EMODnet Thematic Lot n° V – Biology

CINEA/EMFAF/2022/3.5.2/SI2.895681

Start date of the project: 10/05/2025 (24 months)

EMODnet Phase V

Deliverable 2.2: Summary on consortium data flows

The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) 2021/1139 of the European Parliament and of the Council of 7 July 2021 establishing the European Maritime, Fisheries and Aquaculture Fund.

Disclaimer

Document info

Title (and reference)	D2.2 Summary on consortium data flows
WP title (and reference number)	WP2: Data Management
Task (and reference number)	Task 1: Maintain and improve a common method of access to data held in repositories.
Authors [affiliation]	Cyril Radermecker (VLIZ), Leen Vandepitte (VLIZ), Joana Beja (VLIZ)
Dissemination level	Public
Submission date	12/11/2025
Deliverable due date	09/11/2025

Contents

Contents

D2.2: Summary on consortium data flows

Towards the end of Phase V (2023-2025) and into Phase V extension (2025-2027), the EMODnet Biology data management team started reshaping the overall data publication approach. These changes involve several 'behind-the-scenes' adaptations and do not affect the way that any data provider – within or outside of the EMODnet Biology project and consortium – needs to document, format and quality control their data. Globally, it is becoming increasingly easier and more straightforward for data holders to widely share their data via a single access point and to ensure the traceability and integrity of their data.

Within the EMODnet Biology project, this translates to providers being able to make their own choice on which Integrated Publishing Toolkit (IPT) instance they want to use, no longer being dependent on the dedicated IPT offered by the EMODnet Biology data management team or having to set up or manage an IPT within their own institute or organisation.

To accommodate these global changes, the EMODnet Biology harvesting processes are being re-routed, so data from other IPT instances can easily be picked up as soon as the data management team is alerted to either newly available datasets or updates on pre-existing datasets. This new approach tremendously lowers the risk of duplicating these data in some of our long-standing international collaborations and data flows, for example with the Ocean Biodiversity Information System (OBIS) and the Global Biodiversity Information Facility (GBIF), which has led to occasional confusion on the level of version-control of a dataset in the past. Some finetuning and further communication in a local, regional, European and global context still needs to happen during this extension phase, as well as re-working our overall data flow schemes.

1 General data flow

The schematic presentation of the general data flow within the EMODnet Biology project is shown in Figure 1. A large part of the data preparation is done by each of the (consortium) data providers, who format the data according to the DarwinCore OBIS ENV format. During formatting, (consortium) data providers also follow the defined quality control (QC) and standardization steps, after which the data are uploaded to the Integrated Publishing Toolkit (IPT) instance of their choice. This can be an IPT at their own institute or organization, or an IPT maintained by another institute or organization, free to use by others. Once the data are on IPT, the EMODnet Biology data management team is informed. They make final checks. Any issues that might arise during this step are resolved in collaboration with the (sub-contracted) data providers and, when necessary, the data originators.

When data are uploaded to an IPT, (consortium) data providers are requested to ensure that this IPT instance is registered with the Global Biodiversity Information Facility (GBIF). In addition, OBIS needs to be added as a network and this last step ensures that the OBIS (Ocean Biodiversity Information System) community is aware of a dataset, and can take the necessary steps in the data flow process.

The final data checks are done using the LifeWatch & EMODnet Biology Quality Control (QC) tool (https://rshiny.lifewatch.be/BioCheck/ and https://github.com/EMODnet/EMODnetBiocheck), publicly accessible. All (consortium) data providers are expected to run their data through this tool, to identify any issues and resolve them before final submission. Once the data passes the integrity, quality, format and visual checks, they are harvested by the data management team, and made available through the EMODnet Portal viewer and associated web services.

Due to established data flows and international collaboration, data are also shared with the OBIS and GBIF.

EMODnet Biology users can access the data through the EMODnet Biology services (EMODnet viewer and/or through webservices, as described in the EMODnet web service documentation page), as this gives access to the most complete version of the data in a Europe-focused context. Some partners do not use the IPT route, but deliver their data through web services. Even when data arrive through a different route, the quality control and standardization steps remain similar to those described above.

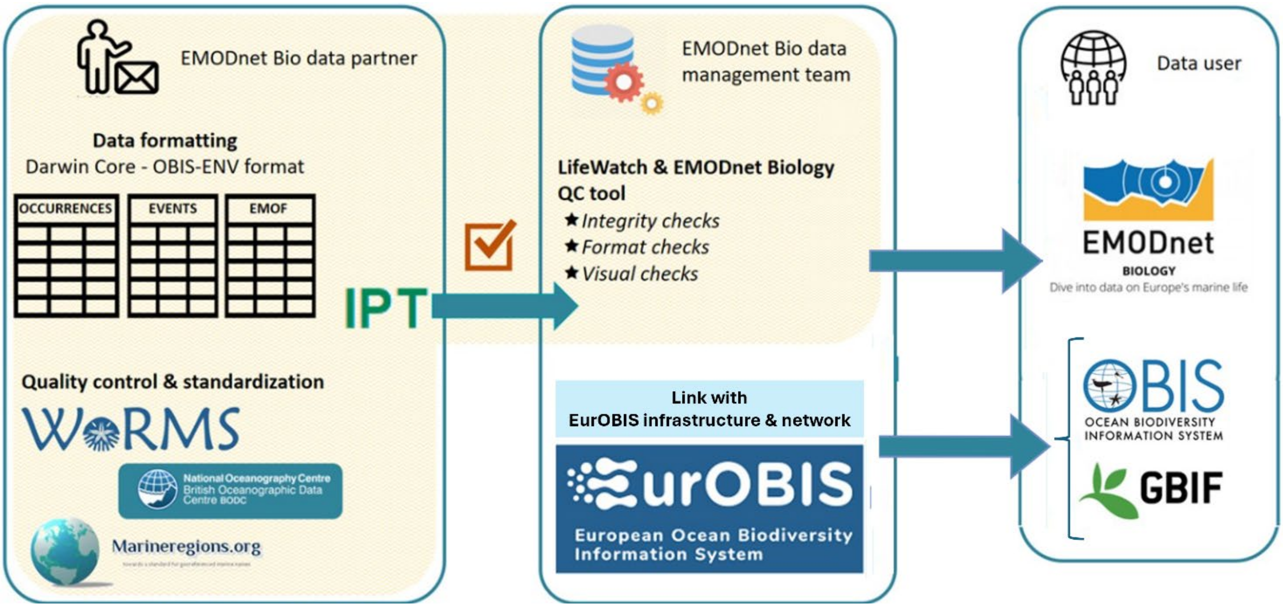

Figure 1: Generalized Data Flow Presentation. The highlighted area represents the area of collaboration between the EMODnet Biology datacontributors and the data management team. The data flow to OBIS and GBIF and the following visibility of the data in OBIS and GBIF depends on their harvesting cycles The EMODnet Biology harvest cycle is performed on a 3-month cycle.

In the change from Phase V to Phase V.2, two new data providers joined WP2 – Data Management: The Cyprus Marine and Maritime Institute (CMMI) and Institut Català de Recerca per a la Governança del Mar (ICATMAR). ICATMAR and the EMODnet Biology team have had previous contacts on a possible data collaboration during Phase V (December 2023), although it took until this phase to formalize this collaboration intent through a sub-contract. Although CMMI is involved in several other collaborations with several divisions within the Flanders Marine Institute (VLIZ), it is the first time they are directly connected to EMODnet Biology through a sub-contract, where they will make available three datasets.

In the evolution towards the extension period of Phase V (2025-2027), a long-standing data collaboration with the Instituto Português do Mar e da Atmosfera (IPMA) and the Marine Information Service (MARIS) within WP2 has ended and they will not contribute with data provision. IPMA has left the EMODnet Biology consortium and MARIS remains involved in the project, but their activities will solely focus on specific tasks included in WP5 - technical infrastructure.

For all other (coonsortium) data providers involved in WP2, the established data flow mechanisms of the previous Phase(s) are continued, and - where relevant or needed - will be fine-tuned during this extension period.

Information on the data flow for these providers can be found in D2.4: Evaluation, definition and implementation of partner-specific data flows published in Phase III and D2.2. Technical implementation of data flows for the new project partner(s) published in Phase IV and D2.2.1. Summary on consortium data flows published in Phase V. Most of the WP2 data providers now deliver data through the Integrated Publishing Toolkit (IPT), one of the preferred and most automated and quality-controlled ways to make data available to the data infrastructure supporting EMODnet Biology and EurOBIS (the European OBIS node).

2 New consortium data providers(s)

This section lists the new institute(s)/organization(s) that will develop work related to data delivery and accessibility within the extension period of Phase V of the EMODnet Biology project, identified as WP2 – Data Management.

2.1 The Cyprus Marine and Maritime Institute (CMMI)

The Cyprus Marine and Maritime Institute (CMMI) is based in Cyprus. It has ten Research and Innovation Centres that are involved in marine technology, marine robotics, maritime digitalization, marine observation, marine and coastal ecosystems, marine biotechnology and aquaculture, the maritime human element, marine cultural heritage, marine and maritime policies and regional cooperation. By partnering with CMMI through a sub-contract, it was agreed that at least 3 datasets will become part of the data flow to EMODnet Biology. The datasets to be delivered concern megafauna (invertebrates, macroalgae, fish, benthos) with a temporal coverage from 2021 to 2026. The data are expected to be made available through IPT, as is the case for many of our consortium data providers.

2.2 Institut Català de Recerca per a la Governança del Mar (ICATMAR)

The Catalan Institute of Research for the Governance of the Sea (ICATMAR) is a collaboration organism between the Directorate-General for Maritime Policy and Sustainable Fisheries of the Catalan Government and the Spanish National Research Council (CSIC). ICATMAR is carrying out a monitoring programme for both recreational and commercial fishing along the Catalan coast, and since 2023 a programme of observation, analysis and prediction of the physical characteristics of the sea. The dataset to be delivered concerns a bottom trawl survey with a temporal coverage from 2019 to 2024. Similar to the majority of our other sub-contracting data partners, the data are expected to be made available through IPT.

3 Summary overview of all WP2 data providers & their data delivery mechanism(s)

The table below provides an overview of all current WP2 (sub-contracted) data providers, the agreed data sharing mechanism and the number of new and/or updated datasets they will provide during this extension period. Three organizations cannot provide an exact number of datasets that will be delivered (VLIZ, HCMR, MBA) as they provide a continuous data flow to EMODnet Biology based on their own networks and the initiatives they are involved in.

Table 1: Overview of all current WP2 consortium data providers (IPT = Integrated Publishing Toolkit; TBD = To be determined). All institutes except OGS (which is a full partner) are sub-contractors to EMODnet Biology.

WP2 data providing institute	Acronym	Data delivery mechanism	New datasets	Dataset updates
Aarhus University, DCE-Danish Centre for Environment and Energy	Aarhus University	IPT	1	1
Alfred Wegener Institute – PANGAEA	AWI-PANGAEA	IPT	6	-
Cyprus Marine and Maritime Institute	CMMI	IPT	3	-
Deltares	Deltares	IPT	4	-
Hellenic Centre for Marine Research	HCMR	IPT	10+TBD	TBD
Institut Català de Recerca per a la Governança del Mar	ICATMAR	IPT	1	-
International Council for Exploration of the Sea	ICES	IPT	-	7
Instituto Español de Oceanografía	IEO-CSIC	IPT	1	2
Institut français de recherche pour l'exploitation de la mer	IFREMER	IPT	1	1
Instituto de Hidráulica Ambiental de la Universidad de Cantabria	IHCantabria	IPT	3	3
Institute for Agricultural and Fisheries Research	ILVO	IPT	3	9
Havforskningsinstituttet Institute of Marine Research	IMR	IPT	1	5
Marine Biological Association	MBA	IPT	TBD	TBD
National Institute for Marine Research and Development "Grigore Antipa"	NIMRD	IPT	4	-
Royal Netherlands Institute for Sea Research	NIOZ	IPT	5	-
Istituto Nazionale di Oceanografia e di Geofisica Sperimentale Trieste	OGS	IPT	8	-
Swedish Meteorological and Hydrological Institute	SMHI	IPT	2	15
Suomen ympäristökeskus - Finnish Environment Institute	Syke	Web services	1	3
Flanders Marine Institute	VLIZ	IPT	TBD	TBD
Ukrainian Scientific Centre of Ecology of the Sea	UkrSCES	IPT	2	-
	Total		56	46