



# DATA MANAGEMENT PLAN



Funded by  
the European Union





Published by:

Helsinki Commission – HELCOM  
Katajanokanlaituri 6 B  
00160 Helsinki, Finland

[www.helcom.fi](http://www.helcom.fi)

Information and views expressed in this publication are the authors' own and might vary from those of the Helsinki Commission or its members.

All rights reserved. Information included in this publication or extracts thereof, with the exception of images and graphic elements that are not HELCOM's own and identified as such, may be reproduced without prior consent on the condition that the complete reference of the publication is given as stated above.

This document is a publication under work package 2 "Data" of the PROTECT BALTIC project.

For bibliographic purposes this document should be cited as:  
"Protect Baltic Data Management Plan – HELCOM (2024)"

Available at: <https://protectbaltic.eu/resources>

© Baltic Marine Environment Protection Commission – Helsinki Commission (2024)

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them.

Layout: Paul Truth

Photo credits: cover photo © Adobe Stock

## Acronyms

API	Application Programming Interface
BioDiv DB	<a href="#">HELCOM Biodiversity Database</a>
BioLib	<a href="#">Biological Library</a>
CSW	Catalogue Service for the Web (Open Geospatial Consortium)
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	Ecosystem Component
EMODnet	<a href="#">European Marine Observation and Data Network</a>
ES	Ecosystem Service
Esri	Environmental Systems Research Institute, commercial GIS company
FAIR	Findability, Accessibility, Interoperability, and Reusability
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
GEMET	<a href="#">GEneral Multilingual Environmental Thesaurus</a>
GIS	Geographic Information System
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
ICES	<a href="#">International Council for the Exploration of the Sea</a>
INSPIRE	<a href="#">Infrastructure for Spatial Information in Europe</a>
ITIS	Integrated Taxonomic Information System
MADS	<a href="#">HELCOM Map and Data Service</a>
MPA	Marine Protected Area
MPA db	<a href="#">HELCOM Marine Protected Areas database</a>
MSP	Maritime Spatial Planning
OBIS	<a href="#">Ocean Biodiversity Information System</a>
WoRMS	<a href="#">World Register of Marine Species</a>
WDPA	The World Database on Protected Areas
WP	Work Package

## DATA MANAGEMENT PLAN

PROJECT	
Project number:	101112866
Project acronym:	PROTECT BALTIC
Project name:	Enabling comprehensive effective and efficient protection and restoration measures for a resilient Baltic Sea ecosystem

DATA MANAGEMENT PLAN	
Date:	14/02/2024
Version:	1.0

## Contents

1.	Introduction .....	4
2.	Data summary .....	4
2.1.	Data types and formats used in the project .....	5
2.2.	Purpose of the data generation and re-use.....	6
2.3.	Expected size of the data .....	6
2.4.	Origin of the data .....	6
2.5.	Data use outside the project.....	8
3.	FAIR data .....	8
3.1.	Making data findable, including provisions for metadata .....	8
3.2.	Making data accessible .....	10
3.3.	Making data interoperable .....	12
3.4.	Increase data re-use.....	13
4.	Other research outputs.....	14
5.	Allocation of resources .....	14
5.1.	Cost for making the data FAIR .....	14
5.2.	How the mentioned costs are covered .....	14
5.3.	Responsibilities of data management.....	14
5.4.	Ensuring the long term preservation of the data .....	15
6.	Data security .....	15
6.1.	Provisions for data storage and transfer safety.....	15
6.2.	Safety of long term preservation of data.....	15
7.	Ethics .....	15
7.1.	Ethical and legal issues on data sharing.....	15
7.2.	Consent for data sharing.....	16
8.	Other issues.....	16
8.1.	Use of national/funder/sectorial/departmental procedures for data management..	16

## 1. Introduction

This data management plan (DMP) describes the data management processes and environment in the Horizon Europe's [PROTECT BALTIC](#) project which is often referred as "project" in the document. DMP delineates the used data formats and publication policies. This is a working document and will be updated during the project until the final version which is to be delivered on Q2 of 2028.

The plan presents the types of data used in the project, where the data is stored, and how it is generally used and re-used. It illustrates the common cycle of data and metadata in PROTECT BALTIC's data management environment throughout the project period and afterwards.

The DMP serves as a guide for project personnel in partner organizations and in the HELCOM Secretariat. It demonstrates the agreements and policies that need to be considered when processing the data.

The DMP highlights the general issues to be considered in data management, such as privacy and security aspects like the General Data Protection Regulation (GDPR), ethical points, and the different rules and regulations to be followed.

Data management in PROTECT BALTIC complies with the [HELCOM Data policy](#) which is explained in more detail in Attachment 2 of the [HELCOM monitoring and assessment strategy](#). PROTECT BALTIC has a [Privacy Policy](#) for addressing practices that concern personal information.

Special attention in the DMP is paid to findability, accessibility, interoperability, and reusability (FAIR-principles) for metadata and data.

## 2. Data summary

Document distribution, viewing, processing and storage in the project should be done in HELCOM's Microsoft 365 platform. The platform's applications such as SharePoint and Teams e.g. allow multiple users to view and edit the same file simultaneously and to have strong information security through restricted access and used security services. Spatial and non-spatial data collected for PROTECT BALTIC is primarily hosted in HELCOM databases such as the [HELCOM Map and data service](#) (MADS), [HELCOM Marine Protected Areas database](#) (MPA db), [BASEMAPS](#) - maritime spatial planning database and the [HELCOM Biodiversity Database](#) (BioDiv DB).

Most data used in PROTECT BALTIC is stored in HELCOM's Microsoft 365's SharePoint - collaboration and document management platform, HELCOM databases, or in the databases of other institutional service providers.

Data is continuously re-used for research, modelling, documents, publications, reports, different map products and manuals. Data in HELCOM's databases is updated in data calls and as an outcome of different projects. Part of the data used in PROTECT BALTIC is natural and social sciences-, policy- or legislation literature, which can also be considered as re-use of existing data.

For PROTECT BALTIC there are two data calls issued during the project for collecting up-to-date data which has not been previously available. At the start of the PROTECT BALTIC project, updating part of the existing data in the first data call was considered but the data collected and compiled in HELCOM 's Red List II, HOLAS III and BaltiCheck projects was found to be sufficiently recent for PROTECT BALTIC purposes.

All the data management done by individual researchers and people involved in the project are not in the scope of the data management plan (DMP). The DMP outlines how the data is shared, managed, and published within the data management environment of the PROTECT BALTIC personnel. It outlines what databases are used and how the data is stored and distributed in the



project. Data is constantly re-used, and the majority will remain openly accessible after the project securing its re-use also in the future.

A list of commonly used HELCOM data services, descriptions of their usage and their sources can be found in Table 2. In addition to HELCOM's own data services, there are numerous institutional data providers from which data is used and re-used. A list of services can be found in Table 2's Data sources column.

## 2.1. Data types and formats used in the project

As primary data, PROTECT BALTIC collects biotic and abiotic data from contracting parties in two consecutive data calls. Secondary data is collected from external sources such as registries or surveys which are done by other than project parties. Collected primary and secondary data are used to analyse and create models for new data and to prepare end products.

Primary biotic data consists of species observations and distribution models of benthic species, birds, fish, and decapod crustaceans. Attributes vary between datasets of different species. The scientific name, quantity, time of observation and coordinates are generally collected from all. A tabular format is preferred as there are multiple attributes requested from each species. The rest of the primary data is to be returned as shapefiles or raster maps. Requested abiotic datasets are bottom organic matter content concentration, chlorophyll-a coverage, coloured dissolved organic matter, hypoxia, sediment maps, sedimentation accumulation, spawning and nursery areas of fish and turbidity.

As secondary data sources PROTECT BALTIC uses e.g. ICES, EMODnet (European Marine Observation and Data network) and CDDA (Common Database of Designated Areas) in addition to data that is existing in HELCOM's databases.

Since most of the required data is requested for the first time, there was a lot of leeway given to data providers. It was expected that there were differences in used units, methods, and attributes. The main goal is to gather as much of the required data as possible and then do the complementation and harmonization work at the HELCOM Secretariat when compiling the data.

During the project, multiple data formats are used and generated in different operating systems (e.g. MacOS, Windows, Linux) in different individual work processes. Intermediate files produced in these processes are not in the control of the project. The project will choose the most appropriate format and file types which are managed between project personnel and published for end users.

The project's final products are often PDF files which can be opened with most browsers and with numerous non-commercial desktop software. Documents in commercial formats such as docx and xlsx -files can be transformed into open document file formats (odt, ods respectively) on demand.

Most of the used data is in databases and is generally downloaded in tabular or shapefile formats when needed. Direct user interfaces for SQL-commands are not available but it is possible to make simple queries in MADS and HELCOM's Marine Protected Areas database (MPA db).

Final map products can be viewed in open access map viewers (e.g. MADS and BioDiv DB) or downloaded and utilized in open-source GIS-software such as QGIS.

The most used file formats in the project are listed in *Table 1*.

Data	File format
------	-------------

<b>Tabular</b>	xlsx, csv, xml, dbf, ods
<b>Document</b>	pdf, docx, txt, odf, pptx, odt, rdf
<b>Spatial</b>	shp and geodatabase files with their respective component files
<b>Programming, scripting</b>	html, css, py, js, json, Rdata, Rds
<b>Image (also raster images for spatial data)</b>	tiff, gif, psd, png, ai, eps, jpg, svg
<b>Video, animation</b>	mp4
<b>Other</b>	zip

*Table 1. Commonly used file formats in PROTECT BALTIC*

## 2.2. Purpose of the data generation and re-use

The project application form states: “The main project innovation will be a Protection Optimization Framework, a holistic transboundary decision support package. The Baltic Sea region would come together to develop a transparent, data driven, science-based, step-by-step approach for identifying targets for protection and restoration and develop a shared approach for implementation all based on the latest scientific knowledge.”

The project being science-based denotes that it relies on most accurate and timely data available. Analysis and modelling require biotic and abiotic data which is collected from HELCOM databases and from selected institutional databases. For the data, which is considered deficient, two consecutive data calls are issued.

End- and intermediate products of the project such as maps, models, assessments, guidelines etc. are built on the information based on gathered data.

Project objectives 1 (Enable sound design and planning of spatial protection and restoration measures) and 3 (Enable effective management of spatial protection and restoration measures) are data dependent and extensively utilize collected data for analysis or models. Also, the components of objective 2 (Enable good governance of spatial protection and restoration measures) have either direct or indirect basis on data that has accumulated during the project.

## 2.3. Expected size of the data

GIS-related materials and products are often large in file size. Also, source material datasets used in models and analysis can be excessively large.

The total space required for the data for PROTECT BALTIC can only been roughly estimated to be several terabytes, but the load is divided between the partners’ data management environments. The amount of project-specific data at HELCOM is not monitored.

## 2.4. Origin of the data

A large part of the data used in PROTECT BALTIC are in HELCOM databases such as BioDiv DB, MADS, and MPA db. The mentioned databases consist of existing data collected in earlier projects and of data which is compiled from external sources. Existing data is updated during the project.



BioDiv DB collates species observation data from various sources and data collection activities. In addition to macrospecies observation data made available by HELCOM contracting parties, it compiles data from different external sources, e.g. AquaNIS and ICES NIS observations.

MADS is an online platform which contains spatial data collected within the HELCOM data collection framework and related projects. It provides open access to a wide range of geographic and environmental data related to the Baltic Sea region. For example, datasets on bathymetry, biodiversity, marine protected areas, shipping traffic, and hundreds more. Original sources for the data can be tracked from the metadata which are stored in the [HELCOM Metadata catalogue](#).

HELCOM Hub is an underwater biotope and habitat classification system which provides a framework for defining biotopes in the Baltic Sea. It is used, for example, to compare biotopes in different regions of the Baltic Sea.

MPA db contains marine protected area (MPA) data provided by contracting parties and is continually updated. The database is utilized by designating sites with special nature values as protected areas, and by managing human activities within those areas. Each site has a unique management plan in the database.

In addition to national sources from contracting parties, HELCOM databases contain data from international organizations and research institutions from countries outside the Baltic region. For example, land cover map for MADS is produced by NASA and the University of Maryland.

Some of the base maps are from non-profit organizations such as administrative boundaries from Open Street Maps or can be free of charge distributions of commercial companies like Esri's Digital Chart of the World. Data utilized in the project is generally free excluding AIS data, which is acquired by HELCOM. All data published by HELCOM can be used freely given that the source is cited (following the creative commons license CC-BY).

Data in HELCOM databases accumulates from projects and is collected from external sources. Generally, data originates from field surveys and has then been evaluated and compiled with the expertise of national experts. PROTECT BALTIC utilizes both newly collected data and existing archived environmental data, models, scripts, results from other projects or studies and satellite data. Detailed information of data lineage can be found in each dataset's metadata.

HELCOM data services used in PROTECT BALTIC with respective sources are listed in Table 2.

HELCOM service	Description	Data sources
<b>HELCOM Biodiversity Database</b>	Macrospecies observation data made available by HELCOM Contracting Parties	National databases in contracting parties, other HELCOM databases, AquaNIS
<b>HELCOM Map and data service (MADS)</b>	Online platform that provides open access to a wide range of geographic and environmental data related to the Baltic Sea region	National organizations and research institutions in contracting parties, other HELCOM databases, Commercial companies (e.g. Esri), non-profits (e.g. Open Street Maps), national organizations and research institutions outside contracting parties (e.g. NASA, USGS), international organizations (e.g. UN, European Environment Agency, ICES (-ocean, -dome, -VMS and Logbook),

		AquaNIS, NOBANIS, BIAS, EMODnet, EMSA, EDP: CCM2, Eurostat
<b>HELCOM Marine Protected Areas database</b>	Centralizes relevant information on the Baltic Sea protected areas. The database includes general information on the sites and their management plans, as well as lists of species, biotopes, and biotope complexes	Contracting parties' continually update
<b>HELCOM Metadata catalogue</b>	Service of viewing and downloading + 1000 metadata of datasets stored in HELCOM databases	HELCOM data services
<b>HELCOM Hub</b>	Provides a framework for defining biotopes in the whole Baltic Sea. Defining biotopes by the same split-rules, enables comparison of the biotopes in different regions of the Baltic Sea	EUNIS compatible classification
<b>HELCOM Coastal fish core indicator database (COOL)</b>	Data collected from coastal fish monitoring and used for coastal fish core indicators	Contracting parties' national databases

Table 2. HELCOM data services used in PROTECT BALTIC

## 2.5. Data use outside the project

Companies involved in marine area and coastal planning, decision makers and officials in different levels of national administrations, international organizations, other projects, NGOs, researchers, students, journalists, and individuals who are interested in the planning and state of the Baltic Sea.

## 3. FAIR data

### 3.1. Making data findable, including provisions for metadata

#### *Persistent identifiers*

Data findability will be maximized by applying persistent identifiers (URLs from HELCOM Metadata catalogue) and/or assigning a digital object identifier (DOI) to each dataset/data product and publishing its associated metadata record.

Metadata of all datasets/data products will be stored and published in the HELCOM Metadata catalogue. Where relevant, the data may also be submitted to other repositories, such as EMODnet or OBIS (Ocean Biodiversity Information System), to increase findability. The metadata will adhere

to international standards such as ISO 19115 and INSPIRE (Infrastructure for Spatial Information in Europe).

At the time of writing DMP v.1.0, there is no active DOI service in HELCOM, but the acquisition is being investigated. This applies to publications other than metadata where the unique identifiers are provided consistently.

#### *Metadata types and standards*

Data in the HELCOM Metadata catalogue is EU INSPIRE compliant, which defines the data as rich.

Datasets collected in data calls include metadata templates which are to be filled by the contracting parties. Each metadata in the HELCOM Metadata catalogue contains the following fields: INSPIRE themes, Categories, GEMET - INSPIRE themes, version 1.0, GEMET, Other keywords, Language, Resource identifier, Legal constraints, Contact for the resource, Technical information, Representation type, Coordinate reference system, Format, Lineage, Metadata information, Download metadata, Contact, Metadata language, Identifier, Temporal extent and Provided by. Spatial datasets have an additional Spatial extent -field.

Search words for “GEMET” and “GEMET – INSPIRE themes” -fields comply with GEMET (General Multilingual Environmental Thesaurus) vocabulary specifications.

HELCOM metadata complies with the ISO 19115 standard, EU INSPIRE metadata Implementing Rules (e.g. abstract, lineage etc.), and DCAT-AP (Data CATALOGue vocabulary - Application Profile for data portals in Europe) specification.

Metadata records can be harvested to GIS software, other metadata catalogues or information sources using CSW (Open Geospatial Consortium’s Catalogue Service for the Web) standard. CSW supports a query language which allows clients to formulate more complex queries when required.

All new spatial data published in PROTECT BALTIC through MADS will have a record in the HELCOM Metadata catalogue. New non-spatial data will have metadata when deemed necessary.

HELCOM is a partner in the AquaINFRA project, which will develop seamless data delivery and access service (DDAS) for freshwater and marine data. PROTECT BALTIC’s metadata records will be included in AquaINFRA DDAS via the HELCOM Metadata catalogue, which is part of the DDAS federated metadata search.

#### *Search keywords in metadata*

General search themes, categories and keywords are provided in the HELCOM Metadata catalogue in “INSPIRE themes”, “Categories”, “GEMET - INSPIRE themes, version 1.0”, “GEMET” and “Other keywords” -elements. Mentioned standards and specifications are explained in the previous chapter.

#### *Metadata harvesting for external usage*

In addition to normal unrestricted website access, HELCOM metadata is available through application programming interfaces (API) at:

<https://metadata.helcom.fi/geonetwork/doc/api/index.html> and is already utilized by services such as [data.europa.eu](https://data.europa.eu).

Metadata is based on the ISO 19115 / INSPIRE metadata standard template and downloadable in zip, pdf, xml and rdf -formats. Each metadata record can be linked to a targeted website with a permalink. Additionally, metadata can be viewed with services that support the CSW standard.

Where relevant, the metadata may also be linked to other repositories, such as EMODnet or OBIS, to increase findability.

### 3.2. Making data accessible

#### *Repository:*

HELCOM data is stored in secured and protected Cloud services provided by Advania. Access is restricted and password protected. There is active virus protection and SPAM-filtering for emails in the used platform.

Access to original spatial data is restricted to HELCOM Secretariat's data management personnel.

#### *Background information of the used repositories*

HELCOM's spatial databases are stored as Esri's geodatabases or in an internal network drive accessible from Cloud-based servers. Geodatabases and the non-spatial password protected MPA db relational database are stored in cloud services.

Non-spatial data which is not in BioDiv DB is managed in password protected Microsoft 365 SharePoint document management and storage system.

Selected HELCOM data is available e.g. in EMODnet, and the World Database on Protected Areas (WDPA). Since the data is under the creative commons license CC-BY, the instances where data are published are not all known.

Data storage solutions for the new MPA Portal, which is to be built under WP8 in PROTECT BALTIC, have not yet been decided.

#### *Data identifiers*

When a record is added in the HELCOM Metadata catalogue, it will have a unique id in the Identifier-element which can be accessed with a URL. Metadata contains information where the actual data is located and where it can be found.

#### *Data openness and availability*

All published end products of PROTECT BALTIC are openly available from the project's website.

Software, scripts, and code will be published on [HELCOM Secretariat's GitHub](#) under an open public license, where applicable.

Data openness of the metadata is achieved by ways listed in the chapter [Making data findable, including provisions for metadata](#).

Certain datasets used in PROTECT BALTIC's work may contain classified, sensitive, or restricted information (such as high-resolution bathymetry or accurate position data of endangered species). To manage information according to law, data providers are instructed to prepare descriptions of the restrictions in the metadata accompanying the data submission to HELCOM. Each restricted data is handled on a case-by-case basis, and lawful use confirmed with the PROTECT BALTIC ethic mentor. Some restricted data might be partly usable e.g. by lowering the resolution for publications. Other measures which can be taken include e.g. aggregation of the data layers to make reverse engineering of the sensitive data impossible. It is also possible to have security clearances made for the modellers / data handlers.

#### *Possible embargoes and intellectual rights*

There are no embargoes for the data. The project's results, publications and other end products will be publicly available when finalized. The Consortium Agreement outlines that the project coordinator and partners agree that the products and tools developed under the project will be under the Creative Commons Attribution-Share A like 4.0 (CC BY-SA 4.0) license, except if such

tools and products are protected by copyright law or a patent law already exists or that are further developed during the project (no such tools or products are foreseen under this project).

#### *Data accessibility and protocols*

“Standardized access protocol” is regarded here as common http (Hypertext Transfer Protocol), https (Hypertext Transfer Protocol Secure) or ftp (File Transfer Protocol) protocol of which http is used for HELCOM’s publicly open data services. Ftp connections are generally not used in the project but can be used for data transfers when needed.

The HELCOM Metadata service provides free access and data can be retrieved by identifiers.

Reports and publications are available in PDF format and generally written in English. All published data are accessible through PROTECT BALTIC’s website, HELCOM data services, and other relevant sources. PROTECT BALTIC data can contain links to other publications and websites where openness cannot be guaranteed.

Several HELCOM services have REST APIs which allow client computer connections to data. Services usable for PROTECT BALTIC are mainly MADS and BioDiv DB.

Spatial data can be downloaded in formats that are usable with open-source GIS-software such as QGIS.

Connection solutions for the new MPA Portal have not yet been decided but general access to the data will be open.

#### *Access to restricted data*

Currently, used data is generally not restricted, sensitive or classified. However, it should be noted that there might be a need for data which has limitations in usage and publication. For example, high resolution bathymetric data may have restrictions and publication of exact habitats for endangered species can be limited. In these cases, PROTECT BALTIC will comply with the prerequisites of the data provider.

#### *Access control*

File distribution of files and communication in PROTECT BALTIC takes place mainly in SharePoint and Teams which are services in the Microsoft 365 product family where access is restricted. User accesses are managed in the Microsoft 365 identity and access management where each user has been invited with their personal email address.

#### *Data access committee*

Due to the open data policy and because the project is only marginally using restricted or sensitive data, a data access committee was considered not to be needed.

#### *Metadata openness and availability*

Metadata will be made openly available licenced under a Creative Commons license CC-BY. The difference between CC0 and CC-BY is that, in CC-BY, HELCOM retains the legal right to be acknowledged as the originator of the data.

#### *Data- and metadata availability after the project*

Data produced during the project are stored in HELCOM databases, which contain data from data calls, from other projects and from external sources. This data will be available also after the PROTECT BALTIC project concludes. Metadata related to datasets is available in the Metadata catalogue from where it is freely downloadable and redistributable.

Data which has been created for communication and for sharing information for project purposes will be available for the partners for the lifetime of the project but will be stored in the HELCOM servers for five years after the project.

#### *Documentation of needed special software*

There is no need for special software to access the data. All data can be utilized with open-source software. As an example, text, spreadsheet, and pdf documents can be accessed with LibreOffice software suite and spatial data with QGIS.

Connection to REST APIs requires REST client software.

### **3.3. Making data interoperable**

#### *What data and metadata vocabularies, standards, formats or methodologies are used*

MADS portal provides data in [Open Geospatial Consortium](#)'s WMS and WFS formats. For software connectivity, there are REST APIs available for the [BioDiv DB](#) and [MADS](#) -services.

Metadata templates applied in the metadata catalogue contain fields described in the ISO 19115 metadata standard and EU INSPIRE metadata Implementing Rule. GEMET is used for keywords for categorizing datasets in HELCOM Metadata catalogue. More information on metadata interoperability is available under the section "Metadata types and standards".

Final products are translated into English to ensure broad impact potential. PROTECT BALTIC utilizes generally used file formats, the most common of which are listed in *Table 1*. Most of the files are easily convertible to open document formats when needed.

BioDiv DB contains species observation data. Taxonomy in this database is based on one of the international databases WoRMS, ITIS, BioLib, or other. The used taxonomy is mentioned in the id section of each record. Species with an existing Aphia ID retrieve their taxonomy entries via a customized script to WoRMS.

Basemaps / background maps are either open source or from Esri's freely distributable maps. Sources of used special background data are listed in metadata.

The HELCOM Hub is compatible with EUNIS classification.

#### *Used uncommon ontologies or vocabularies*

PROTECT BALTIC is not producing project-specific ontologies or vocabularies. All used vocabularies are generally used within their respective fields of work.

During the project, the HELCOM Secretariat published a living document "Shared vocabulary of protection related terminology to ensure a consistent approach within HELCOM" to make sure that there is common understanding on used terminology and to avoid internal and later external misconceptions of the discussed topics. This vocabulary is available at:

<https://protectbaltic.eu/glossary> and will be updated throughout the life of the project.

#### *Qualified references to other data*

Project data and metadata can contain qualified cross-references to other data or metadata. Referenced data may also be from previous research of other projects.

Data collected in PROTECT BALTIC WP2's data calls are used as a raw data and data material for the other work packages. Models, maps, and results of analysis can be used as an input for further analysis.

For WP4 (Ecosystem Services and Valuation), raw data received in the data call as well as existing data gathered from prior data calls will be used to create and develop existing models which relate to three interlinked aspects of WP4. These aspects are ecosystem component (EC) location and habitat quality; human activities/impacts, as well as the social and economic value(s) of ecosystem services (ES) provision within the Baltic Sea MPA network.



Regarding data use, the first aspect will build models evaluating how habitat quality influences the provision of ES (Task 4.2). This will be mainly based on the biotic and abiotic data received/available within the project, where EC – e.g., biotic species observations and abiotic characteristics – location will be factored into analysis of ES provision.

Relating to the second aspect, data on the human impacts to these ecological variables will be accounted through its potential affect to ES supply (Task 4.3) – the existing and updated data on human activities will be used by WP3 and WP4 T4.1 to create further models.

The outputs from these two aspects will be used in the social and economic valuation of various ES within the Baltic Sea MPA network, where ES (based on the ecological data) will be spatially mapped to identify key areas for ES production inside and outside MPAs and analyse how well the current MPA network covers different ES categories, their production, and values (Task 4.5). Additionally, data on human preferences regarding different ES and their associated benefits may be evaluated through targeted surveys, as well as expert and stakeholder consultation using primary data gathering methods.

### **3.4. Increase data re-use**

#### *Provided documentation for data validation and re-use facilitation*

Data published in the HELCOM Metadata catalogue complies with standards described in section 2 of this DMP.

Responsibility of the initial data collected in data calls during PROTECT BALTIC is on the data provider. Acquired data is then checked for consistency and cleaned in the quality control process of HELCOM Secretariat. Possible issues in data are communicated with the data providers.

In the process of data harmonization, the used terminology is checked and unified where possible. In cases where the data cannot be harmonized due to too different data collection methods, used units, or calculation methods, data is published on a country-by-country basis. Information of used variables and quality control measures can be found in the metadata.

Products published in PROTECT BALTIC will be openly available from project-related websites and through data- and metadata sharing instances such as EmodNET and WDPA.

For the software code published in GitHub, there is information about the software in the readme files and in respective parts of the software code. Software developers are encouraged to comment the code consistently, provide links to original code when copied from elsewhere and write instructions whenever deemed useful.

#### *Free availability of the data*

Published products in PROTECT BALTIC are following the creative commons license CC-BY, meaning that data is openly reusable, but HELCOM retains the legal right to be acknowledged as the originator of the data. All published data are available primarily from the project's website.

#### *Data usability for third parties after the project*

Most of the produced data is available for common use. Published data are commonly available after the project from the project's website and from HELCOM services such as MADS and the Metadata catalogue. There are also other instances who are distributing data which is collected or published in PROTECT BALTIC. Since project data is mainly openly reusable, not all publishers are known.

### *Data provenance*

All published data with metadata follow the [INSPIRE standard](#) for metadata on geographical datasets. The list of required fields includes lineage.

### *Relevant quality assurance processes*

Data collected in data calls is initially checked with formulas in data collection spreadsheet templates. Contracting parties are encouraged to use the templates to make sure the same naming conventions and classifications are used between countries. Since there are many data which have not been collected before, it is expected that there are differences e.g. between used classifications. For this reason, the data is further analysed manually at HELCOM and harmonized where applicable.

Before entering the data into databases, it is programmatically tested against the restrictions of the database.

## **4. Other research outputs**

Other data outputs of PROTECT BALTIC consist of code, methodologies, algorithms, assessment results, analyses, software, models, methodologies, frameworks, and approaches. All mentioned material can be considered to fulfil the FAIR principles of being findable, accessible, interoperable, and reusable. All material will be made publicly available. Published manuals and guidelines will be prepared and accompany all tools developed in the project to ensure usability and enable further development by end-users.

Software such as the augmented reality application, which allows any user to access the information on any of the Baltic Sea MPAs and the public MPA Portal are all made under the creative commons license CC-BY. The code will be made available from HELCOM's GitHub as described in the section [Making data accessible](#).

All the dynamic results (tools, modelling code, assessment methodologies, frameworks, and approaches) will be developed to be fully scalable and supported with step-by-step manuals, safeguarding their replicability.

## **5. Allocation of resources**

### **5.1. Cost for making the data FAIR**

The costs for making data or other research outputs FAIR in the project consist of data architecture such as physical hard disk space, processor load payments, internet connections and firewalls. In addition, expenses consist of user fees and of licences for commercial software used for GIS-services, virus protection, and data processing.

### **5.2. How the mentioned costs are covered**

In the project budget there are funds reserved for monthly payments of cloud services, and salaries for the data manager and for both front-end and back-end developers. The data produced by the project are stored in HELCOM's existing data infrastructure, so the long-term costs and maintenance are covered by HELCOM.

### **5.3. Responsibilities of data management**

Project personnel are personally responsible for the appropriate and secure management of data on their own behalf. The lead of WP2 is the data manager who is responsible for developing this data management plan and keeping it updated through the project.

## **5.4. Ensuring the long-term preservation of the data**

Data is stored in HELCOM's databases which compile data from multiple sources and projects. Databases are not PROTECT BALTIC-specific which guarantees the preservation and longevity of the data collected in the project. If new database solutions are later used, existing data will be migrated to the new ones. Therefore, there is no defined end date for data storage.

Data published on the PROTECT BALTIC website, will also be stored and available on the HELCOM website which enables the project products to be found also after the project.

Project documentation and work files are stored offline for five years after the project concludes.

## **6. Data security**

### **6.1. Provisions for data storage and transfer safety**

PROTECT BALTIC's data management uses a cloud based, highly secured Microsoft 365 productivity platform, which is provided and maintained by HELCOM. Final products and draft documents are stored in the platform which allows multiple users to edit the same files simultaneously. There is active virus protection and SPAM-filtering for the emails in the platform where all services are behind a firewall.

Access for the services is restricted. Users log in with personal usernames and passwords. The access is controlled with email addresses. Originally, each user has been invited individually by email.

Being compliant with the GDPR, contact information within the project is shared between users with signed consent from all members. Access to contact information is restricted for project personnel and it is not transferred to other entities.

The infrastructure of the used cloud services meets a broad set of international and industry-specific compliance standards, such as ISO, HIPAA, FedRAMP, and SOC, as well as country-specific standards. The service includes the optimized backups of the data.

### **6.2. Safety of long-term preservation of data**

As mentioned in the section on [Allocation of resources](#), PROTECT BALTIC data is stored in databases that are independent of the project. The data will be stored, managed, and utilized in the databases also after the project.

## **7. Ethics**

### **7.1. Ethical and legal issues on data sharing**

In PROTECT BALTIC data calls, there are datasets requested which might contain classified information such as bathymetry or accurate geospatial information of endangered species. Data providers are instructed to present the possible legal constraints and limitations for public access of each dataset in the metadata. As an example of limited data use, restricted data can be used in models, but the resolution of published data will be lowered according to limitations.

Data collected in PROTECT BALTIC data calls are biotic and abiotic measurements, maps, or models. Most of the data generated, utilized, and published within the project are not considered as ethically questionable due to their nature being scientific measurements, statistical data, or legislative information.

If there is personal data required e.g. for workshops, webinars, public surveys, interviews, demos, newsletters or recorded events, an informed consent form will be used to inform and obtain the

written consent of individuals outside of the consortia in compliance with the GDPR. The distribution of documents and e-newsletters takes place in a restricted environment.

Should ethical issues arise, there is an ethic mentor assigned in the project to ensure compliance with fundamental ethical principles present in the Nuremberg Code, European Textbook on Ethics in Research, and EC Ethics Appraisal Procedure as well as the GDPR.

Contact information of project personnel which is subject to GDPR is signed with the consent of each member and access to this information is restricted from others.

## 7.2. Consent for data sharing

When contact information is collected e.g. for the purpose of newsletters, public surveys, interviews, workshops, demos, recorded trainings and webinars, it is done with the consent of the respondents. When filling the form for registration to any of mentioned events, there is a link to the project's [Privacy Policy](#) page. By sending the registration, the registrant agrees to PROTECT BALTIC's data preservation regulations.

A restricted online list of contacts will be removed at the end of the project and stored offline five years afterwards.

## 8. Other issues

### 8.1. Use of national/funder/sectorial/departmental procedures for data management

Data handling and procedures are the responsibility of the partners until they are delivered to the project's data management environment.

Partners are universities, international non-governmental organizations, government agencies or private consulting companies, which generally have restricted intranets for data management. Data deliveries are normally done with emails or ftp-connections.

HISTORY OF CHANGES		
VERSION	PUBLICATION DATE	CHANGE
1.0	14.02.2024	Initial Data Management Plan.



**Funded by  
the European Union**

The PROTECT BALTIC project is funded by the European Union under Grant agreement ID 101112866. This publication was funded by the European Union. Its contents are the sole responsibility of the author(s) and do not necessarily reflect the views of the European Union.