



EMODnet



European Marine
Observation and
Data Network

EMODnet Thematic Lot n° VI – Biology

CINEA/EMFAF/2022/3.5.2/SI2.895681 - *EMODnet Biology*

Start date of the project: 10/05/2023 - (24 months)

EMODnet Phase V

Deliverable 2.2.1: Summary on consortium data flows



Disclaimer

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the EASME or of the European Commission. Neither the EASME, nor the European Commission, guarantee the accuracy of the data included in this study. Neither the EASME, the European Commission nor any person acting on the EASME's or on the European Commission's behalf may be held responsible for the use which may be made of the information.

Document info

Title [ref]*	D2.2.1. Summary on consortium data flows
WP title [ref]*	WP2: Data Management
Task [ref]*	Task 1: Maintain and improve a common method of access to data held in repositories.
Authors [affiliation]	Ruben Perez Perez (VLIZ) ruben.perez@vliz.be Leen Vandepitte (VLIZ) leen.vandepitte@vliz.be
Dissemination level	Public
Submission date	09/11/2023
Deliverable due date	10/11/2023

Contents

D2.2.1: Summary on consortium data flows	4
1 General data flow	4
2 New data partner(s)	6
2.1 Universität Bremen (UniHB)	6
2.2 Centre for Ecology Fisheries and Aquaculture Science (CEFAS)	6
3 Summary overview of all WP2 partners & their data delivery mechanism(s)	7

D2.2.1: Summary on consortium data flows

During the first six months of EMODnet Biology Phase V, the data management team at VLIZ has been in contact with all data partners involved in WP2- Data Management, to discuss the data flow to VLIZ, for integration into the EurOBIS¹ data infrastructure and subsequent publication in EMODnet Biology.

1 General data flow

The schematic presentation of the general data flow within EMODnet Biology is shown in Figure 1. A large part of the data preparation is done by each of the EMODnet Biology data partners, where they format the data according to [DarwinCore OBIS ENV format](#). During formatting, partners also follow the defined quality control and standardization steps, after which the data are uploaded to the [Integrated Publishing Toolkit \(IPT\)](#). Once the data are in IPT, the EMODnet Biology data management team is informed, and final checks are made, issues arising during this step are resolved in collaboration with the data partners and, when required the originators.

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>), accessible to the data partners, who are recommended to run their data through this tool, to identify any issues and resolve them before final submission. Once the data pass the integrity, format and visual checks, they are uploaded into EurOBIS, the data infrastructure behind EMODnet Biology. Upon addition to the EurOBIS database, they will be made available to the users through EMODnet Biology, more specifically as a layer in the EMODnet Central Portal viewer ([EurOBIS Database Observations](#) under the Occurrence data sub-theme). Due to established data flow and collaboration, data are also shared with the Ocean Biodiversity Information System (OBIS)² and the Global Biodiversity Information Facility (GBIF)³, which represent alternative ways for users to access the data.

EMODnet Biology users are recommended to access the data through the EMODnet Biology services, as this gives access to the most complete version of the data. Some partners do not use the IPT route, but deliver their data through web services, although data arrive through a different route, the quality control and standardization steps remain similar to those described above.

¹ European Biodiversity Information System (www.eurobis.org)

² Ocean Biodiversity Information System (www.obis.org)

³ Global Biodiversity Information System (www.gbif.org)

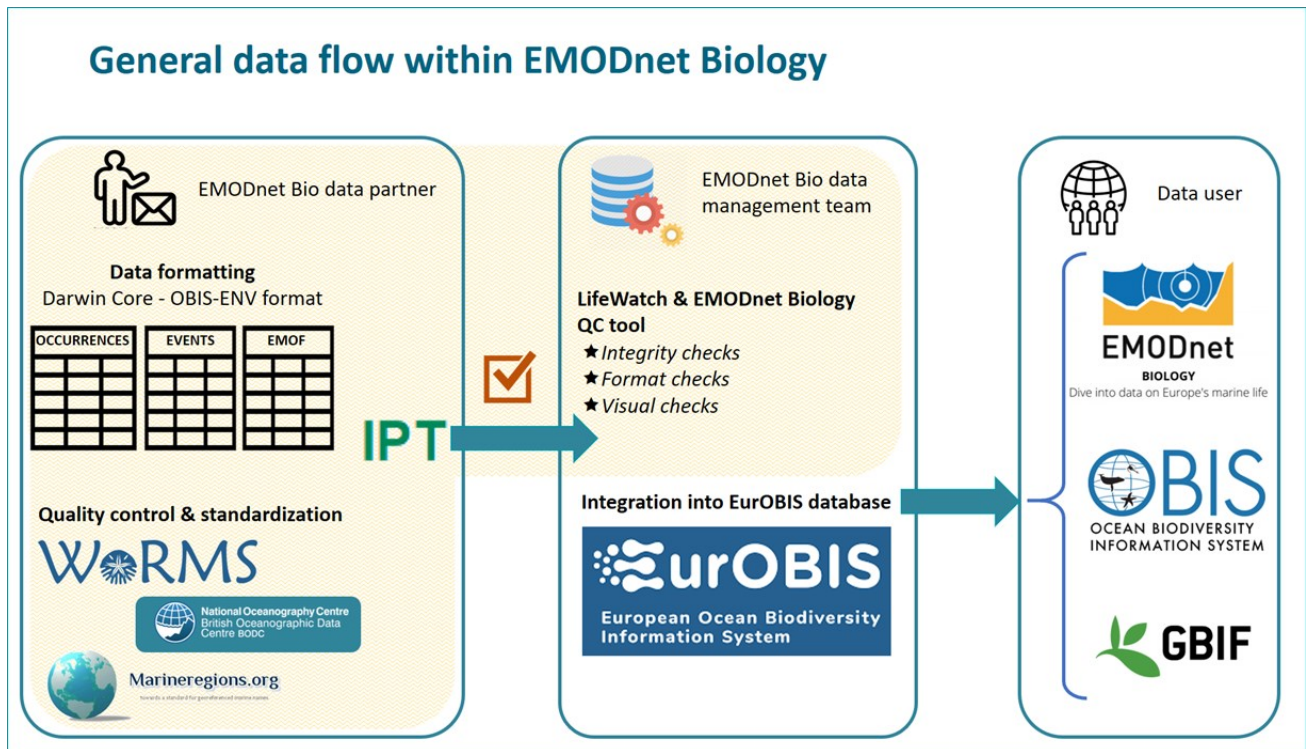


Figure 1: Generalized Data Flow Presentation. The highlighted area represents the area of collaboration between the EMODnet Biology data partner and the data management team. From within the EurOBIS database, data are made available to EMODnet Biology, OBIS & GBIF simultaneously, although visibility of the data in OBIS and GBIF depends on their harvesting-cycles, which don't always align with the ones in EMODnet Biology.

In the change from Phase IV to Phase V, two new data partner was added to WP2 – Data Management: Universität Bremen (UniHB)/Pangaea and CEFAS.

These partners are not completely new to EMODnet Biology, nor to data management and how access to marine biological data can be accomplished and improved. Both organisations have been connected to one or more of the previous EMODnet Biology Phases, as partners in the Preliminary Phase and through ad hoc delivery of data.

Additionally, IFREMER, which was a partner in the previous Phases, did not join the consortium in Phase V.

For all other partners involved in WP2, the established data flow mechanisms of the previous Phase(s) are continued, and - where relevant or needed - will be fine-tuned in the course of this Phase. Information on the data flow of these partners 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](#). Most of the partners 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 EurOBIS data infrastructure.

2 New data partner(s)

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

2.1 Universität Bremen (UniHB)

The Universität Bremen is one of the institutions running PANGAEA, an information system and data publisher for geoscientific and environmental data. PANGAEA is jointly operated by the Centre for Marine Environmental Sciences (MARUM), UniHB and the Helmholtz Centre for Polar and Marine Research at the Alfred Wegener Institute (AWI). In the past, EMODnet Biology has had collaborations and agreements with both MARUM and AWI. We are now closing the circle within Phase V, by partnering with UniHB and the agreement that at least 5 major datasets from PANGAEA will become part of the data flow to EMODnet Biology. All datasets to be delivered concern megafauna (seabirds, marine mammals and other megafauna, thereby also addressing one of the data gaps within the EMODnet Biology consortium. The data are expected to be made through IPT, as is the case for the majority of our partners.

2.2 Centre for Ecology Fisheries and Aquaculture Science (CEFAS)

Due to its geographic location and established data flows within the United Kingdom, CEFAS collaborates with MBA with respects to their data flow into EMODnet Biology. The procedure is as follows: CEFAS provides the datasets as excel files to MBA, who will then process them further and will make them available via the IPT. Note that CEFAS was not initially listed as a data provider in the proposal, but following the project's kick off meeting, it was deemed relevant to include them also in this WP. For the reasons presented above, they are not included in the table in the next section, although it is expected that CEFAS will deliver a few datasets during this Phase.

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

The table below provides an overview of all current WP2 partners, the agreed data sharing mechanism and the number of new and/or updated datasets they will provide during this phase. Four partners cannot provide an exact number of datasets that will be delivered (VLIZ, HCMR, MBA & MARIS) as they provide a continuous data flow to EMODnet Biology based on their own networks and initiatives they are involved in. Several data partners have also indicated they might have access to more datasets than what they have currently inventoried, so the numbers in the table below should be seen as the minimum number of submissions for Phase V.

Table 1: Overview of all current WP2 partners (IPT = Integrated Publishing Toolkit; SDN = SeaDataNet; TBD = To be determined)

Partner	Partner acronym	Data delivery mechanism	New datasets	Dataset updates
Suomen ympäristökeskus - Finnish Environment Institute	SYKE	web services	-	3
Universität Bremen	UniHB	IPT	5	-
Royal Netherlands Institute for Sea Research	NIOZ	IPT	3	-
Instituto de Hidráulica Ambiental de la Universidad de Cantabria	IH Cantabria	IPT	1	2
Havforskningsinstituttet Institute of Marine Research	IMR	IPT	-	5
National Institute for Marine Research and Development "Grigore Antipa"	NIMRD	IPT	2	-
Swedish Meteorological and Hydrological Institute	SMHI	IPT	2	12
Ukrainian Scientific Centre of Ecology of the Sea	UKRSCES	IPT	6	-
Aarhus University, DCE-Danish Centre for Environment and Energy	Aarhus University	SDN	-	1
Instituto Português do Mar e da Atmosfera	IPMA	IPT	-	5
Deltares	Deltares	IPT	5	-
International Council for Exploration of the Sea	ICES	web services	1	7
Instituto Español de Oceanografía	IEO	IPT	2	2

Partner	Partner acronym	Data delivery mechanism	New datasets	Dataset updates
Istituto Nazionale di Oceanografia e di Geofisica Sperimentale Trieste	OGS	IPT	4	4
Institute for Agricultural and Fisheries Research	ILVO	IPT	2	7
Flanders Marine Institute	VLIZ	IPT	1+ TBD	TBD
Hellenic Centre for Marine Research	HCMR	IPT	6 + TBD	TBD
Marine Biological Association	MBA	IPT	1 + TBD	TBD
Marine Information Service	MARIS	SDN	TBD	TBD
	Total		41	48