholders.

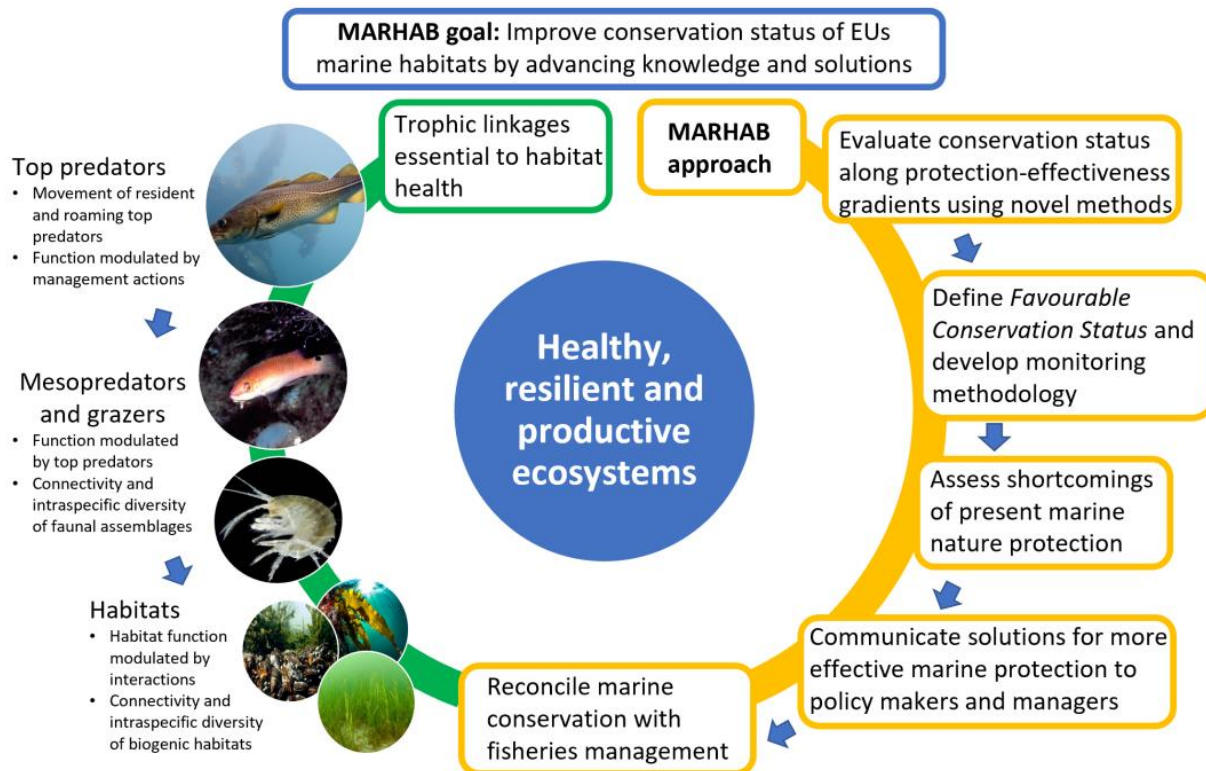


Figure 1: MARHAB's theory of change. Centre bubble: the vision of the project is to contribute to achieving healthy, resilient, and productive ecosystems. Left side of circle: example of linkages between trophic groups involved in ecosystem interactions essential to maintain habitat health. Right side of circle: MARHAB's approach to bringing about our vision statement.

1.2 The project's Open Science strategy

MARHAB will fully align with the Open Science principles of Horizon Europe. The project's Open Science strategy will be mainly based on:

- ensuring the open access obligations whilst properly managing the intellectual property rights of the beneficiaries, implementing a findable, accessible, interoperable, and reusable (FAIR) data management, as described in this document, and the principle of open by default for every data set that is not leading to commercial exploitation
- ensuring the sharing of research outputs as early and widely as possible
- engaging national and international stakeholders and citizens in the implementation of the responsible research and innovation foreseen in the project
- cooperation with other Horizon Europe projects contributing to the biodiversity and ecosystem services destination to protect and restore biodiversity in EU and beyond.

2. Data used and generated in MARHAB

The project will generate and use a wide range of data from different fields of research. This section describes these data, with indications of their provenance and further use within the project. This section will follow the WP structure in MARHAB (Figure 2).

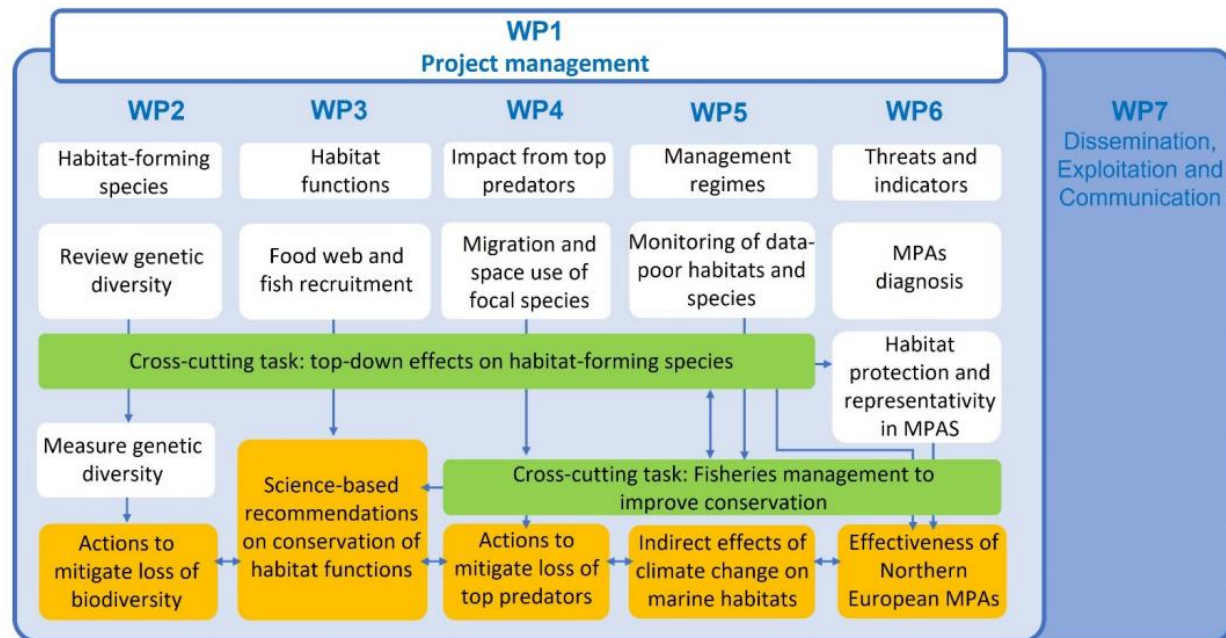


Figure 2: Diagram of the relationship between WPs in MARHAB and the key steps and cross-cutting activities in the project.

To facilitate having an overview and follow up of the datasets generated in the different WPs, an overview table has been set up internally in the project. In it, the project participants provide their input on various aspects of their datasets, by answering a series of questions, summarized in Table 1. At this early stage of the project, several fields remain unclear, and the number and granularity of datasets is to be decided; the table will be a living document, reviewed and refined throughout the duration of the project. Table 2 shows the current response from the template form. It includes WP objectives, dataset descriptions, types, and utilities.

Partner	<i>Institutional partner</i>
WP/ Task / Deliverable	
Contact person and e-mail	<i>Of the person responsible for the dataset</i>
DATA SUMMARY	
Dataset name / brief description	

Preexisting data or wholly generated in MARHAB?	<i>Generated in MARHAB / Generated from pre-existing data</i>
If already existing, DOI or link	
Purpose, target user	<i>State the purpose of the data collection/generation, and which other WPs will use it, if applicable</i>
Observation or model?	
Data source	<i>Instrument, sensor, model short name, sampler...</i>
Expected size	
Level of processing and description	<i>Describe if it's raw, calibrated, postprocessed... Describe calculations, averaging, interpolation performed. QC procedures, etc. ...</i>
Variables and units	
Spatial extent and resolution	<i>Resolution particularly important for gridded and map data</i>
Time span and resolution	
Metadata provided	<i>If some standard is followed, specify</i>
TIMING	
When will be ready to share with partners	<i>By month</i>
When will be ready to publish	<i>Made publicly available. If applicable</i>
INTEROPERABILITY	
File format	
Will you use standard vocabularies? Which?	
If you need to use your own vocabulary, will you provide mappings?	
ACCESSIBILITY	
Will the data be open access?	
If it cannot be published, provide the reason	<i>Specify the reasons: ethical rules of personal data, intellectual property, commercial, privacy-related, security-related</i>
Which repository will you use?	
Are there any ethical or legal concerns?	
Which license will you use?	

Table1: Template for the dataset information form to be filled in by project participants.

WP	WP Objectives	Partner	Dataset name and description	Type	Data utility
WP2 –Intraspecific diversity of habitat-forming species and their inhabitants	1) Provide estimates of genetic condition and genetic connectivity among declining habitat-forming species and their associated fauna. 2) Establish the link between genetic diversity and connectivity of declining habitat-forming species with connectivity of mobile fauna depending on the above habitat-forming species 3) Link genetic estimates of connectivity with other types of data to gain an understanding of the different scales of connectivity, and how they depend on the life-histories of the organisms. 4) Combine connectivity estimates of declining habitat-forming species and their inhabitants with estimates from mobile fauna to identify habitats and geographic areas that are in critical need of restoration.	UGOT	Database on published information on genetic diversity and connectivity for focal habitat forming species and associated fauna along the European Atlantic coast.	Compilation of pre-existing sources	General research use
			Genetic (Axiom SNP array) data from 6 species at 6 different locations, minimum of 160 individuals	New observations from tissue samples	Input for WP3
			Maps/output of a biophysical model on transport and genetic structure at different spatial scales	New model output	Input for WP3
			Maps of genetic diversity health indices	Derived from new data	Input for WP7, policy makers
WP3 – Habitat functions – from point observations to management	1) Describe and quantify habitat functions, using sandbanks and bivalve reefs as selected case studies. 2) Identify pressures and define Favourable Conservation Status in the context of habitat functions. 3) Investigate current conservation status of the functions provided by these selected habitats and propose ways to improve it by improving the ecological potential.	DTU	Compilation of data from field studies about sand banks and bivalve reefs	New observations from video	Input for WP3 and WP5
			Compilation of data from field studies about sand banks and bivalve reefs	New observations from dredging	Input for WP3
			Compilation of data from literature review about sandbanks and biogenic bivalve reefs	Compilation of pre-existing sources	Input for WP3 and WP5
			Quantification of habitat function protection	Derived from MPA databases and fisheries records	Input for WP3, WP4 and WP6
WP4 – Movement, connectivity and ecosystem impact of	1) Obtain knowledge on movement patterns and habitat use of marine top predators in Skagerrak, including fish and marine mammals.	IMR	Telemetry data on movements and diving behaviour of marine top predators (fish and marine mammals) in the Skagerrak	New observations	General research use

top-predators across habitats and seascapes	2) Identify overlapping areas of high use by top predators with human stressors. 3) Identify habitats and geographic areas in need of conservation and restoration of top predators, in order to maintain or restore their ecological functions in maintaining healthy habitats		Telemetry data on movements and diving behaviour of marine top predators (fish and marine mammals) in the Skagerrak	Pre-existing satellite tags observations	General research use
			Maps of habitat suitability for top predators in the Skagerrak	Derived from the telemetry data	Input for WP6
			Maps of overlapped top-predator space use and human activities (fishing effort, other)	Derived from the telemetry data	Input for WP6
WP5 – Integration of fisheries and habitat management to improve conservation status of marine habitats	1) Understand the cascading effects from large predatory fish to habitats, and their typical species 2) Assess how climate change may impact the potential to restore keystone predator populations and their associated ecosystem functions, and improve the conservation status of marine habitats. 3) Broaden the scientific basis for how an adaptation of fisheries management can be used to improve the conservation status of marine habitats.	SLU	Estimations of predation pressure and surface cover of habitat-forming species	New observations from video	Input for WP3
			Environmental variables	New observations from sensors	Input for WP3
			Abundance estimates of top- and mesopredatory species	Derived from observations	
			Estimation of geographical shifts in distribution and suitable habitats	Model output	
WP6 – Threats, indicators, and shortcomings to achieve favorable conservation status	1) Assessing conservation efforts: Conduct a diagnosis of Scandinavian MPAs (Marine Protected Areas). 2) Assessing whether MPAs can appropriately curb local threats/ potential MPA effectiveness. 3) Assessing habitat representativity in different protection levels 4) Informing MPA design and management	CCMAR	Dataset with the list of Northern MPAs, their levels of protection and main features and threats; Associated sources and literature searched. Representativity maps.	Compilation of pre-existing sources	General research use

Table 2: Dataset information provided by the project participants.

3. Data Flow

“Data flow” can be understood as who is using data from whom, or the series of “places” where the project data is to be hosted.

The technical infrastructure for data storage and data sharing between project participants during the project's lifetime is not decided on and may consist of several solutions. Since much of the work within each WP is performed by a single partner institution, the data sharing within each WP can be performed using each partner's internal system and procedure for data sharing. Eventually when sharing is necessary across WPs, a common file sharing service might be necessary. One solution is then for IMR to set up an ownCloud service for all project participants (see section 5). OwnCloud is an open-source alternative to the commercial file sharing services, and enables secure storage, collaboration and sharing. Note that ownCloud is not a suitable repository for archiving or publishing data after the end of the project.

The long-term stewardship of data, however, will fall on and be guaranteed by existing, well-established repositories. The main repository for long-term archival and publishing will be the Norwegian national Infrastructure for marine data, the Norwegian Marine Data Centre ([NMDC](#)), which is coordinated by the Norwegian Marine Data Centre (NMD) at IMR, certified by CoreTrustSeal. NMD data management is based on the FAIR principles.

Data from WP4, coordinated by IMR, will be long-term archived and published at NMDC. It is possible but not mandatory for other MARHAB partners to use the NMDC infrastructure. The partners can use the repositories they prefer and/or are required to use (e.g. requirements to use national repositories). All partners are responsible for ensuring ‘their’ data is timely archived and published according to FAIR principles.

4. Ensuring FAIRness

The FAIR Guiding Principles for scientific data management and stewardship (Wilkinson *et al.* 2016) are a series of general principles aimed at increasing the reuse of scientific data with a particular emphasis on machine actionability. They cover aspects of Findability, Accessibility, Interoperability and Reusability at the metadata and data level. The principles are as follows:

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communication 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

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be Reusable:

R1. meta(data) are richly described with 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 detailed provenance

R1.3. (meta)data meet domain-relevant community standards

FAIR data exist in a continuum, with different degrees of FAIRness and room for improvement (see Figure 3). Not all fields of research present the same level of FAIR maturity. Fields that generate large amounts of data by “machines”, where global collaboration is at the basis of research, or require significant infrastructure, tend to be more FAIR. Genome research (GenBank), remote sensing, weather forecast, or climate modelling (CMIP) are examples of these kinds of disciplines. The principle of Interoperability is particularly challenging to achieve. It requires, among other aspects, wide community agreements on standards and a certain level of governance to maintain them.

4.1 Making data Findable

The datasets made publicly available by MARHAB will receive a globally unique, Persistent Identifier (PID) from the data repository. Digital Object Identifiers (DOI) are the most common PIDs, although the handle system (hdl) is used by some repositories. Other types of PID exist, but they are rarely used for research data. PIDs are globally resolvable, point to a single digital object and are independent of location. Most of the long-term repositories use DOIs by default, and it will be the preferred PID in MARHAB.

All data will be characterized by a full set of metadata precisely describing the nature and characteristics of a dataset. The data will be richly described with query metadata to improve findability. To the extent possible, standard metadata fields and keyword vocabularies will be used instead of descriptions and free text. The MARHAB project will be acknowledged via general keywords, project metadata fields or included in the description, depending on the options offered

by the data repository. If applicable, version numbers will be provided for provenance. If the data is associated with a publication, the PID of the publication will be included, too.

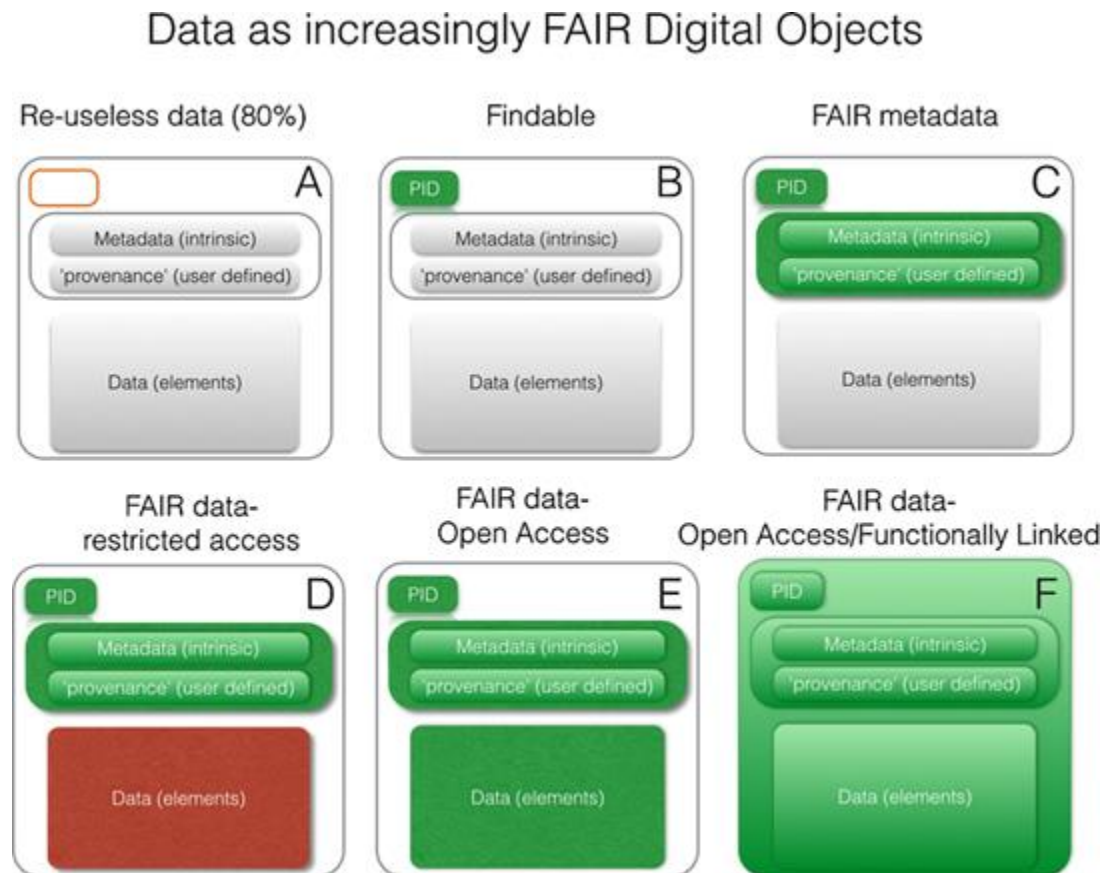


Figure 3: Varying degrees of FAIRness. From Mons et al. (2017).

4.2 Making data Accessible

Following the H2020 and Horizon Europe guidelines on research data, MARHAB datasets will be made public without restriction by default, unless well justified. However, some preliminary results will be shared within the consortium first for quality assessment and analyses.

Data will be deposited in repositories that comply with the FAIR principles. Repositories that do not require registration are preferred. The machine accessibility of the data and metadata will be granted by the repositories. If data needs to be pulled out, a strong case must be made, and the landing page of the dataset will be substituted by a “tombstone” page that describes the data and the reason for withdrawal. Different versions of the same dataset will link to each other.

4.3 Making data Interoperable

The major principle in data interoperability is that data exchange, reuse, and the potential for re-combinations with different datasets from different origins is maximized and feasible at machine level. In MARHAB, interoperability will be tackled by appropriate, community-agreed formats and controlled vocabularies where they can be found.

MARHAB is a multi-disciplinary project, where the different fields have reached variable levels of interoperability maturity. At a minimum, file formats not exclusive to commercial software, descriptive metadata and an internal, controlled list of terms are the basic requirements for data publication.

Table 3 summarizes some of the existing formats and controlled vocabularies, and to which type of data they apply.

Item	Name, link, other info	Governance
Sensor models	SDN L22: http://vocab.nerc.ac.uk/collection/L22/current/	BODC NVS
Institution	EDMO: https://edmo.seadatanet.org/ ROR: https://ror.org/	SeaDataNet
Variable names	SDN P01: http://vocab.nerc.ac.uk/collection/P01/current/ SDN P09: http://vocab.nerc.ac.uk/collection/P09/current/ Climate and Forecast (CF) standard names: https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html	BODC NVS CF
Units	SDN P06: https://vocab.nerc.ac.uk/collection/P06/current/	SeaDataNet
Quality Flag Scheme	SDN L20: https://vocab.nerc.ac.uk/collection/L20/current/ WOCE QC flags	SeaDataNet WOCE
Keywords	Global Change Master Directory keywords (GCMD): https://www.earthdata.nasa.gov/learn/find-data/idn/gcmd-keywords	NASA
Time	ISO8601	ISO
Country	ISO3166	ISO
PI	ORCID https://orcid.org/	ORCID
Species	World Register of Marine Species (WoRMS): https://www.marinespecies.org/	LifeWatch
Discovery metadata	Attribute Convention for Data Discovery (ACDD): https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery	ESIP
Atmosphere, ocean data	NetCDF-CF: https://cfconventions.org/	CF
Ocean mooring data	OceanSites (and derived): http://www.oceansites.org/data/index.html	OceanSITES
Map data	GeoTIFF: http://www.opengis.net/doc/IS/GeoTIFF/1.1 GeoJSON: https://geojson.org ESRI shapefiles: https://www.esri.com/content/dam/esrisites/sitecore-archive/Files/Pdfs/library/whitepapers/pdfs/shapefile.pdf	OGC IETF Esri
Biological data	DarwinCore Archive (DwCA): https://dwc.tdwg.org/	TDWG
Genomics		

Table 3: A selection of controlled vocabularies, identifiers, and data format standards available

4.4 Making data Reusable

MARHAB data will be published with the machine-readable Creative Commons license CC BY 4.0 (Attribution 4.0 International) and adhere to the 'Open' definition as defined by the Open Knowledge Foundation. CC-BY specifically reflects the academic practice of source citation and allows for a less restrictive reuse, including the use in larger data synthesis. In turn, datasets generated externally and used in MARHAB will be adequately cited, following the FORCE11 Joint Declaration of Data Citation Principles (FORCE 11, 2014). Unless appropriate reasons can be stated, the usage license should not be more restrictive than CC-BY: NC (Non-Commercial), and/or SA (Share-Alike). While seldom used, CC0 (public domain) can be used if the dataset creators choose to.

Any software products or tools developed during the project will be published under open source licenses (GNU, MIT, among others) and archived to public repositories allowing process-based assessment methods to be validated, repeated, and reused. <https://choosealicense.com/> contains information about the different licenses and their conditions and level of restrictions.

5. Data Security

The data will be safely stored both during the duration of the project, and after the project has ended.

During the project, the data is stored in the originating data collecting system for partners, and possibly also in a centralized file sharing service. A good option for such centralized sharing would be for IMR to set up an ownCloud file sharing service. ownCloud is an open-source alternative to the commercial file sharing services, and enables secure storage, collaboration and sharing. The ownCloud service would be hosted on IMR internal servers. The data stored on the ownCloud would be included in IMRs backup routines which are managed by the IMR IT Department and follow ISMS standards (based on ISO 27001/2).

After the project has ended, the data will be safely archived in international, and preferably certified, repositories for long-term preservation and curation. The main repository for long-term archival and publishing will be the Norwegian national Infrastructure for marine data, the Norwegian Marine Data Centre ([NMDC](#)), which is coordinated by the Norwegian Marine Data Centre (NMD) at IMR. NMD has been an IOC/IODE National Oceanographic Data Centre since 1971, is a regular member of the [ISC World Data System](#), and is certified as a trusted repository according to [CoreTrustSeal](#). Other data repositories might be used, but this is not decided on at this stage in the project.

6. Ethical aspects

To ensure that the ethics issues are adequately addressed, the project will appoint a team member as Ethics Manager. The Manager will support the consortium on all ethics matter and will prepare a report to be submitted at the end of each reporting period.

The fulfillment of Open Access requirements for data in Horizon 2020 and subsequently in Horizon Europe is predicated in the general principle of “as open as possible, as closed as necessary”. In MARHAB some data must opt out of open access and remain restricted during and after the project lifetime. The reasons fall in two principal categories:

- Data sensitive to commercial operations and/or owned by external institutions.
- Personal data used for administration purposes (e.g., event organization) will be managed following GDPR guidelines, and not be published, since it is not a relevant research output.

For some external data sources, like the AIS data, the data owner decides on the level of restrictions when it comes to publishing. Such restrictions can for instance be related to the allowed granularity of the data. Dialogue will be kept between MARHAB and the data owner(s).

The tagging of fish and marine mammals (T4.1) requires a series of permits from the different food authorities (Mattilsynet in Norway). These will be acquired, and the fieldwork will be carried out in a way that minimizes stress to the animals.

7. List of available datasets

A list of available datasets will be added here in later versions of this document. The list of available datasets will also be made available on the [project's website](#).

8. References

- FORCE 11 Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA; 2014 <https://doi.org/10.25490/a97f-egykh>
- Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. Information services & use, 37(1), 49-56. <https://doi.org/10.3233/ISU-170824>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9. <https://doi.org/10.1038/sdata.2016.18>