Deliverable 2.1 Overview of the available biodiversity datasets and platforms relevant for planning

Overview of the available biodiversity datasets and platforms relevant for planning

Grant Agreement number	101060707
Project title	MSP4BIO: Improved science-based maritime spatial planning to safeguard and restore biodiversity in a coherent European MPA network
Deliverable title	Overview of the available biodiversity datasets and platforms relevant for planning
Deliverable number	D2.1
Deliverable version	1.0
Contractual date of delivery	31/07/2023
Actual date of delivery	03/07/2023
Document status	Final
Document version	1.0
Online access	No
Diffusion	Public
Nature of deliverable	Report
Work Package	2
Partner responsible	VLIZ
Contributing Partners	UTARTU, HELCOM, CCMS, UAC, NIMRD, SYKE, CNR, SEASC
Authors	Lawrence Whatley, Lennert Schepers, Jonne Kotta, Robert Szava-Kovats, Francisco R. Barboza
Contributors	Inne Withouck, Isabelle Rombouts, Fien De Raedemaecker, Débora Gutierrez, Helena Calado, Ana C. Costa, Camila Pegorelli, Javier Garcia Sanábria, Javier Garcia Onetti, Maria de Andres, Margarita Stancheva, Hristo Stanchev, Alina Spinu, Claire Boudy, Neil Alloncle, Marcello Magaldi, Roberta Sciascia, Andrea Barbanti, Mauro Randone, Kemal Pınarbaşı, Nerijus Blazauskas, Ivana Lukic, Ivana Stojanovic
Reviewers	Hristo Stanchev, Kemal Pınarbaşı
Approved by	Ivana Stojanovic

Project Officer	Victoria Beaz Hidalgo
Abstract	This deliverable presents an analysis of the availability of datasets, data platforms, tools, and models required by MSP4BIO. Data were compiled and the availability for each test site and data requirement (desideratum) was analysed. Data availability varied considerably between test sites and desiderata. The analysis found several data gaps, which serve as guidance for data collection by other work packages.
Keywords	Data availability, data gaps, biodiversity, MSP, MPA, data platforms, test sites
Suggested Citation	

Table of contents

List of Figures

List of Tables

Acronyms

BPNS	Belgian Part of the North Sea
DST	Decision Support Tool
EBSA	Ecologically or Biologically Significant Marine Area
ESE	Ecological-Socio-Economic
FAIR	Findable, Accessible, Interoperable, and Reusable (data)
FTP	File Transfer Protocol
GA	General Assembly
MPA	Marine Protected Area
MSP	Maritime Spatial Planning
MSP4BIO	Improved Science-Based Maritime Spatial Planning to Safeguard and Restore Biodiversity in a coherent European MPA network
OECM	Other Effective Area-based Conservation Measures
UNESCO	United Nations Educational, Scientific and Cultural Organization
WoRMS	World Register of Marine Species
WP	Work Package

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

1. Executive Summary

This deliverable presents a quantitative and qualitative analysis of the availability of data required by the MSP4BIO project. The overall aim of the MSP4BIO project is to support the implementation of the EU (European Union) Biodiversity Strategy (EUBS) 2030, the Convention on Biological Diversity (CBD) post-2020 framework, as well as the EU Green Deal, by mainstreaming biodiversity into policy decisions on different governance levels, and by developing an integrated socio-ecological management of the marine ecosystems. A wide range of ecological, socio-economic, and climate-related data is required by the project's six test sites, mostly existing open-access spatial data. These data requirements (or desiderata) are described in detail in this deliverable, along with the methodology used to compile data and analyse availability.

In total, 339 datasets, data platforms, tools and models were compiled from a wide variety of sources. Data availability varied considerably between the test sites, being higher in the Belgian Part of the North Sea and Baltic Sea and lower in the Western Black Sea and the Azores. Significant data gaps were found for several desiderata, including regional climate change projections and larval behaviour data. The quality and accessibility of the data were also highly variable. D2.1 therefore provides MSP4BIO partners with an overview of the availability of the data that they require and indicates data gaps, setting the scene for data collection.

2. Introduction

The MSP4BIO project aims to develop and demonstrate the ways in which knowledge-based MSP can become a vehicle for the protection and recovery of ecosystems across European seas. By building on and integrating existing knowledge and results from multiple origins, including other relevant projects and initiatives, the project will develop an integrated and flexible socio-ecological framework for the management of coastal, offshore, and deep-sea ecosystems in times of accelerated changes. This framework will identify an improved set of biodiversity and climate-related prioritisation criteria for MPAs and EBSAs based on the best available scientific knowledge and will link this environmental knowledge with socio-economic considerations. New socio-economic criteria will also be developed. The project uses a participatory approach to co-develop ecosystem service trade-off scenarios to prioritise MPA designation and assess the suitability of spatial and strategic management measures from ecological and socio-economic perspectives.

MSP4BIO uses existing data as much as possible. The project takes advantage of existing marine data infrastructures and platforms such as Copernicus, the Group on Earth Observations (GEO - GEOSS and GEO BON), The European Marine Observation and Data network (EMODnet), the European node of the Ocean Biodiversity Information System (OBIS), and national/regional data platforms. In addition, required local and regional information will be identified and accessed with support from the test sites. The data will be mainly used by WP3 (mostly ecological and climate change-related data), and WP4 (mostly socio-economic data) for the development of the modules of the Ecological-Socio-Economic (ESE) framework. A large amount of data of different types and from different sources is therefore required. This deliverable will outline the relevant available data and highlight persisting data gaps.

2.1. Overview of test sites

The ESE management framework will be tested and fine-tuned in 6 test sites in 5 European Sea Basins (Fig. 1). These are: 1. Northwest Mediterranean (France/Italy), 2. Gulf of Cadiz (Spain), 3. Belgian part of the North Sea (BPNS), 4. Western Black Sea (Bulgaria/Romania), 5. Baltic Sea, and 6. The Azores (Portugal). Data requirements and availability vary between the test sites.
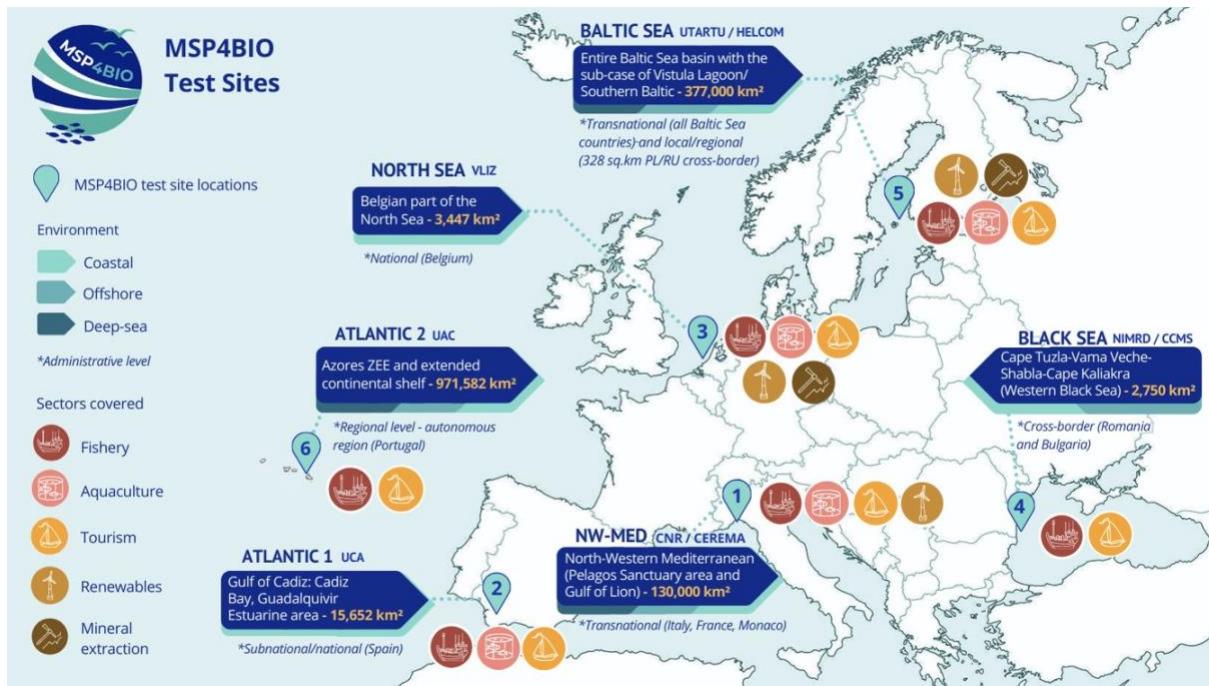
Figure 1: Locations and characteristics of the six test sites

2.2. Role of this deliverable

This deliverable provides an in-detail review of the availability of the data required to develop and test the ESE management framework and identifies persisting knowledge gaps. It provides information on data compiled and sources identified in the test sites and data platforms and infrastructures that provide relevant products for the development of the activities planned in MSP4BIO. The knowledge gaps identified during this process provide a roadmap for future data collection both within the MSP4BIO project and beyond.

3. Methodology

3.1. Key definitions

This section will define the key vocabulary used in this deliverable, namely data, tools, models, data sources, and metadata. The classification of the data into different types will also be outlined with examples.

3.1.1. Data

The foundation of scientific research is the gathering and analysis of recorded observations and measurements in the form of data. Data are becoming increasingly important because they are a primary intellectual output of research and their reuse is valuable for future research and subsequent studies. The International Organization for Standardization (ISO) defines data as a re-interpretable representation of information in a formalised manner suitable for communication, interpretation, or processing (ISO/IEC,

2015). The [Intergovernmental Oceanographic Commission (IOC)](#) of UNESCO further differentiates the definitions of data (observable, raw 'values', either numerical or nominal, that result from research or monitoring activities) and information (processed data and/or interpreted results) (IODE, 2006). MSP4BIO will mainly re-use existing data, rather than collecting new data. These data may be raw data (e.g., presented in the form of tables), or data products (e.g., map layers presenting the predictions of species distribution models).

3.1.3. Data sources

A data source is the data originator or provider offering data and information on their platform. It can be an international organisation, a government agency/department, a research institute, a private company, or a repository (e.g., EurOBIS, SeaDataNet, EMODnet, etc.). A good understanding of the available data sources in the marine domain is essential to review available data and gaps.

3.1.4. Metadata

Metadata is defined as 'data about data', or information that describes, explains, or locates an information resource. Metadata is essential to inform us of the origin, producer, restrictions, format, etc., of a given dataset or data product. The data's FAIRness (findability, accessibility, interoperability, and reusability) is considerably increased by good quality metadata, typically in a standard format such as ISO/TS 19139-1:2019 for geographic information datasets or INSPIRE Directive 2007/2/EC for spatial datasets, enabling interoperability, increasing the quality of data, and facilitating greater use of data.

3.2. Data requirements

WP2 received a list of guidelines for the collection of information (desiderata) from WP3 and WP4 in months 5 (December 2022) and 4 (November 2022) respectively. Since some of these desiderata were rather broad, an interactive session was carried out during the first General Assembly (month 8, March 2023), in which partners from WPs 3, 4, and 5 were asked to provide a more detailed description of the desiderata. The original, broad desiderata and the more specific desiderata are listed in Table 1 below. The numbering of the desiderata follows that of the original list, hence there is no desideratum 1 or 7.

Table 1: List of data-related desiderata from WPs 3 and 4

WP	Task	Broad desiderata	Specific desiderata
3	3.1	**2: Scientific knowledge and data feeding ecological and environmental criteria considered in area-based conservation and restoration measures**. The scientific knowledge, data and bibliographic resources underlying the general ecological and environmental criteria	Point data (e.g., occurrences of species) should be transferred to seamless species occurrence maps.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

		defined for the identification of EBSAs and the designation and implementation of MPAs and other area-based conservation and restoration measures in strategic framing directives and documents of national, regional and pan-European/international reach.	
3	3.2	**3: Regional and IPCC climate change projections**. Available climate change models or future projections at regional level: specifically, water temperature, oxygen, salinity, acidification, ice cover; if available, other potential climatic stressors. Also, information on time frame of the projections, resolution, uncertainty, models used, data availability, data sources.	Prediction of valuable habitats (e.g., seagrass, kelp) under different climate change scenarios.
			Sub-regional projections where possible/available.
			Possible links between highly protected areas and climate refugia.
3	3.1	**4: Data feeding new criteria for area-based conservation measures (defined in T3.1 and T3.2)**. E.g., other species relevant for conservation (keystone habitat forming, invasive, etc.) and ecological processes.	Marine ecological/migratory corridors.
			Species uniqueness and/or rarity.
3	3.3	**5: Spatially defined information on nature values/assets (existing nature value maps)**. Existing georeferenced layers for relevant nature values (e.g., occurrence, abundance and/or biomass distribution of relevant species, habitat-forming species, biodiversity indexes, species of commercial interest, top predators, spatial predictions of relevant ecosystem processes and associated services such as primary production, nutrient sequestration, carbon stocks and sequestration).	Localisation of areas that support ecological functionalities.
			Occurrence, abundance, and/or biomass distribution of habitat-forming species.
			Occurrence, abundance, and/or biomass distribution of species of commercial and conservation interest.
			Databases/data on functional and specific traits.
			Marine and coastal ecosystem services case studies.
			Health state of ecosystem/species engineers.
3	3.3	**6: Human uses/activities and expected developments/changes in case studies**. A list and description of human uses/activities	Aquaculture development.
			Windfarm developments in the North Sea.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

			currently developed or expected to be developed in the test sites (e.g., wind farm development plans, aquaculture, fishing), and management actions in place or planned to mitigate the consequences of these activities/uses (e.g., plans and targets for the reduction of nutrient loads and the effects of eutrophication).	Windfarm development in the NW Mediterranean.
				Land use change in the coastal zone.
				Multi-use areas - case studies.
				Coastal protection structures including beach nourishment.
				Shipping density, anchorage areas, pipelines, cables, geophysical survey
3	3.3	**8. Physical/dynamic data**.		Hydrodynamical information (i.e., marine current data) from different sources.
				Transport data (lagrangian trajectories) from drifters and floats.
				Larval behaviour data (settling velocities, pelagic larval duration, location depth, timing/frequency of spawning etc.).
				Connectivity data.
3	3.3	**9. Existing dispersion models**.		LTRANS
				ARIANE
				pyGNOME
				CMS
				PARCELS
				PARTRAC
				Ichthyop
4	4.4	**MSP databases**.		Portugal/The Azores' MSP
				Mediterranean MSP (France and Italy)
				Spain's MSP
				Baltic Sea MSP database
				Belgium's MSP
				Black Sea MSP database

3.3. Data compilation

The data compilation was conducted by WP2 (scoping and gap analysis). In T5.1, the test site leaders were asked to compile a list of relevant datasets, data platforms, tools,

and models covering their test site and to provide the respective metadata and links in a standardised table. The provided datasets/data platforms mainly covered national or regional geographical scales. The information was combined into a single metadata table. The metadata were recorded in different columns for each dataset or data platform. These columns are summarised in Table 2, below.

Table 2: Information included in the T2.1 metadata table

Column	Description
Pilot Site	Pilot site(s) covered by the dataset (text).
Pilot Site Coverage (6 columns)	Specification of test site(s) covered by the dataset. There are 6 columns, one per test site, to indicate coverage per test site.
Dataset name	Name of dataset.
Type	Broad-scale type of data: biotic / abiotic / anthropogenic impacts / social-economic-cultural / climate / spatial / other (Table 3).
Subtype/Variables	More specific description of data subtype and/or variables.
Ownership	Owner of the data / platform.
File format	File format of downloadable data.
Accessibility	Accessibility of data: open access or to be requested.
Access type	Means of accessing the data.
Scale	Geographical scale of data coverage: test site / sea basin / European seas / other.
Spatial coverage	A more specific description of the geographical coverage of the data.
Spatial resolution	Spatial resolution (for raster data only).
Temporal coverage	Temporal coverage of the data, with start date and end date where relevant.
Temporal resolution	Temporal resolution of the data recorded as start date and end date. "Once" indicates that the dataset covers one moment in time.
Time series available?	Availability of a time series.
Future scenario available?	Availability of future projections/predictions.
Object type	Whether the entry was a dataset/database, data platform/catalogue, model, or tool
Link/metadata page	Link to download the data and/or consult the metadata.
Data platform	Platform hosting the data, where relevant.
Notes	Additional information about the data, where relevant.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

Relevant WP3 Desiderata	Desiderata from WP3 and WP4 for which the dataset may be relevant (Table 1)
Des. 2-9, WP4 MSP (8 columns)	Relevance for each desideratum (Y/N) from WP3 and WP4 in 8 separate columns, used for filtering the table (Table 1)

Other potentially relevant datasets, mainly at the European or global scale from services such as EMODnet and Copernicus, were added to the metadata table by T2.1 partners. Two interactive sessions, one with all MSP4BIO partners at the GA, and one within the VLIZ data centre, were held to share ideas of datasets fulfilling WP3 and WP4's requirements (Table 1). Selected models and tools which were relevant to the project (e.g., dispersion models, fish stock assessment tools) were added, and some relevant datasets found on the GEOSS portal and the metadata catalogue produced by the Mission Atlantic project were also included. These datasets, databases, data platforms, models, and tools were recorded as separate entries in the metadata table. A metadata table entry could therefore be a record of an individual dataset, multiple datasets, or a derived data product.

After compiling the metadata table, it was screened for missing or incorrect metadata. Gaps in the metadata were filled by verifying the original data sources, and corrections were made if necessary. The table was also screened for duplicate entries. New datasets and platforms can and will be added to the metadata table throughout the duration of MSP4BIO, so it should be considered a living document. The metadata table will also be shared openly with sister projects so that they can benefit from it and provide feedback.

3.4. Data categorisation

The data were categorised by type, spatial and temporal coverage, resolution, accessibility, and other characteristics. The data type categories used were abiotic, biotic, anthropogenic impacts, social/economic/cultural, climate, spatial, or other (Table 3). The purpose of this was to match the system developed by MSP4BIO for the categorisation of existing ecological and environmental criteria used in the prioritisation of EBSAs and the design and management of MPAs and OECMs (T2.2). These data type categories are listed in Table 3 with examples, below.

Table 3: Categorisation of data types used in T2.1 based on T2.2's categorisation of criteria used in the description of EBSAs, MPAs, and OECMs

Data type	Definition	Examples of datasets
Biotic	Data related to living organisms.	Species occurrence, biogenic habitat maps, plankton abundance
Abiotic	Data related to non-living elements of the environment that influence the way organisms and ecosystems function.	Hydrodynamic data, bathymetry, meteorology, seabed characteristics, water chemistry

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

Anthropogenic Impacts	Data related to the presence of anthropogenic activities that might generate effects/pressures on biotic/ecological elements.	Marine litter, contaminants and pollution, fishing pressure
Climate	Data related to climate projections, climate impacts, climate change or climate mitigation.	Climate projections, the effects of climate change on other factors
Social, Economic, and Cultural	Data related to social, economic or cultural values, including ecosystem services with social, economic or cultural value.	Vessel traffic, leisure and tourism, aquaculture, fisheries economic data, underwater cultural heritage
Spatial	Purely spatial data related to the designation of areas such as MSP, MPAs, and marine regions, including location, spatial coverage, size, and connectivity.	MSP databases and zones, MPA databases, marine regions and boundaries
Other	Any data which does not fall into the categories above.	Data platforms with multiple data types, models

The data to be re-used in the project also cover different geographical scales (Table 2, row "**Scale**"). For the purposes of T2.1, these scales were categorised as:

1. Test site-specific* (e.g., the Belgian Part of the North Sea case study - BPNS)
2. Sea basin (e.g., North Sea)
3. European seas
4. Other (including oceanic scale and global datasets).

*The geographical scale of the data covering the Baltic Sea test site was categorised as "sea basin".

The **accessibility** of the data was classified as either open access (including data that could only be downloaded after registering on the platform for free), to be requested (including cases where access to the data must be purchased), or other (e.g., data platforms with both open access and restricted data). The temporal coverage was recorded as start date and end date (both of which could be in the past, present, or future) in the metadata table. For datasets covering one moment in time, the date was recorded in the start date column and "once" was indicated in the end date column. Each entry was categorised as a dataset or database, a data platform or catalogue, a model, or a tool. Finally, the data were also classified by the WP3/WP4 desiderata for which they may be relevant (Table 1).

Data entries were also classified according to the **access type** in the following five categories (multiple categories possible):

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

1. Web map service: online services that display and provide georeferenced information. For reasons of simplicity, feature map services were also included under this category.
2. Downloadable using web-based API: data that can be accessed and downloaded using web-based application program interfaces (API).
3. Downloadable from FTP: data that can be retrieved from a file transfer protocol (FTP) server based on clients' specific requests.
4. Downloadable from a data repository (e.g., Zenodo): data that can be downloaded from specific open-access data repositories that focus on the sharing of scientific data and research outputs.
5. To be requested: data that cannot be openly accessed and require a formal request to a research or governmental institution to be accessed. A dataset was also assigned to this class when the website was not available or operational.

These categories were defined based on the most common strategies used for providing access to data generated and stored by research institutions, governments and European programmes. To classify the data according to the access type, the websites provided by the test site leaders and other contributing partners were visited and revised in detail. The classification was performed considering both the stability of the websites (i.e., if the website was available and operational during the revision of the provided metadata and access type classification) and the information provided in the websites. If available for a given dataset, multiple access types were indicated.

3.5. Data gap screening

The numbers of entries per data type category, test site, accessibility category, desideratum, and sea basin were extracted and compared by using the filter feature in the metadata table. A desideratum for which potentially relevant data was lacking at a particular test site was defined as a data gap. Since the metadata table contained both individual datasets and data platforms hosting multiple datasets, the number of entries per desideratum and/or test site was not considered to accurately reflect data availability. Therefore, a qualitative approach to data gap screening was also taken, in which the quantity, quality, and accessibility of the available data were considered in addition to the stability of the source platform and the coverage of the test site. The results of the quantitative and qualitative analysis are presented in sections 4 and 5, and supplementary tables are provided in the Annexes (section 8). Sections 4 and 5 present the results of the analysis carried out on 31/05/2023. Since more entries will be added to the metadata table over the project's lifetime, some of the results are likely to change, although we expect overall changes to be minimum.

4. Data sources

Data were derived from a wide variety of sources, 149 in total, with more than half of the metadata table entries (178) coming from a source with ≤3 entries (Figure 2, Annex 1). The data platform with the greatest number of entries was EMODnet (23 entries), followed

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

by Copernicus Marine (20 entries) and the French government platform Directive Cadre Stratégie pour le Milieu Marin (18 entries).
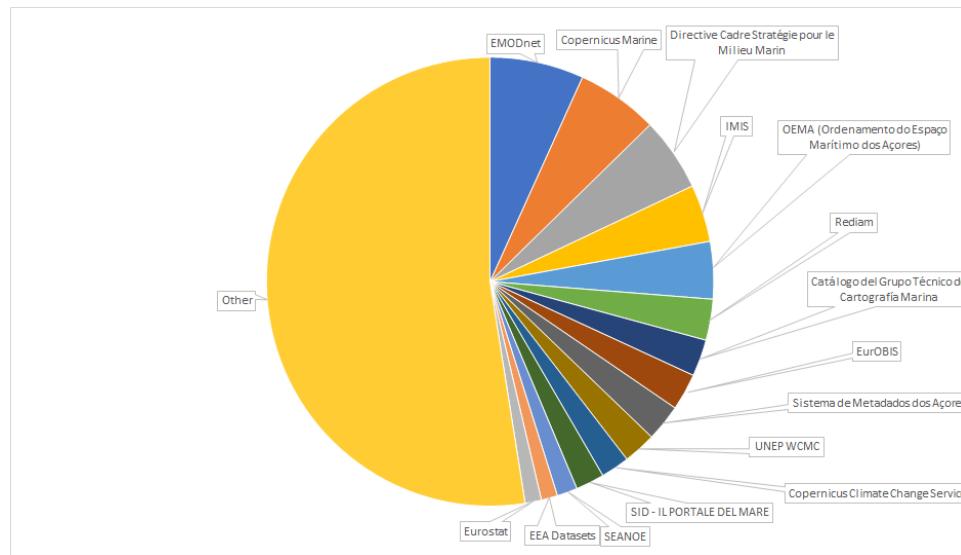
Figure 2: Number of metadata table entries by data source. Sources with ≤3 entries are included in "Other".

5. Data availability, accessibility, and gaps

159 entries were compiled from test site leaders and 180 were subsequently added, resulting in a total of 339 entries in the metadata table. 296 of these entries were datasets and databases, 31 were data platforms or catalogues giving access to multiple datasets, 5 were tools, and 7 were models. Despite the relatively small number of data platforms and catalogues, they are of great importance to the overall data accessibility because they can give access to hundreds of datasets. Furthermore, since the datasets hosted on such platforms and catalogues are often harmonised and pass through the same quality control procedure, they may be of greater interest to MSP4BIO partners. Therefore, a simple comparison of the number of entries in the metadata table may not capture the true availability of the data.

By data type, there were:

- 115 (33.9%) biotic entries
- 52 (15.3%) abiotic entries
- 34 (10.0%) anthropogenic impacts entries
- 36 (10.6%) social/economic/cultural entries
- 14 (4.1%) climate entries
- 65 (19.2%) spatial entries
- 23 (6.8%) other entries

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

These results confirm that there is a lack of data related to socio-economics, a finding echoed by the European Commission's MSP data study (European Commission, 2017), and climate change. In all the data type categories except "other", the entries were mainly datasets or databases (Figure 3, Annex 2). The "other" category contained models, data platforms and catalogues with multiple data types, and datasets which did not fall into any of the other categories.
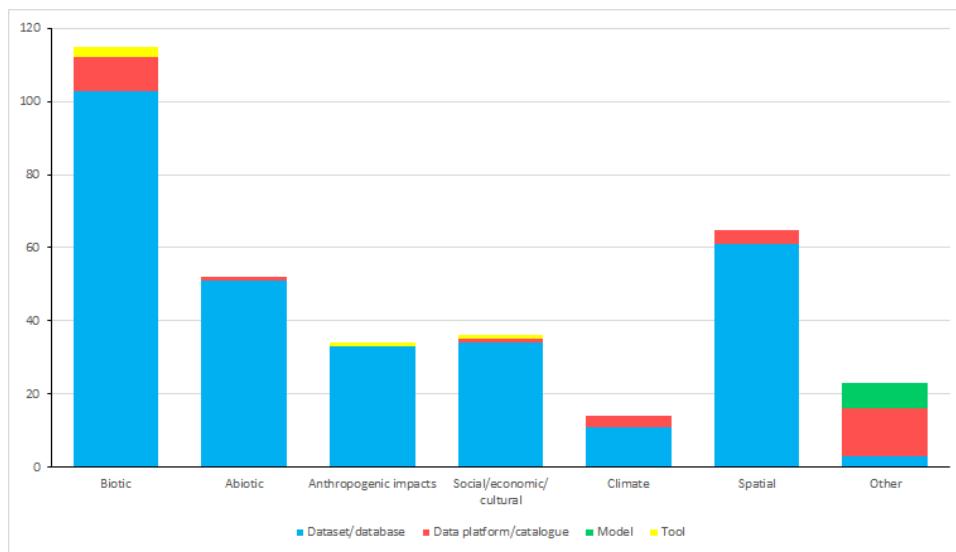
Figure 3: Breakdown of metadata table entries by data type and object type

In total, 283 (83.5%) of the entries were initially classified as open access, while 53 (15.6%) were classified as "to be requested". 3 entries (0.9%) contained both open access and restricted or unpublished data.

Regarding accessibility, most of the data entries were openly accessible, the provided websites were available and operational, and they offered one or multiple strategies for downloading the indicated data. However, out of the 283 entries initially classified as "open access", 32 were not accessible during the final revision, mainly due to problems in the websites (which were down or under maintenance). For precautionary reasons, these 32 data entries were classified as "to be requested" in the access type as an indication of potential problems associated with the stability of the websites, although access to these 32 entries may still be open in the future. Most of the data entries that provided information at the European scale were openly available; less than 3% needed to be formally requested from the institutions that produced the data. When looking at data entries covering single sea basins, the percentage of data entries that needed to be requested was 19%, being most of them related to the Mediterranean Sea and its transition to the Atlantic Ocean, and the Black Sea. At the test site level, the percentage of data entries that need to be requested was almost 29%, of which almost 90% were associated with four case study sites (Northwest Mediterranean, Gulf of Cadiz, Western

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

Black Sea, and the Azores). Problems related to the availability of the websites during the revision and restrictions imposed by local government and research institutions explained these values.

5.1. Data availability by test site

132 (38.9%) of the entries covered multiple test sites, while the remaining 207 (61.1%) covered just one test site. Per test site, the following entries in the metadata table were available (see also Annex 3):

- 167 entries for the Northwest Mediterranean
- 159 entries for the Gulf of Cadiz
- 148 entries for the BPNS
- 152 entries for the Western Black Sea
- 120 entries for the Baltic Sea
- 155 entries for the Azores

However, these numbers do not capture the true availability of the data at each test site because they include both individual datasets and data platforms or catalogues hosting multiple datasets. Furthermore, they do not discriminate between open access and restricted data. Some particularly significant data platforms and catalogues included in the metadata table (i.e., those hosting many datasets and covering multiple test sites and/or desiderata) are outlined in Table 4, below.

Table 4: Selected data platforms and catalogues included in the metadata table

Data platform/ catalogue	Test sites covered	Data hosted	Number of datasets hosted	Link
Bio-ORACLE	All test sites	Geophysical, biotic, climate and environmental data for surface and benthic marine realms in the present or future under different RCPs	18 predictors and three time periods. Version 3, expected in late 2023, will include data for 18 variables and 6 SSPs.	https://bio-oracle.org/downloads-to-email.php
Global Biodiversity Information Facility (GBIF)	All test sites	Species occurrence records from a variety of sources	85093 (including non-marine species)	https://www.gbif.org/

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

OBIS / OBIS Mapper	All test sites	Species occurrence records from a variety of sources	4783	https://obis.org/, https://mapper.obis.org/
One Shared Ocean	All test sites	Climate and future impacts, ecosystem health, fish and fisheries, governance, pollution, productivity, socio-economics	136	http://onesharedocean.org/data
OSPAR Data and Information Management System	Gulf of Cadiz, BPNS, Azores	Biological diversity and ecosystems, environmental impacts of human activity, hazardous substances, eutrophication, offshore industry, radioactive substances	813, including 692 spatial datasets	https://odims.ospar.org/en/
ICES Datasets	Gulf of Cadiz, BPNS, Baltic Sea, Azores	Biological communities, catch statistics, contaminants and biological effects, fish trawl surveys, fish eggs and larvae, fish stomach, plankton, Vulnerable Marine Ecosystems, ocean hydrochemistry, ocean climate	21	https://www.ices.dk/data/dataset-collections/Pages/default.aspx
Mediterranean Platform on Biodiversity	Northwest Mediterranean	Biodiversity, physical-chemical features, MPAs and OECMs	251, some of which do not cover the test site	http://data.medchm.net/en/catalogue
Marine cartography technical group catalogue	Gulf of Cadiz	MPAs, environmental monitoring facilities, bathymetry, geology, habitats, oceanographic geographical features, industrial facilities,	4942	http://www.infomar.miteco.es:8080/geonetwork/srv/eng/catalog

		species distribution, sea regions		.search#/home
Black Sea Database	Western Black Sea	Biodiversity, statistics, eutrophication, contaminants, hydrography, litter, energy (noise)	Not specified	http://blackseadb.org/
HELCOM Map and Data Service	Baltic Sea	Indicators and assessments, monitoring, human activities, pressures, red listed species and habitats, biodiversity, shipping, background information	991	https://maps.helcom.fi/website/mapservice/
BASEMAPS	Baltic Sea	MSP input data (Administrative borders, aquaculture, fishing areas, installations and infrastructures, maritime transport, nature protection, military training, raw material extraction, scientific research, cables and pipelines, tourism and recreation, underwater cultural heritage) and output data (planned areas)	Not specified	https://basemaps.helcom.fi/
OEMA GeoPortal do Mar	Azores	Geology, oceanography, biodiversity, marine zoning, MSP, MPAs, infrastructure, cultural heritage, human uses and activities	109	https://sigmar.dram.azores.gov.pt/#/viewer/openlayers/geoportal

Since the two Baltic Sea data platforms listed in Table 4 (HELCOM map & data service, BASEMAPS) host many good-quality datasets, data availability at this test site is greater than it may appear from the number of entries. In contrast, data availability may be lower

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

in the Northwest Mediterranean and Western Black Sea than the numbers of metadata table entries suggest because separate datasets from each country (France/Italy and Bulgaria/Romania respectively) were often recorded in the table. In single-country test sites such as the BPNS and the Azores, such data were more likely to be recorded in one dataset. Figure 4 gives an overview of the metadata table entries by test site.
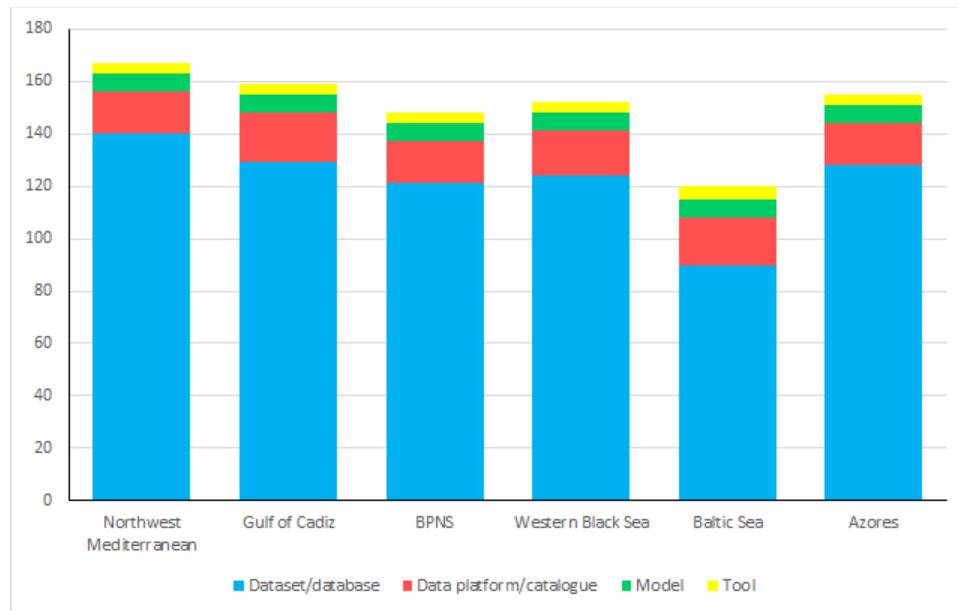
Figure 4: Breakdown of metadata table entries by test site coverage and object type.

Table 5 gives an overview of data availability by test site and desideratum derived from a qualitative analysis of the listed datasets and platforms. Data availability was defined as:

- High (green, Table 5): a sufficient quality and quantity of data to fulfil all or most of the specific desiderata
- Medium (yellow, Table 5): a sufficient quality and quantity of data to fulfil some of the specific desiderata
- Low (red, Table 5): data quality and quantity are insufficient to fulfil most or all of the specific desiderata

Table 5: Qualitative analysis of data availability by test site and desideratum. Green, yellow, and red indicate high, medium, and low data availability respectively.

Test site	Desideratum							
	2	**3**	**4**	**5**	**6**	**8**	**9**	**WP4**
	MPA / OECM criteria	Climate change projections	New MPA / OECM criteria	Nature values - spatial data	Human uses / activities	Physical / dynamic data	Dispersion models	MSP databases

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

Northwest Med.	Global and European data available	Projections available at global but not regional scale	Global and European data available, data availability limited in Italy	Global, European, and test site-specific data available	Data availability is high in France but limited in Italy	Data available from Copernicus Marine Service. Transport data available.	All requested dispersion models are available	MSP data available from France but not Italy
Gulf of Cadiz	Global and European data available	Projections available at global but not regional scale	Global and European data available, test site-specific data limited	Global and European data available, test site-specific data limited	European data available but test site-specific data is limited	Data available from Copernicus Marine Service	All requested dispersion models are available	Available
BPNS	Global and European data available	Projections available at global but not regional scale	Global, European, and test site-specific data available	Global, European, and test site-specific data available	European data available but test site-specific data is limited	Data available from Copernicus Marine Service	All requested dispersion models are available	Available
Western Black Sea	Global and European data available	Projections available at global but not regional scale, poor coverage of Copernicus Climate Change products	Global and European data available, test site-specific data limited	Global, European, and test site-specific data available	Data availability is high but differs between Romania and Bulgaria	Data available from Copernicus Marine Service	All requested dispersion models are available	MSP data available from Bulgaria but not Romania
Baltic Sea	Global and European data available	Projections available at global scale, few projections available at regional scale	Global, European, and test site-specific data available	Global, European, and test site-specific data available	Data widely available from HELCOM data platforms	Data available from Copernicus Marine Service	All requested dispersion models are available	Available
Azores	Global and European data available	Projections available at global but not regional scale, poor coverage of Copernicus Climate Change products	Global and European data available, test site-specific data limited	Global and European data available, test site-specific data limited	Data widely available from OEMA geoportal	Data available from Copernicus Marine Service	All requested dispersion models are available	Available

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

5.1.1. Northwest Mediterranean

Data availability in the Northwest Mediterranean varies by desideratum and within the test site; data availability was generally higher in the French part of the test site than the Italian part. For example, the French MSP spatial data are openly available, but MSP data are unavailable in Italy because the country has not yet completed its MSP. Data availability for desiderata 2, 4, 5, and 6 was greater in France than in Italy due partly to the data hosted by the French DCSMM data catalogue, although not all was open access. Access to a similar data platform from Italy (SID – Il Portale del Mare) was restricted, which limited the data availability in the Italian part of the test site. Some potentially relevant Mediterranean-wide datasets and platforms were included, such as SPA/RAC's Mediterranean Platform on Biodiversity and the GFCM's fisheries statistics in addition to Mediterranean datasets from Copernicus and EMODnet.

5.1.2. Gulf of Cadiz

Data availability in the Gulf of Cadiz varies greatly, but in general is more limited than in some of the other test sites. Many high-quality datasets are hosted on the Spanish Marine Cartography Technical Group's Catalogue and Andalucia's Environmental Information Catalogue, although only 210 and 20 respectively cover the Gulf of Cadiz (4.0% and 0.3% respectively of datasets hosted on these platforms). Some high-resolution datasets from scientific articles were included (e.g., seagrass and macroalgae cover in the Bay of Cadiz) which must be requested from the authors. In addition to European-scale datasets from platforms such as EMODnet, this site is covered by some Atlantic-wide datasets, e.g., from OSPAR and ICES. Data availability for desiderata 4, 5, and 6 was somewhat limited in this test site due to a lack of test site-specific data relevant for certain specific desiderata.

5.1.3. Belgian Part of the North Sea

Since the BPNS is a highly studied region, data availability in this test site is generally high. Many of the open access datasets covering the BPNS are available on platforms such as EMODnet, EurOBIS, and IMIS. Along with the Gulf of Cadiz and the Azores, some Atlantic-wide datasets cover this test site such as those available from OSPAR and ICES, in addition to many datasets covering the (wider) North Sea, including data on contaminants, plankton, and cetaceans. There were many datasets specific to the BPNS, mostly open access biotic datasets; restricted access was only found for a small number of datasets (19%). A large amount of data on plankton and benthic fauna is available in the BPNS, unlike the other test sites. Data availability for desideratum 6 (human uses and activities) was somewhat limited, namely for aquaculture, land use change, and multi-use areas, although Belgium's MSP spatial data indicates zones designated for certain activities.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

5.1.4. Western Black Sea

Data availability is often limited in the Western Black Sea and differs between Bulgaria and Romania. The cross-border nature of this test site means that data which is available in one country might be unavailable in the other, similar to the Northwest Mediterranean. Data availability is generally higher in Bulgaria due to the inclusion of datasets such as those found on its MSP platform (Romania's MSP has not yet been completed). In both countries, however, access to some datasets is restricted. Data availability in Romania is likely to improve slightly when data from the SIMSHAB species and habitat monitoring project is made available, probably in late 2023. There were few test-site-specific general data platforms for marine data such as those found in some of the other test sites; the Black Sea Database and MARSPLAN-BS Geoportal offer some useful datasets, although data on the latter must be requested. The availability of data for desiderata 3 (regional and IPCC climate projections) and 4 (data feeding new criteria for area-based conservation measures) was low but data availability was higher for desideratum 6 (human uses and activities).

5.1.5. Baltic Sea

The availability of data in the Baltic Sea test site is high, mainly thanks to the large number of high-quality open access datasets available on HELCOM's data platforms (Map and Data Service, BASEMAPS, the Biodiversity Database, and the MPA Database). Many of these datasets cover the EEZs of multiple countries unlike some datasets in the other cross-border test sites (Northwest Mediterranean and Western Black Sea). These datasets are particularly relevant for desiderata 2, 4, 5, and 6, and are supplemented by wider scale datasets from platforms such as EMODnet. One Baltic-specific dataset was found for desideratum 3 (regional and IPCC climate change projections).

5.1.6. Azores

Data availability in the Azores test site was variable, and some data gaps were found. A wide variety of relevant data can be found on the OEMA GeoPortal do Mar platform, although the data here can only be downloaded in tabular format and lack metadata. Data availability in the Azores is supplemented by a number of datasets from organisations covering the Atlantic, such as OSPAR and ICES. Data availability was low for desideratum 3 (regional and IPCC climate change projections), since many Copernicus Climate Change Service datasets do not cover this test site. Data availability was also somewhat limited for desiderata 4 (data feeding new criteria for area-based conservation measures) and 5 (spatially defined information on nature values/assets) but is high for desideratum 6 (human uses and activities). Datasets hosted by the Azores' Metadata System were originally found to be open access, but access was subsequently lost and the platform currently appears to be down.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

5.2. Data availability by desideratum

A total of 127 entries (37.5%) were matched to multiple desiderata, 158 (46.4%) to one single desideratum, and 54 (15.9%) to no desiderata. Table 6 shows the total number of entries matched to each desideratum and the number covering each test site per desideratum. A low number of entries per desideratum and/or test site does not necessarily indicate a data gap. One good-quality dataset or database may be sufficient for the purposes of the project; HELCOM's BASEMAPS database for example is likely to fulfil WP4's desideratum of MSP databases for the Baltic Sea test site.

Table 6: Overview of data availability (number of metadata table entries) by test site and desideratum. Red indicates a low number of table entries and green indicates a high number. Desiderata are numbered according to Table 1.

Desiderata		Test Sites						TOTAL
		NW Med.	Gulf Cadiz	BPNS	W Black Sea	Baltic Sea	Azores	
	2. MPA / OECM criteria	38	41	38	37	36	34	62
	3. Climate change projections	17	16	16	10	13	11	26
	4. New MPA/OECM criteria	50	48	48	38	35	39	87
	5. Nature values – spatial data	70	66	67	54	55	60	111
	6. Human uses / activities	40	41	33	43	31	50	115
	8. Physical / dynamic data	18	20	15	17	13	15	37
	9. Dispersion models	7	7	7	7	7	7	7
	WP4 MSP databases	1	1	1	2	1	11	17
TOTAL		167	159	148	152	120	155	339

5.2.1. Data availability for desideratum 2 (Scientific knowledge and data feeding ecological and environmental criteria considered in area-based conservation and restoration measures)

62 entries were identified as being potentially relevant for this desideratum. Spatial data on the extent of MPAs and OECMs and data feeding the ecological and environmental criteria (e.g., the presence of species or habitats warranting a site's designation) were generally available. These data are normally on a global or European scale covering multiple or all test sites, but are sometimes only available in a non-interoperable format.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

For instance, data on Ramsar sites and OSPAR MPAs are only available as PDF factsheets. See below for a list of relevant datasets:

- Natura 2000 sites: descriptive and spatial data, monitoring data on the status of habitats and bird populations.
- European Red List of habitats.
- ICES' Vulnerable Marine Ecosystems (VMEs).
- Ecologically or Biologically Significant Marine Areas (EBSAs).
- Ramsar sites.
- Important Bird Areas (IBAs).
- Important Marine Mammal Areas (IMMAs).
- IUCN Red List spatial data.
- OSPAR MPAs.
- Key Biodiversity Areas (KBAs).
- EMODnet essential ocean variables: hard coral cover, macroalgal canopy cover, seagrass cover, coralligenous and other calcareous bio-concretions (Mediterranean).
- HELCOM MPAs.
- Nationally designated areas (CDDA)

Species occurrences were widely available from platforms such as OBIS, GBIF, and HELCOM's biodiversity database. However, these were more often point data than seamless species occurrence maps (a specific desideratum, Table 1). Due to the global or European scale of most of the relevant datasets, data availability is similar between the test sites.

5.2.2. Data availability for desideratum 3 (Regional and IPCC climate change projections)

A total of 26 entries were identified as being potentially relevant for this desideratum. Climate change projections are available on a global scale from platforms such as Bio-ORACLE, Copernicus Climate Change Service, the ESGF portal, and One Shared Ocean (sometimes unavailable), but few regional and sub-regional projections (one of the specific desiderata) are available. One exception is a model simulation dataset for the Baltic Sea. Regional data is available from the IPCC, although the IPCC's regions may be too large to be relevant for the project. Copernicus also provides data products concerning the effect of climate change on variables such as eutrophication, fish abundance, ocean fronts, marine biogeochemistry, surface waves, and offshore wind farm performance. These data products do not cover all test sites, however; data availability is generally lower in the Western Black Sea and the Azores and higher in the BPNS. Data availability for predictions of valuable habitats under climate change appears to be low, although GlobTherm may be relevant for habitat-forming species. No data were found on the links between highly protected areas and climate refugia.

5.2.3. Data availability for desideratum 4 (Data feeding new criteria for area-based conservation measures)

For this desideratum, 87 entries were identified as being potentially relevant. In general, data regarding large, charismatic organisms such as mammals and birds were more widely available than data related to smaller-bodied organisms such as invertebrates and microorganisms. Data availability for this desideratum is higher in the French part of the Northwest Mediterranean site, the BPNS, and the Baltic Sea, and lower in the Italian part of the Northwest Mediterranean, Gulf of Cadiz, Western Black Sea, and the Azores. However, data availability will also depend on the new criteria to be developed in T3.1 and T3.2.

Habitat maps were available from platforms such as EMODnet seabed habitats and OSPAR habitats, and occurrence data of habitat-forming species such as seagrasses, corals, and macroalgae was widely available from platforms such as EMODnet Biology and the UN-WCMC's Ocean Data Viewer. General species occurrence data can be found on platforms such as OBIS and GBIF, and ranges of rare and threatened species are available from the IUCN (this data platform is also relevant for the specific desideratum "Species uniqueness and/or rarity", Table 1). The geographical coverage of many of these datasets is global or European, but some smaller scale datasets were found such as the *hábitats de interés comunitario* series in the Gulf of Cadiz and the Belgian marine mammal strandings database in the BPNS. For the specific desideratum "marine ecological/migratory corridors" (Table 1), the MiCO system covers the migratory habitats of large vertebrates and the European Tracking Network data may also be of interest.

5.2.4. Data availability for desideratum 5 (Spatially defined information on nature values/assets)

111 entries were identified as being potentially relevant for this desideratum. Data availability was relatively similar between test sites due to the number of global (e.g., UN-WCMC) and European (e.g., EMODnet) scale datasets but was generally higher in the Northwest Mediterranean and Baltic Sea and lower in the Gulf of Cadiz and the Azores. There was some overlap with desideratum 4, namely habitat maps (EMODnet, EUSeaMap21, OSPAR), occurrence data of habitat-forming species (seagrasses, corals, macroalgae, multiple species), and occurrence data of species of commercial (e.g., ICES DATRAS, FishStatJ) and conservation interest (e.g., IUCN, Natura 2000 reporting for the Birds and Habitats Directives).

A limited amount of data on areas supporting ecological functionalities (a specific desideratum, Table 1) was found, although there are datasets on biomass production of low-mid trophic levels and fish, and the spawning and nursery grounds of selected commercially important fish species in the French part of the Northwest Mediterranean and the Baltic Sea. Several databases on functional traits are available. Marine Species Traits is a general database which can be visualised using the Lifewatch Data Explorer, and there are also specific databases for bacteria and archaea, copepods, marine invertebrates, elasmobranchs, fish, birds, and mammals. The availability of ecosystem

services case study data was low, apart from a map database in the BPNS. Examples of spatial data on multiple ecosystem services can be found in the Baltic Sea on HELCOM's Map and Data Service. For data on ecosystem engineers, the previously mentioned datasets concerning habitat-forming species and species of conservation interest may be relevant, along with a bioturbation classification database.

5.2.5. Data availability for desideratum 6 (Human uses/activities and expected developments/changes in case studies)

115 entries were identified as being potentially relevant for this desideratum. Most of these entries covered a single country, although some were larger in scale such as the datasets from EMODnet Human Activities. Data availability therefore varied between test sites and depended on the relative importance of different human activities at each site. In the Northwest Mediterranean, data from France was widely available and open access but Italian data were often restricted. For the Gulf of Cadiz, the Marine Cartography Technical Group and Environmental Information Network of Andalucia have catalogues with some relevant datasets, but data availability here is more limited. Apart from Belgium's MSP data there are few specific datasets relevant to desideratum 6 for the BPNS, although there is good coverage of data from EMODnet and OSPAR. Data availability is relatively high in the Western Black Sea, although more data was available in Romania than Bulgaria. The MARSPLAN-BS II webapp offers cross-border spatial data to view but not download. Data availability for the Baltic Sea is high thanks to the good quality data on human impacts and pressures found on HELCOM's Map and Data Service, MSP database, and Baltic Sea Impact Index tool. Data availability is also high in the Azores; spatial data regarding human activities can be found on the OEMA geoportal, while data on coastal land use is found on the Azores' Metadata System.

Many datasets which are relevant for the specific desiderata can be found on EMODnet Human Activities (e.g., shellfish and marine finfish aquaculture locations, wind farms). Relevant datasets (e.g., MSP databases, see section 5.2.8.) are also often available at the national or regional level for several test sites. MSP spatial data may also be relevant for the identification of multi-use areas, although some more specific datasets are available including the spatial intersection of hydrocarbon exploitation and exploration activities and protected areas in the Romanian Black Sea. However, data on coastal land use change was not found, and data on coastal protection were only available in the Northwest Mediterranean, BPNS, and the Azores. No data on beach nourishment were found.

5.2.6. Data availability for desideratum 8 (Physical/dynamic data)

36 entries were identified as being potentially relevant for this desideratum. Many of these were data products from Copernicus Marine Service, such as observations of temperature and salinity, currents and sea level, and surface wind and stress, available at the global or regional scale. Thanks to these data products from Copernicus, the availability of hydrodynamical data was relatively high and uniform across the test sites,

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

although additional site-specific current data are available in the Northwest Mediterranean and the BPNS. A global dataset of eddy trajectories was also identified.

Global transport datasets from floats and drifters were identified. Additional drifter datasets covering the Northwest Mediterranean and Western Black Sea were also found, although access to the Black Sea dataset was restricted. High Frequency Radar data on radial surface velocities were found in the Gulf of Cadiz. The availability of larval behaviour data was limited, although the functional trait databases (e.g., Marine Species Traits, see also section 5.2.4.) contain some relevant data. Larval data can also be sourced from ICES for the Atlantic sites and Baltic Sea and Medtrix for the French part of the Northwest Mediterranean. Connectivity data are limited to MiCO.

5.2.7. Data availability for desideratum 9 (Existing dispersion models)

Metadata and links for all 7 of the dispersion models listed in the specific desiderata (Table 1) were included in the metadata table. These models are not test-site-specific.

5.2.8. Data availability for WP4's desideratum (MSP databases)

MSP spatial data were available at all test sites except the Italian part of the Northwest Mediterranean site and the Romanian part of the Western Black Sea site because Italy and Romania have not yet completed their MSPs. MSP databases are usually country-specific, but a transboundary database exists for the Baltic Sea. There is also a transboundary MSP database for the Western Black Sea including Bulgarian and Romanian data (developed by the MARSPLAN BS-II project), but the data are not available to download. For each test site the data can be found here:

- Northwest Mediterranean (France only)
- Gulf of Cadiz
- Belgian Part of the North Sea (data for the previous MSP are also available)
- Western Black Sea (Bulgaria only)
- Baltic Sea
- Azores (the data can be downloaded but metadata is not available)

The reason a high number of entries is given for the Azores compared to the other test sites (Table 5) is because the different layers were listed separately in the original table sent by T5.1, whereas for the other test sites each MSP database is listed as one entry.

5.3. Spatial/temporal patterns in data availability

Per sea basin, the following entries were available:

- 226 completely or partially covering the Northeast Atlantic Ocean
- 121 completely or partially covering the Baltic Sea
- 152 completely or partially covering the Black Sea
- 187 completely or partially covering the Mediterranean Sea
- 161 completely or partially covering the North Sea

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

The number of entries may not accurately reflect the data availability in each sea basin, however, because some entries are data platforms containing multiple datasets. Despite the low number of entries for the Baltic Sea, for example, data availability is high due to the large number of datasets hosted by HELCOM's Map and Data Service, MSP database, MPA database, and Biodiversity Database. Conversely, coverage of the Northeast Atlantic Ocean is only partial for many of the datasets. The quality of the data also differs between sea basins; many datasets from the Mediterranean and Black Sea only cover the EEZ of one country and access to many was restricted, particularly in Italy and Romania.

A similar pattern of data availability is seen on Europe-wide data platforms such as EMODnet. On EMODnet Biology, for example, data availability is greatest in the Atlantic and Mediterranean regions and lowest in the Baltic Sea and Black Sea regions (Figure 5). However, many of the datasets from the Atlantic and Mediterranean do not cover the MSP4BIO test sites in these regions, whereas all the EMODnet Biology datasets covering the Baltic Sea may be relevant because this test site covers the entire sea basin.
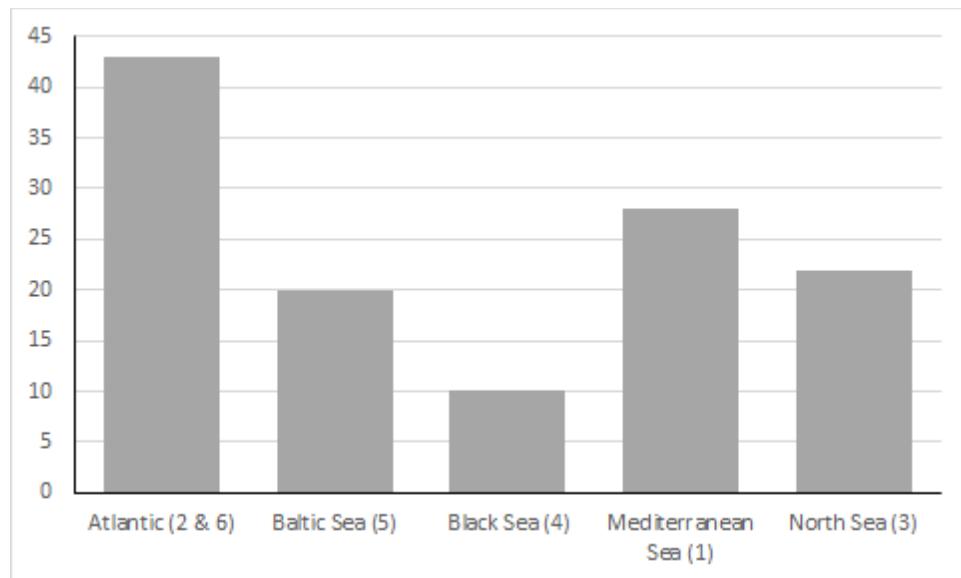
Figure 5: Number of datasets available from EMODnet Biology by EMODnet Region (EMODnet Biology Reporting Tool, 2023). Numbers in brackets indicate the test sites covered by each region (1. Northwest Mediterranean, 2. Gulf of Cadiz, 3. BPNS, 4. Western Black Sea, 5. Baltic Sea, 6. Azores).

A time series was available in 129 entries, while 134 entries covered one specific moment in time. A total of 16 entries included future projections. The number of metadata table entries was analysed for the following time periods: pre-1900, 1900-1999, 2000-2023 (not including the present), present, and future (Annex 4). Datasets could be counted in multiple time periods, depending on their temporal coverage. The highest data availability

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

was found for the period of 2000-2023, and data availability for the present was also relatively high (Figure 6). Conversely, data availability is low for the pre-1900 period and the future.
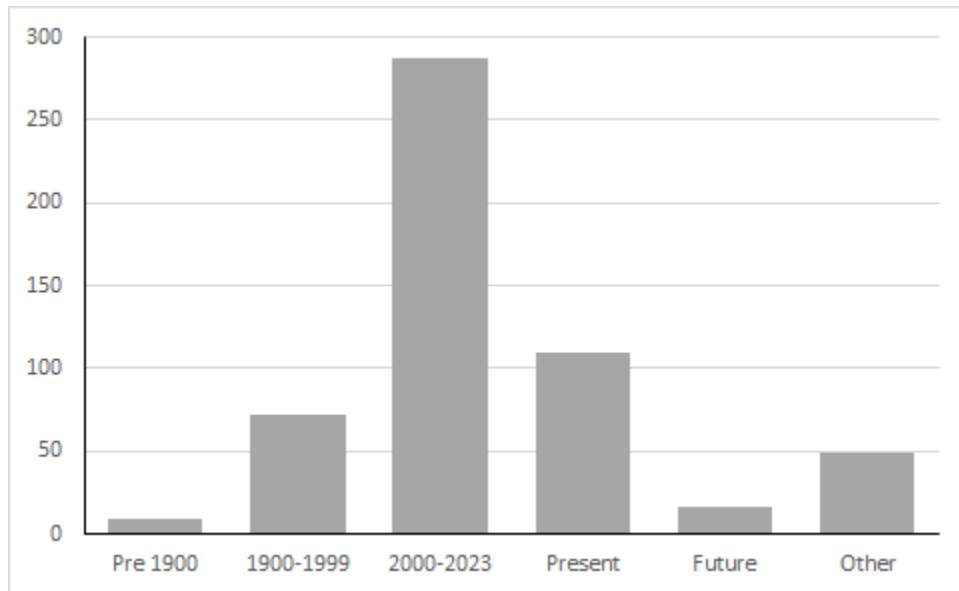
*Figure 6: Number of metadata table entries by time period covered, partially or completely. "Other" includes models, tools, and data platforms with multiple datasets covering several time periods.*

5.4. Data availability by resolution/geographical scale

By geographical scale, the following entries were available (see also Annex 5):

- 144 (42.5%) at the test site or sub-test site scale
- 70 (20.6%) at the sea basin scale
- 39 (11.5%) at the European scale
- 86 (25.4%) at "other" scales (mainly global)

Of the entries classified as "other", 77 had global coverage while the remaining 9 covered multiple sea basins (e.g., the Northeast Atlantic and North Sea). These large-scale datasets and platforms included global distribution layers for certain taxonomic and functional groups, dispersion models, and databases of taxonomy, species traits, physical data, and human activities. The European scale data mainly originated from EMODnet, Copernicus and EU platforms. Sea basin scale data came from a wide variety of sources, whereas test site scale data were mainly sourced from national data platforms and portals. 69 entries only partially covered a test site; these were mainly located in the cross-border sites of the Northwest Mediterranean or Western Black Sea and covered the EEZ of one country.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

The spatial resolution of 58 entries, mainly raster or vector datasets from a variety of sources, was recorded in the metadata table. Like geographical scale, the spatial resolution varied considerably. However, it was difficult to compare the spatial resolution of many entries because they used different units (e.g., km, degrees).

6. Conclusion

The aim of this deliverable is to provide an overview of available data at each test site and for each desideratum, thereby identifying data gaps which will set the scene for data collection. The metadata of more than 300 datasets, data platforms, models, and tools from a wide variety of sources were listed in a table. The entries were classified by data type based on categories developed in T2.2 and linked to relevant desiderata from WP3 and WP4. The resulting T2.1 metadata table is a living document and a tool which MSP4BIO partners can use to find data that is relevant for their tasks; datasets relevant for a specific test site and/or desideratum, for example. T2.1 is now considering creating a more user-friendly interface, such as a data catalogue, to help partners find and filter relevant data.

The test sites with high data availability were the Baltic Sea, due largely to the datasets hosted on HELCOM's data platforms, and the BPNS. However, even at these sites there might be instances where certain data is unavailable. Data availability was lower in the Western Black Sea and the Azores, partly because access to some of the datasets at these sites is restricted. Data availability was high for some desiderata (e.g., MPA and OECM spatial data, species occurrence maps, dispersion models, MSP databases), and low for others (e.g., regional and sub-regional climate change projections, ecological/migratory corridors, transport and larval behaviour data). This information will help guide partners in WP3, WP4, and WP5 as they collect data in specific test sites.

At future stages of the project, platforms such as EMODnet, SeaDataNet, and Copernicus Marine and Climate Change Services may be potential sources of relevant data for some of the data gaps as more data and data products are published. It may also be possible to fill some data gaps by requesting access to some of the restricted datasets identified in T2.1, although it is not guaranteed that the owners will grant access. Searching on open access repositories such as Zenodo may also yield useful data. Another possibility is to search in the literature and directly contact the authors of papers with potentially useful data.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

7. References

Copernicus Climate Change Service. https://climate.copernicus.eu/

Copernicus Marine Service. https://marine.copernicus.eu/

https://emodnet.ec.europa.eu/enEMODnet Biology Reporting Tool. Retrieved 2 May 2023, from https://rshiny.emodnet-biology.eu/reporting/

European Marine Observation and Data Network (EMODnet). https://emodnet.ec.europa.eu/en

European Ocean Biodiversity Information System (EurOBIS). https://www.eurobis.org/

European Commission, Executive Agency for Small and Medium-sized Enterprises, Schulz Zehden, A., Ramieri, E., Cahill, B. (2017). *MSP data study – Evaluation of data and knowledge gaps to implement MSP*, Publications Office. https://data.europa.eu/doi/10.2826/25289

Integrated Marine Information System (IMIS). https://www.vliz.be/en/imis

IODE (2006). Marine Data Management: We can do more, but can we do better? Retrieved from https://www.iode.org/index.php?option=com_content&view=article&id=3&Itemid=33

ISO/IEC. (2015). *ISO International Standard ISO/IEC 2382:2015 Information technology - Vocabulary*. Geneva, Switzerland: International Organization for Standardization (ISO). Retrieved from https://www.iso.org/standard/63598.html

Pınarbaşı K., Galparsoro I., Borja A., Stelzenmüller V., Ehler C.N., Gimpel A. (2017). Decision support tools in marine spatial planning: Present applications, gaps and future perspectives. *Marine Policy*, 83:83-91. https://doi.org/10.1016/j.marpol.2017.05.031.

Stamoulis K.A., Delevaux J.M.S. (2015). Data requirements and tools to operationalize marine spatial planning in the United States. Ocean & Coastal Management, 116:214-223. https://doi.org/10.1016/j.ocecoaman.2015.07.011.

8. Annexes

8.1. Annex 1

Number of entries in the metadata table by data source.

Source	Number of entries
EMODnet	23
Copernicus Marine	20
Directive Cadre Stratégie pour le Milieu Marin	18
IMIS	14
OEMA (Ordenamento do Espaço Marítimo dos Açores)	14
Rediam	10
Catálogo del Grupo Técnico de Cartografía Marina	9
EurOBIS	9
Sistema de Metadados dos Açores	9
UNEP WCMC	8
Copernicus Climate Change Service	7
SID - IL PORTALE DEL MARE	7
SEANOE	5
EEA Datasets	4
Eurostat	4
Sources with ≤3 entries each	178

8.2. Annex 2

Number of entries in the metadata table by data type and object type.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

		Data type						
		Biotic	Abiotic	Anthropogenic impacts	Social / economic / cultural	Climate	Spatial	Other
Object type	Dataset / database	103	51	33	34	11	61	3
	Data platform / catalogue	9	1	0	1	3	4	13
	Model	0	0	0	0	0	0	7
	Tool	3	0	1	1	0	0	0
TOTAL		115	52	34	36	14	65	23

8.3. Annex 3

Number of entries in the metadata table by test site and object type.

		Test site					
		Northwest Mediterranean	Gulf of Cadiz	BPNS	Western Black Sea	Baltic Sea	Azores
Object type	Dataset / database	140	129	121	124	90	128
	Data platform / catalogue	16	19	16	17	18	16
	Model	7	7	7	7	7	7
	Tool	4	4	4	4	5	4
TOTAL		167	159	148	152	120	155

8.4. Annex 4

Number of entries in the metadata table by time period covered.

Time period	Number of entries
Pre-1900	9
1900-1999	72
2000-2023	287

Present	109
Future	16
Other	49

8.5. Annex 5

Number of entries in the metadata table by geographical scale.

Geographical scale	Number of entries
Test site	144
Sea basin	70
European seas	39
Other	86 (of which 77 global)

8.6. Annex 6

Methodology and Workflow of Data Compilation and Harmonisation Post-T2.1

Although T2.1 ended in month 12, data scoping, compilation, and harmonisation efforts continued throughout the project in support of other WPs. These developments are outlined here. The methodology and workflow of T2.1 until month 12 can be found in D2.1. T2.1 continued to add relevant datasets and platforms to the compilation after month 12, resulting in a total of 404 entries by the end of the project.

Alignment with MSP Data Framework

The Technical Expert Group on Data for MSP published the Maritime Spatial Planning Data Framework (MSPdF) (Abramic et al., 2023) in month 11 of MSP4BIO. The MSPdF is a conceptual framework that structures MSP input data into seven thematic clusters, which differ from the thematic classification detailed in D2.1 (see D2.1, section 3.4), although there is some overlap. Each entry in the data compilation was consequently classified into one of the seven MSPdF clusters. When an entry could be placed in more than one cluster (e.g. data platforms with many datasets), this was indicated with the tag "multiple".

Alignment with WP3

T3.1 created four clusters of criteria for the inclusion of ecological functioning in the prioritisation, designation, and management of ABMTs which are described in D3.2: biological/ecological traits, functional/taxonomic diversity, trophic ecology, and connectivity. These were added to the data compilation along with three themes derived

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

from T3.2: climate change, species distributions, and species population trends. Each entry in the compilation was then assessed for its relevance to the seven additional thematic classifications, which was recorded in additional columns (*Table 7*). Along with the inclusion of the MSPdF clusters in the data compilation, this allowed partners to search for data using multiple classification systems.

Table 7: Columns added to the T2.1 metadata table (data compilation) after submission of D2.1

Column	Description
MSPdF Cluster	Relevant cluster according to the Maritime Spatial Planning Data Framework (MSPdF). Eight options: 1. Marine & Coastal Environment; 2. Marine & Coastal Conservation and Designated Sites; 3. Oceanographic Characteristics and Climate; 4. Coastal Land Use and Planning; 5. Operative Maritime Activities and Maritime Spatial Planning; 6. Socio-Economic Information; 7. Governance Information; Multiple.
Ecological Traits (cluster)	T3.1 criteria cluster outlined in D3.2. Two options: Y/N.
Connectivity (cluster)	T3.1 criteria cluster outlined in D3.2. Two options: Y/N.
Trophic Ecology (cluster)	T3.1 criteria cluster outlined in D3.2. Two options: Y/N.
Taxonomic Diversity (cluster)	T3.1 criteria cluster outlined in D3.2. Two options: Y/N.
Climate Change	Theme derived from T3.2. Two options: Y/N.
Species Distributions	Theme derived from T3.2. Two options: Y/N.
Species Population Trends	Theme derived from T3.2. Two options: Y/N.

Data Compilation App

An RShiny app (https://msp4bio.vliz.be/) was created by VLIZ to help partners search through the data compilation, using the desiderata, topics, and themes listed in D2.1 (Tables 1 & 3), the MSPdF clusters, and the clusters and themes from WP3 (*Table 7*). The compilation can also be filtered by test site and accessibility. Once the desired filtering options are selected, a table of relevant datasets is displayed with some basic metadata (dataset name, description, accessibility, spatial coverage, and temporal coverage). More metadata can be seen by clicking on the row in the table of the dataset of interest, and the dataset names are clickable links to the dataset's download or

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

metadata page. The app also allows the public to explore the compiled data and has been added to the MSP4BIO website.

Broken Links

Some links to the metadata or download pages in the data compilation did not work, either because a broken link had been originally provided to T2.1 or because the link had stopped working after being included in the compilation. Some of these broken links have subsequently been updated (see Table 8), while others appear to be permanently offline, e.g. One Shared Ocean (http://onesharedocean.org/data).

Table 8: Summary of updated links since submission of D2.1

Dataset / platform	Previous link	Updated link
Birds Portal - Bird observations in Romania	https://www.monitorizare-pasari.ro/date	https://monitorizareapasarilor.cndd.ro/
ESGF Portal	https://esgf-index1.ceda.ac.uk/projects/esgf-ceda/	https://esgf-ui.ceda.ac.uk/search
Vulnerable Marine Ecosystems	https://vme.ices.dk/download.aspx	https://www.ices.dk/data/data-portals/Pages/vulnerable-marine-ecosystems.aspx
Benthic occurrences, habitat maps, and species traits	https://www.eurobis.org/imis?module=dataset&dasid=6620	https://github.com/EMODnet/EMODnet-Biology-Benthic-Habitats-Occurrences-Traits
European Red List of habitats	https://forum.eionet.europa.eu/european-red-list-habitats/library	https://www.eea.europa.eu/en/datahub/datahubitem-view/de2276d8-e295-4cd7-89c9-88812065db87
FishBase	https://fishbase.mnhn.fr/search.php	https://www.fishbase.se/search.php

Broken links remain in the data compilation and will be fixed when possible and necessary in follow-up work in other projects, including BLUE CONNECT.

Future Developments

The data compilation app will remain online without maintenance, and the database will be archived and made publicly available on Zenodo. T2.1's methodology, table structure, and lessons learnt were adapted for a similar data scoping task in the BLUE CONNECT

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning

project (BLUE CONNECT T4.1), and some datasets were transferred to the BLUE CONNECT data compilation.

D2.1 Overview of the available biodiversity datasets and platforms relevant for planning