



# EMODnet



European Marine  
Observation and  
Data Network

## EMODnet Biology

**EASME Identifier**

**Start date of the project: 19/04/2017**

**EMODnet Phase III**

**D2.2: Data standardization and formatting of a subset of the data that is needed for the data products**

**Reporting period (if applicable): 19/04/2018**





## Disclaimer<sup>1</sup>

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: Data standardization and formatting of a subset of the data that is needed for the data products
WP title [ref]*	WP2: Data access to marine biological data
Task [ref]*	
Authors [affiliation]	Leen Vandepitte, Daphnis De Pooter, Thomas Lanssens, Paula Oset Garcia, Simon Claus
Dissemination level	Public
Submission date	27/04/2018
Deliverable due date	19/04/201

\*[ref] refers to the corresponding abbreviated name of the Deliverable (or WP, or Task...), if appropriate

---

<sup>1</sup> The disclaimer is needed when the document is published

# Contents

<b>1 Overview .....</b>	<b>4</b>
1.1 Institute of Oceanography and Fisheries IOF.....	6
1.2 Institute of Agricultural and Fisheries research (ILVO): .....	6
1.3 Centre for Environment, Fisheries and Aquaculture Science (CEFAS) .....	7
1.4 Marine Biological Association of the UK (MBA).....	7
1.5 Instituto Português do Mar e da Atmosfera (IPMA).....	8
1.6 National Institute for Marine Research and Development "Grigore Antipa" (NIMRD) .....	8
1.7 Instituto Español de Oceanografía (IEO).....	9
1.8 Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER) .....	9
1.9 Royal Netherlands Institute of Sea Research (NIOZ).....	9
1.10 Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS) .....	10
1.11 Finnish Environment Institute (SYKE) .....	11
1.12 International Council for the Exploration of the Sea (ICES) .....	11
1.13 Deltares.....	11
1.14 Institute of Marine Research (IMR).....	12
1.15 Swedish Meteorological and Hydrological Institute (SMHI) .....	12
1.16 The Sir Alister Hardy Foundation for Ocean Science (SAHFOS).....	12
1.17 Flanders Marine Institute - VLIZ .....	12
<b>2 Future work .....</b>	<b>Error! Bookmark not defined.</b>

# 1 Overview

Today, already 37 EMODnet Biology datasets have been delivered by the data partners, of which 7 datasets are updates of data that was already previously available in EurOBIS. These 37 datasets represent a little over 5 million distribution records (5.053.908), of which 70% (3.560.466) passes the quality control (QC) procedures.

The QC procedures are the following:

1. taxon is matched with World Register of Marine Species ([www.marinespecies.org](http://www.marinespecies.org))
2. taxon is at genus or (sub)species level
3. latitude & longitude are different from zero
4. latitude & longitude are within possible boundaries ( $-90 < \text{lat.} < +90$  &  $-180 < \text{lon.} < +180$ )

A record needs to comply to all these QC-steps to pass the quality control procedure.

Most of the records that do not pass the QC, are due to taxa that have not (yet) been matched to the World Register of Marine Species (WoRMS), the taxonomic backbone for EurOBIS. The Data Management Team (DMT) is still working with the providers to sort these out, as it mostly concerns 'dubious matches', which means that the scientific name is in use for more than one species and the authorship of the scientific name needs to be added to be able to distinguish between the different species, and to make correct assumptions when mapping these names to WoRMS. A number of the records ( $\pm 1$  million records) have not been identified to genus or species level – from the ICES datasets – which also contributes to the number of records not passing the QC (=step 2).

Other issues with the data are linked to the use of non DarwinCore parameters, which may need more standardization by using BODC vocabularies. The DMT is committed to helping the providers with this. This task has largely been delayed due to time constraints, as too many datasets being delivered too close to the deadline date. These remaining issues will be picked up in the coming months and the data providers are expected to help the DMT with this.

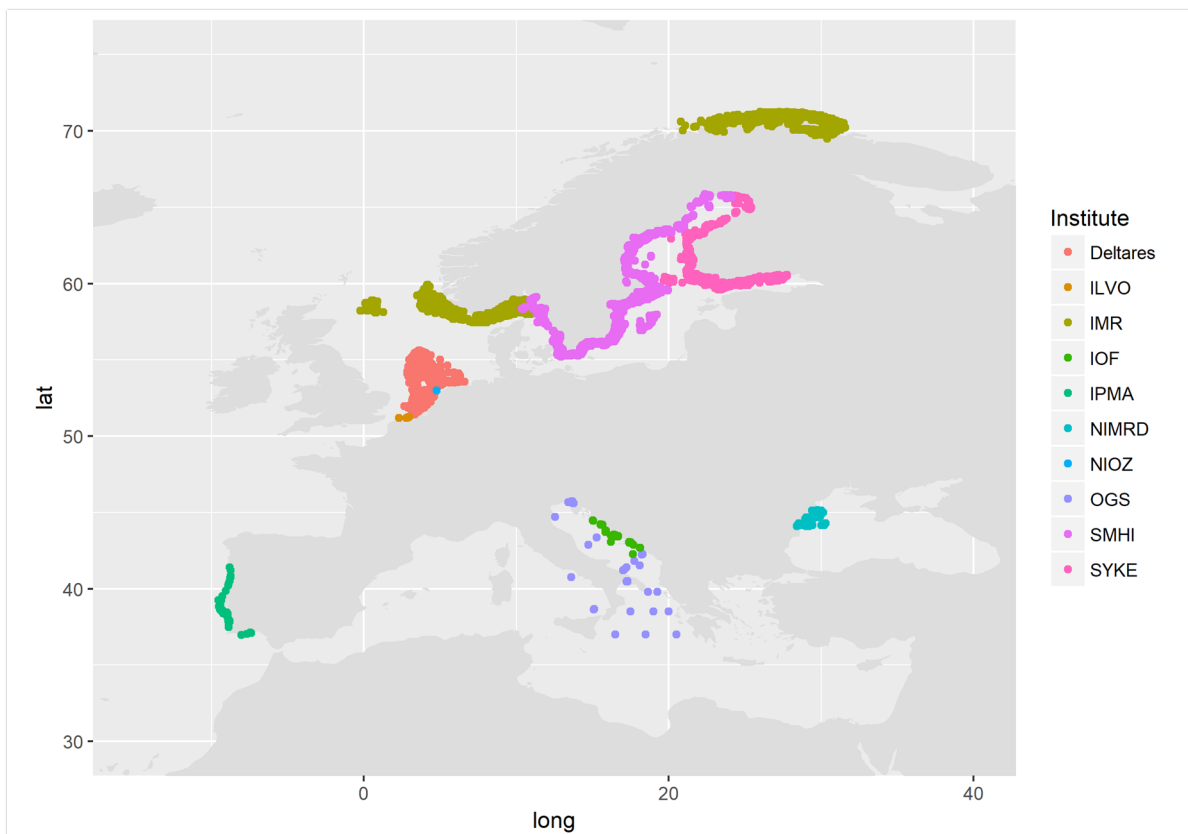
In some cases, there are structural issues with the data, which still need to be resolved. In these datasets, the data format does not yet fully comply with the OBIS format guidelines. This will also be dealt with in the coming months, and the data providers are expected to assist in this and to fully comply with the needed data format for EurOBIS and EMODnet Biology.

Based on the inventory D2.1, all datasets from WP2 have been subdivided into 4 categories:

- (1) Only benthic data
- (2) Benthic data in combination with non-benthic data
- (3) Non-benthic data
- (4) Datasets that are already in the system, but need an update

We have now 20 out of the 35 identified priority 1 datasets, being published (= 57%), 9 out of 19 priority 2 datasets published (=47%) and 8 out 54 priority 3 datasets. (=14%). In total we have 35% of the datasets been made available.

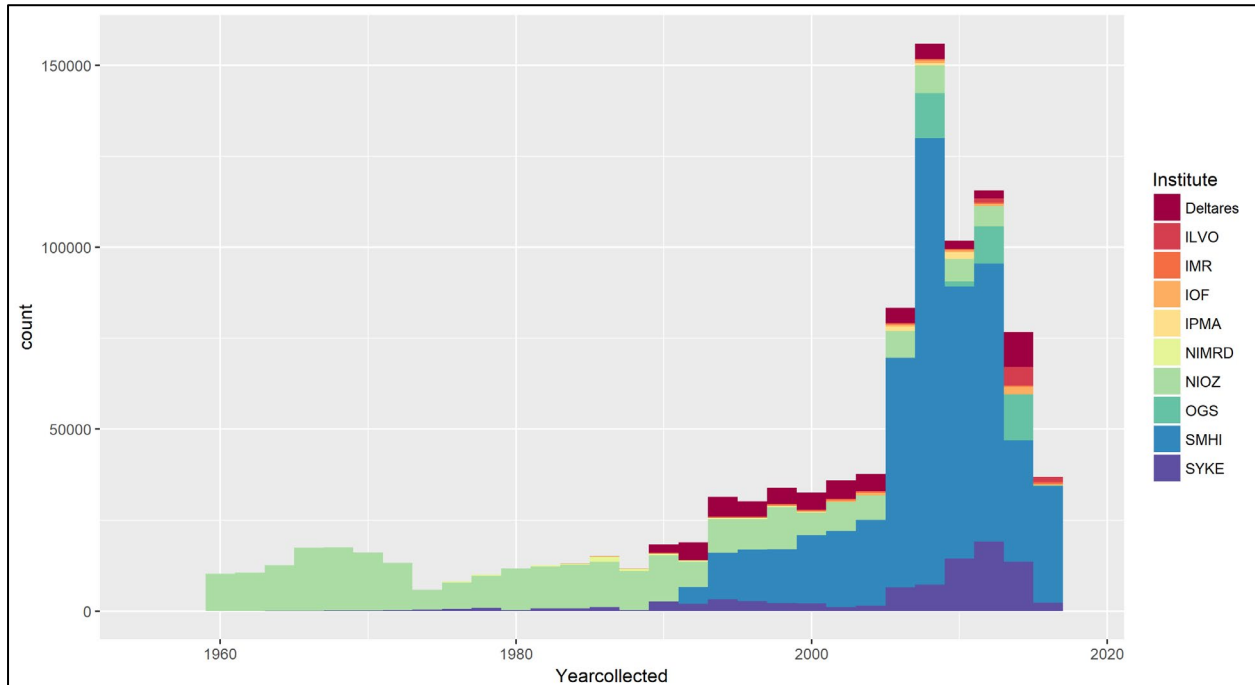
The majority of the datasets (33) have been delivered through IPT, the preferred way of data delivery to EurOBIS.



Map showing the location of the distribution records that have been provided by the EMODnet data partners – through 30 datasets - and that are already available through EurOBIS and the EMODnet Biology Data Portal.

#### Overview of the time range of the delivered data, subdivided per partner

In the past year, an additional 97 datasets have been made available through EurOBIS, as part of the daily activities of the EurOBIS Data Management Team. Although these datasets are not delivered through EMODnet Biology, they do make a valuable contribution to the project. It concerns 1.5 million distribution records, of which 87% (1.3 million) passes the QC procedures.



Overview of the data that have become accessible through the Portal, per partner

## 1.1 Institute of Oceanography and Fisheries IOF

Harvested

- Phytoplankton (Priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5718>

Not yet harvested

- Benthos (Priority 1):

Metadata information still missing

- MEDITS-Croatia: Demersal species from the MEDITS (Mediterranean International Trawl Survey) project in the Croatian territorial waters (Priority 2):  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5771>

## 1.2 Institute of Agricultural and Fisheries research (ILVO):

Harvested

- Subtidal macrobenthos monitoring in function of a foreshore suppletion at the Belgian coast, period 2013-2016 (Priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5701>
- Subtidal hyperbenthos monitoring in function of a foreshore suppletion at the Belgian coast, period 2013-2016 (Priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5708>
- Subtidal epibenthos and demersal fish monitoring in function of a foreshore suppletion at the Belgian coast, period 2013-2016 (Priority 2)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5712>

### 1.3 Centre for Environment, Fisheries and Aquaculture Science (CEFAS)

None of the datasets were provided. Datasets are to be provided to EMODnet Biology through MBA. However, MBA was time restricted because of internal restructuring.

### 1.4 Marine Biological Association of the UK (MBA)

MBA was time restricted because of internal restructuring. Datasets have been made available on their IPT but none of them was suitable for a harvest as they do not yet formatted according to the OBIS guidelines. A discussion regarding the granularity of datasets is ongoing (for example the MSFD dataset consists of 106 individual contributions. Discussion is whether these contributions should be provided as separate datasets.)

Harvested

- 2005-Ongoing UK MarLIN Shore Thing timed search results (priority 3)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=1884>

Not yet harvested

- 2010-ongoing UK Offshore Marine Conservation Zone (MCZ) Survey Data  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5670>
- 1817-1983 Paul F. Clark North East Atlantic Crab Atlas  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5685>
- 2005-Ongoing UK MarLIN Shore Thing timed search results  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=1884>
- DASSH Data Archive Centre expert sighting records  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=1885>

- DASSH Data Archive Centre Academic surveys  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=1890>
- DASSH Data Archive Centre volunteer sightings records  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=1891>
- 1844-2016, The Marine Biological Association, Verified Marine records from Indicia-based surveys, United Kingdom.  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5656>
- 1800-2016 Department for Environment Food & Rural Affairs (Defra), Marine Strategy Framework Directive (MSFD) Collation of invasive non-indigenous species data United Kingdom.  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5666>
- 1844-1995 Ivor Rees North Wales Marine Fauna Ad-hoc sightings shore and ship-based surveys.  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5688>
- 1800-2010 National Marine Aquarium (NMA) United Kingdom Marine Fish Recording Scheme.  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5693>

## 1.5 Instituto Português do Mar e da Atmosfera (IPMA)

### Harvested

- Benthic fauna collected in the Arrábida Marine Protected Area (SW Portugal) from 2007 to 2009 (Priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5780>
- Benthic macrofauna of the Ericeira coast (central Portugal) collected in May 2001 (Priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5781>
- Macrozoobenthos of marine waters in mainland Portugal collected in March and September 2010 (Priority 1).  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5782>
- Benthic fauna of the Southwest Alentejo and Vicentine Coast Natural Park (SW Portugal) collected in August 2011 (Priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5783>

## 1.6 National Institute for Marine Research and Development "Grigore Antipa" (NIMRD)

### Harvested

- Macrozoobenthos collected from the longitudinal profiles in the Romanian marine waters between 1986-1990 (Priority 1):  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5650>



- Macrozoobenthos data collected in the East Constanta sector of the Romanian marine waters between 1977 and 1999 (Priority 1):  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5661>
- Macrozoobenthos data collected in the Northern part of the Romanian littoral (Danube mouths) between 1977 and 1999 (Priority 1):  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5663>

## 1.7 Instituto Español de Oceanografía (IEO)

Not harvested

- MEDITS-Spain: Demersal and mega-benthic species from the MEDITS (Mediterranean International Trawl Survey) project on the Spanish continental shelf between 1994 and 2010 (Priority 1).  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5760>

## 1.8 Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)

Not harvested

- RSL: Lagoon Monitoring Network of Languedoc-Roussillon (Priority 2):  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=4411>

## 1.9 Royal Netherlands Institute of Sea Research (NIOZ)

The data provider at NIOZ does not have direct access to the datasets listed in WP2. Datasets are maintained at a different department. Discussions are ongoing in NIOZ internally on whether the data can be made available for EMODnet Biology. We harvested the one dataset which has been available on the NIOZ IPT for 3 years now. However this dataset is not formatted according to OBIS guidelines and has missing data.

Not Harvested

- NIOZ Eems Dollard Macrozoobenthos (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5762>
- NIOZ Epibenthos cruises 1974 (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5764>

- NIOZ Wadden Sea Macrozoobenthos Balgzand 1971-present (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5767>
- NIOZ Western Wadden Sea subtidal macrozoobenthos survey (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5769>

Harvested

- NIOZ Kom Fyke Mokbaai (priority 3)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5776>

## 1.10 Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS)

Harvested

- Microphytobenthos monitoring in the Mar Piccolo of Taranto. 2013-2014 (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5735>
- Microphytobenthos monitoring in the Trieste harbour, North Adriatic Sea (Port Authority) in 2015 (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5737>
- Microphytobenthos in mussel farms - North Adriatic Sea (SosteMiTS Project) 2008-2009 (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5740>
- Microphytobenthos in Panarea Island. ECO2 Project (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5743>
- Phytoplankton monitoring in the Trieste harbour, North Adriatic Sea (Port Authority), March 2015 (priority 2)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5736>
- Phyto- and microzooplankton in mussel farms - North Adriatic Sea (SosteMiTS Project) 2008-2009 (priority 2)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5739>
- Microzooplankton monitoring in the Mar Piccolo of Taranto. 2013-2014 (priority 2)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5734>
- Phytoplankton collected in the Po River Delta - North Adriatic Sea (RITMARE Project) 2013-2014 (priority 3)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5747>
- Phytoplankton in the South Adriatic Sea (MEDGES Project) 2013 (priority 3)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5748>
- Phytoplankton monitoring in Grado, GO - North Adriatic Sea. 2016 (priority 3)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5744>
- Phytoplankton in the South Adriatic Sea (ADREX Experiment) 2014 (priority 3)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5750>
- Phytoplankton monitoring in the Trieste harbour - North Adriatic Sea (FERRIERA). 2008-2014.  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5778>

Not Harvested

- Microphytobenthos collected in the Po River Delta - North Adriatic Sea (RITMARE Project) 2013-2014 (priority 1)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5752>

## 1.11 Finnish Environment Institute (SYKE)

Harvested

- Finnish Baltic Sea benthic monitoring, POHJE database  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5725>

## 1.12 International Council for the Exploration of the Sea (ICES)

Harvested

- ICES Phytobenthos community dataset  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5755>
- ICES Phytoplankton Community dataset  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=4424>
- ICES Zoobenthos Community dataset  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=4425>
- ICES Zooplankton Community dataset  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=4431>

Not Harvested

- Vulnerable Marine Ecosystems  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5757>

## 1.13 Deltares

Harvested

- Benthos collected in the Dutch North Sea between 1991 and 2015  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5759>

## 1.14 Institute of Marine Research (IMR)

### Harvested

- Northern shrimp (*Pandalus borealis*) in the Norwegian Deep and Skagerrak, 1984-2017  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5729>
- Red king crab survey data from Finnmark Northern Norway in the period 1994 -2016  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5732>

### Not Harvested

- *Lophelia pertusa* reefs in Norwegian seawaters  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5730>

## 1.15 Swedish Meteorological and Hydrological Institute (SMHI)

### Harvested

- SHARK - National Epibenthos monitoring in Sweden since 1992  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=2455>
- SHARK - Regional monitoring and monitoring projects of Epibenthos in Sweden since 1994.  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=5727>

## 1.16 The Sir Alister Hardy Foundation for Ocean Science (SAHFOS)

### Harvested

- Continuous Plankton Recorder (Phytoplankton)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=785>
- Continuous Plankton Recorder (Zooplankton)  
<http://www.emodnet-biology.eu/data-catalog?module=dataset&dasid=787>

## 1.17 Flanders Marine Institute - VLIZ

In the past year, an additional 97 datasets have been made available through EurOBIS – the European Ocean Biogeographic Information System -, as part of the daily activities of the EurOBIS Data Management Team. Although these datasets are not delivered through EMODnet Biology, they do make a valuable contribution to the project. It concerns 1.5 million distribution records, of which 87% (1.3 million) passes the Quality Control procedures.