

## **Biology Preparatory Action of EMODnet: draft final report**

### **1. Introduction & Objectives**

### **2. Description & content of the system**

#### 2.1. The biological data portal: System & functionalities

- 2.1.1. Data Portal System
- 2.1.2. Portal functionalities

#### 2.2. Data content

- 2.2.1. Methodology (Data management including standardization & quality control)
- 2.2.2. Inventory of available data and gap analysis
- 2.2.3. Inventory of national monitoring programs per country
- 2.2.4. Overview of data per species group

### **3. Analysis & Lessons learned**

- 3.1. Main barriers to the provision of data
- 3.2. Challenges to rendering data interoperable
- 3.3. Challenges to producing contiguous data
- 3.4. Fitness for purpose (measuring ecosystem health)
- 3.5. Improving accuracy, precision and coverage
- 3.6. Performance of portal technology

### **4. Monitoring effectiveness of portal**

- 4.1. Intensity of use
- 4.2. Possible improvements on ease of use

### **5. Recommendations for the overall EMODnet**

- 5.1. Sustainability
- 5.2. The model for governance by actors in the system
- 5.3. Availability of standard procedures facilitating data flow
- 5.4. Future activities for the biological project

## **1. Introduction & objectives**

The Final Report describes the activities from month 1 to month 24 of the Lot N°4 – Biology (SI2.531562) of the Service Contract No MARE/2008/03 on the Preparatory Actions for European Marine Observation and Data Network. This report will list the activities carried out, challenges faced, lessons learned, an analysis of performance and recommendations for the future.

### **Background**

Marine biological data are often the result of projects with a limited temporal and spatial cover. Taken in isolation, datasets resulting from these projects are only of limited use in the interpretation of large-scale phenomena. Individual studies are restricted in the amount of data they can generate; but by combining the results from many studies, massive databases can be created that make analyses on a much-enhanced scale possible. Such data have never been of greater importance for Europe, considering the European Union's ambitious Marine Strategy Framework Directive to protect more effectively the marine environment across Europe. The Marine Strategy Framework Directive (MSFD) states that, by 2012, Member States shall make an initial assessment of their marine waters and marine biodiversity, taking account of existing data where available. A large number of marine biological data are already assembled and archived in large data management systems located across the EU but an integrated and coordinated approach is still lacking. By taking into account this fragmentation of systems and data networks, hampering easy access to marine biological data in Europe, EMODnet Biology built its project and data system.

The overall objective of the biological project was to assemble fragmented and inaccessible marine data into interoperable, publicly available data streams. By building the marine biological data portal we define appropriate processes and best technology for of a final operational European Marine Observation and Data Network as well as provide first components of a final system.

### **Objectives**

The main objective is the development of an online marine biological data portal allowing the access and download of marine biological data across Europe. Other objectives of the biology preparatory action of EMODnet are:

- Complete the inventory of existing holdings of marine data in collaboration with the consortium partners, representing national and regional marine data centres, such as MarBEF, SeaDataNet, ICES, WCD-MARE/PANGAEA, GBIF and OBIS
- Performance of gap analyses to determine the shortcomings in data quality (accuracy and precision) and geographical and taxonomical coverage. Expert members of the consortium will review data and report on this topic in the final report
- Propose a strategy plan on the sustainability of the EMODnet biological portal, which should assist in implementing collaboration and governance arrangements to ensure long term investments.

## 2. Description & content of the system

### 2.1. The biological data portal: System & functionalities

#### 2.1.1. Data Portal System

The biological data portal aims to visualize and distribute fragmented marine biological data for complete maritime basins. In order to be accessible both for specialists and for non specialist, the system is developed to be intuitive and easy to use. The architecture of the system, partly based on the European Ocean Biogeographic Information System (EurOBIS), is developed to meet the final objective of EMODnet - that is to become an integrated and inter-operable network of systems of European marine observations and data communications.

#### Network of distributed data systems

The biological data portal architecture (fig 1) allows storing and integrating marine biological data and species observations in different ways. Depending on the needs and technical background of the local data providers, the EMODnet biological data portal can function as a warehouse and archive of marine biological data where data providers can upload or submit biological data to or data providers can make use of a distributed database system. This leaves the maintenance and update of the databases in the hands of their owners and developers. The distributed system makes use of the Distributed Generic Information Retrieval (DiGIR) protocol and is fully platform independent. Also other web services can be supported by the EMODnet Biological data portal (as is the case with the ICES data centre).

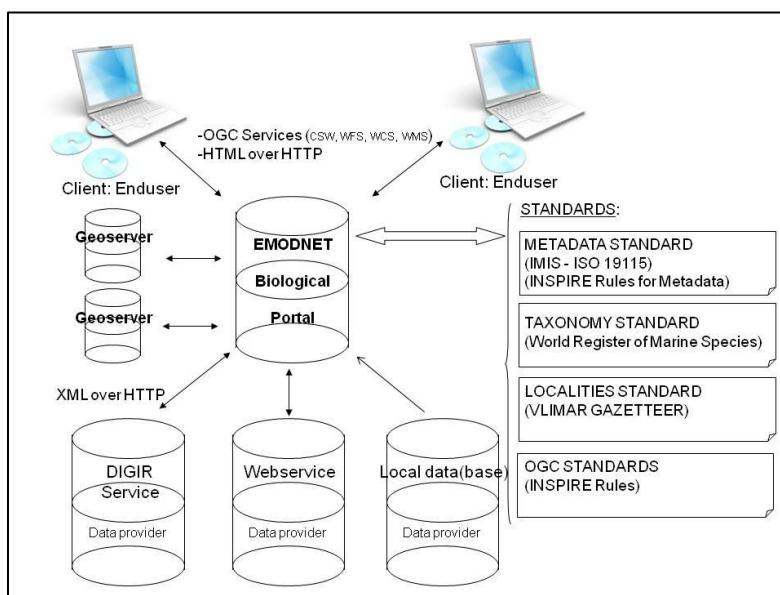


Figure 1: Architecture EMODnet marine biological data portal

### OGC Complaint data system

The biological data portal is also fully OGC compliant, allowing compatibility with OGC compliant data servers (like Geoserver). This allows integrating and visualizing species observations served from different data providers and databases, and OGC compliant geographic maps. Through OGC compliancy, the marine biological data portal can visualize also data products developed in the other lots of EMODnet.

### Standardized data system

In order to integrate marine biological observation data, taxonomic standardization is a key element. The taxonomic standardization allows to detect and filter out spelling mistakes of species names occurring in the contributing datasets, solve issues or ambiguities related to the nomenclature of a species and to search and browse data for aggregated groups. Therefore all species names are matched with the World Register of Marine Species (WoRMS) and its European component, the European Register of Marine Species (ERMS). WoRMS is currently the most authoritative list of names of all marine species globally, ever published. WoRMS is a contribution to the Catalogue of Life, the Encyclopedia of Life, the Global Biodiversity Information Facility and the Census of Marine Life.

The data scheme used to integrate the species observation data is the data scheme of EurOBIS and is based on the Darwin Core standard that is used by the Global Biodiversity Information Facility (GBIF) and the Ocean Biogeographic Information System (OBIS). This biogeographic data scheme is able to handle information and data of annual, seasonal, and spatial distribution of species composition, abundance and biomass in the water column and on the sea-bed. The EurOBIS data scheme is also OGC compliant. Metadata are standardized using the currently most accepted metadata standard for geographic information ISO 19115. Relevant dictionaries developed under the Seadatanet project were used in the metadata standards.

### System integrating data with different levels of resolution

The EMODnet marine biological data portal allows integrating data and information on different levels of precision or resolution. The portal can integrate data on three different levels. First of all the metadata can be submitted to the dataportal. The information on when, where, what, why and how biological data was collected allows to have an idea of the availability of the data, without having direct access to the raw data. By submitting metadata to the marine biological dataportal, the dataset description will become part of the EMODnet Bio data catalog. Although there is no direct access to the data through the dataportal, the user can have an idea of the type and spatio-temporal cover of the dataset. He could possibly request direct access to the data by contacting the data provider. The second level of resolution is the aggregated or contiguous data. These data can be uploaded as geographic data maps (OGC compliant data products) or as derived parameter values (seasonal or monthly means, annual anomalies, abundances of higher taxonomic groups....). Visualizing aggregated biological data products allows a higher precision and more information of the data but will not provide the raw data to the user. The highest level of resolution is the raw

monitoring data. These data have exact geographic coordinates and an exact temporal indication, possible with abundance and biomass information. These data can be freely downloaded in the highest precision possible.

### 2.1.2. Portal Functionalities

Main functionalities of the EMODnet Biological Data Portal include a metadata catalog and the querying, viewing, downloading and submitting of the data. There are also online instructions, a monitoring and a feedback mechanism available. The data portal, operational 24 hours a day, 7 days a week is available from the project web-site at <http://bio.EMODnet.eu>, providing also information on the progress of the project, partner information, documents, minutes and reports.

#### Viewing a catalogue of the data available

The metadata catalogue provides an inventory of all available datasets. This catalogue is ISO19115 compliant and can contain general information on the dataset (type of dataset -monitoring or research-, the access constraints, the version, the keywords or citations of the dataset, a general description or abstract), information on the geographic, taxonomic and temporal cover of the dataset, parameters collected, who collected the data, point of contact and information on the precision and resolution of the data. If available, information on the sampling methodology and a link to the online dataset is also provided. The metadata provides in most cases information on the precision of the data and how it has been processed. Users can search the data catalogue for a dataset name or use the advanced search option, allowing to search on multiple criteria, including the availability of the dataset in the Biological EMODnet data portal. If the raw data of the dataset are available through the data portal, a direct (deep) link in the metadata description will allow direct access to the data in the EMODnet portal.

The screenshot shows the 'Data catalog' search interface of the EMODnet Pilot Portal For Biology. The header includes the EMODnet logo (European Marine Observation and Data Network) and the title 'Pilot Portal For Biology Data Discovery and Access Service'. A navigation bar contains buttons for Home, Partners, Workshops, Documents, Data portal, Data catalog, and Submit dataset. The main content area is titled 'Data catalog' and shows 'Datasets (442)'. It features a search bar with the placeholder 'Search in all fields: [ ] Title or word in abstract'. Below the search bar are dropdown menus for 'Theme:' and 'Type:'. A series of input fields for advanced search criteria include: Title, Institute, Person, Word in abstract, Begin date (with a 'before' dropdown and a date field [MM/YYYY]), End date (with a 'before' dropdown and a date field [MM/YYYY]), Archived (checkbox), Thesaurus term, Taxonomic term, Other term, and Geographical term, each followed by a '[ search ]' button. An 'Availability:' dropdown is at the bottom. A 'search' button is located at the bottom left, and a 'show full list' link is at the bottom center. There are also 'less options' and 'help' links.

Figure 2: View and selection of the data catalog

## Data querying

Users can search and select the data portal (fig 3) for biological data through different data modules. The different modules represent different levels of precision of the data and include:

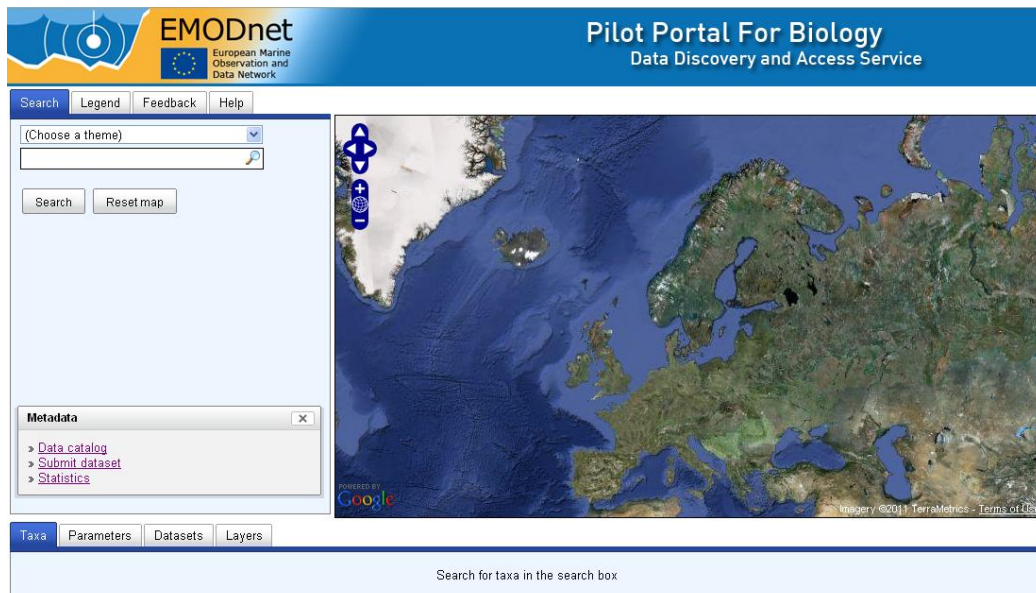


Figure 3: Entry page of the biology data portal

- ✓ **Taxa:** List of species observations in EurOBIS. For each species or taxon, the scientific name, authority & year of publication, the common name (if available), its unique Aphia-ID and the number of records are listed. The user can plot the data, ask for a table containing extra information on the species observation (lat, long, time, number of observations, citation, depth, sex and link to metadata) or go to the corresponding taxon page in the World Register of Marine Species (WoRMS).
- ✓ **Parameters:** information on aggregated biological parameters. This aggregated data can include yearly or monthly means, annual anomaly's... of abundance or biomass data of specific species or species groups. The data module is operational, but at the moment there is limited parameter information available.
- ✓ **Datasets:** List of datasets containing the specified taxonomic information. For each dataset, the full dataset name and the number of available records within that dataset is listed. The user can directly link to the corresponding metadata page, where all relevant metadata of a dataset is listed (citation, responsible persons and institute, abstract, measured parameters, geographical and temporal scope, ...). If the data are freely available, all records from a dataset can either be plotted on a map or be listed in a table for further download.
- ✓ **Layers:** List of all OGC compliant GIS layers corresponding to the search criteria. These are pre-defined and described GIS layers, containing derived or aggregated data. They can be

seen as data products: the raw data has been aggregated and recalculated to something that is easily understood and with high relevance to scientists or policy makers.

The data selection starts with the taxonomic query (fig 4). After selection of a species group (phytoplankton, zooplankton, angiosperms, macro-algae, benthos, birds, sea mammals, reptiles, fish and pigments) or using a specific species name or dataset name, the user gets data results from four different modules. The different tabs list the number of results. For example selecting the zooplankton species group returns currently 115 taxa, 2 parameters, 56 datasets and 1 aggregated data layer. For the taxa and datasets module, the number of records is listed indicating respectively the number of observations of the species and the number of records in the dataset.

ScientificName	Authority	Common name	AphiaID	RecordCount	Display
Acartia longiremis	(Liljeborg, 1853)		104257	13,151	
Acartia negligens	Dana, 1849		104259	1	
Aetideus armatus	(Boeck, 1872)		104275	1,033	
Anomalocera patersoni	Templeton, 1837		104722	690	
Branchiostoma lanceolatum	(Pallas, 1774)	Lancelet,	104906	908	

Fig 4: Start the data selection with the search for data on species group

After the initial taxonomic query, different options – represented by different icons can be selected:

- the selected data will be plotted on the portal (see further)
- the raw observation data will be listed in a *data table* for further treatment
- the selected observation data can further can be filtered, by temporal, spatial (bounding box coordinates) or min, max depth parameters
- the selected data can be downloaded (see further)
- link to the species in the World Register of Marine Species





link to the Biological EMODnet data catalogue (see previous section)



close the additional *data table*

## Data visualization

The species observations data can be plotted on the data portal (fig 5). The data visualization follows main INSPIRE implementing rules for data visualisation. Different layers can be added, selected and deselected on the data portal. There is a zooming and panning option. Clicking on a data point (species observation) allows the user to retrieve the attributes (map features) of the observation. Attributes are the catalogue number (unique number of species observation); scientific name, year, month, day, latitude and longitude, minimum and maximum depth of observation, sex, observed individual count, institute collecting the data, the unique LifeScienceID (standardized quality controlled scientific name) and a link to the metadata from the observation.

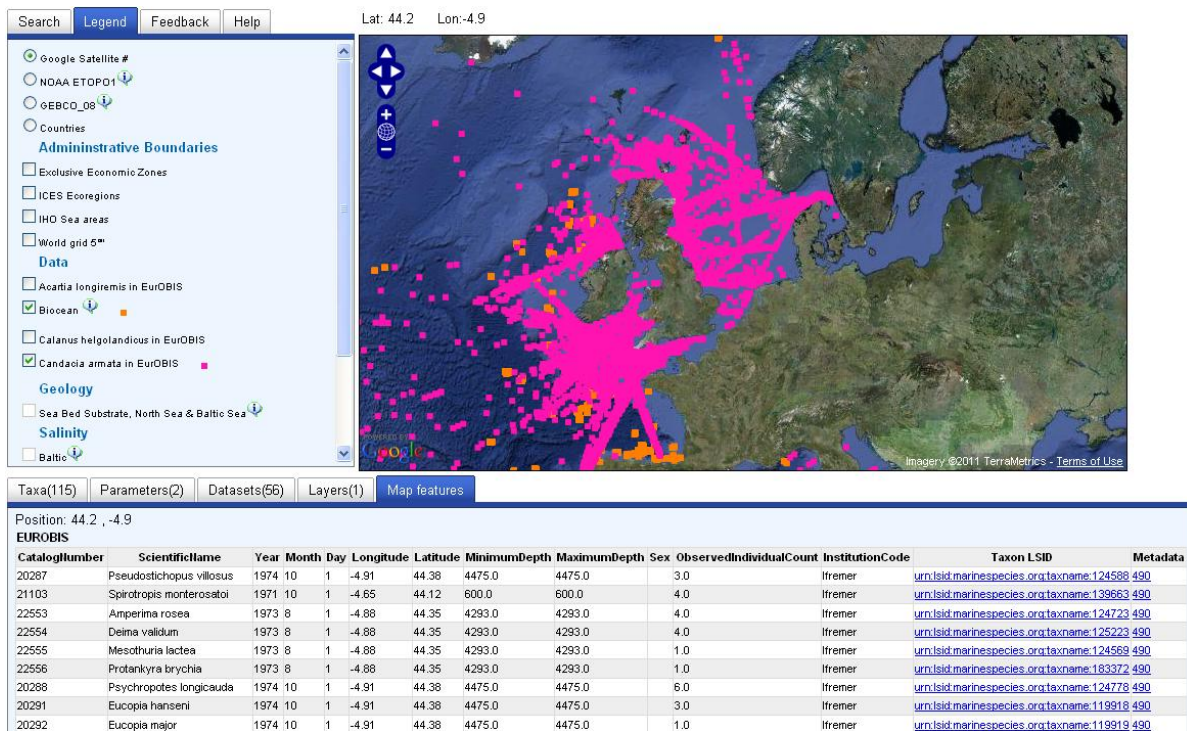


Fig 5: Data visualization and table of attributes of a species observation

The GIS platform of the portal, based on the open source module Open Layers, allows browsing and manipulating of geographic maps through different web browsers. The biological data can be plotted using several background views: a Google satellite view, the ETOPO1, a 1 arc-minute global relief model of Earth's surface that integrates land topography and ocean bathymetry or the GEBCO08 Grid, a global 30 arc-second grid bathymetric maps. Both bathymetric maps are available on the portal through the OGC compliant Web Mapping Service (WMS). Communications between the biological Geoserver and other external GIS servers run through Web Mapping Services (WMS) and Web Feature Services (WFS). These protocols allows that external GIS layers, for example with abiotic data like sea bed substrates, salinity or temperature provided through other portals, can be accessed through the biological data portal interface. Also marine administrative boundaries like the Exclusive



Economic Zones, the ICES Ecoregions or the International Hydrographic Organization Sea areas can be added as a baseline view.

### **Understanding the precision of the data and how it has been processed**

The precision of the data is assessed by three main criteria: the taxonomic precision, the temporal precision and the geographic precision. The taxonomic precision ranges from (sub)species level to kingdom level. Through matching with the taxonomic register, the level of taxonomic precision can be determined. Also information on the individual specimen level can be determined by matching to the World Register of Marine Species. The temporal precision, ranges from data with information on the year, the season, the month, the day or the time of observation. The geographic precision varies from the exact coordinates, to information on locations (for example observed in the North Sea). When only information on the location is available in the literature, the geographic name is translated into a geographic centroid coordinate with a certain precision, corresponding to the radius of the corresponding polygon. More information on the taxonomic, temporal and geographic precision is described under the methodology section. The data with different precision are processed in the EMODnet marine biological data portal, in different ways as discussed in the data system section. Raw observation data (with high precision) aggregated data and metadata can all be made accessible through the portal.

### **Data downloading**

The data can easily be downloaded from the biological data portal. The raw monitoring data or observation data can be downloaded after selection of a specific taxa or a specific dataset, possibly with extra spatial, temporal or depth selections. Since all observation data stored in the EurOBIS database is freely available, users can download the information after supplying some information (name, organization, email, country and purpose of download). This information is only be used to monitor the usage and the downloading of the data. Users can download the data as tab delimited files which can be used in most data handling and statistical programs. For smaller downloads (<4000 records), the data can also be downloaded as an excel file.

The GIS layers, or aggregated data can be accessed through the Web Mapping Service (WMS) available at <http://geo.vliz.be/geoserver/web/>. An automated download interface through the portal is not yet available (To be realized before May). Before download, users need to agree with the terms of use which are currently formulated as follows:

“If data are extracted from the EMODnet Data Portal for secondary analysis resulting in a publication, the appropriate source should be cited:

- Online raw data (background data) should be cited as follows: EurOBIS Data. European node of the Ocean Biogeographic Information System. Available online at <http://www.eurobis.org> Consulted on 2011-03-12.
- If any individual datasources of EurOBIS constitutes a significant proportion of the records used in the secondary analysis (e.g., more than 10% of the data are derived from this source), the individual data source should also be cited.

- If any individual datasource of EurOBIS constitutes a substantial proportion of the records used in the secondary analysis (i.e. more than 25% of the data are derived from this source, or the data are essential to arrive at the conclusion of the analysis), the manager/custodian of this dataset should be contacted. It may be useful to contact the data source directly in case there are additional data that may strengthen the analysis or there are features of the data that are important to consider but may not have been apparent from the metadata.
- Online data products and GIS maps (foreground data) should be cited as follows: EMODnet Biological Data Products. Available online at <http://bio.EMODnet.eu> Consulted on 2011-03-12.

The data may not be redistributed without the permission of the appropriate data owners. If data are extracted from the EMODnet Data Portal for redistribution, please contact us at [bio@EMODnet.eu](mailto:bio@EMODnet.eu)”

### Help & feedback function

Feedback forms are available at the project website and at the biological data portal. Users need to provide their name, email address and their feedback. The results are discussed in the section of ‘monitoring of use’. Also the number and amount of downloads is monitored (is it?), together with number of unique visitors and the number of hits. The help tab on the data portal provides a manual describing and illustrating the usage and functionalities of the portal. However lots of effort was put in the design of an intuitive, easy of use data portal.

Finally, a help desk function is available at [bio@EMODnet.eu](mailto:bio@EMODnet.eu).

The image shows two versions of a feedback form. The left version is from the project website and features a navigation bar with 'Search', 'Legend', 'Feedback', and 'Help' tabs. The form text reads: 'The EMODnet Biological data portal is currently under testing phase. Please provide us feedback on the functionalities and usability of the portal.' It includes input fields for 'Name \*', 'E-Mail \*', and a larger text area for 'Your Feedback \*'. A 'Submit' button is at the bottom, along with a link: '» [Participate in the Online Survey](#)'. The right version is from the portal, with a dark background and white text. It has the same fields: 'Name \*', 'E-Mail \*', and 'Your Feedback \*'. A prominent red 'SEND' button is located below the text area. At the bottom, it says: 'Please provide any feedback on the project website or data portal'.

Fig 6: Feedback forms available from the portal and the project website

## **2.2. Data content**

### **2.2.1. Methodology (Data management including standardization & quality control)**

#### **General data management**

Each dataset which can possibly contribute to EMODnet is thoroughly described at VLIZ, making use of the Integrated Marine Information System (IMIS). These metadata descriptions, ISO19115 compliant, include – amongst others – information on the spatial and temporal coverage of the dataset, keywords, included taxonomic groups, data quality and user constraints as well as information on the conditions of use, the measured parameters, involved and responsible persons, how and why the data were collected and possibly a list of publications that made use of the data or are describing the data. Data providers are asked to check the online metadata of their dataset and can make changes or additions at any given time.

When data can be shared within EMODnet, there are two ways of contributing the data. When it concerns (smaller) datasets from institutes or researchers, these data can be sent directly to VLIZ (e.g. as an Excel or Access file). A copy of this dataset is then locally stored at the Marine Data Archive (MDA), to prevent corruption or loss and the data is integrated in the EurOBIS database. This is mostly the case for research (PhD) datasets for which no further data collection is planned and the dataset can thus be seen as finished. For large datasets such as monitoring data (e.g. CPR) or large data centres or institutes (e.g. ICES), a different procedure is followed. In these cases, DiGIR (Distributed Generic Information Retrieval) or other web services are set up. This implies that the data are still stored and managed at the data providing institute and a copy is stored in EurOBIS/EMODnet. Agreements are then made between the data provider and EMODnet on the update frequency. This can vary from monthly to yearly updates or a notification can be sent to EurOBIS/EMODnet when major updates have taken place at the host institute. 29 out of the 277 datasets - but accounting for the highest proportion of the data (large monitoring datasets from ICES, PANGAEA, NBN, CPR) are still stored and managed at the data providing institute but are regularly updated on EurOBIS. For the other 248 small-scale research datasets, EurOBIS can therefore more be considered as the main storage repository.

The European Ocean Biogeographic Information System is a database system developed at the Flanders Marine Institute (VLIZ) in the framework of the European Marine Biodiversity and Ecosystem Functioning Network of Excellence (MarBEF NoE) in 2004. It is a distributed system in which individual datasets go through a series of quality control procedures before being integrated into one large consolidated database. EurOBIS is available online ([www.eurobis.org](http://www.eurobis.org)), all available biogeographical data – with a focus on taxonomy, temporal and spatial distribution – can be consulted freely. EurOBIS shares its data with OBIS – the international Ocean Biogeographic Information System – which in turn shares its content with GBIF, the Global Biodiversity Information Facility.

The EurOBIS database consists of a standard list of 74 data fields, the OBIS Schema version 1.1, which is an extension of the Darwin Core 2 (<http://www.iobis.org/node/304>). The OBIS Schema is the content standard used by OBIS and is designed for marine biodiversity data, specifically to records the capture or observation of a particular species at a certain location and time. It can also be used to document specimens from museum collections and literature data. The Scheme lists 74 data fields, of which 7 are mandatory and an additional 15 are classified as highly recommended. All other data fields are optional.

### Quality control

When data can be made open-access and can be shared within EMODnet and integrated into the EurOBIS database, the datasets go through a set of quality control procedures.

### Taxonomy

All received taxon names are matched to the European Register of Marine Species (ERMS), which is included in the World Register of Marine Species (WoRMS: <http://www.marinespecies.org/>). The use of a standardized taxonomic register is imperative when integrating data from different biological datasets. It allows ruling out any spelling variation or spelling mistake and makes it possible to link synonyms to their currently accepted name. The originally delivered taxon name is always safeguarded, so data providers can keep track of their taxa and their currently accepted names. If the taxon name cannot be matched to WoRMS, or in case of doubt, the data provider is consulted and asked for feedback.

<b>1. Level: Quality control of taxonomy</b>		
Species name matches with ERMS database?		
NO	YES	
Spelling error or synonym of an existing ERMS record?		Record was used as correct
NO	YES	
Identification uncertain (cf), not identified to species level (sp.; spp.) or two different species names were given	Correction of spelling or mapping to actual ERMS name	
Taxonomy was reduced to the first common higher and certain classification level		

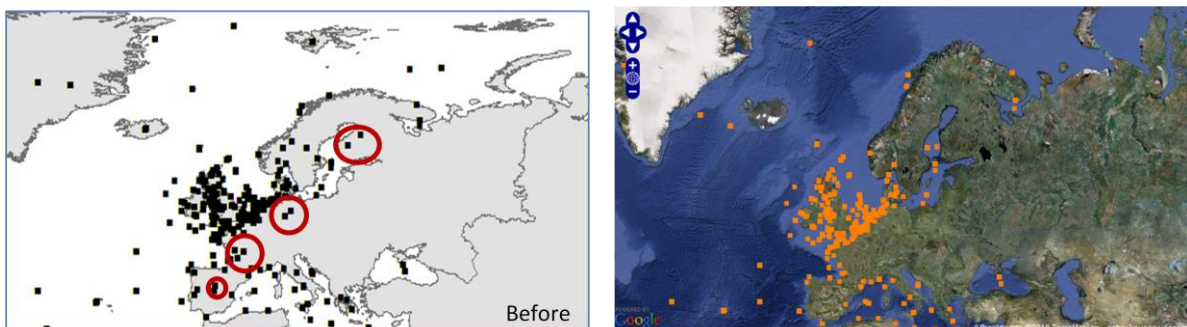
### Geography

For each dataset, all sampling locations are plotted on a map, to check for odd locations. If there is any doubt or if errors are suspected, these are communicated with the data provider so corrections can be made. During the testing and monitoring phase, distribution records without geographic coordinates were identified. After communication with the data providers, almost 40 000 distribution records were updated with correct coordinates.

<b>2. Level: Quality control of geographic information</b>			
Transformation of all given geographic information into WGS 84 decimal degrees			
Checking the correctness of georeference by plotting and comparison with meta data			
Correct location after conversion and plotting?			
NO		YES	
Latitude and Longitude were switched?			
NO		YES	
Signs for North/South/West/East were wrong?		Latitude and Longitude data were switched as correction	Record was used as correct
NO	YES		
3% of the data were delivered with missing or useless coordinates	northern Latitude = + southern Latitude = - eastern Longitude = + western Longitude = -		

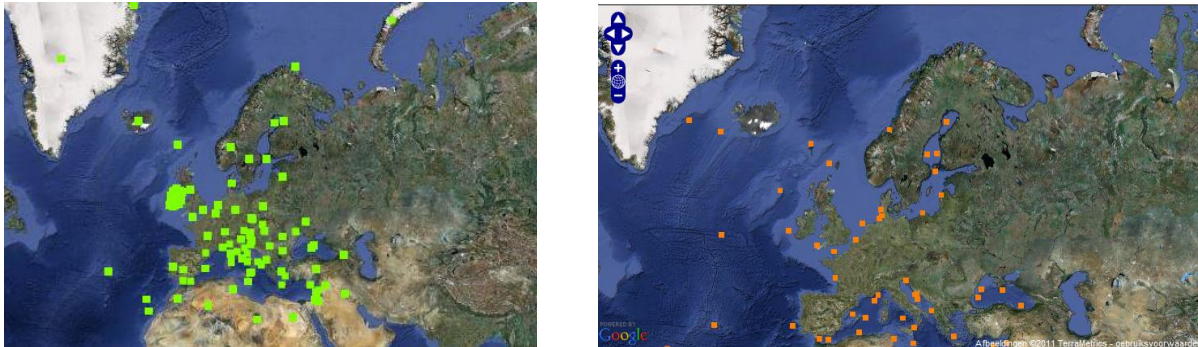
In a number of literature datasets, distributions of marine species were linked to a country instead of a sea-area. These distribution records have a lower precision than exact coordinates. As the land-based distribution names create confusion, these records are looked at in detail and adapted where relevant. So far, this has been the case for two datasets: the Taxonomic Information system for the Belgian Coastal area (Tisbe) and Algaebase. The followed methods for both datasets are briefly discussed.

For Tisbe, all land-based coordinates were filtered and all coordinates situated within reasonable distance from the coastline (20 km), were treated as 'marine'. This buffer was included, as not all coastline maps have the same resolution or precision and literature sometimes says 'Oostende' with the city coordinates when they actually want to refer to the beach. When coordinates were situated in the centre of a country, the literature source was checked, together with the actual species habitat. E.g. if a species was linked to the centre of the United Kingdom, but it was clearly a marine species, it is now linked to the Exclusive Economic Zone of the UK. The maps below show the results of the quality control procedure and adaptations.



Overview of the Tisbe dataset before (left) and after (right) the quality control actions on land-based coordinates

For Algaebase, a similar method of working was followed. Algaebase documents the appearance of a marine species as being present in a certain country, giving rise to land-based coordinates for marine species. By matching the land-information to the adjacent Exclusive Economic Zone, marine species are now appearing in the right environment.



Overview of Algaebase dataset before (left) and after (right) the QC on land-based coordinates for marine species

These methods give rise to non-exact, approximate distributions represented as point-locations. In the future, these coordinates will be made visible in a different way compared to exact coordinates, so users know they are dealing with less exact information and they can decide for themselves whether these records can be included in what they want to accomplish with the data.

### Duplicate records

To avoid possible duplication of data within EMODnet and EurOBIS and to increase the transparency of the systems, larger datasets – mostly the result of integration of several smaller datasets – are being splitted into their component datasets. This has already been done for the ICES DATRAS database and PANGAEA, avoiding that a dataset contributing to ICES DATRAS will be added again, not through ICES but through the original and local provider. A first check for possible duplication is based on the dataset title and description. When duplicate delivery is suspected, the data management team will check with the local provider to see if these data are indeed also delivered to a national or regional data centre which already contributes its data to EMODnet and EurOBIS.

Because duplicates cannot always be identified on the metadata level, a second mechanism has been developed: queries are run at regular time-intervals to identify duplicate records which were not filtered out manually by the title-description approach. These records will be retained from the portal, to avoid confusion.

A duplicate record is defined as *“an identical taxon collected at an identical location (latitude-longitude) at an identical time (year-month-day)”*.

### Sex

The OBIS Scheme can also capture the sex of an individual in a non-standardized way. To document this, a tentative list of values which should be used to indicate the sex is provided. As part of the quality control of EurOBIS, all sex-related information has been converted into the proposed values.



The development of such a standard sex-vocabulary is in progress with the scientific community. As soon as this list is available, EurOBIS will implement this.

### Life stage

Very mixed information is available in the life stage field (life stage, sex, size, combined information ...). As part of the quality control procedures of EurOBIS, information that is not related to the life stage of a specimen has been transferred to the appropriate field. In contrast to the sex field, no tentative list for life stages is available. There is not yet a way to document this in a standardized manner, making this information easy-to-use and giving it added-value for the user. OBIS has recognized this problem and will take the lead in the development of a standardized vocabulary of the different life stages. This is currently a work in progress and EurOBIS will adopt and implement the list as soon as it becomes available.

### Sample Size

When abundance data is delivered to EurOBIS / EMODnet, it is always checked whether the sample size is also documented. Without an indication of the sample size, the abundance information is not useful: without sample size, scientists cannot standardize and compare different datasets (see also: standardization of abundance data). For a number of abundance-datasets, the sample size is still missing. We are currently in the process of trying to complete this information.

### **Standardization of abundance data**

Within EMODnet / EurOBIS, 175 of all available datasets (277) contain abundance information, i.e. state how many individuals were found at a certain time on a certain location. Comparing this abundance data is however not straightforward; additional information on sample size is necessary in order to make a sensible comparison between datasets or to perform calculations on data combined from different datasets. Next to sample size, sampling equipment or gear can also provide valuable insights on the (in)comparability of samples.

By setting standards for sample sizes (e.g. m<sup>2</sup> for benthos and liter for plankton), conversion factors can be calculated and original abundances can be re-calculated to a standardized area or volume (see example table). Users will only be able to plot comparable data, e.g. only benthos data gathered by a Van Veen grab or other selected gear. To be even more complete, information on 'trapping efficiency' should be documented, but preliminary searches have already indicated that this information is very hard to find and this exercise will not be given priority.

By re-calculating original abundance values to standard sample sizes, the creation of reliable density maps becomes possible. These maps will be available through the Portal. The originally provided abundance data will also remain available through the Portal.

Dataset	Original sample size	Standard sample size	Conversion factor	Sampling gear
A	0.5 m <sup>2</sup>	m <sup>2</sup>	2	Van Veen grab
B	0.8 liter	liter	1.25	WP2 net
C	10 cm <sup>2</sup>	m <sup>2</sup>	1000	Box corer
D	...	...	...	...

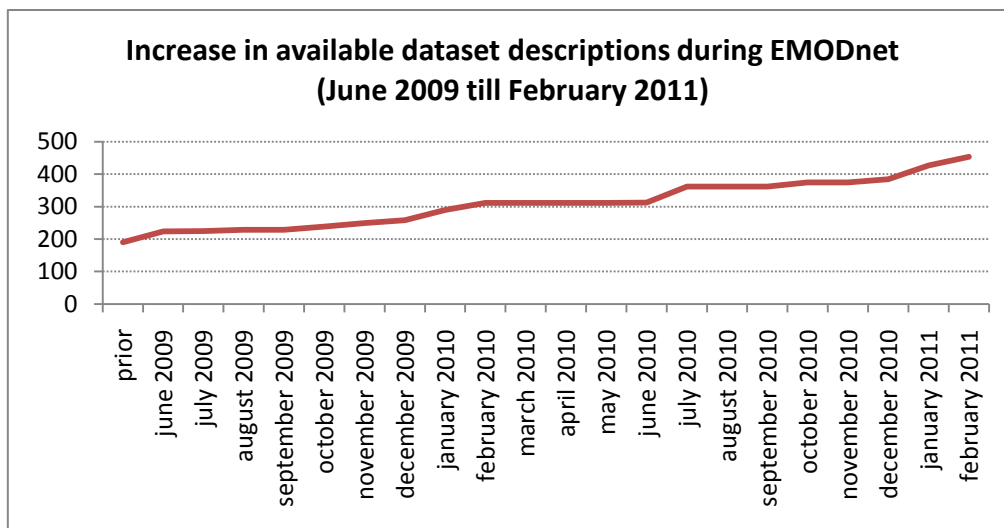
*Example extract of conversion table to re-calculate abundances to a standardized area or volume*

The documentation of the sampling size and sampling gear is a work in progress. So far, this information has already been retrieved for 139 datasets (out of 175 with abundance information).

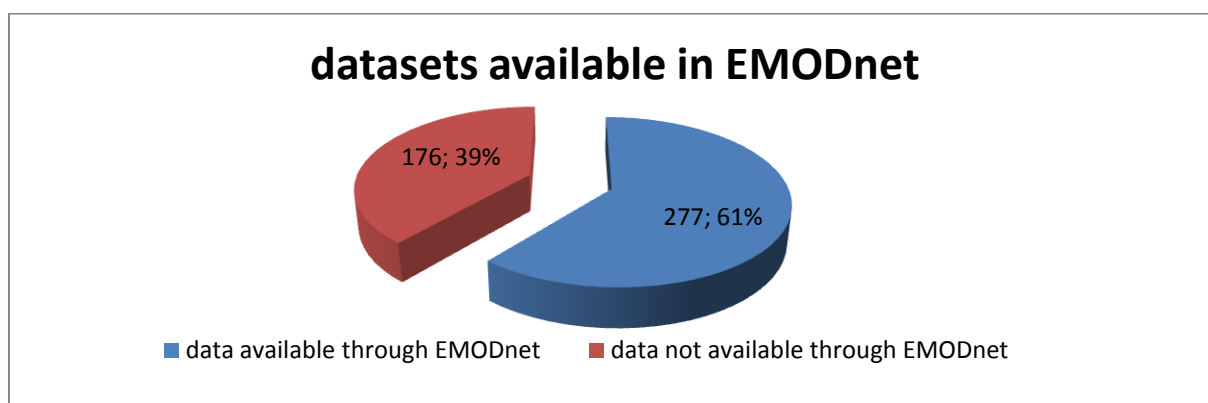
## 2.2.2. Inventory of available data and gap analysis

### Inventory of existing datasets

With the start of EMODnet, a list of known European marine datasets was compiled, based on the information available within EurOBIS (European Ocean Biogeographic Information System, [www.eurobis.org](http://www.eurobis.org)), the FP6-MarBEF NoE (Marine Biodiversity and Ecosystem Functioning Network of Excellence, [www.marbef.org](http://www.marbef.org)) and FP5-Biomare (Implementation and networking of large scale, long-term marine biodiversity research in Europe, [www.biomareweb.org](http://www.biomareweb.org)). This list was used as a baseline to send out questionnaires to the partners and partner networks of the EMODnet Biology Lot. Each partner or partner network was asked to (1) review the existing list, (2) to make additions if datasets were missing and (3) to inform us whether the data could contribute to EMODnet Biology and how this could best be arranged. Over 100 questionnaires were sent out, with an average reply of 30 % (see table below). From the replies, 262 datasets were newly described, bringing the total amount of described datasets within EMODnet Biology to 453 (<http://bio.EMODnet.eu/data-catalog>) (graph). A number of institutes have agreed to deliver the metadata of their marine datasets in the near future. New dataset descriptions are still being added to the catalog. An overview of the datasets identified during the test and monitoring phase can be found in Appendix I. 97 of these 262 described datasets were added to EurOBIS and the data transfer or quality control procedures are in progress for an additional 67 datasets. 7 datasets were made available as data product, implying that their GIS-layers are available in the Portal. The total number of datasets now available through EurOBIS is 277 (situation early March 2011). Given the rather low response to our questionnaire – on average 30% - we suspect that there is still a lot of (research) datasets known to the involved researchers that remain undocumented within EMODnet.



Increase in the available number of dataset descriptions during EMODnet. The graph indicates the number of dataset descriptions known prior to the start of EMODnet ('prior') and the growth until February 2011.



The number and percentage of datasets that are available through EurOBIS/EMODnet and those of which the metadata is available (based on available data in March 2011)

Partners	Dataset descriptions	Datasets in EurOBIS	# available records
OBIS	14	14	434 432
OBIS Seamap	19	19	1 178 132
GBIF	25	25, in progress	to be defined
ICES	11	11	6 187 964
IBSS	18	2	151 610
PANGAEA	39	39	1 713 388

OBIS = Ocean Biogeographic Information System; OBIS Seamap = Biogeographic Information System Spatial Ecological Analysis of Megavertebrate Populations; GBIF = Global Biodiversity Information Facility; ICES = International Council for the Exploration of the Sea; IBSS = Institute of Biology of the Southern Seas; PANGAEA = Publishing Network for Geoscientific and Environmental Data

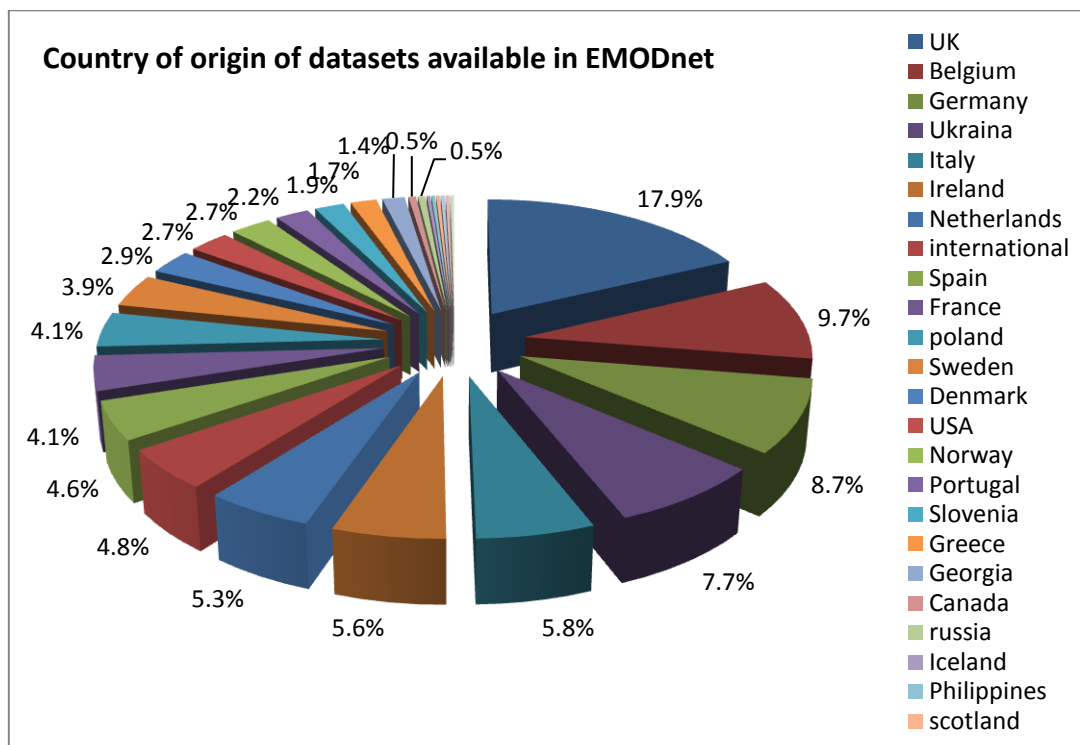
Networks	People contacted	Replies	Dataset descriptions	Datasets in EurOBIS
MARS/Seadatanet Networks	116	36 (= ± 30%)	100	6
Others (EurOBIS)	/	/	36	16

MARS = The European Network of Marine Research Institutes and Stations; SeaDataNet = Pan-European Infrastructure for Ocean and Marine Data Management; Others = other contacts established before and during EMODnet, but not related to any of the partners or networks that have been contacted

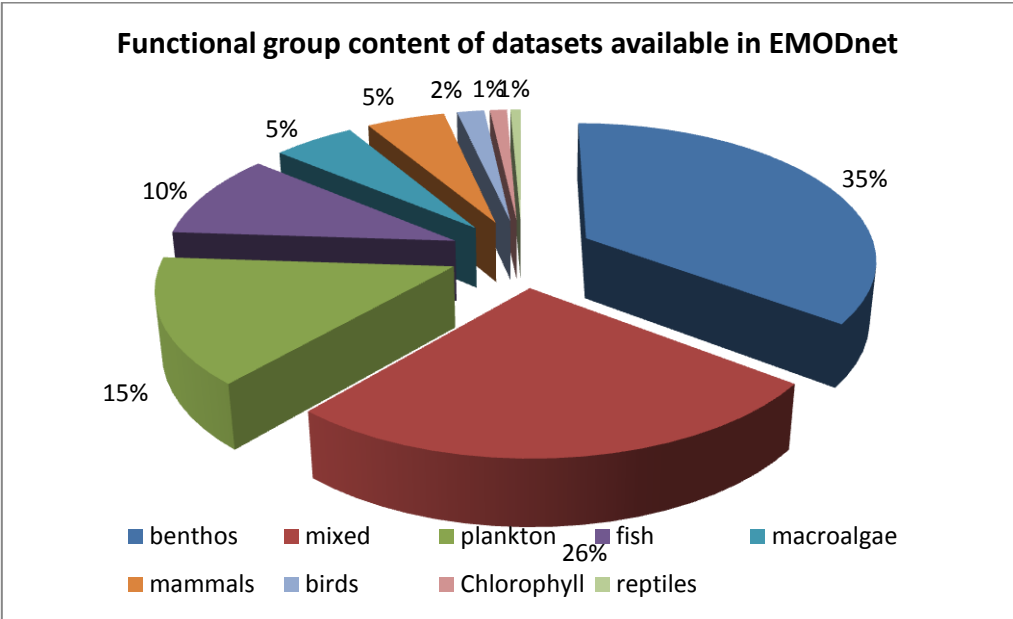
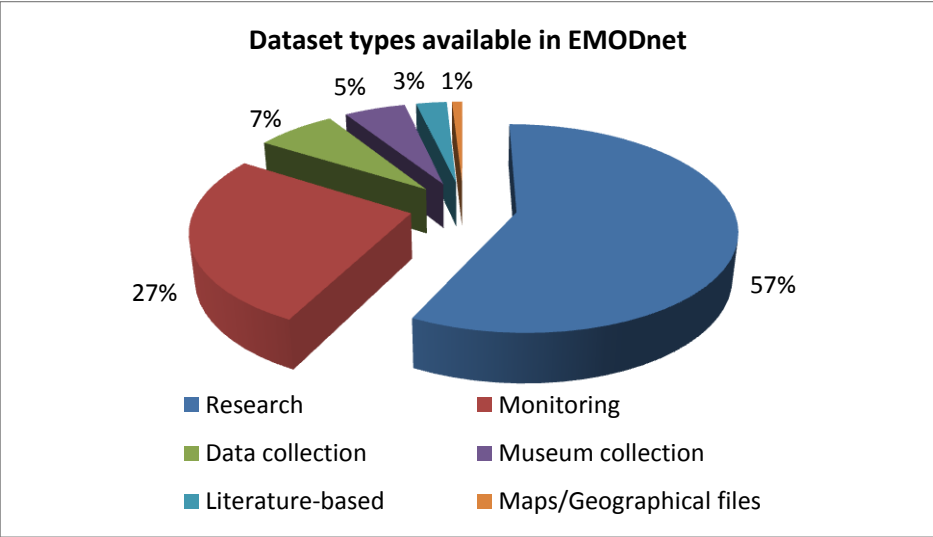
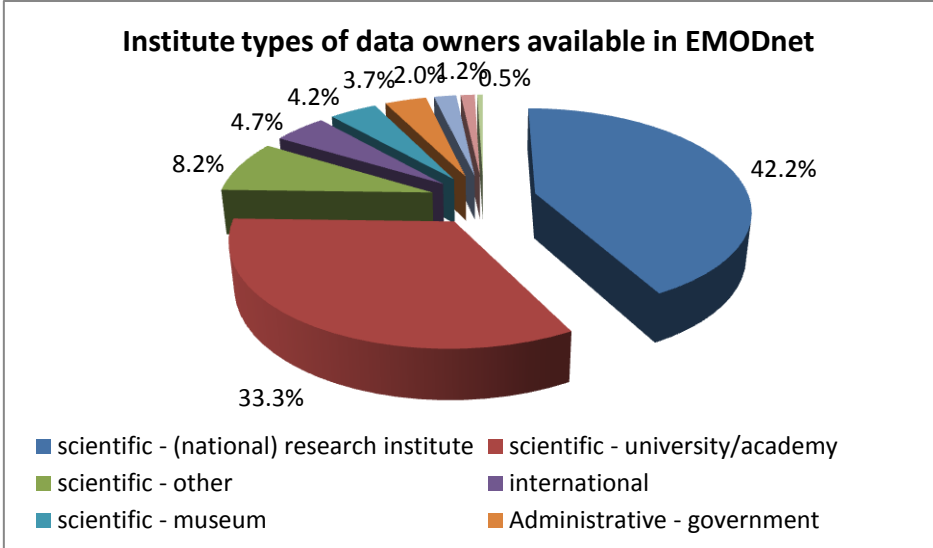
## Basic analyses of biological metadata

Some basic analyses on the metadata show us that most of the datasets in EMODnet originate from the UK (18%), Belgium (10%) and Germany (9%). All European countries are represented, as well as a number of non-European countries that have collected data within European marine waters. The indication 'international' refers to datasets with multiple origin, spread around the world. The institute type of the original data holding institutes have been defined and a subdivision has been made accordingly. The majority of the available dataset descriptions originate from a science-related institute, either national research institutes (42%), universities and academies (33%) or other scientific institutes. In some cases, the distinction between a governmental institute and a national research institute was rather vague, as national research institutes can be considered as governmental institutes. All datasets from the data catalog have received a 'data type' label. The majority of the datasets originate from research (58%), followed by monitoring data (27%). Research data include data from short-term research projects (theses, PhD, generally less than 5 years of data), whereas monitoring data are data from long-term actions (over 5 years). Some datasets have been categorized as 'maps/geographical files' (1%), indicating that EMODnet does not have access to the data files, but derived products – such as GIS layers – have been made available.

Each dataset receives keywords based on its functional group content: benthos, plankton, macroalgae, birds, mammals, reptiles, fish and chlorophyll. The majority of the datasets available in the data catalog contain benthos data (35%). 25% of all datasets has a 'mixed' content, indicating that they contain data on two or more defined functional groups. Datasets with data and information on reptiles are very few, easily explained by the fact that reptiles within European marine waters are limited to a few species of sea turtles.



*Division of available dataset description according their country of origin (March 2011)*

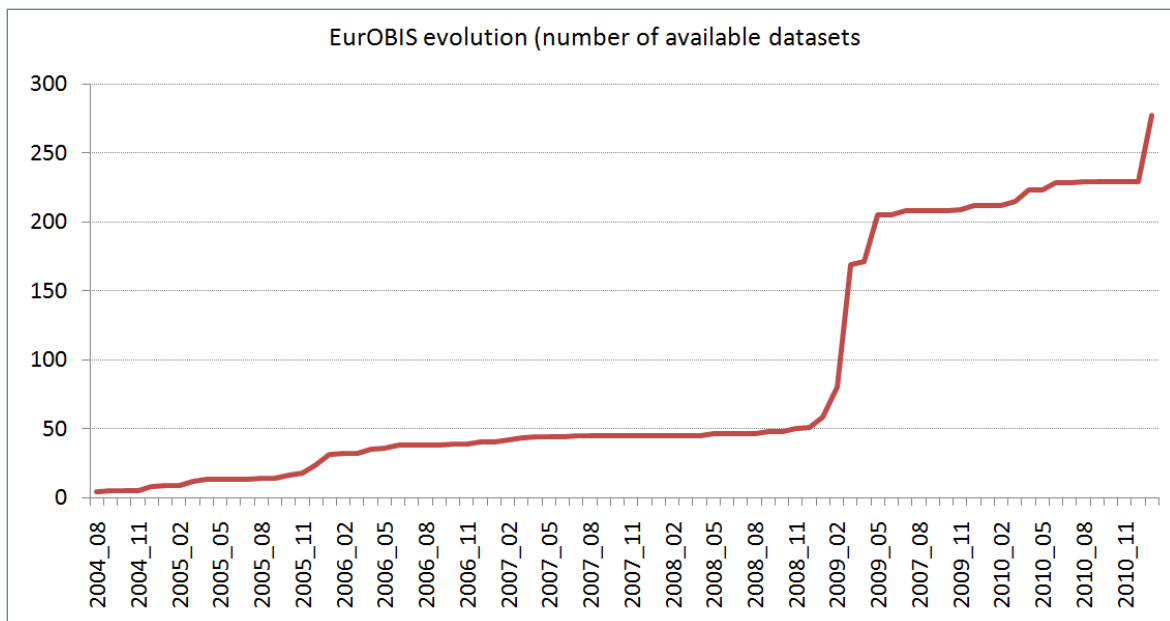


*Division of available datasets according to institute type of the original data owner, data type and functional group (March 2011).*

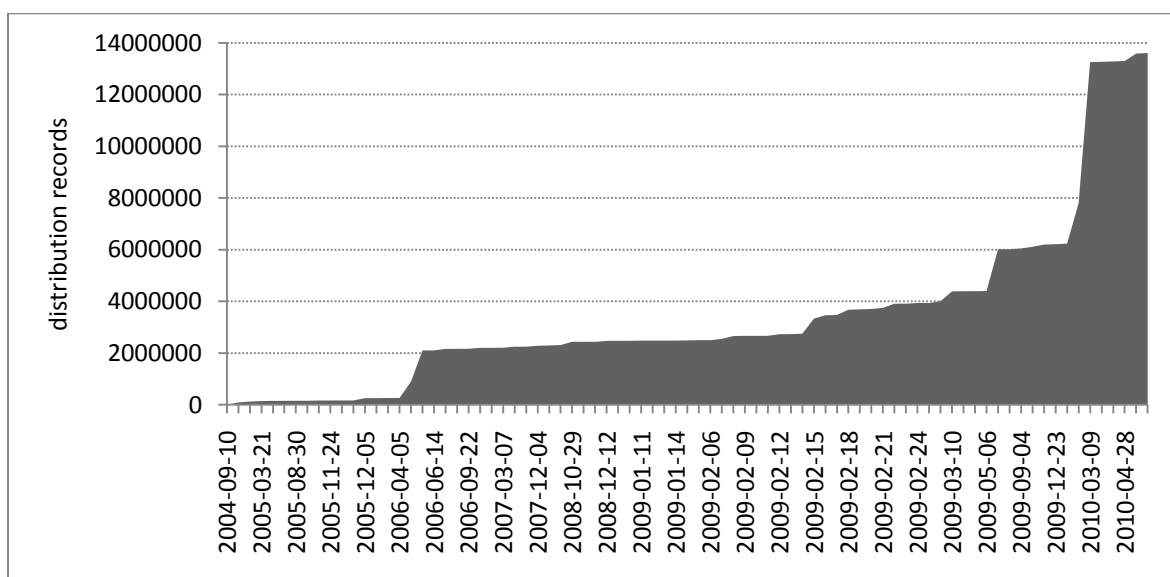
## Content and gaps of the European Ocean Biogeographic Information System EuroOBIS

The main goal of a thorough gap-analysis of the content of EuroOBIS was to identify gaps in data availability on both spatial and temporal scale and to list possible applications and (mis)-uses of the system. During the first 2 years of EMODnet, the gap analysis has been repeated three times – summer 2009, summer 2010 and February 2011 - in order to document the evolution in the filling of the gaps and to determine the gaps that still remain and are less easy to deal with. The results and conclusions of the gap analysis of summer 2010 have been incorporated into a manuscript, which was submitted to the scientific journal *Hydrobiologia*. The manuscript is currently in press.

Since the start of EMODnet, the number of available datasets and distribution records has grown enormously, as shown in the following graphs.



Evolution of the number of available datasets within EuroOBIS (February 2011)



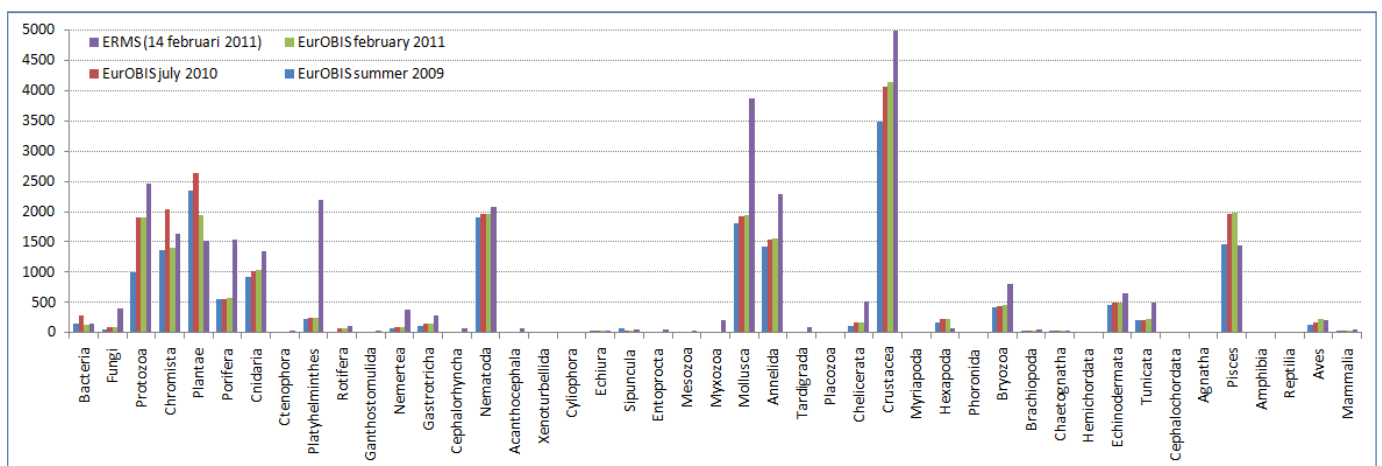
Evolution of the number of available distribution records in EuroOBIS (February 2011)



## Taxonomic coverage

Since the start of EMODnet in 2009, 2609 new species names have been added to EurOBIS, bringing the total number of available species to 21 252, or 67 % of the species documented to appear in European marine waters based on the European Register of Marine Species ERMS.

The strongest absolute growth in available species is documented for Protozoa, Crustacea and Pisces (fig). The most remarkable however is that – thanks to the data collection effort during EMODnet – we were able to capture one species for both the Xenoturbellida and the Mesozoa, previously undocumented species in EurOBIS. For three groups (Gnathostomulida, Cycliophora and Phoronida) no additional species were documented during EMODnet, although more species are known to occur in European marine waters. For plants and Chromista, there is a fluctuation visible over the whole period: this can be explained by the thorough quality control on the Algaebase dataset where records of terrestrial algae or algae only appearing in freshwater lakes were excluded. The large differences between ERMS and EurOBIS in number of European species per taxonomic group clearly indicate a (significant) gap in the taxonomic coverage of the system. For 14 of the defined taxonomic groups, less than 50 % of the species documented in ERMS are represented in EurOBIS, amounting to less than 25 % for 7 of these groups (Fungi, Platyhelminthes, Nemertea, Cephalorhyncha, Acanthocephala, Mesozoa and Tardigrada). Compared to the summer of 2010, there has been a shift in the top ten of most commonly documented species within EurOBIS. The species with the highest number of distribution records is still common dab (*Limanda limanda*), but the top ten now consists of 9 fish species and 1 bird species (*Fulmarus glacialis*) compared to 6 fish species and 4 bird species in 2010. There are currently 19 species with more than 100 000 distribution records in EurOBIS. This top 10 shows that – although commercial fisheries data is not a part of EMODnet – lots of data of fish (monitoring and research) is collected.

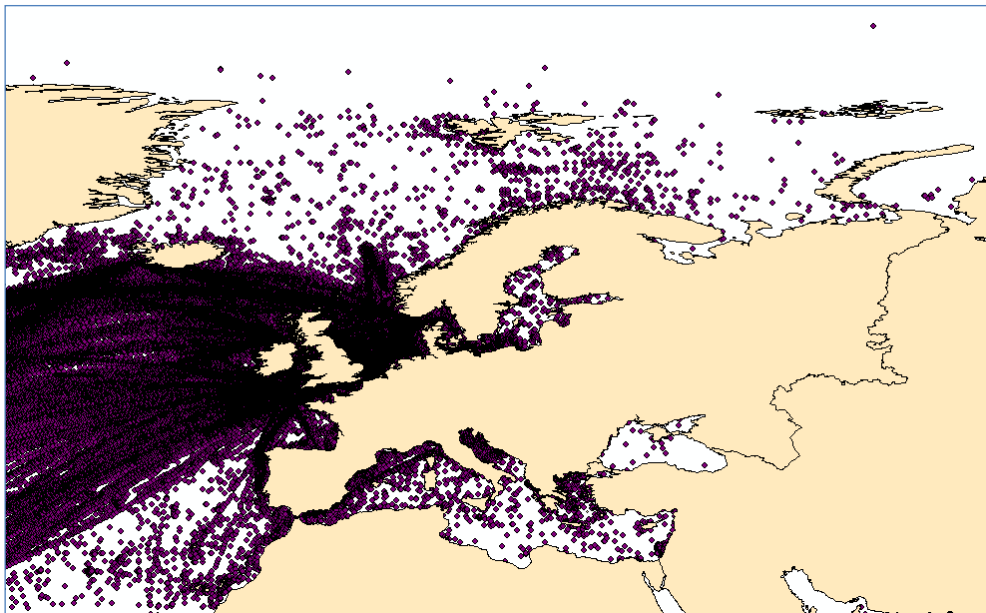


Comparison of the documented species in EurOBIS at three different times (summer 2009, July 2010 and February 2011) and the species known to occur within European marine waters based on the European Register of Marine Species (ERMS).

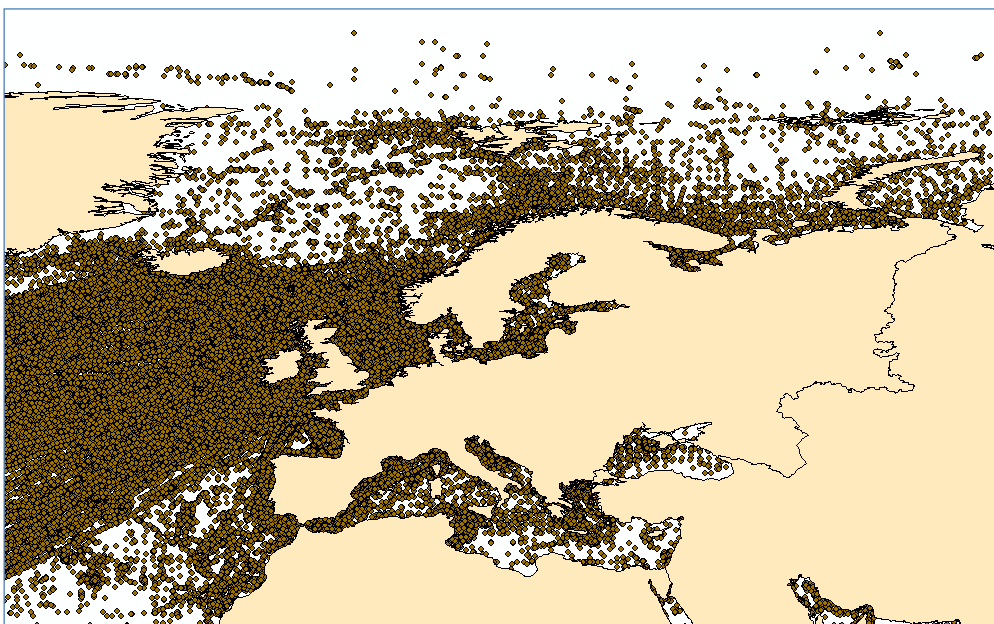
## Geographical coverage

### A. Sampling locations (= distinct lat-long from eurobis database)

Comparing the documented sampling locations just before the start of EMODnet (April 2009) and now (February 2011) shows a vast improvement of the geographical coverage of the European marine waters. Whereas the Arctic region was previously under-documented compared to other regions, this has now greatly improved: the geographical gap in the White Sea has been addressed, as has the gap in the Barentsz Sea. More to the south, there is a denser coverage of the Black Sea and the Mediterranean, although the African side of the Mediterranean does remain largely undocumented.



*Overview of the sampling locations documented in EurOBIS in April 2009*

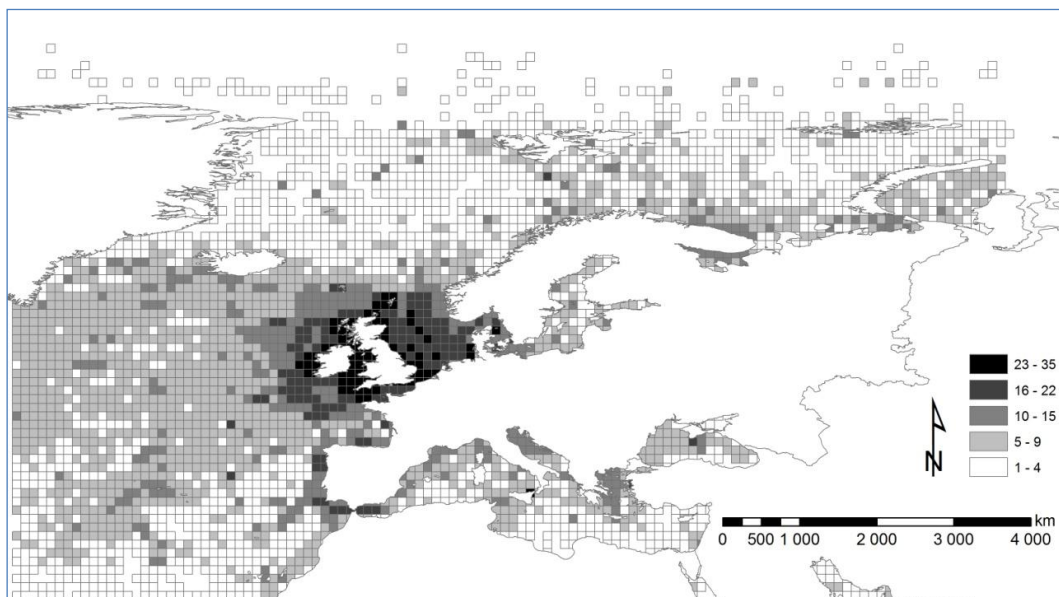


*Overview of the sampling locations documented in EurOBIS in February 2011*

## B. Major taxonomic groups per grid cell (1x1°)

In this analysis, the number of higher taxonomic groups (as defined in the graph under 'taxonomic coverage') per grid cell of 1 by 1 degree have been counted and plotted on the European map. The map shows that the North Sea, English Channel and North East Atlantic regions are very well documented within EurOBIS, whereas the Arctic has hardly any data and diversity – in number of higher taxonomic groups - represented. Differences in number of distribution records and number of higher taxonomic groups are related to data gathering efforts, which differ strongly between regions, depending for example on their accessibility.

It's important to realize that this map does not represent the general state of biodiversity across European marine waters, but it should be seen as a proxy for the general biodiversity and data coverage so far available within Europe. As we assume that all major taxonomic groups (45) are present all over the European marine waters, then it's clear from the map that there is much room for improvement in e.g. the Mediterranean area (especially the African part) and the most northern regions. When the number of major taxonomic groups is lower than the defined total of 45, this identifies a gap.



*Number of higher taxonomic groups per grid-cell of 1°x1° (situation summer 2010)*

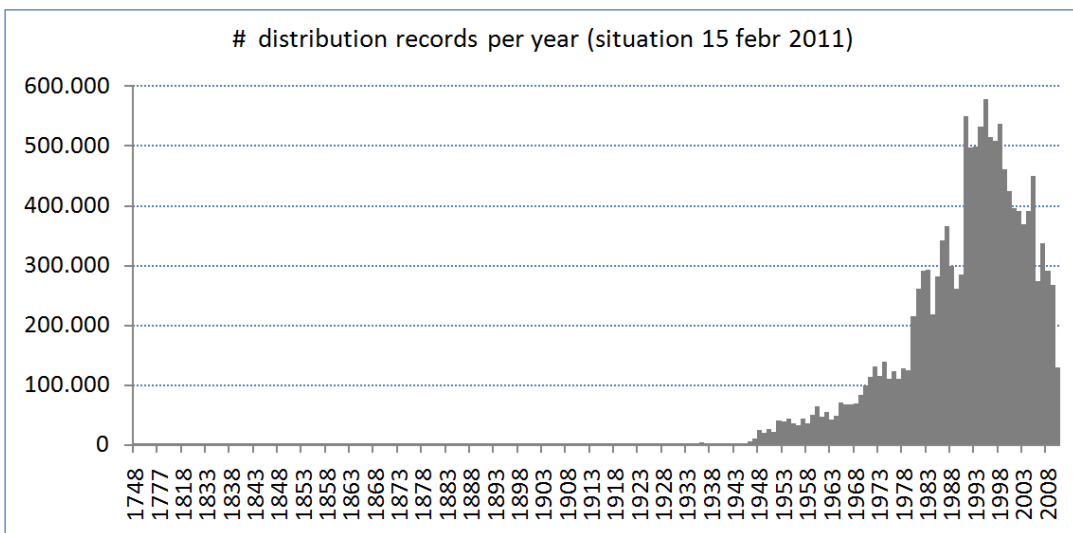
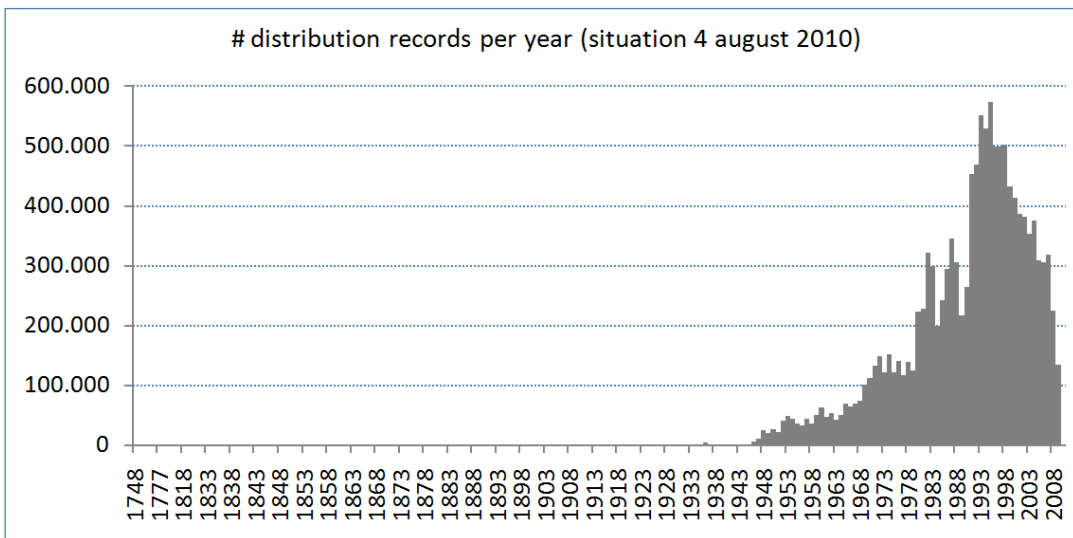
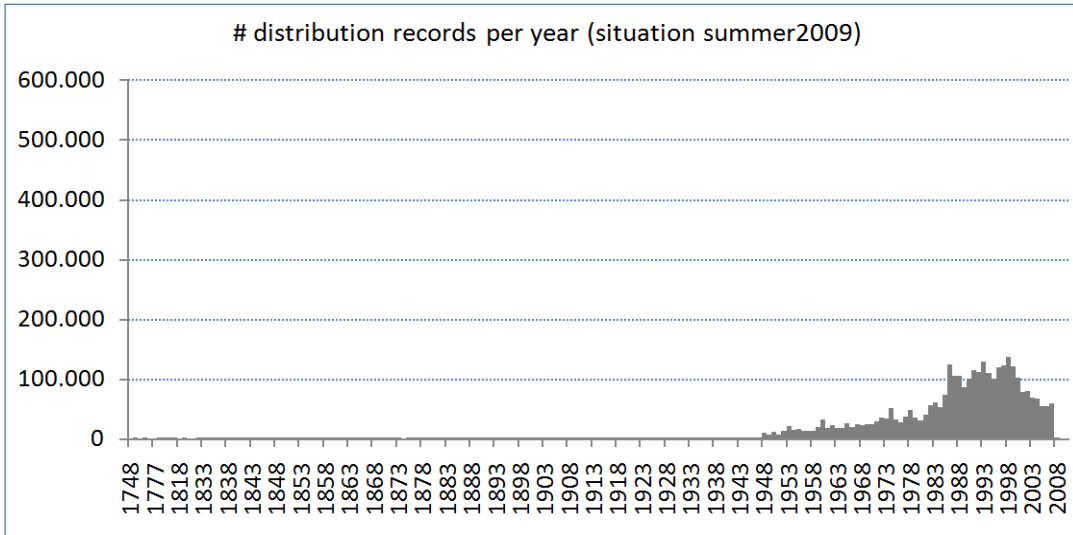
### Temporal coverage

The majority of the available distribution records have a time indication, making them suitable for temporal analyses. In February 2011, this amounts to 13.8 million records or 96 % of the totally available distribution records which can be used in such analyses. The missing time indication for 4 % of the records might be due to the fact that they have been collected from literature or from specimen collections where the collection date is not known (anymore) or unreadable on the labels.

The three graphs illustrate the amount of distribution records available for a given year for each of the performed gap analyses. In each gap analysis, the period post-1950 is clearly distinguishable from the pre-1950 period in available records, indicating that older (or historical) data is not easily mobilised and incorporated into public database systems such as EurOBIS and EMODnet. In all cases, the pre-1950 data constitute less than 0.7 % of the total amount of time-referenced records. These 'historical' data do however represent an important component in the global representation of species distributions: these data can give scientists insights on the biodiversity of species prior to human impacts and can help establish a baseline. The loss of this valuable knowledge can be counteracted by investing in so-called 'data-rescue' actions which focus on tracing and identifying historical paper-based datasets and providing the possibility to digitize them. A second, but less marked gap is situated in the post-2000 period. Although some improvement in available data post 2000 is visible in February 2011, there is still a significant fallback compared to the nineties which can be explained by the time-lag in making data available. Post-2000 data are relatively recent and researchers are not prone to make their data available in such an early stage of their research or data processing. Moreover, species identifications – for example of meiofauna (smaller than 1 mm) - can be very time-consuming, also adding to this time-lag in data availability.

	summer 2009	August 4th, 2010	February 15th, 2011
<b># time referenced records</b>	3 266 816	13 073 541	13 805 132
<b># records prior to 1950</b>	32 957	86 711	87 355
<b># records 1950 - 2010</b>	3 235 253	12 986 830	13 717 777
<b># records 2000 - 2010</b>	572 228	3 203 858	3 724 237

*Random indications of number of time-references records at three different timings during EMODnet*



Evolution of the number of distribution records per year in EurOBIS:  
summer 2009 (top graph), August 2010 (centre graph) and February 2011 (bottom graph)

## Species abundance

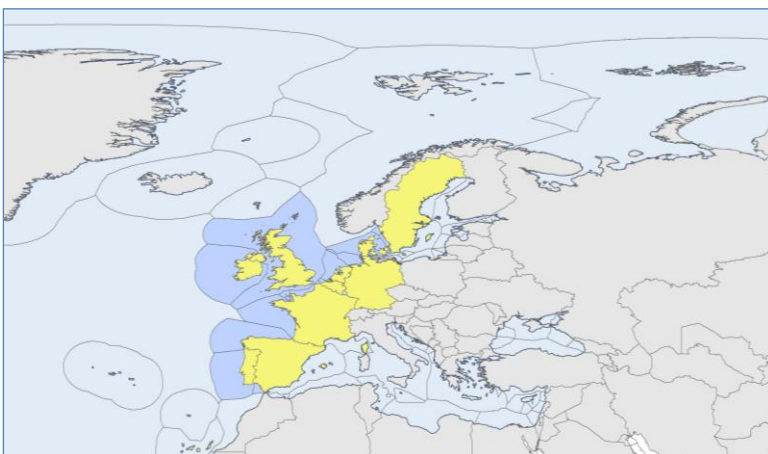
All distribution records in EurOBIS represent the presence of a taxon at a given location and a given time. These presence data can be used in applications or data products in need of geographical taxon occurrences or in documenting the species richness. During the first two years of EMODnet, there has been a spectacular growth of number of distribution records that also contain abundance information, indicating how many species were observed at a certain location at a certain time. In February 2011, more than half of the distribution records contain abundance information compared to less than 5% in the summer of 2009. Although this is a major improvement compared to the start of EMODnet, there is still a lot of work to do, as the use of the available abundance data is not always straightforward.

	<b># records with abundance</b>	<b>Total # records</b>	<b>Relative proportion</b>
Summer 2009	522 974	10 782 964	4,85 %
Summer 2010	1 933 003	13 600 744	14,2 %
February 2011	7 719 707	14 360 293	53,8 %

*Indication of the total number of records available within EurOBIS and the absolute and relative number of records containing abundance information at three different times during EMODnet*

### **2.2.3. Inventory of national monitoring programs per country (to add)**

Next to contacting the EMODnet Biological Network to complete our inventory on marine biological datasets through questionnaires, an additional search for long-term biological monitoring data was performed. This search had two approaches: identifying both national marine biological monitoring data and regional marine biological monitoring data. This additional search focused on the assigned geographical area, being the Bay of Biscay, Iberian coast and the Greater North Sea, including



Kattegat and English Channel. All countries bordering these sea areas were contacted. We specifically targeted biological monitoring series with data on benthos, plankton, birds, mammals, reptiles and algae. The results so far are listed in the following table, followed by a more elaborate description per country. An overview of the regional instances that were contacted is also given.

*Focus area where national marine biological monitoring programs identified*



## National monitoring

Country	Groups	Temporal scope
Sweden	benthos, plankton, mammals	1971 - present
Denmark	benthos, plankton, algae, reptiles, mammals	1976 - present
Germany	benthos, plankton, algae, birds, mammals	1974 - present
Netherlands	benthos, plankton, birds, plants, mammals, bacteria	1956 - present
Belgium	benthos, birds, mammals	1979 - present
United Kingdom	benthos, plankton, birds, algae, reptiles, mammals	1970's - present
Ireland	benthos, plankton, birds, algae, mammals	1970's - present
France	benthos, plankton	1980's - present
Spain	plankton	1987 - present
Portugal	(benthos), phytoplankton, shellfish	?

*Overview of countries bordering the assigned geographical area, indicating the functional groups they are monitoring on a national level and the temporal scope this is covering*

### Sweden

#### ○ Marine biological data of Sweden – SHARK

Responsible: Swedish Meteorological and Hydrological Institute

- Monitored groups:
  - Zooplankton: since 1979, 672 visits, 51800 measurements
  - Phytoplankton: since 1983, 300 visits, 78000 measurements
  - Macrozoobenthos soft sediment: since 1971, 3130 visits, 200000 measurements
  - Marine macrophytes: since 1992, 1439 visits, 15000 measurements
  - Grey seals: since 1989, 6274 visits, 12700 measurements
  - Ringed seals: since 1995, 7400 visits, 7400 measurements
  - Harbor seals: since 1988, 5700 visits, 5800 measurements
  - Chlorophyll (hose): since 1982, 420 visits, 3870 measurements.
  - Chlorophyll (bottle): since 1978, 19000 visits



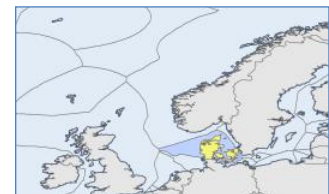
Metadata at <http://bio.EMODnet.eu/component/imis/?module=dataset&david=2136>

### Denmark

#### ○ The Danish national database for marine data – MADS

Responsible: National Environmental Research Institute (NERI)

- Monitored groups:
  - Macrozoobenthos soft sediment: since 1979, 700 sites, 1000 species
  - Zooplankton: since 1979
  - Phytoplankton: since 1979
  - Macro-algae (on stone reefs): since 1979
  - Chlorophyll: since 1976



Metadata at <http://bio.EMODnet.eu/component/imis/?module=dataset&david=2234>

There's also monitoring activities carried out by NERI on reptiles (loggerhead turtles) and mammals (seals, harbour porpoises and minke whales). These data are however not integrated into the Danish national database for marine data.

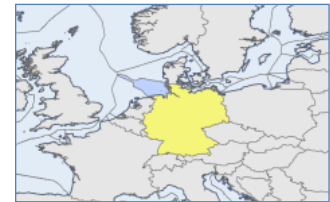
## Germany

### ○ **German marine monitoring programme – BLMP**

Responsibles:

- Bundesamt für Seeschifffahrt und Hydrographie: Deutsches Ozeanographisches Datenzentrum (DSH/DOD)
- Federal Environmental Agency (UBA-QA)
- Monitored groups (North Sea: since 1974; Baltic Sea: since 1979):
  - Macrophytes
  - Macrozoobenthos
  - Phytoplankton
  - Zooplankton
  - Birds
  - Fish
  - Mammals

Metadata at <http://bio.EMODnet.eu/component/imis/?module=dataset&dased=2323>

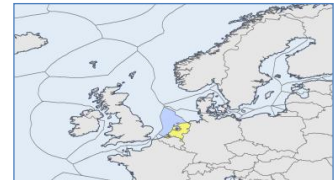


## The Netherlands

### ○ **Dutch national MWTL monitoring**

Responsible: Ministry of Transport, Public Works and Water Management (Rijkswaterstaat)

- Monitored groups:
  - Bacteria: since 1956
  - Macrobenthos: since 1991, 100 sites, 300 species
  - Zooplankton: since 1948, 4 sites
  - Phytoplankton, since 1990, 100 sites, 400 species
  - Waterbirds: since 1988
  - Coastal breeding birds: since 1979, 20 species
  - Sea birds: since 1984, 6 flights per year, 38 species
  - Sea mammals: since 1984, 6 flights per year, 6 species
  - Bivalves: since 1993 (WOT Shellfish Monitoring: Responsible IMARES)
- Partly online available (DONAR)



## Belgium

### ○ **Benthos monitoring**

Responsible: ILVO Fisheries Institute

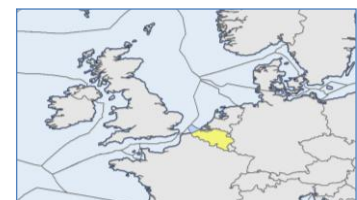
- Macrobenthos, epibenthos
- From 1979 onwards: 100 sites, 600 species

### ○ **Sea bird monitoring**

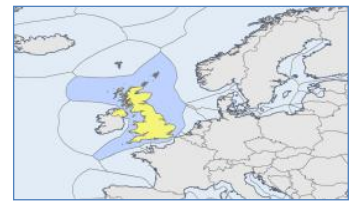
Responsible: Research Institute for Nature and Forest (INBO)

- Sea birds : from 1992 onwards: 40 ship days per year

Metadata at <http://bio.EMODnet.eu/component/imis/?module=dataset&dased=155>



## UK



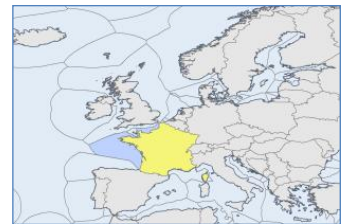
Several monitoring programs are in place in the United Kingdom, Ireland and Scotland. All the existing programs are listed at the website of the United Kingdom Directory of Marine Observing Systems UKDMOS (<http://www.ukdmos.org>). This website provides an easy search interface to the full inventory of all monitoring activities of the UK in the marine environment and functions as a central access point to all this information. The processing of the information available at this metadata website is still ongoing.

## Ireland



Ireland has national monitoring programmes in place for benthos and phytoplankton. Additionally, there is an Integrated Marine Programme (IME) that aims to support and develop Ireland's national and international deep sea research activity. A number of Irish datasets are also listed on the UKDMOS website. Additionally, the Marine Institute of Ireland makes Irish metadata available online. (<http://www.marine.ie/home/publicationsdata/data/MarineDataOnline.htm>).

## France



### ○ **Coastal environment monitoring database – QUADRIGE<sup>2</sup>**

Responsible: Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)

- Phytoplankton (REPHY): since 1987
- Microbiologie (REMI): since 1989
- Benthos [to add] REBENT

Metadata at <http://bio.EMODnet.eu/component/imis/?module=dataset&dasid=2259>

### ○ **Deep sea benthic fauna database – BIOCEAN**

Responsible: Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)

- Benthos: since 1967

Metadata at <http://bio.EMODnet.eu/component/imis/?module=dataset&dasid=490>

### ▪ **Observations et previsions cotières - PREViMer**

Responsible: Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER)

- Chlorophyll data

## Spain



### ○ ***Oceanographic time series of the Instituto Español de Oceanografía - RADIALES***

Responsible: Ministerio de Educación y Ciencia; Instituto Español de Oceanografía (IEO)

- Plankton, CTD and Chlorophyll data from 8 transect in Spanish waters

Metadata at <http://bio.EMODnet.eu/component/imis/?module=project&proid=3259>

## Portugal



No specific national marine monitoring programs currently seem to exist in Portugal. The monitoring of Portuguese transitional and coastal waters involves a number of different institutions. Discussions are ongoing on the establishment of some research group which will specifically focus on the monitoring of the marine biodiversity in Portugal, but nothing has started yet.

## **Regional monitoring**

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

A number of the above mentioned national monitoring programs submit their data to the ICES data centre and these data are thus part of the ICES DOME and DATRAS datasets which are already available within EMODnet Biology. To increase transparency, the DATRAS dataset has been split up into its child datasets (see table). For DOME, this subdivision is still in progress: about 60 institutes have delivered their data to the ICES DOME database, the coupling between institute and their submitted (national monitoring) data still has to be made. DATRAS contains data on fish, collected through scientific activities (not commercial data), whereas DOME is a collection of benthos and plankton data.

<b>DATRAS child dataset</b>	<b>Temporal coverage</b>	<b># distribution records</b>
French Southern Atlantic bottom trawl survey	1987 - present	181 980
Baltic International trawl surveys	1991 - present	704 092
Beam trawl surveys	1985 - present	390 181
Irish ground fish survey	1992 - present	170 984
North Sea international bottom trawl survey	1965 - present	2 984 722
Northern Ireland survey	1992 - present	167 873
Scottish Western Coast Via groundfish survey	1981 - present	408 363
Rockall survey ICES VIb	1985 - present	35 208

Metadata of DATRAS at <http://bio.EMODnet.eu/component/imis/?module=dataset&dasid=2141>

Metadata of DOME:

- DOME community: <http://bio.EMODnet.eu/component/imis/?module=dataset&dasid=2157>
- DOME biota: <http://bio.EMODnet.eu/component/imis/?module=dataset&dasid=2159>

## HELCOM

HELCOM has coordinated the Baltic Sea monitoring since 1979 and biological variables have been monitored under the HELCOM COMBINE Monitoring Programme. All HELCOM contracting parties submit their biological monitoring data to ICES, the data consultant and the data is thus freely available through ICES. Within the ICES databases, these data are available in the DOME databases. As mentioned earlier, breaking up the DOME database into its component datasets has not yet been done. As soon as this is implemented, the different providers and their national monitoring data will become visible, consultable and downloadable through the EMODnet data portal.

## Black Sea Commission

Early contacts with the pollution monitoring and assessment officer of the Black Sea Commission, Violeta Velikova have learned us that the Black Sea Commission annually collects biological data from the Black Sea coastal states (Bulgaria, Georgia, Romania, Russia, Turkey and Ukraine) on the following groups: plankton, benthos and fish. They however only collect data on a higher taxonomic level, not on species level. This is in essence aggregated information and is preferably not incorporated into EurOBIS / EMODnet. EMODnet prefers to make the 'raw' data on species level available, as they are collected by the different countries. The Black Sea Commission is however not allowed to deliver the raw data to third parties (in this case EMODnet).

Through one of the project partners – Institute of Biology of the Southern Seas, IBSS – a regional action was set up to contact the different coastal countries and to start an inventory and the gathering of the available data. This action is called the 'Black Sea mini-data-grant program and will be discussed under the 'analysis' chapter.

## Trilateral Monitoring and Assessment Program (TMAP)

Since 1978, The Netherlands, Denmark and Germany have been working together on the protection and conservation of the Wadden Sea covering management, monitoring and research, as well as political matters. In 1982, a Joint Declaration on the Protection of the Wadden Sea was agreed upon in which the countries declare their intention to coordinate their activities and measures for the protection of the Wadden Sea. Within TMAP, several biological parameters are measured: macrozoobenthos, phytoplankton, breeding birds, migratory birds, beached birds and seals. Most of the TMAP parameters are part of already existing or planned monitoring programs in the three countries and cover the requirements of the EC Directives and other international agreements.

It will be key to document which institutes contribute to the Trilateral Monitoring and Assessment Program of the Wadden Sea and to see whether the data from these institutes already contribute directly or indirectly (through e.g. ICES) to EMODnet / EurOBIS.

#### 2.2.4. Overview of data per species group, including 3 selected species per group

Within the tender, eight groups of species – categories – were defined for which data should become available through the Portal:

- Phytoplankton
- Zooplankton
- Angiosperms
- Macro-algae
- Invertebrate bottom fauna (benthos)
- Bird communities
- Sea mammals
- Reptiles (if appropriate for the marine basin in question)

According to the tender specifications, three species or species groups should be selected for each of these eight categories. Per category, the ‘top-3’ of most abundant species – in number of distribution records - will be presented, with an indication of the number of available distribution records, the time-frame in which these data are available and an indication of the number of datasets that contain data on that specific species. Although the tender requests three species which will reflect the completeness of the monitoring programme (or all data available within EurOBIS / EMODnet), we have chosen not to do so and only show the three most abundant species. During the data product workshop, extensive discussions have taken place on the sense and non-sense of selecting a few species for the European marine waters. The general conclusion was that the selection of a species will largely depend on the investigated area (as also indicated in the tender), but will also be influenced by the initial goal of the question one wants to answer. If for example one wants to inform on reef-building benthic species, such species should be selected and one will not be interested in the pre-selected (non-reef-building) benthic species.

Next to the top-3 per category, some maps and specific data products will be shown and explained.

	<b># datasets described</b>	<b># datasets online</b>	<b># data products</b>
Phytoplankton	58	16	0
Zooplankton	56	18	1
Angiosperms	6	5	2
Macro-algae	44	32	0
Benthos	219	174	5
Reptiles	8	6	2
Sea birds	16	12	3
Sea mammals	31	19	1
Chlorophyll	19	1	2

*Overview of the defined categories, the number of datasets described within the data catalog, the number of datasets available online through the portal and the number of relevant data products*

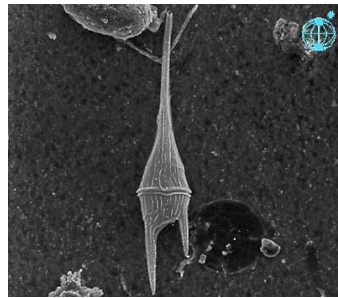
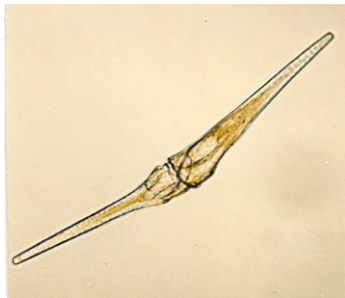


## Phytoplankton

The data catalog holds 58 dataset descriptions of datasets that contain phytoplankton data. For 16 of those datasets, the data are available online through the Portal. The most significant dataset is the Continuous Plankton Recorder (CPR) from the Sir Alister Hardy Foundation for Ocean Science (SAHFOS, UK), containing 632 473 presence records on phytoplankton. The second largest dataset with only phytoplankton data is Réseau de Surveillance phytoplanktonique (REPHY) from Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER, France), containing over 250 000 distribution records.

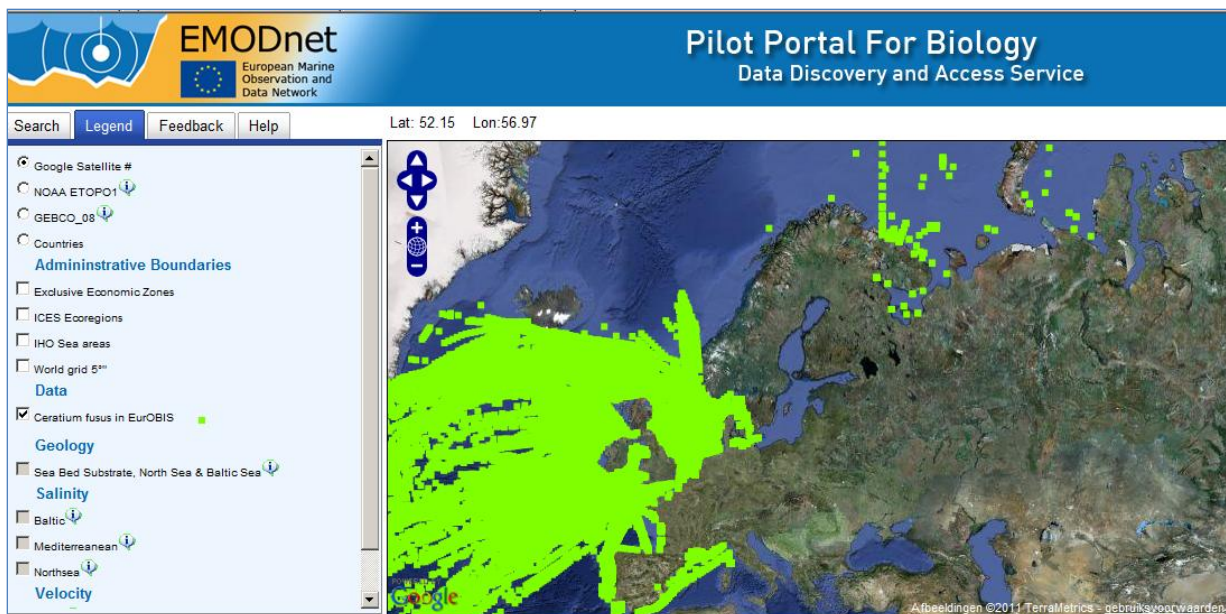
	Common name	# records	Time-frame	# datasets	
	<i>Ceratium fusus</i>	/	59 267	1900 - 2009	11
	<i>Ceratium furca</i>	/	42 255	1921 - 2009	10
	<i>Ceratium tripos</i>	/	32 228	1921 - 2009	9

*Top-3 of phytoplankton species based on the number of available distribution records*



*Ceratium fusus* (left), *Ceratium furca* (middle), *Ceratium tripos* (right).

Pictures taken from [www.marinespecies.org](http://www.marinespecies.org)



Map showing the sampling locations where *Ceratium fusus* has been recorded (data available in March 2011)



## Zooplankton

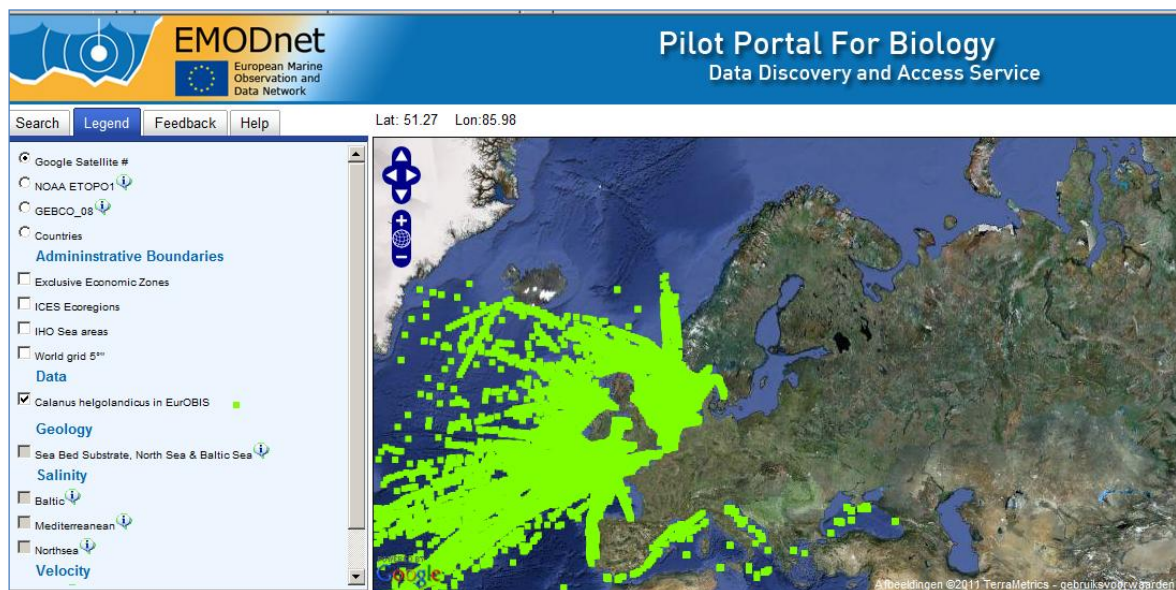
56 datasets containing zooplankton information have been described in the data catalog, of which 18 have made their data available through the Portal. Also here, the most significant dataset is the Continuous Plankton Recorder (CPR) from the Sir Alister Hardy Foundation for Ocean Science (SAHFOS, UK), containing 1.2 million presence records on zooplankton. Another – purely – zooplankton dataset is the ‘historical zooplankton records from the Black Sea’, made available by the Institute of Biology of the Southern Seas (IBSS, Ukraine) and represents over 65 000 distribution records.

	Common name	# records	Time-frame	# datasets
	<i>Calanus helgolandicus</i>	/	1958 - 2008	12
	<i>Temora longicornis</i>	/	1946 - 2009	16
	<i>Pseudocalanus elongatus</i>	/	1906 - 2009	11

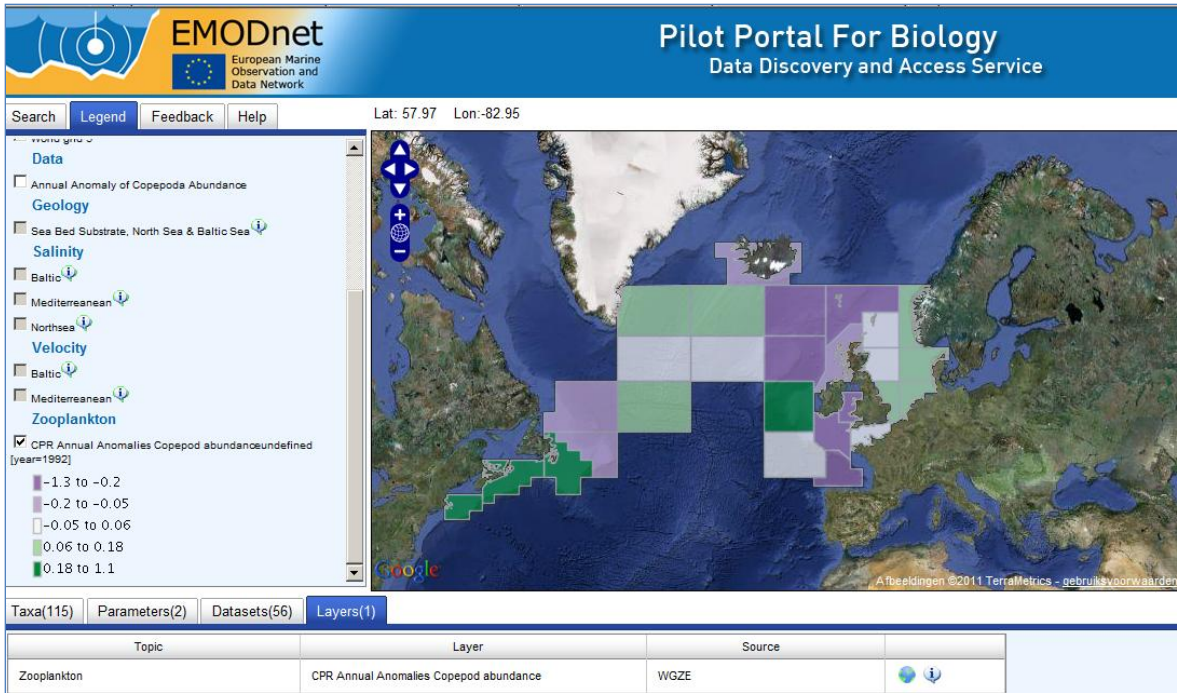
*Top-3 of zooplankton species based on the number of available distribution records*



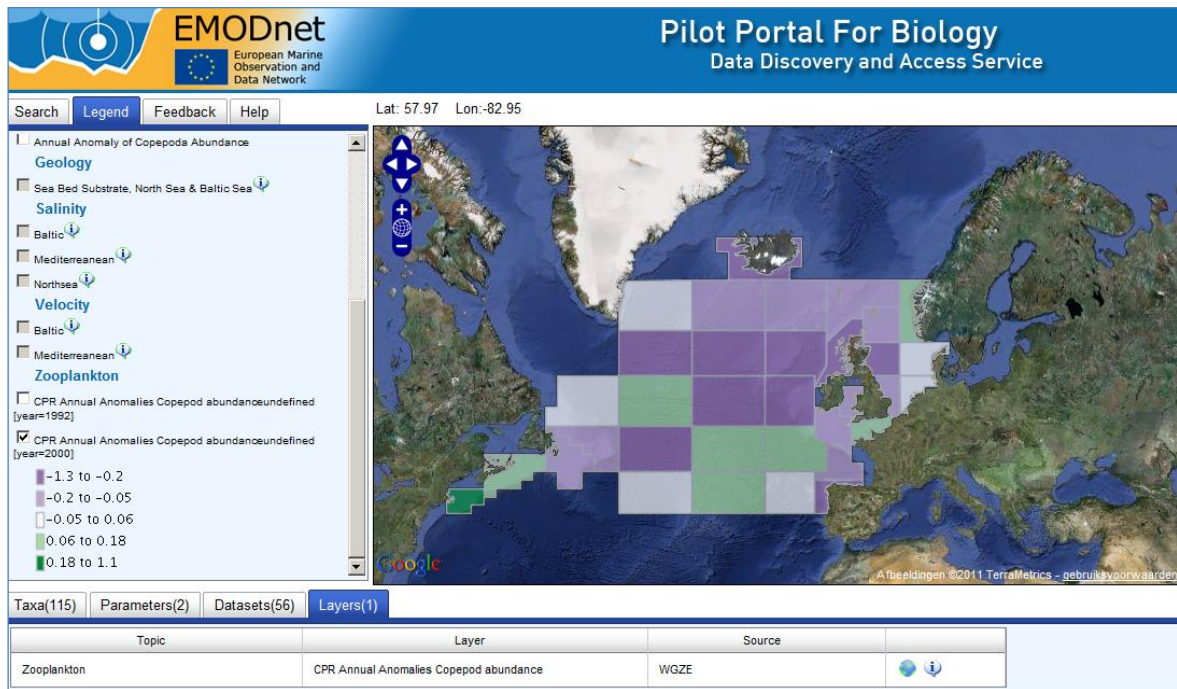
*Calanus helgolandicus* (left), *Temora longicornis* (middle), *Pseudocalanus elongatus* (right).  
Left and middle pictures from [www.marinespecies.org](http://www.marinespecies.org); right picture from EOL.



Map showing the sampling locations where *Calanus helgolandicus* has been recorded (data available in March 2011)



Data product showing the annual anomalies of the copepod abundances in the European marine waters in 1992



Data product showing the annual anomalies of the copepod abundances in the European marine waters in 2000



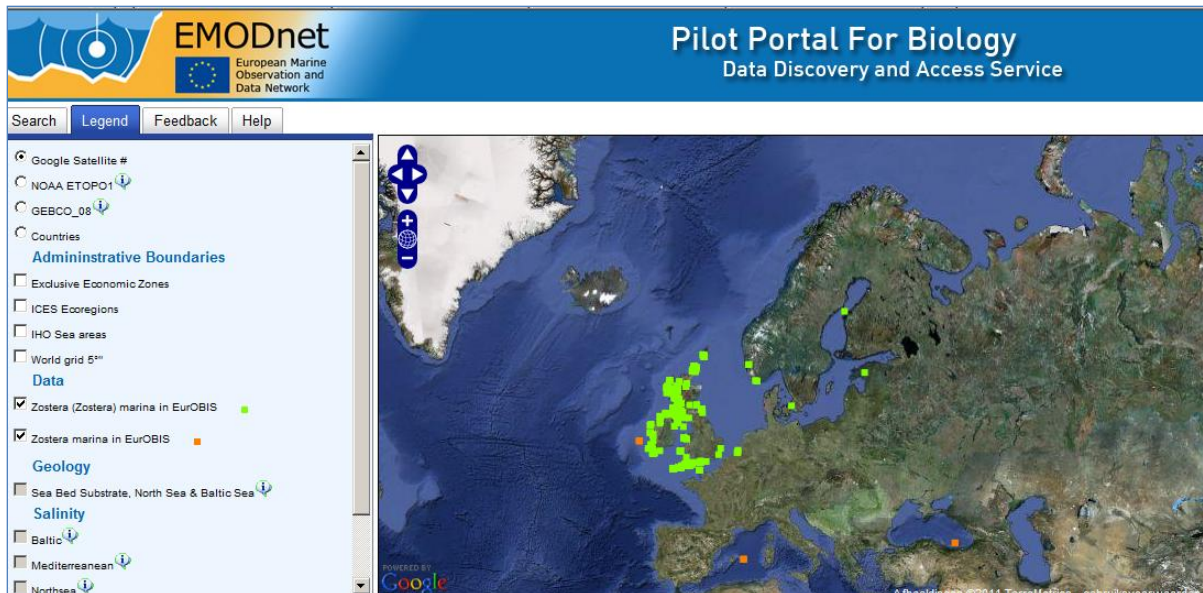
## Angiosperms

There are 6 dataset descriptions available on angiosperms, of which 5 are online available through the Portal. None of the datasets deals exclusively with angiosperms, there's always other categories represented within the datasets. In total, EurOBIS holds information on 55 angiosperm species, representing some 1 500 distribution records. For angiosperms, two data products have been developed so far: (1) an impression of the number of angiosperm species per grid cell in the European marine waters, and (2) an indication of the sampling effort, defined as the count of all unique sampling locations, taking into account the moment they were visited (e.g. one station visited four times a year is counted four times).

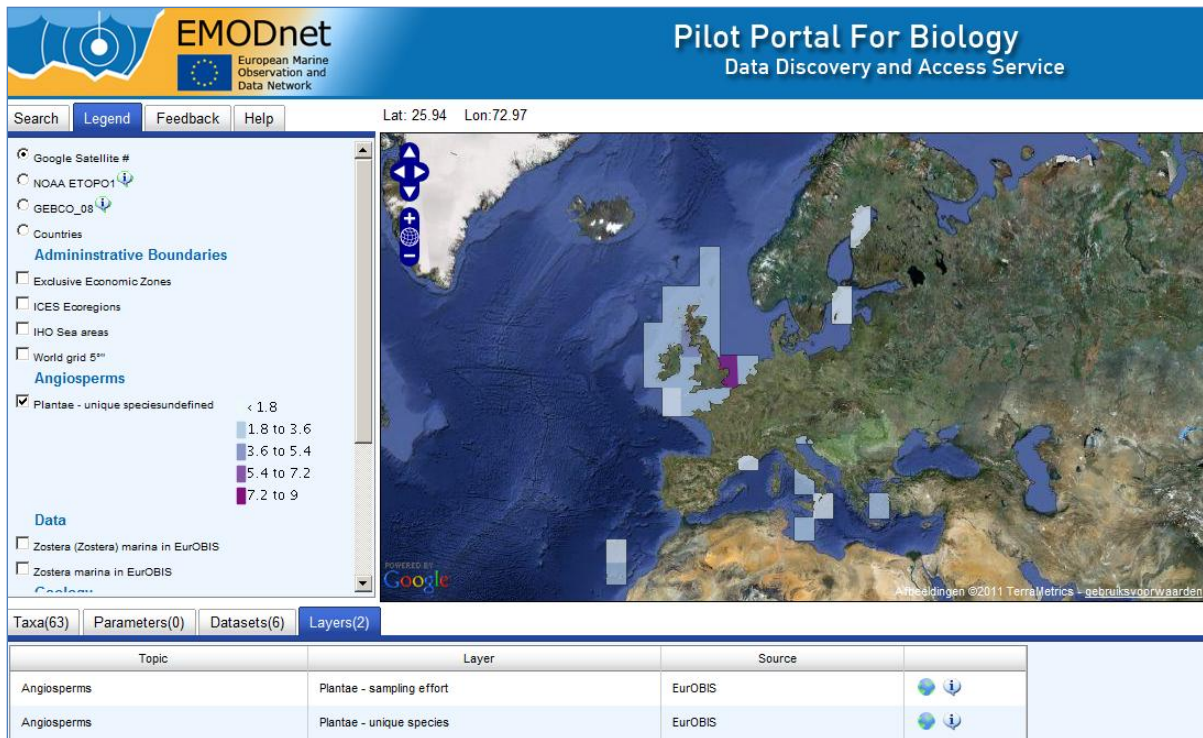
	Common name	# records	Time-frame	# datasets
<i>Zostera (Zostera) marina</i>	eelgrass	544		15
<i>Posidonia oceanica</i>	Neptune grass	408		3
<i>Armeria maritima</i>	sea pink	113		4



*Zostera (Zostera) marina* (left), *Posidonia oceanica* (middle), *Armeria maritima* (right).  
Left & middle picture from Algaebase, right picture from [www.marinespecies.org](http://www.marinespecies.org)



Map showing the sampling locations where *Zostera (Zostera) marina* in European marine waters (available data in March 2011). *Zostera marina* and *Zostera (Zostera) marina* are synonymized. To get an accurate idea of its distribution, both species have to be plotted (orange & green)



Visualisation of the data product 'unique number of angiosperm species per 3 by 3 degrees grid cell (Mercator projection)', based on the data available in EuroBIS on June 3<sup>rd</sup>, 2009

## Macro-algae

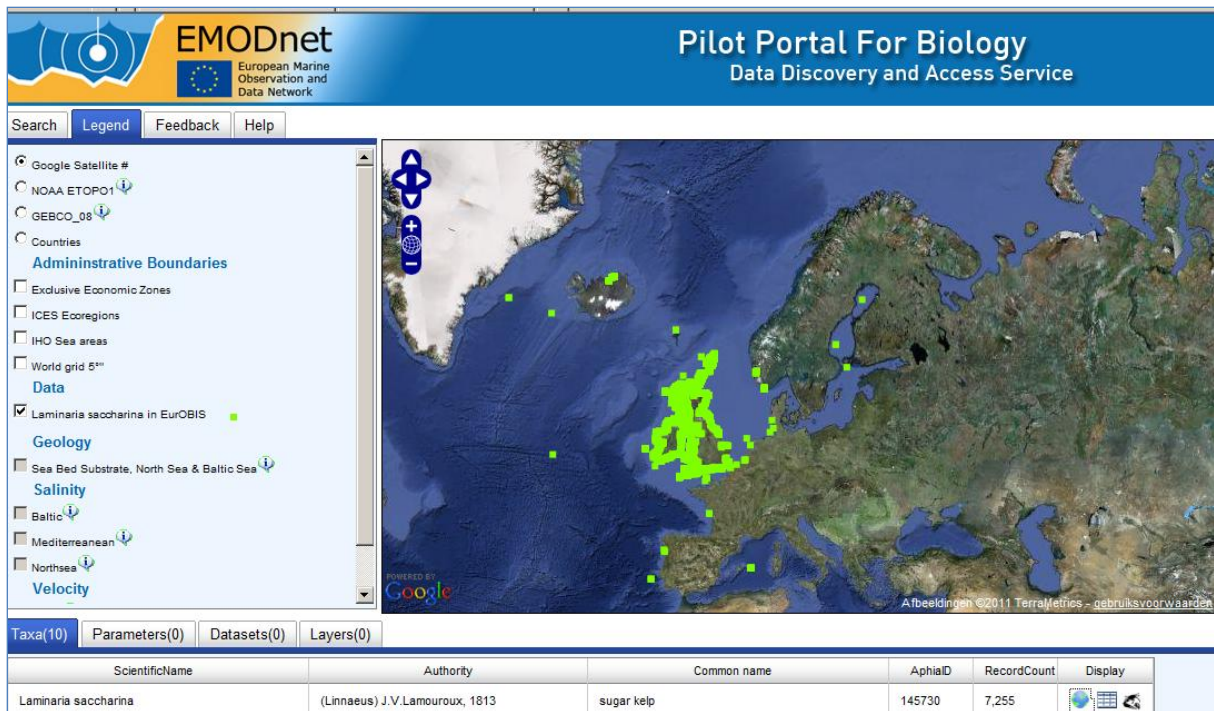
A total of 44 dataset descriptions with relations to macro-algae have been described in the data catalog. The data of 32 of those datasets are online available through the Portal. A number of these datasets deal exclusively with algae. The largest of those datasets is Algaebase, managed by the National University of Ireland and containing over 115 000 algal distribution records. The second dataset contains seaweed data for Great Britain and Ireland in is co-managed by the British Phycological Society and the UK National Biodiversity Network and also represents over 100 000 records. There are some smaller datasets on algae from the Icelandic coast, making a major contribution to the geographical coverage of this category.

	Common name	# records	Time-frame	# datasets
<i>Laminaria saccharina</i>	sea-belt / sugar kelp	7 255	1850 - 2007	18
<i>Laminaria hyperborea</i>	cuvie	7 152	1888 - 2007	24
<i>Corallina officinalis</i>	coral weed	6 467	1858 - 2007	27



*Laminaria saccharina* (left), *Laminaria hyperborea* (middle), *Corallina officinalis* (right)  
Left and middle pictures from [www.marinespecies.org](http://www.marinespecies.org); right picture from Algaebase.





Map showing the sampling locations where *Laminaria saccharina* has been recorded (data available in March 2011)

### Invertebrate bottom fauna (benthos)

This category includes both macrobenthos (larger than 1 mm) and meiobenthos (smaller than 1 mm).

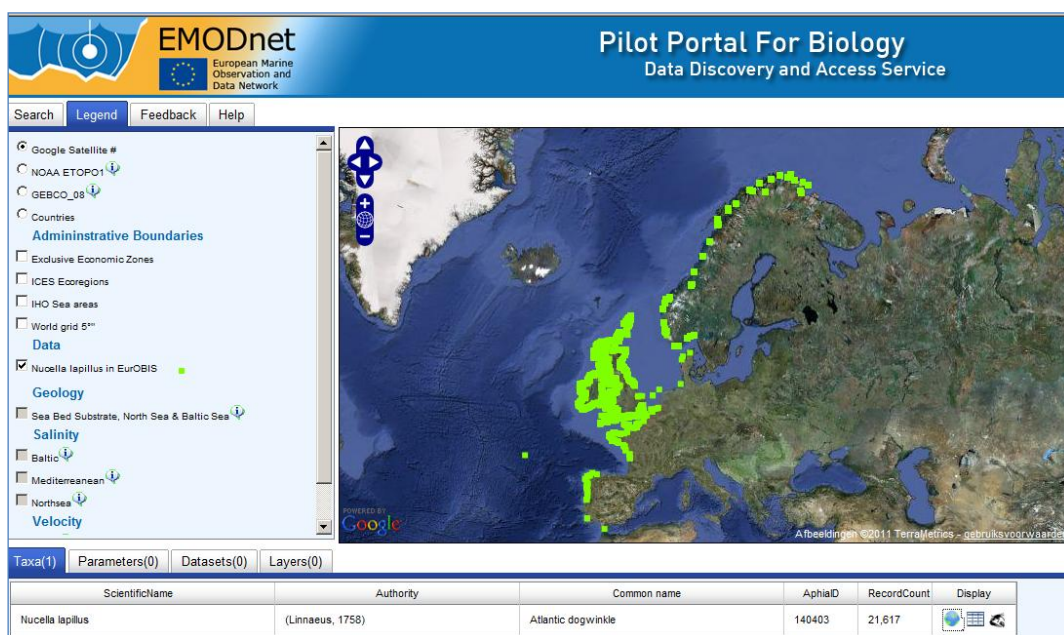
There are currently 219 datasets described in the EMODnet data catalog which contain data on invertebrate bottom fauna or benthos. 174 of those are available online, through the Portal. For 129 of those 174 available datasets, the focus lies exclusively on benthos, the other datasets have a mixed origin. The largest dataset is the 'marine benthic dataset (version 1) commissioned by UKOAA' that was delivered by the Joint Nature Conservation Committee (JNCC, UK), representing almost 204 000 distribution records on benthos. The second largest dataset is 'BIS dataset of the south-western part of the Netherlands (1985-2004), provided by the Centre for Estuarine and Marine Ecology of the Netherlands Institute of Ecology (NIOO-CEME), representing over 136 000 benthic distribution records.

So far, 5 data products related to the benthos have been made available through the Portal. The number of unique species per grid cell has been visualized for Crustacea and Echinodermata, as has the sampling effort for Echinodermata. Additionally, the occurrence of the species *Echinocardium cordatum* (heart-urchin or sea-potato) from one specific dataset has been visualized, followed by a map showing all sampled stations of that dataset. In the future, such actions should make it possible to create 'absence' maps of species. This is however not yet fine-tuned and will need a lot of additional work before reliable presence-absence maps can be created. It will also only be possible to create such maps based on monitoring datasets where you know they have looked for the species but not found it.

	Common name	# records	Time-frame	# datasets
<i>Nucella lapillus</i>	dogwhelk	21 617	1854 - 2009	22
<i>Mytilus edulis</i>	blue mussel	21 586	1986 - 2010	55
<i>Nephrops norvegicus</i>	Dublin bay prawn	19 239	1929 - 2010	35



*Nucella lapillus* (left), *Mytilus edulis* (middle), *Nephrops norvegicus* (right). All pictures from [www.marinespecies.org](http://www.marinespecies.org)



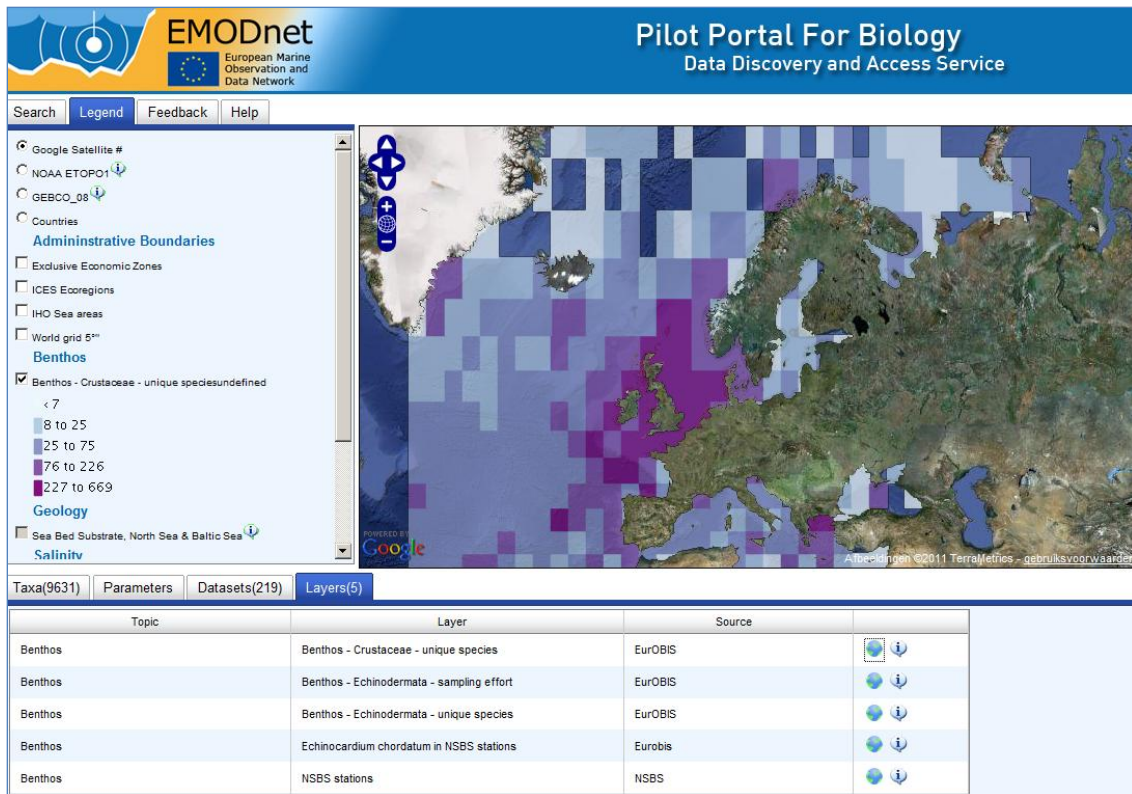
Plotting of all sampling locations where *Nucella lapillus* or the dogwhelk has been documented, based on the data available in EurOBIS in March 2011



When comparing the distribution of the dogwhelk *Nucella lapillus* based on actual observations present in EurOBIS / EMODnet and the general habitat map of the species, we can observe that the data available are matching the generally known European distribution of the species very closely. From this – and on the condition that enough distribution information is available in EurOBIS / EMODnet, we can conclude that the actual observations are a reliable proxy for the distribution patterns of a species within the European marine waters.

General habitat map of the dogwhelk *Nucella lapillus*, from the OCEANA website ([www.oceana.org](http://www.oceana.org)). The dog whelk inhabits middle and lower rocky shores along the Northwestern and Northeastern Atlantic





Visualisation of the data product 'unique number of Crustacea species per 3 by 3 degrees grid-cell (Mercator projection)' based on the data available in EuroBIS on June 3<sup>rd</sup>, 2009



Plotting of all sampling locations of the dataset 'North Sea Benthos Survey - NSBS' (left) and plotting of all the sampling locations where Echinocardium cordatum was found (right). Combining both maps lead to a presence-absence map of that particular species in the North Sea

## Bird communities

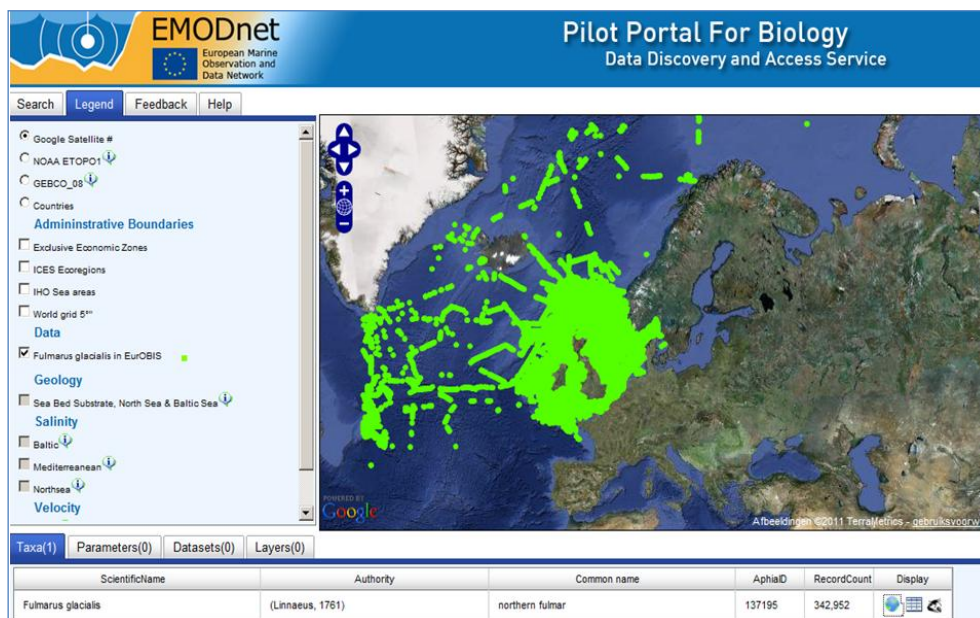
There are currently 16 datasets documented which contain data on birds. For 12 of those, the data can be consulted through the Portal. The available datasets represent 263 taxa, of which 256 are species. The northern fulmar is best documented within EurOBIS. Three data layers (products) have been developed so far on the bird communities, based on the data available in EurOBIS on June 3<sup>rd</sup>, 2009: number of species per grid-cell of 3 by 3 degrees, the sampling effort (defined as the count of all unique sampling locations, taking into account the time of sampling) and the ES50. Hurlbert's Index or ES50 is calculated as the number of distinct species to be present in a random sample of – for example – 50 individuals from a certain area. In this case, the area is set to grid cells of 3 by 3 degrees. All three data products are available on the Portal.



	Common name	# records	Time-frame	# datasets
<i>Fulmarus glacialis</i>	northern fulmar	342 952	1966 - 2008	8
<i>Uria aalge</i>	common guillemot	197 767	1971 - 2008	10
<i>Rissa tridactyla</i>	black-legged kittiwake	144 423	1965 - 2008	9



*Fulmarus glacialis* (left), *Uria aalge* (middle), *Fissa tridactyla* (right). All pictures from [www.marinespecies.org](http://www.marinespecies.org)

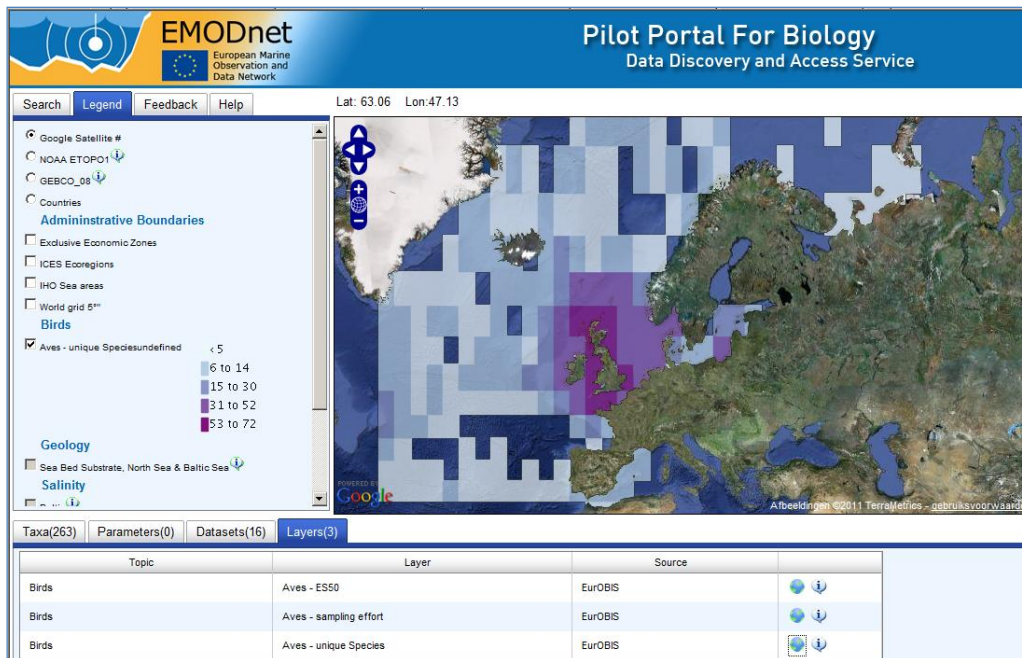


Plotting of all sampling locations where *Fulmarus glacialis* or the northern fulmar has been documented, based on the data available in EurOBIS in March 2011

A comparison of the actual distribution of the northern fulmar *Fulmarus glacialis* based on the information available in EurOBIS / EMODnet and the general habitat map of the species (from [www.oceana.org](http://www.oceana.org)) shows that the observations from EurOBIS / EMODnet fall completely within the areas indicated on the generalized habitat map and confirm the fact that the species does not seem to be present in e.g. the Baltic Sea. The general habitat map also indicates that observations of this species more north still seem to be lacking in EurOBIS / EMODnet.



General habitat map of the northern fulmar *Fulmarus glacialis*, from the OCEANA website ([www.oceana.org](http://www.oceana.org)). The northern fulmar occurs in the North Atlantic, North Pacific and the ice free areas of the Arctic



Visualisation of the data product 'unique number of bird species per 3\*3° grid-cells (Mercator projection)' based on the data available in EurOBIS on June 3<sup>rd</sup>, 2009

## Sea mammals

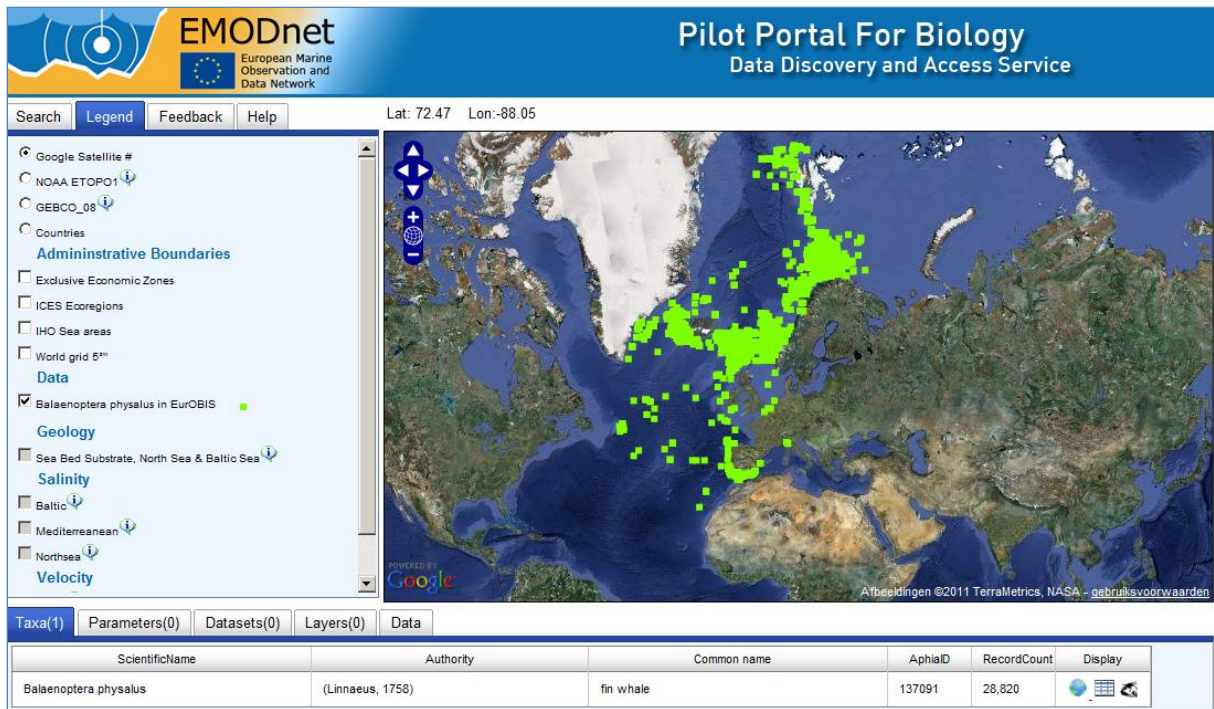
There are currently 16 datasets documented which contain data on sea mammals. For 19 of those datasets, the data is also available online through the Portal, representing over 77 000 distribution records of 44 species. 12 of those datasets deal exclusively with marine mammal data, while the others are of mixed origin. The largest dataset in number of records is managed by the Sea Mammal Research Unit (SMRU) of the University of St. Andrews and contains almost 9 500 distribution records on a number of grey seals that were intensively monitored using Argos Satellite Relay Data Loggers (SRDL). The top-3 of sea mammals consists of two whale species (fin whale and sperm whale) and the grey seal. One data product has been made available so far, a map showing the number of documented species in a 3 by 3 degrees grid-cell in the European marine waters. This map (data product) was created based on the data available in EurOBIS on June 3<sup>rd</sup>, 2009.

Scientific name	Common name	# records	Time-frame	# datasets
<i>Balaenoptera physalus</i>	fin whale	28 820	1971 - 2006	8
<i>Physeter macrocephalus</i>	sperm whale	14 963	1967 - 2004	8
<i>Halichoerus grypus</i>	grey seal	10 408	1980 - 2008	13

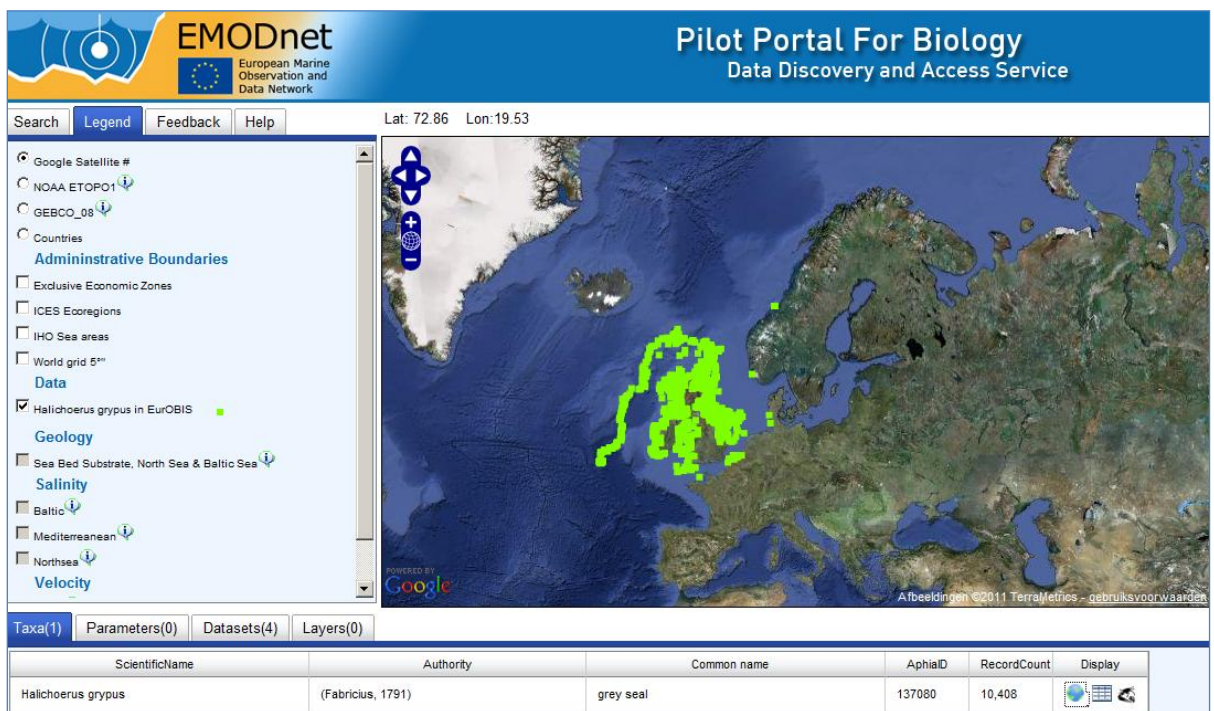


*Balaenoptera physalus* (left), *Physeter macrocephalus* (middle), *Halichoerus grypus* (right)  
All pictures from [www.marinespecies.org](http://www.marinespecies.org)





*Plotting of all sampling locations where Balaenoptera physalus or the fin whale has been documented, based on the data available in EurOBIS in March 2011*

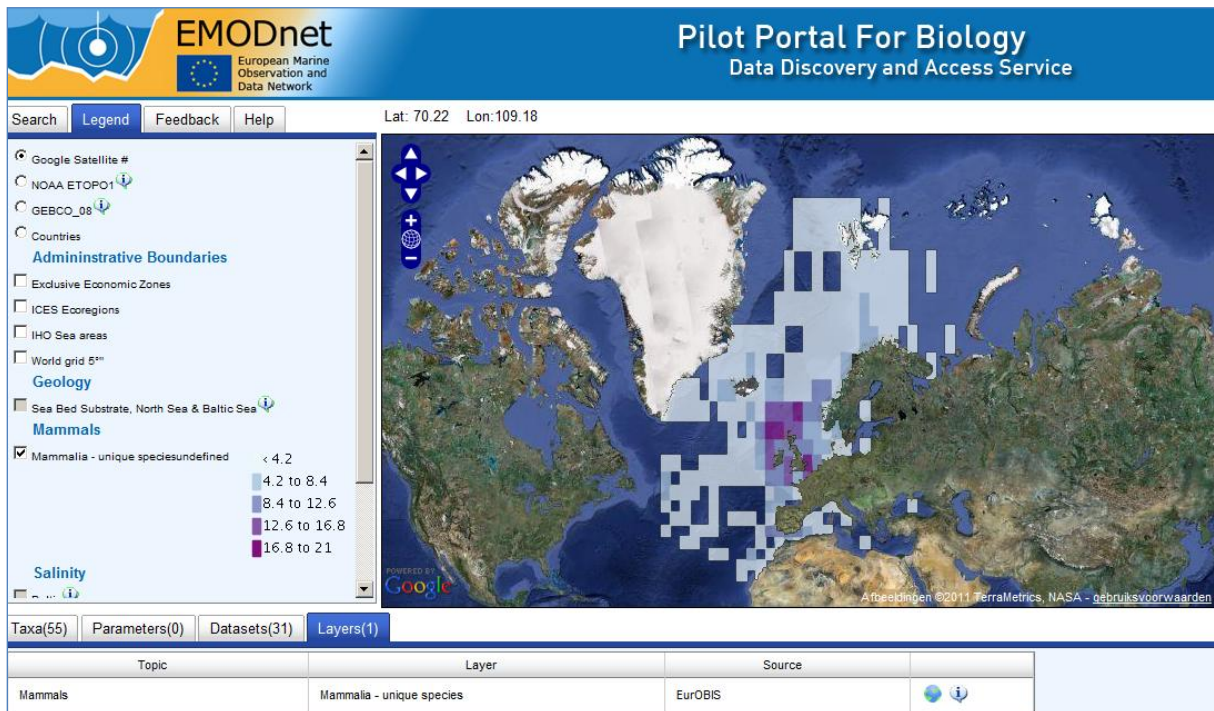


*Plotting of all sampling locations where Halichoerus grypus or the grey seal has been documented, based on the data available in EurOBIS in March 2011*

Comparing the actual distribution of the grey seal *Halichoerus grypus* – based on the information available in EurOBIS / EMODnet and the range map from the IUCN Red List website ([www.iucnredlist.org](http://www.iucnredlist.org)) shows us that there are some important areas where this species occurs which are not yet documented within EurOBIS. The most striking gaps are the Wadden Sea, the Norwegian Sea and the Baltic Sea, indicating that additional efforts should be undertaken to identify relevant datasets from those areas and to convince data providers to contribute their (monitoring) data to EurOBIS / EMODnet. Possible contributors to this gap could be the Trilateral Monitoring Program of the Wadden Sea, the Marine biological data of Sweden (SHARK) and the Danish national database for marine data.



Range map of the grey seal, from the IUCN Red List website



Visualisation of the data product 'unique number of mammal species per 3 by 3 degrees grid-cells (Mercator projection)' based on the data available in EurOBIS on June 3<sup>rd</sup>, 2009

## Reptiles

Up till now, eight datasets containing information on the distribution of reptiles have been documented in the data catalog. For six of these datasets, the data have been made available to EMODnet / EurOBIS. Within European marine waters, only five reptile species have been documented so far, representing almost 4 500 distribution records in the system. Two datasets are exclusively documenting reptile distributions: (1) 'Marine turtles' provided by the National Biodiversity Network Trust (UK) and (2) 'Islas Canarias (Proyecto Aegina: juvenile loggerheads',



delivered by the Instituto Canario de Ciencias Marinas (ICCM, Spain). Both datasets make up for almost 4 500 distribution records or almost 100 % of all reptile-related distribution records.

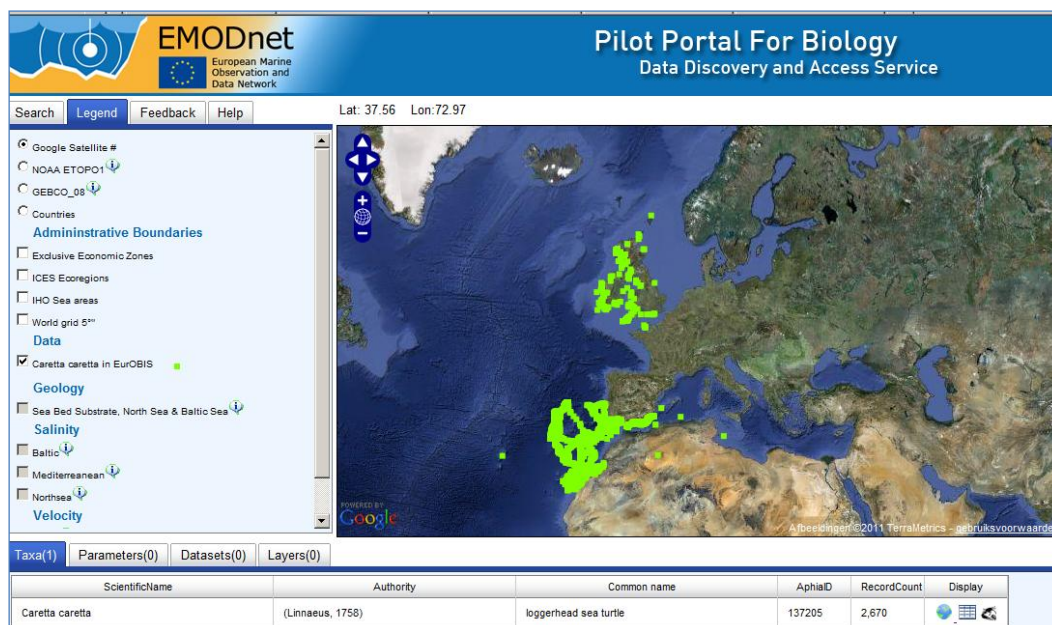
Two data layers or data products are available concerning reptile data:

- Unique number of reptile species present in each 3 by 3 degrees grid cell (Mercator projection) in the European marine waters, based on the data available in EurOBIS on June 3<sup>rd</sup>, 2009.
- Sampling effort in each 3 by 3 degrees grid cell in the European marine waters. Sampling effort is defined as the count of all unique sampling locations, taking into account the sampling moment (e.g. one location visited four times a year will be counted as 4 different sampling events).

Scientific name	Common name	# records	Time-frame	# datasets
<i>Caretta caretta</i>	loggerhead sea turtle	2670	1816 - 2008	4
<i>Dermochelys coriacea</i>	leatherback turtle	1753	1756 - 2007	5
<i>Lepidochelys kempii</i>	Kemp's ridley sea turtle	35	1913 - 2007	2



*Caretta caretta* (left), *Dermochelys coriacea* (middle), *Lepidochelys kempii* (right). All pictures from [www.marinespecies.org](http://www.marinespecies.org)



Sampling locations of the Loggerhead sea turtle *Caretta caretta* within European marine waters, based on the available EurOBIS data (March 2011)

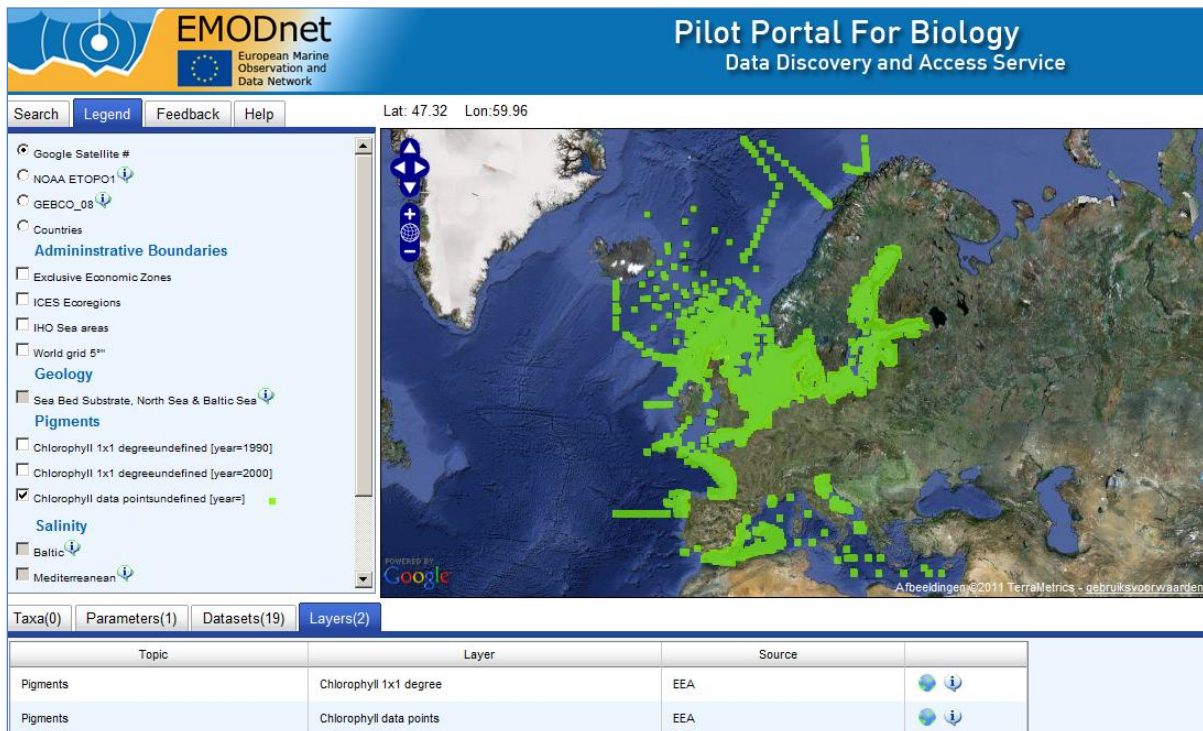
## Pigments

Chlorophyll a data are considered to be very important, as they can be a proxy for the presence and abundance of phytoplankton in a certain area. It is however not always very clear if this should be seen as a chemical or a biological parameter. Due to its relation with the presence of phytoplankton (chlorophyll can be seen as an indicator for phytoplankton biomass), it was decided that the Biological Lot will be responsible for the documentation and gathering of existing chlorophyll data from European marine waters.

The most comprehensive database of chlorophyll a data is stored at the European Environment Agency (EEA): **Waterbase – transitional, coastal and marine waters**. EEA has agreed on making Waterbase available through EMODnet. VLIZ has made two data layers from this database:

- An overview of all data points from which chlorophyll data is available, selection of a certain year is possible.
- An aggregated map on 1°x1° grid cells of the chlorophyll values. This map has an additional filter, so data can be plotted for a given year and season.

Additionally, another 19 datasets documented within EurOBIS / EMODnet have chlorophyll data available.







The screenshot shows the EMODnet Pilot Portal For Biology interface. The header includes the EMODnet logo and the text "Pilot Portal For Biology Data Discovery and Access Service". Below the header, there is a search bar and navigation tabs for "Search", "Legend", "Feedback", and "Help". The current location is displayed as "Lat: 47.32 Lon:59.96".

The left sidebar contains a legend with various categories and layers. Under the "Pigments" category, the following layers are listed:

- Chlorophyll 1x1 degreeundefined [year=1990]
- Chlorophyll 1x1 degreeundefined [year=2000]
- Chlorophyll data pointsundefined [year=]

The main map area shows a satellite view of Europe with numerous green dots representing sampling locations. The map is powered by Google and includes a copyright notice for TerraMetrics.

At the bottom, there is a table showing the layers loaded in the portal:

Topic	Layer	Source	
Pigments	Chlorophyll 1x1 degree	EEA	 
Pigments	Chlorophyll data points	EEA	 

Visualization of all sampling locations of chlorophyll within the Waterbase database of EEA

### 3. Analysis - Lessons learned (to add)

#### 3.1. Main barriers to the provision of data

If we analyze the amount of data that became available during the biological preparatory action (fig.), it is fair to say that a significant number of European marine biological data is now available on the biological EMODnet portal. However the number of datasets increased more moderately (fig). This non linearity between the number of observations and the number of datasets can be explained by the addition of a few big data collections to the system, mainly provided by data collating centers, which were, in most cases, partner in the pilot project (e.a. ICES, OBIS, IFREMER, PANGAEA, EEA...).

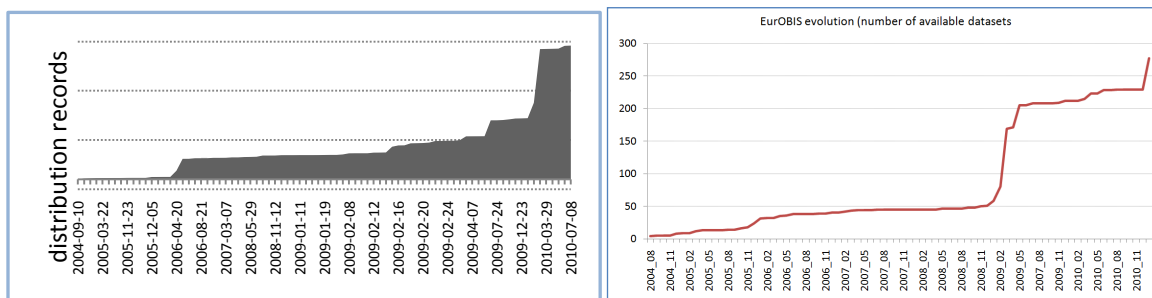


Fig: Number of distribution records vs. number of datasets available

Other large national marine biological data centers and large biological data systems from our focus area (MADS, NERI (DK), SHARK, SMHI (SV), BLMP, DSH (GE), DONAR, RWS (NL), UKDMOS (UK)) were contacted and requested to share their data through EMODnet. Although there was initial interest and in some cases the metadata was shared with EMODnet, data was finally not transferred to EMODnet. The main reason for not sharing the data was of legal nature, by the lack of permission from the environmental agency to share the data on the internet. Sometimes the request to share data within EMODnet was passed to a higher level, or a national panel – but without any further feedback.

The institutes – managing national databases – that have been positive about sharing their data, want to share their data by making use of web-services, which are in several cases not yet developed on their end (work in progress). Once it is explained that they can also send their data to EurOBIS /EMODnet and these web-services are – in a first phase - not really necessary to make their data available through the EMODnet system, no further response is heard. We suspect this has to do with the fact that – by sending a physical copy – they feel they have no further control of their data. This is however not true: data within EurOBIS / EMODnet can be updated at any given time. From this point of view, using web services or sending a copy of the data does in essence not make a difference to the way data are made available. Making use of web services does facilitate the transfer of data from a provider to EMODnet, but is in no way a prerequisite to publish data through EMODnet.



Several datasets resulting from scientific studies, with a small temporal and spatial scope were already available in EurOBIS and accessible for EMODnet. Most of these datasets were gathered within the EU-Network of Excellence MarBEF, a network consisting of 94 European marine research institutes. Through the organization of thematic research projects within MarBEF, several scientific research data became available. Main reason for not sharing scientific data is that data is not published yet, or there are not enough institutional resources to transfer the data into the requested format. Within the EMODnet biological pilot, actions are currently performed to cope with these limitations, trying to mobilize more biological research data from Marine Research Institutes. These actions take into account experiences gathered during the MarBEF project.

- Primo, we send out questionnaires to possible data providers, mainly from the scientific community. From the 116 persons contacted, there was an average response of 30%. Metadata of 106 marine biological datasets was collected through the questionnaires, which can be considered as a good result. Six datasets were actually delivered to the system. We can conclude that working through specific questionnaires is feasible in order to receive metadata information, but becomes more difficult to get the actual data. Specific intensives (see further) will be organized to try to mobilize these datasets into the system.
- Secundo, a set of workshops with marine researchers are and will be organized. In the first year of the project a data products workshop was organized to 1) discuss the marine biological (monitoring) data availability in Europe and gaps and 2) to define a set of derived data products (for example target species maps) relevant for private bodies, public authorities and researchers. The workshop focused on different species groups. 56 participants attended the workshop, including representatives from DG MARE, DG Research, OSPAR, ICES, HELCOM, Black Sea Commission, marine biology, ecology and data management experts. The represented countries were multiple: Greece, France, UK, Italy, Belgium, Netherlands, Germany, Sweden, Ireland, Russia, Ukraine, US and New Zealand. The workshop revealed the striking fact that scientists, policy makers, the private sector and educators are all in need of similar data products concerning marine biological data. All these user groups would like easy access to species distribution maps and visualization of trends in their appearance and their sensitivity and/or vulnerability to certain factors (e.g. oil pollution). There is also a need to classify species in a non-taxonomical way, by attributing them certain characteristics or functionalities ('attributes'). This will make it possible to group species being listed as 'red list species', harmful algal bloom species (HAB), to ascertain if species are under some sort of protection-status (e.g. Annexes of the Habitat and Bird Directive), or just being able to distinguish between benthos and plankton species. The possibility to easily calculate biodiversity indices is also mentioned as being significant.

Based on these outcomes, a data analysis workshop will be organized in the fall of 2011. During this workshop, a number of hypotheses related to the distribution and patterns of European marine biodiversity will be tested, leading to the creation of some of the identified data products. During this workshop, the EurOBIS data (over 14.3 million distribution records) will be used as a baseline, but participants will be urged to include additional marine biological datasets not yet included in EurOBIS. These additional scientific datasets will be

used to increase the data coverage and strengthen the data analysis. Upon permission, these new data will be included in EurOBIS (presence data and/or abundance data). In order to attract scientists to participate in these common analyses and to provide data to the system, we aim to publish the results of this workshop in a high impact journal. The created data products will be visualized through the EMODnet data portal. It is too early to comment on the success of this initiative.

- Tertio, a mini grant data program is set up in the Black Sea Area. In close collaboration with the Institute of Biology of the Southern Seas of the National Academy of Sciences of Ukraine (IBSS), VLIZ has set up a small data grant program to make scientific datasets – still largely on paper - from the Black Sea available through EurOBIS and EMODnet.

During the summer of 2010, IBSS has coordinated a detailed request within Ukraine and other Black Sea countries to inventory marine datasets from the Black Sea Region. This inventory identified 15 datasets representing more than 2.5 million distribution records that can contribute to EurOBIS and EMODnet. The metadata of these datasets has already been described and is available through the EMODnet data catalog. As these data are largely available in paper format, the mini-grants will be distributed to finance the digitization process and the quality control of the data. Contracts have been drawn up in December 2010 and January 2011, clearly stating that all data digitized during this initiative will become freely available online. More specific, the abundance and biomass data will become available through EurOBIS for all data collected more than 5 years ago. For the more recent data (collected less than 5 years ago), only presence data will become available. This gives the scientists the opportunity to analyse their data before sharing them with the scientific community.

Digitization of the first datasets started in January 2011 and the first data are expected to go online by the summer of 2011. Communication on the identification and contribution of more datasets from Turkey, Bulgaria, Russia and Ukraine is in progress. The biological content of the contributing datasets is rather diverse: nine datasets deal with phyto- or zooplankton, 3 give data on Cetacea, 1 on benthos, 1 on seagrasses and 1 dataset on macro-algae. VLIZ is so far investing 37.350€ in this initiative. Given the preliminary successful results of this mini-data-grant program in the Black Sea, conversations will be held with people from the Mediterranean region to start a similar initiative to inventorize and digitize existing datasets so they can contribute to EMODnet and EurOBIS.

Another barrier for making data available is determined by the temporal cover of the dataset. The overview graph of the number of available distribution records shows that it is mostly historical data and very recent data that seem to form a problem to be made available to EurOBIS / EMODnet.

- For the most recent years, this might be explained by a time-lag. Going from the collected samples to a ready-to-use data format is not a trivial task: the identification of species is a manual work which can be very time-consuming, especially when microscopes are needed for the identifications. Once the data matrix is ready, scientists still need time to analyze the data and publish their findings, what can take another 6 months to 2 years time. A time lag of 3 to 5 years thus seems reasonable, although scientists should become convinced of the advantages of sharing/publishing their data as quickly as possible.
- The data gap before 1950 can be explained by the fact that these older data are mostly only available on paper. Transferring data from paper to a digital format is a very time-consuming matter, which not every institute can afford to do (both in terms of money and staff-time).

### **Plan to overcome these barriers**

We conclude that the strategy to involve large data collating centers in the project turned out to be a successful strategy. Large data providers like the ICES datacenter, the PANGAEA datacenter, OBIS and IFREMER transferred data by using the DIGIR webservice (OBIS, PANGAEA), setting up an own webservice (ICES), or by sending a copy of their monitoring data to EMODnet. Datamaps and dataproducts from other EMODnet pilots are made available through the OGC compliant Web Mapping Service (WMS). The possibility to contribute data to this system, by using a set of a technical challenging web services, or by sending the data as a hard copy, has the advantageous that different data providers can make their data available in their most convenient way.

Several institutes managing big national marine biological datasets were positive about the idea but did not contribute data to the system. Reasons for not including varied, but 'need to have an official permission' occurred several times. Trying to involve these national data centres officially in the project could possibly improve the willingness to cooperate. Therefore, EU could for example officially ask these institutes for cooperation, or they could be involved in the project as a legitimate project partner. Through a good cooperation between the OSPAR working group ICG-COBAN (Intersessional Correspondence Group on the Coordination of Biodiversity Assessment and Monitoring) and the biological project of EMODnet, the national bodies, responsible for marine biological data monitoring could be further involved in the project and national monitoring data could be mobilized into the system.

Data collected within the framework of scientific studies, could be mobilized by involving scientists into the project. Besides the general networking to involve scientists (through presentation at scientific seminars), specific data workshops and a data grant program is organized. At the moment, the work through a small data grant program to mobilize scientific data from the Black Sea seems to be successful. The organization of scientific workshops, to mobilize more data into the system still needs to prove its success. Similar activities carried out during the EU MarBEF Network of Excellence have proven to be successful. So far, not much effort was put in order to mobilize data from private sector.

### **3.2. Challenges to rendering data interoperable (including measurement techniques, standards, nomenclature ...)**

The main challenges to render the data interoperable are discussed in the methodology section. The biggest effort is spent on the taxonomic and geographic quality control. Through the European Register of Marine Species (ERMS), the taxonomic quality control is feasible. It is however a constant and huge effort to maintain the taxonomic register up to date, and to match the incoming datasets with the European Register of Marine Species. Other main quality control actions were performed to remove duplicate records, work on the sampling size and the standardization of abundance data, work on standardization of sex, life stage and sample size.

The standard list used for EurOBIS consists out of 74 data fields, the OBIS Schema version 1.1, which is an extension of the Darwin Core 2 (<http://www.iobis.org/node/304>). The OBIS Schema is the content standard used by OBIS and is designed for marine biodiversity data, specifically to records the capture or observation of a particular species at a certain location and time. It can also be used to document specimens from museum collections and literature data. The Scheme lists 74 data fields, of which 7 are mandatory and an additional 15 are classified as highly recommended. All other data fields are optional. An update of the scheme, to include extra biological parameters, could be envisaged.

### **3.3. Challenges to producing contiguous data**

In order to produce contiguous data over a maritime basin from fragmented, inhomogeneous data need to be made interoperable. This is done through the standardization, described in section on methodology. It is a challenge to create the same contiguous data in the different sea basins. For example, a lot of data was available for the North Sea, less for the Bay of Biscay and only few datasets from the Iberian Coast have been identified. This makes it difficult to compare the data granularity species observations between the different sea basins. To produce sound temporal contiguous maps (variation of number of species in time), more monitoring data are required.

### **3.4. Fitness for purpose (measuring ecosystem health)**

Biodiversity and biogeographic information are essential to measure and study the ecosystem health of maritime basins. As a test for fitness for purpose, already a few maps with diversity parameters of the different ICES Ecoregions were created, based on the biogeographic contiguous data from EurOBIS. Measuring the ecosystem health, based on the data will be further analysed during the planned data analysis workshops. Some of the data could also be used to calculate parameters that could support indicators for GES descriptors 1 (biological diversity), 2 (alien species), 4 (abundance) and 6 (health of benthic communities).

### **3.5. Improving accuracy, precision and coverage**

As discussed in the previous section we could improve the coverage of the data through the proposed mechanisms to include monitoring programs, and to mobilize scientific datasets by organization of data products workshops and grant mobilizing programs. Quality controls can possibly increase the precision of the data.

### 3.6. Performance of portal technology

At the moment, there are no serious difficulties encountered with respect to the portal performance. The system is continuously monitored. The used technology of the portal has allows a high performance. Mapping and downloading of less 100,000 records goes relatively fast. The increasing performance of the portal is made possible through the installation of a Geoserver. The communication between the databases and client goes through the Ajax protocol. The section on monitoring and feedback indicates that the user friendliness of the system was highly appreciated with the different users.

## 4. Monitoring effectiveness of portal

The first version of the biological EMODnet portal was demonstrated and discussed during the data product workshop (02/2020). There was a consensus amongst workshop participants that the look and functionalities of the prototype were meeting the requirements. Since then improvements on the performance of the portal were made and some extra functionalities have been developed. In March 2011, the prototype of the portal was launched to several relevant stakeholders across Europe.

### 4.1. Intensity of use

Since 2011, we start monitoring the usage of the data portal. This monitoring will continue throughout the maintenance phase of the project. We can show for now only very preliminary results. Figure show the number of visitors, visits, hits of the website at March 14 2011. We see a clear increase in hits from March 2011, when the portal was pre-launched.

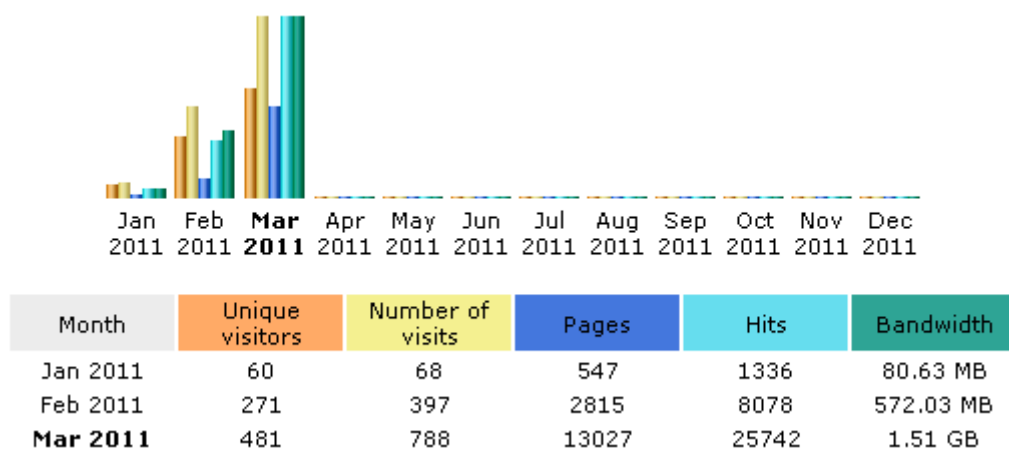


Fig: Number of unique visitors, visits, pages and hits of the biological data portal (17/03/2011)

The table one shows us the number of data downloads that was performed since January 2011. We see that at this moment, 24 datafiles have been downloaded through the EMODnet portal. Person, institute, email, purpose of download, download date and what was downloaded are monitored. At this moment it is early to draw conclusions, but we see that the data have been used for different purposes so far (PHD studies, scientific analysis, biogeographic research, marine training and a few testing downloads). A more profound analysis on the data usage will follow later. The feedback is listed in table two. At the moment 14 feedback comments and one extensive review (not listed here) were received. A detailed analysis on possible improvements of use will be presented once the period for feedback has ended (end of march).

#### 4.2. Possible improvements on ease of use

(to add-after all feedback is collected and analysed)

- Base the search on species attributes. Search on functional groups (benthos, plankton) is already possible, but this search will be expanded to e.g. Red List species, species listed in annexes of e.g. the Bird or Habitat Directive. All these attributes will be stored in the World Register of Marine Species (WoRMS, [www.marinespecies.org](http://www.marinespecies.org)) and linked to EMODnet (this is a starting exercise, not sure if it will be feasible in the proposed time frame).
- Inclusion of the visualization of the temporal component of data. This will give a clear indication of the observations over time.
- Data should be visualised by abundance: the size of the dot should be in relation to the plotted value (abundance or biomass). The more individuals, the larger the dot. This is optional and not sure if it will be feasible in the proposed time frame. It is also dependent on the amount of abundance/biomass data available and the success of the data standardisation of these data (see further).
- An email alert system through rss-feeds will be included in the portal allowing the users to see what new metadata have been included and what new datasets have been integrated into the EurOBIS system.
- Offer more possible file-formats when downloading data from the Portal (txt, xml, kmz, ...)

organisation	Email	Country	Purpose	download_date	data
NULL	<a href="mailto:santiago.alvarezfernandez@wur.nl">santiago.alvarezfernandez@wur.nl</a>	Netherlands	PhD	2011-01-24 13:10:40.517	Dataset: Historical hyperbenthos data (1987-2001) from the North Sea (35153)
NULL	<a href="mailto:santiago.alvarezfernandez@wur.nl">santiago.alvarezfernandez@wur.nl</a>	Netherlands	PhD	2011-01-24 13:11:14.360	Dataset: Historical hyperbenthos data (1987-2001) from the North Sea (35153)
nioz	rob.louws@nioz.nl	Netherlands	Informative	2011-02-01 14:37:40.423	Dataset: macrobenthos in the Dutch Sector of the North Sea 1991-2001 (4663)
nioz	rob.louws@nioz.nl	Netherlands	Informative	2011-02-01 14:38:19.343	Dataset: macrobenthos in the Dutch Sector of the North Sea 1991-2001 (4663)
Netherlands Environ Assessment Agency	rick.wortelboer@pbl.nl	Netherlands	Scientific purposes (Governmental institution)	2011-02-02 15:16:24.733	Dataset: Seasearch Marine Surv...
Netherlands Environ Assessment Agency	rick.wortelboer@pbl.nl	Netherlands	Scientific research (Governmental institute)	2011-02-02 15:17:44.443	Dataset: Seasearch Marine Surv...
OBIS	Brook.Herlach@iobis.org	USA	Integrating into iOBIS	2011-02-03 18:42:31.920	Dataset: Biocean (29954)
NULL	brook.herlach@iobis.org	USA	Integration into iOBIS	2011-02-03 18:43:20.300	Dataset: Biocean (29954)
NULL	brook.herlach@iobis.org	USA	Integration into OBIS	2011-02-03 18:43:52.637	Dataset: Biocean (29954)
NULL	brook.herlach@iobis.org	USA	Integration into iOBIS	2011-02-03 18:44:28.043	Dataset: Biocean (29954)
NULL	brook.herlach@iobis.org	USA	Integration into OBIS	2011-02-03 18:45:03.780	Dataset: Biocean (29954)
AWI	hannes.grobe@awi.de	Germany	validation routines	2011-02-08 09:36:50.230	Chlorofyl a in Åµg/l in water(123929)
VLIZ	ward.appeltans@vliz.be	Belgium	Test	2011-02-21 11:47:21.680	Observations of Tisbe furcata (31)
University of Vigo	ritagonzalez@uvigo.es	Spain	For classes	2011-02-22 12:37:34.493	Dataset: Taxonomic Information System for the Belgian coastal area (22150)
Malone O'Regan	mpurcell@morce.ie	Ireland	Research	2011-02-28 13:17:26.770	Dataset: BioMar - Ireland: ben...
Marine Institute	liam.caffrey@marine.ie	Ireland	Test	2011-03-04 11:16:57.850	Dataset: iziko South African Museum - Shark collection (162)
PANGAEA	spesant@marum.de	Germany	Portal Evaluation	2011-03-04 16:30:33.470	Dataset: PANGAEA - Data from Glacial Atlantic Mapping and .. (GLAMAP2000) (16689)
PANGAEA	spesant@marum.de	Germany	Portal Evaluation	2011-03-04 16:46:19.033	Dataset: PANGAEA - Data from Benthic biology and geochemistry (BENGAL) (1937)
PANGAEA	spesant@marum.de	Germany	Portal Evaluation	2011-03-04 16:47:41.050	Dataset: PANGAEA - Data from Benthic biology and geochemistry (BENGAL) (1937)
university of auckland	<a href="mailto:m.costello@auckland.ac.nz">m.costello@auckland.ac.nz</a>	new zealand	Biogeographic research	2011-03-05 08:25:45.897	Observations of Amphipoda (6828)
IODE trainer	m.brown.nsb@gmail.com	USA	marine data mgt. training	2011-03-09 13:23:00.943	Observations of Abramis brama (358)
IODE	m.brown.nsb@gmail.com	USA	marine data mgt. training	2011-03-09 13:27:47.213	Observations of Abramis brama (358)
IPIMAR	ernesto@ipimar.pt	Portugal	Testing for MODEG	2011-03-10 14:48:08.910	Observations of Trachurus trachurus (72025)
IPIMAR	ernesto@ipimar.pt	Portugal	Testing for MODEG	2011-03-10 14:48:44.240	Observations of Trachurus trachurus (72025)

Table 1: Monitoring of the downloads of the EMODnet portal



Name	Email	Message
Gisbert Breitbach	<a href="mailto:gisbert.breitbach@hzq.de">gisbert.breitbach@hzq.de</a>	I would advice to add examples of use cases into the manual. For me as a non-biologist without specific questions it is not quite clear how I could use the portal meaningfully.
gaynor evans	<a href="mailto:gaev@bodc.ac.uk">gaev@bodc.ac.uk</a>	Really effective portal. Intuitive, easy to use and fast. The reliance on using google maps is surprising as they are not really open source.
Keith Hiscock	<a href="mailto:khis@MBA.ac.uk">khis@MBA.ac.uk</a>	All possible efforts needed to obtain additional records/datasets as records are very patchy. Perhaps some targetted requests to scientists in, for instance, Svalbard. A facility is needed to display the presence of a species in a geographical area but without a precise latitude and longitude for a single site - then regional fauna and flora lists can be incorporated quickly. A facility is needed to 'challenge' suspicious records (for instance, Eunicella verrucosa in the Isle of Man) and to feed those challenges back to the data holder/provider. I will use EMODnet for contextual information on European distribution of a species but will use searchnbn.net for UK as I can 'dig deeper' int the records.(I have not explored the site in detail.)
Dr. Melckzed eck K. Osore	<a href="mailto:babaalmasi@yahoo.co.uk">babaalmasi@yahoo.co.uk</a>	So far so good - but the zoom seems too sensitive even at rather low scale. Otherwise excellent view
James G Wilson	<a href="mailto:jwilson@tcd.ie">jwilson@tcd.ie</a>	Search for 'benthos' and 'Dublin' only gave 2 records of Dublin Bay Prawn. If that's all there is, then it would help to have some indication of limits.
Mark Costello	<a href="mailto:m.costello@auckland.ac.nz">m.costello@auckland.ac.nz</a>	Very nice, works reasonably quickly from home computer; Some comments: colours for species richness and observations might avoid blue as that is also used for the sea; not ommediately obvious that one can pan the map and clicking the map causes it to ozoom (pointer does not change when over map); Maybe one should have indication of what GEBCO and ETOPO are (e.g. bathymetry); What are 'countries' - do not show national boundaries on land, and are not EEZ equivalent for sea (including territorial waters); Statistics says it includes species and observations but not clear which is given - seems to be only number of observations; Bottom tabs menu option did not show a search box, and what one can search on (e.g. what are Layers of); Now I think I see I need to have top tab Search on first, then to use Bottom tabs. Aora typica does not occur in Europe but is a valid species from south PAcific, it has been confused with Aora gracilis in Europe; thus A typica in Europe is almost certainly all A graciliis - not sure how you flag this to readers; I'll keep playing around with it.
Alessandro Saccà	<a href="mailto:asacca@uni-me.it">asacca@uni-me.it</a>	There should be more concordance in the species names. For example, I wanted to know the distribution of the copepod Acartia clausi, and for doing this I had to make separate searches for: A. clausi, A. clausii and A. (Acartiura) clausi, since they are managed as different species by the database. Fortunately, the signals on the map do not disappear unless you do not press the "Reset map" button, which is very useful for other purposes too.
M. Sonnewald	<a href="mailto:msonnewald@senckenberg.de">msonnewald@senckenberg.de</a>	The search and map display works pretty fast, but I would suggest to implement the automatic display of the associated data when moving the mouse over a data point in the map.
ivan cantani	<a href="mailto:ivan.cantani@unibo.it">ivan.cantani@unibo.it</a>	it's a optimal tool for scientists and researchers,easy to use and graphically ok.
Henry Vallius	<a href="mailto:henry.vallius@gtk.fi">henry.vallius@gtk.fi</a>	The data portal seems to work well after a short trial. Alopex lagopus, Rodentia and other land living mammals appear in MARINE Observation and Data Network. Is this a mistake or is there a good reason for that? I can understand that one might see arctic foxes on sea ice, but you could also see elks, lynx or wolves as well but they are not included in the data set. It seems that it is too difficult to open some data sets (biota and fish at least). Are the data sets too large so that I would need to zoom in before trying to access the data?
Carol Ogborne	<a href="mailto:Carol.ogborne@gov.bc.ca">Carol.ogborne@gov.bc.ca</a>	friggin awesome site. I wish my organization had the \$\$ to do this here. Can I ask how long and how much this type of a portal cost to develop. Thank you
Murray Brown	<a href="mailto:m.brown.nsb@gmail.com">m.brown.nsb@gmail.com</a>	I've just spent some time playing with your new website for BIO data, and want to tell you how great it is. I'm very impressed by the flexibility and rubustness of the graphical interface, especially as it relates to the "moveable map". This is a significant achievement, and you should all be very proud of it.
PERGEN T-MARTINI	<a href="mailto:pmartini@uni-v-corse.fr">pmartini@uni-v-corse.fr</a>	I saw with a great interest your portal, but according to my experience in seagrasses, it seems that there are some mistake in species distribution (for instance, Posidonia oceanica species on french atlantic coasts), and i have found difficulties to identify the source for this wrong signalisation. The only link is to algaebase or WoRMs but without indication of the document (or author) responsible of this citation. A direct link between the plot on map and the source of information will be very useful Best regards
Mirko Hauswirth	<a href="mailto:mirko.hauswirth@bfv-vilm.de">mirko.hauswirth@bfv-vilm.de</a>	Great tool! Do you think about providing a WMS-Service? Congatulations!

<p>Stéphane PESANT</p>	<p><a href="mailto:spesant@marum.de">spesant@marum.de</a></p>	<p>I am overall very pleased with the portal. Please find attached a list of comments that could improve the portal. I hope you will find them useful. Do not hesitate to contact me if you need clarifications There is often confusion about EMODnet and EurOBIS throughout the portal. Although the backbone of both EMODnet and EurOBIS is the same, and even though it was first developed for EurOBIS, we should not really refer to EurOBIS as a backbone for EMODnet. On the other hand, when EurOBIS is a provider of layers, then of course it should be identified as the source. The links to Metadata » Data catalog » Submit dataset, currently in a box at the bottom of the Search space could be reformatted as higher level navigation buttons above the map. The link to Metadata » Statistics should not be a higher level navigation button, but could be displayed only in the Legend space. On the Map, you can click on individual data points and show its metadata in the Map Features section. However, that section does not provide the citation for the data point. As explained in other comments below, the InstitutionCode is not sufficient. Regarding the mapping process, every time the map is updated, it would be good to resize (zoom) the map to the geographic extent of the entire data records being mapped. In the Map Features section it would be useful to be able to delete features or to compile them in a single table instead of generating a new table for each feature In the Map Features section it would be easier to read if we rounded off the values for depths, lats and longs to perhaps two or three decimal places... for display purposes only... not modifying the raw data. In the Map Features section the taxonomic information is given as Isid, whereas in the Data section, the taxonomic information is given as AphiaID. We should be consistent throughout the portal. In the Legends space, it would facilitate the navigation to be able to expand/collapse the categories of features (e.g. Administrative Boundaries). This would also apply to additional categories corresponding to the Themes. In the Legends space, it would facilitate the navigation to group the first three layers under the category Background Map. Also, I am not sure what the layer Country# is supposed to show. If it is meant to display country boundaries, it should be moved to the category Administrative Boundaries In the Legends space, it would facilitate the navigation to group the categories Geology, Salinity and Velocity under a common category Environmental Conditions (in prep). In the Legends space, it would be useful to have info bubbles for the ALL layers, including the thematic ones. Also, the info bubbles for the Salinity and Velocity layers are dead links. In the Legends space, the option Number of species and observations in EurOBIS appears twice. I suspect there should be one for number of species and one for the number of observations. Also, it is not clear which spatial polygon system was used to compute/display the statistics... although we can find out that it is the IHO Sea areas. The title should be Number of species per IHO Sea Areas and Number of observations per IHO Sea Areas In the Search space, it would be useful to have a thingy (e.g. wheel, dots, message) showing that the system is processing the request and thus clearly indicates to users that they should wait as the Layers, Datasets, Parameters and Taxa sections are being filled. In the Search space, thematic searches lead to results in sections Taxa, Parameters, Datasets and Layers. However, free text search (e.g. chaetoceros, chlorophyll, ICES, picophytoplankton) will return results from either Taxa, Parameters, Datasets or Layers sections, based on text matching. It would be useful to "fill in" all sections based on the results of the free text search. For example, if you search for "coscinodiscus" you get 28 records in the Taxa section, but the Datasets section is empty. It should in fact contain several datasets. Currently you can figure out this list for each of the 28 record by displaying data in the Data section and clicking on citation, but this is really not practical. When searching using the term "Pigments", the Parameters section allows plotting locations on the Map, to view values in the Data section and to download values, and the Datasets section provides more information on the source of the data. However, there is no way to link data to sources or to map, view or download subsets of all sources. Linking data to sources is essential. In the Datasets section, several datasets have no records listed, e.g. about 40/218 in the case of Benthos datasets... this is the case for example of PANGAEA and ICES data. If a dataset has no records, it should probably not be listed. But, knowing that both PANGAEA and ICES have lots of data on Benthos, I suspect this is a granularity issue. This should be resolved so that the actual datasets and their corresponding record numbers are displayed in the dataset list. In the Map Features and the Data section, there is a column "InstitutionCode" and "Institution", which very often have for value an initiative or a data centre/publisher which is not an institution. I would suggest to use the column title "Source" in both cases, as is done in the Layers section. In the Data section, a link is provided at the top to see a compiled list of citations for the data included in the Table. The citations were previously displayed for each line of the data table (this is really the preferred way), but I understand that it was increasing the time to generate the table online. However, the citations for each line of the data table must be included in the download version... otherwise we completely lose the capacity link data to its citation (authorship) or to the sampling and analysis methods that vary among value. Citing the InstitutionCode (i.e. in most cases the data centre/publisher) or providing a link to the EMODnet data catalogue is not enough. Unique identifiers are attributed to data for that purpose (e.g. DOIs at PANGAEA, CDIAC, BODC) and must be used. PANGAEA cannot distribute data if this</p>
----------------------------	---	--

		<p>requirement is not fulfilled by dissemination portals.</p> <p>In the Data section, the column "Observed Ind. #" is misleading (and in many cases wrong) because in many cases the units are not simply "numbers". This field was meant to describe a number of specimens in a jar and not abundances, but in many cases the information displayed in that column is abundance with units "numbers" "per area" or "per volume". Similarly, when one downloads the data from the portal, no units are given for that field, and yet this information is provided to EMODnet (at least by PANGAEA). This is a known issue and should be given a high priority. A pragmatic way to solve the issue is to include a column for units in the Data section and in the download tables.</p> <p>In the Data Catalogue, dataset descriptions show urls where other replicates/versions of the dataset are available, but there is no url linking to the dataset being described. This url must be added.</p> <p>In the Data Catalogue, we should use the wording "replicated" and "replicates/versions" instead of "copied" and "copies" when referring to replicated datasets. The latter wording could be misinterpreted.</p> <p>The first time I tried to download data, Firefox blocked pop-ups required for the download to happen... it took me three attempts to figure it out... perhaps this is unavoidable, but I thought of mentioning it as some users might get the impression that the download function does not work.</p>
--	--	--

Table 2: Feedback on the data portal

## **5. Recommendations for the overall EMODNET**

Based on the work carried out during the first two year of the preparatory action of biology, we can formulate some general recommendations for the overall EMODnet and a few specific recommendations for the biological work of EMODnet.

### **5.1. Sustainability**

The biological project welcomed the communication Marine Knowledge 2020, in which a common architecture and governance model for marine observations and data across the EU is proposed. We consider the thematic approach, initiated in the preparatory actions and proposed through the creation thematic assembly groups, as very positive. Data and monitoring programs are in most cases thematically organized, and could best be assembled using the same thematic approach. Essential is a sustainable EU funding to maintain the data management (including quality control & standardization) of the thematic data. A general remark on the proposed architecture is the fact that many biological data are not available or archived at national data centres, these data risk to be not envisaged within the proposed architecture. The proposal of a secretariat, which could give technical support for an overall EMODNET, including the development of a portal that could access all different thematic data is good and will even be essential for the user in order to find, access, assemble and apply data efficiently and rapidly.

### **5.2. The model for governance by actors in the system**

The governance model of EMODNET should aim at involving all relevant actors; these include the large data collating centers, national organizations in charge of the monitoring of the marine environment, the private industry and the scientific research community. Besides the EMODNET system developers, a formal membership with data providers could optimize access to the large monitoring data collections through EMODNET. These data providing institutes could become part of a sort of 'general assembly' of the EMODNET thematic assembly groups. The model should be flexible enough to involve also the 'small dataproviders'. The proposed model from the biological pilot to mobilize data through small data grants or by the organization of thematic workshops to introduce new data into the system and to create derived dataproducts are promising. A regional approach to include these 'local' datasets is the most appropriate approach (cfr. Black Sea data grant program).

Important for the future biological thematic assembly group will be the (operational) link between the upcoming Marine Strategy Framework Directive (MSFD), and the data monitoring that will be required through this directive, and the accessibility of these European biological marine monitoring data through EMODNET.

### **5.3. Availability of standard procedures facilitating data flow**

The biological project built its portal and project upon the EurOBIS/OBIS standards and procedures for data flow. This data flow includes the data providers, the integration, standardization and quality control and open access and free redistribution of quality controlled

data. This was a successful approach for the biological data. We points on the importance of the freedom of use for publicly funded data, one of the main EMODNET principles. It is in that context that the biological data portal tries to provide the data open, freely available and accessible. In addition to the existing EurOBIS/OBIS standards, the interoperability between the biological portal and other data systems (including other lots), was made possible through the implementation of OGC standards.

#### 5.4. Future activities for the biological project

Based on the experiences gathered during the biological project, specific priorities for a biological thematic assembly group in 2011-2013 can include:

- Use and focus on data & information identified during the gap analysis that is being performed by the biological lot.
- Focus on data products identified during the biological data products workshop (e.a. species attributes). Tagging and collecting data from relevant species using specific attributes - functional groups, HAB's, invasive species, red list or protected species, species relevant to Marine Framework Strategy Directive, Habitat Directive, Bird Directive.
- Taxonomic quality control is key in order to integrate biological data. Further effort in updating the European Register of Marine Species by taxonomic editors, by increasing the taxonomic coverage (including lower organisms, bacteria...).
- Increase number of abundance and biomass data, standardize abundance and biomass data.
- Incentives to mobilize data providers through hosting and organization of biological thematic workshops, aiming at the identification and integration of biological data at European scale.
- Explore the possibilities to work with genomic data (explore the possibilities, explore scalability of the system, explore how to add biological data generated by new technologies)
- Optimize data capture processes by the development of new IT tools
- Integration with other data portals (physical data, chemical data...)