



EMODnet



European Marine
Observation and
Data Network

EMODnet Biology

EASME /EMFF/2016/006

Start date of the project: 19/04/2019 - (24 months)

EMODnet Phase III

D2.4: Evaluation and definition of partner-specific data flows, and the technical implementation of partner-specific data flows



Disclaimer¹

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	D2.4: Evaluation and definition of partner-specific data flows, and the technical implementation of partner-specific data flows
WP title	WP2: Data access to marine biological data
Task	T1: A common method of access to data held in repositories
Authors [affiliation]	Leen Vandepitte (VLIZ)
Dissemination level	Public
Submission date	31/10/2019
Deliverable due date	31/10/2019

¹ The disclaimer is needed when the document is published

Contents

D2.4: Evaluation and definition of partner-specific data flows, and the technical implementation of partner-specific data flows 4

1 Overview of each data partner, and the mechanism(s) that will be used to provide their data to EuroBIS, the database behind the EMODnet Bio portal... 4

1.1	Marine Biological Association (MBA).....	4
1.2	Centre for Environment, Fisheries and Aquaculture Science (CEFAS)	5
1.3	International Council for the Exploration of the Sea (ICES).....	5
1.4	Swedish Meteorological and Hydrological Institute (SMHI).....	5
1.5	Istituto Nazionale di Oceanografia e Geofisica Sperimentale (OGS).....	5
1.6	Finnish Environment Institute (SYKE).....	5
1.7	Instituto Español de Oceanografía (IEO)	6
1.8	Institute of Marine Research (IMR)	6
1.9	Institute of Agricultural and Fisheries research (ILVO)	6
1.10	Instituto Português do Mar e da Atmosfera (IPMA).....	6
1.11	Institute of Oceanography and Fisheries (IOF).....	6
1.12	Marine Information Service (MARIS)	6
1.13	Royal Netherlands Institute of Sea Research (NIOZ).....	7
1.14	Aarhus University	7

D2.4: Evaluation and definition of partner-specific data flows, and the technical implementation of partner-specific data flows

During the first six months of the second phase of the EMODnet Biology 3 project, the data management team at VLIZ has been in close contact with all partners involved in WP2, to discuss how the data transfer to VLIZ can be optimised, and to identify possible problems in the data processing.

This second phase will partly bring new data to EMODnet Biology, but will largely focus on expanding the currently available datasets, so that they include related data such as e.g. abundance, biomass and abiotic measurements related to the species observations.

The closing meeting of the first phase was also seen as the kick-off of the second phase. During this meeting (May 2019), the data management team sat down with each WP2 partner separately, to discuss in depth the data that had already been delivered – with a focus on how they can be extended and improved – and the data that could additionally flow to EMODnet Biology. As the focus of this Deliverable is specifically on the data flow mechanisms, the other topics will not be discussed here.

1 Overview of each data partner, and the mechanism(s) that will be used to provide their data to EurOBIS, the database behind the EMODnet Bio portal

Below, we list all partners and the data delivery mechanism of their choice. In addition, we will mention whether the work only concerns updates of already available datasets, or if new datasets will be delivered. Updates should be seen as either (1) more recent records being added, or (2) more data added to the already existing records, e.g. specific biotic/abiotic measurements or (3) a combination of the above.

1.1 Marine Biological Association (MBA)

Recently, MBA has become the official OBIS node for the United Kingdom. This implies that MBA takes responsibility within the OBIS-network to capture all UK-related marine biological datasets (including Ireland and Scotland) and will make them available to OBIS. For this data flow, MBA will make use of the standard data exchange mechanism used within the OBIS network, more specifically, the Integrated Publishing Toolkit (IPT).

As MBA will thus make use of IPT, and is guaranteeing a good quality of the gathered datasets, the EurOBIS data management team can easily harvest datasets from the MBA IPT. Within EurOBIS, all UK datasets are of interest, but a primary focus will of course be on the UK datasets that have been defined within the scope of the EMODnet Biology project.

As SAFHOS was recently merged with MBA, MBA now also takes responsibility for the Continuous Plankton Recorder (CPR) data. Due to this merge, the CPR-data will also flow to EurOBIS/EMODnet Biology through IPT, which is a major improvement on the previous data exchange model for the CPR data.

MBA has also negotiated that the CPR data are more openly available, implying that the abundance data can now be shared, whereas in the past we only had access to presence-data.

1.2 Centre for Environment, Fisheries and Aquaculture Science (CEFAS)

CEFAS collaborates with MBA to let their data flow into EMODnet Biology. CEFAS will provide the datasets as Excel files to MBA, who will then process them further.

In addition to the datasets that were contractually agreed in the first phase, CEFAS will provide an additional 2 datasets, relating to long-term time-series.

1.3 International Council for the Exploration of the Sea (ICES)

So far, ICES data came to the data management team as data dumps. This has proven to be a challenge in the past, so several improvements will be made to the procedures.

A renewed mapping will be made between the ICES and the EurOBIS database, concurring to the Darwin Core (DwC) format used by EurOBIS/EMODnet Biology. Once this mapping is in place, ICES can still work with customized data dumps, but import these data will become much easier for the data management team.

By the end of 2019, ICES will have a new data portal in place, including new data services that will make the data exchange even more efficient.

1.4 Swedish Meteorological and Hydrological Institute (SMHI)

SMHI will start using the IPT for the data exchange which will greatly enhance and speed up the process. In parallel, they will convert the datasets that were delivered during the previous EMODnet Biology projects to EventCore, thus capturing more data and information, and they will also start making use of the BODC vocabularies, which in its turn will improve standardisation possibilities through the EMODnet Bio portal.

1.5 Istituto Nazionale di Oceanografia e Geofisica Sperimentale (OGS)

OGS is a long-term IPT user, thus no enhancements are required on their data delivery mechanism. The OGS data manager will however work on improving the metadata of the currently available datasets, and – where possible and relevant – transfer these to the EventCore format. In addition, a number of new datasets could be provided. OGS informed us that all Italian data from the Marine Strategy Framework Directive (MSFD) have been made public, and could also flow to the EMODnet Biology portal. The only constraint for these data to become available through the portal is the time needed to process them.

1.6 Finnish Environment Institute (SYKE)

The phytoplankton and macrozoobenthos data from SYKE are provided through an API and are updated every night. An agreement has been made with the VLIZ data management team that we will harvest/load the SYKE data every 6 months.

The zooplankton data will be manually updated. SYKE also has a line-based macroalgae dataset that is part of a monitoring program. Contribution of this dataset will however largely depend on the possibility to digitise the data within the project resources.

1.7 Instituto Español de Oceanografía (IEO)

IEO will work on scripts to transform the data from within their local database to a DwC compatible format. In parallel with this, they will deal with existing data problems and will put a focus on standardisation to DwC terminology.

IEO will provide annual updates of their data, whilst taking into account a three-year moratorium.

1.8 Institute of Marine Research (IMR)

Currently available datasets from IMR in the EMODnet Biology portal will be transformed to the DwC EventCore format. Through this action, data linked to the species observations will also become available. This can include e.g. abundance or biomass data. Some updates – in number of records – will be given for the zooplankton data, and the possibility to deliver some extra datasets will be looked in to. The IPT was already in use as a data exchange mechanism for IMR, and will continue to be used.

1.9 Institute of Agricultural and Fisheries research (ILVO)

The datasets provided in the previous EMODnet Biology phases will be transformed to the DwC EventCore format, thus bringing more linked data to the portal. In addition, the data flow will be optimised by making use of the IPT.

1.10 Instituto Português do Mar e da Atmosfera (IPMA)

For IPMA, environmental data will be added to the previously delivered datasets, thus transferring these data to EventCore.

IPMA already makes use of the IPT for data exchange, so no improvements can be made on the data delivery mechanism. In addition to the transfer to EventCore, existing problems with the data provided will be addressed and two new datasets can possibly be delivered in this second phase. Due to a change in staff at IPMA, the ongoing data activities might be slowed down in the coming months.

1.11 Institute of Oceanography and Fisheries (IOF)

In this second phase, a new colleague will start working on the data management at IOF. New possible datasets include phyto and zooplankton from different stations and temporal ranges, but will require a digitisation.

There are additional datasets on benthos and mammals (BlueWorld) but the possibility for the data to be share needs to be double-checked internally first. Continuous efforts are being coordinated with DGMARE to follow-up on the MEDITS datasets that are still locked.

1.12 Marine Information Service (MARIS)

MARIS currently has a filter installed for EMODnet Biology datasets on the SeaDataCloud system. This will be updated, so that data formatted in the BIO-ODV format are easily accessible to VLIZ. A notification system will be set in place, to alert VLIZ on new data. MARIS aims to have this system in place before the end of 2019.

Improvements will be made on the metadata level, making sure that the metadata from the SeaDataCloud datasets reach EMODnet Biology's standards.

1.13 Royal Netherlands Institute of Sea Research (NIOZ)

NIOZ data will be delivered through the IPT. In order to share more data with EMODnet Biology, NIOZ will contact their colleagues and check if they would be willing to share additional datasets.

1.14 Aarhus University

Since the start of EMODnet Biology, Aarhus has delivered their datasets to the data portal through SeaDataNet. Aarhus will continue to use this data delivery mechanism.

VLIZ will evaluate the current data format and provide feedback to Aarhus on this, in order to allow Aarhus to modify and improve their script that transforms the data from their local database into SDN format.