

A quick guide to publish biodiversity data in EurOBIS and EMODnet Biology

Introduction

What type of data can you submit?

Where can you submit your data?

1. EMODnet Biology online submission form
2. EMODnet Data Ingestion portal
3. Integrated Publishing Toolkit (IPT)

Requirements

1. Metadata
2. Data standards
3. Data quality checking
4. Digital Object Identifier

Introduction

This guide will discuss how to make data available through EurOBIS. Data in the EurOBIS database are synchronized with EMODnet and will flow to OBIS and GBIF. There are several ways to share your biodiversity data and make it available in EurOBIS and EMODnet Biology. In any case, data and metadata need to be formatted following certain accepted standards to allow efficient discovery of datasets and data interoperability. This document gives a broad overview with the key aspects on how to publish your data, pointing to more detailed guidelines for each of the steps.

What type of data can you submit?

Geographical scope	European and non-European, provided they are collected by European institutes. Water column and seabed.
Temporal scope	All.
Data type	Raw data. Data products. Dataset descriptions: when dataset cannot become publicly available (yet).
Parameters	Taxon name, a geographical location and and observation date. Observation: presence only, abundance, biomass. Other parameters, including abiotic (e.g. body length or sediment composition).
Data origin	Smaller scale, one-off scientific research data. Larger scale and/or long-term monitoring. Museum collections and literature data.
Taxonomic scope	All different groups of marine species (phytoplankton, zooplankton, macro-algae, angiosperms, benthos, birds, mammals, reptiles and fish).
Data accessibility	CC licence required.

Where can you submit your data?

1 EMODnet Biology online submission form

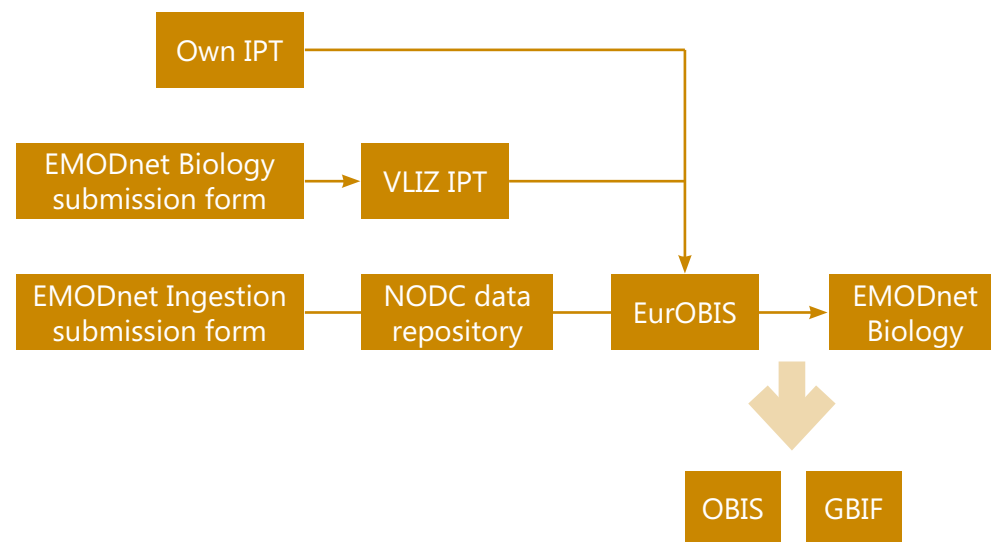
The EMODnet Biology portal provides a simple [submission form](#) for users who want to contribute their biodiversity data. Some basic mandatory metadata fields have to be filled in. The dataset will be handled by VLIZ in collaboration with the data provider and will be made available in EMODnet Biology and EurOBIS if desired.

2 EMODnet Data Ingestion portal

[EMODnet Ingestion](#) is a web portal to facilitate submission of marine data of any discipline. A straight-forward online submission form and a help-service will help the user fill in the metadata and submit their data in the most appropriate format. The user will be put in contact with the corresponding National Oceanographic Data Center (NODC) in charge of processing the data and implementing the pathway from Ingestion to the EurOBIS and EMODnet Biology portals.

3 Integrated Publishing Toolkit (IPT)

The GBIF [Integrated Publishing Toolkit \(IPT\)](#) is a freely available open source web application that makes it easy to share biodiversity-related information. The user is guided by the application through a series of steps to format their data, fill in the necessary metadata and publish the dataset. If you have a good number of datasets or you will need to publish or update datasets regularly, you might consider setting up your own IPT by following these [guidelines](#).



Did you know?



In the [website](#) of the Ocean Teacher Global Academy (OTGA) (a UNESCO-led project) you can access different resources and online training in the field of Ocean Sciences.

There is a specific training section about [OBIS](#). Take a look at the latest courses:

- [OBIS-INDEEP training & workshop \(2016\)](#)
- [Marine Biogeographic Data Management \(contributing and using OBIS\) \(2015\)](#)

Requirements

There are some requirements for your dataset to be published in EurOBIS or EMODnet Biology:

- Your dataset needs to be described: a minimum set of **metadata** has to be provided.
- Your data has to be organized and formatted according to the Darwin Core **standard** (DwC).
- Your data will have to undergo some essential **quality checking**.
- Optionally, VLIZ can assign a **Digital Object Identifier** (DOI) to your dataset.

This might seem quite challenging but, in practice, all these steps will be assisted by an online tool or facilitated by an institution (e.g. NODC) that will handle your dataset with your collaboration. Take a quick look at each step:

1 Metadata: describing your data

Metadata will help other users to better understand the content of your data, it will extend data longevity and facilitate data discovery and reuse ([example](#)).

There is some [essential information](#) that is always required and/or recommended. In practice, you will just need to fill in an online form with a few mandatory and non-mandatory fields. The most important ones, together with some standard advice on how to fill them are listed below:

a) Choose a **dataset title** that is descriptive, meaningful and concise. By reading its title, a user should know what to find in the dataset:

Benthic data from the Southern Irish Sea from 1989-1991
Zoobenthos of the Kyklades (Aegean Sea) from a survey in 2009

b) Provide a good **abstract** that can help potential users to understand if the data may be of their interest. This should be a short description indicating data type or origin, taxa or functional groups included, when and where data was collected, etc:

Distribution of dinoflagellates causing harmful algal bloom in the Mediteranean sea collected from literature. The data contains only HABS that appeared prior to 2004.

c) Give a **resource contact**, a person or organization that can be contacted in case there are questions related to the dataset: name and institute of the data provider, an email address and/or a telephone number.

d) Provide a **citation** for your dataset (the equivalent of a publication reference). It should contain the author(s) of the dataset, year of publication, dataset title, publisher and identifier (if available). It can also mention the version and resource type:

Belmonte, G. (2010) Zooplankton - Crustacea from the Taranto Seas. University of Salento, Department of Biological and Environmental Science and Technologies, Laboratory of Zoology and Marine Biology (LZMB), Italy. <http://dx.doi.org/10.14284/57> Bio-environmental research group; Institute of Agricultural and Fisheries research (ILVO), Belgium; (2016): Macrobenthos monitoring at long-term monitoring stations in the Belgian part of the North Sea between 1979 and 1999. <http://dx.doi.org/10.14284/201>

e) Add a few **keywords** that enhance dataset discovery: functional and or taxonomic group (crustacea, zooplankton), geographical area (Mediterranean), etc.

f) Specify the usage of the dataset or a **data license**: i.e.: under which conditions the dataset can be used. We request the use of [Creative Commons licenses](#).

g) Give the date or the **temporal scope** covered by your dataset, preferably in the YYYY-MM-DD format (or [ISO 8601](#) compliant).

h) Similarly, describe the spatial or **geographical extent** of the data by listing the area(s) or location(s) where data was collected (e.g. South Atlantic Ocean, Belgian part of the North Sea). You can use marineregions.org to find adequate geo-units.

i) Provide an overview of the **taxonomic scope** present in the dataset (e.g. Crustacea, Pisces). You can look for internationally accepted taxa at marinespecies.org.

2 Data standards

Standardization facilitates data discovery, integration, sharing and interoperability. EurOBIS (and OBIS), EMODnet Biology and IPT use the [Darwin Core](#) standard. The Darwin Core is a body of standards with a list of [defined terms](#) that allow your data to be understood and used by anyone.

The list of Darwin Core terms might seem overwhelming. This is because Darwin Core is a general schema designed to fit all biodiversity purposes, but not all terms are needed by each specific research community. Therefore, you have to keep in mind that:

- **Only a handful of fields are strictly necessary.** An overview of the mandatory and recommended fields of the OBIS schema, together with an explanation on their meaning and how to fill them in is available at the [OBIS manual](#). A spreadsheet [template](#) is available for download.
- If you submit your data via EMODnet Ingestion or EMODnet Biology you will be assisted by the correspondent NODC or VLIZ respectively.
- If you publish your data using IPT you will use the Darwin Core mapping tool, which simply allows you to match or link the fields in your source with the appropriate Darwin Core terms.

Important:

Additional information such as biometrics of the observed taxa (e.g.: body length) or even non-biological parameters (water temperature, sediment grain-size, etc) are supported by the Darwin Core schema and (Eur)OBIS (see [De Pooter et al., 2017](#)), and therefore welcome.

3 Data quality checking

Before any data is made available through EurOBIS or EMODnet Biology, it goes through a series of quality control procedures. This step improves the quality of the data source and the overall EurOBIS and EMODnet databases by checking that:

- All the necessary metadata and data fields are filled in.
- All taxon names present are matched to their accepted standard name in the World Register of Marine Species (WoRMS).
- All supplied coordinates are given in WGS84 and have possible values (-90 to 90, -180 to 180), and all supplied dates are in ISO 8601 format.
- The correct units are giving for abundance and/or biomass (if these parameters are supplied).

After submitting your data, these will be integrated into a database where all these QC tests will be performed in a semi-automatic way. If problems are encountered, you will be contacted by either the corresponding NODC or VLIZ to give feedback on how to improve the data quality.

However, you can prepare your dataset and do some quality control on your own using the [online verification tools](#) developed by VLIZ:

- Taxonomic match: the [WoRMS taxon match](#) will compare your taxon list to the taxa available in the World Register of Marine Species (WoRMS). A manual is available on the page.
- LifeWatch data validation and QC services: the LifeWatch portal offers useful [web services](#) than can help you validate your data. Check the [manual](#) to see how can you:
 - Plot the coordinates on your file with the *Show on map* tool.
 - Check if the coordinates values are possible (global values, in the marine environment) with the *Check OBIS file* tool.
 - Check if the date values provided are correct with the *Check OBIS file* tool.

4 Digital Object Identifier (DOI)

A Digital Object Identifier (or DOI) is a character string used to uniquely identify an object. Metadata describing the object is stored in association with the DOI name, including an URL which leads to where the object can be found.

In science DOIs have been widely used by publishers of peer reviewed journals for over 10 years to uniquely identify a specific article. Associating DOIs to scientific publications has not only **increased the traceability** of the cited literature but also simplified the maintenance of citation indexes which serve today to assign academic credit to scientists for their work.

As is true for claims based on information from other publications, in scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited. Moreover there is growing international support for the idea that **dataset citations should also lead to academic credit**.

VLIZ can aid you in the process of assigning a DOI to your dataset. For more information, you can read the [VLIZ DOI guidelines](#) or contact data@vliz.be.