Building a European Marine Observation and Data Network ‘EMODnet’: Biological Data Analysis Workshop

VLIZ Special Publication 56

25 – 27 October 2011

Hellenic Centre for Marine Research
GR-Heraklion
Building a European Marine Observation and Data Network ‘EMODnet’: Biological Data Analysis Workshop

VLIZ Special Publication 56

SUGGESTED CITATION


CONTACT INFORMATION

Flanders Marine Institute - VLIZ

Innovocean site
Wandelaarkaai 7
B-8400 Oostende
Belgium

Phone: ++ 32 (0)59 34 01 59
Fax: ++ 32 (0)59 34 21 31
E-mail: simon.claus@vliz.be

ISSN 1377-0950
Contents

1. Introduction ..................................................................................................................................................................3

2. Data availability for the workshop ..........................................................................................................................4

   2.1. The European Ocean Biogeographic Information System (EurOBIS) ..........................................................4

   2.2. QC procedures of EurOBIS .....................................................................................................................................7

   2.3. European Seabirds at Sea Database (ESAS) ......................................................................................................8

   2.4. SCANS – Small Cetaceans in the European Atlantic and North Sea datasets .................................................9

   2.5. Other available data resources ..........................................................................................................................10

       Seabed substrate data .............................................................................................................................................10

       Seabed habitat maps ............................................................................................................................................10

       Geographical boundaries ........................................................................................................................................11

       FAO distribution maps ...........................................................................................................................................12

       World Ocean Database ........................................................................................................................................13

       Human impacts ......................................................................................................................................................13

       Delivering Alien Invasive Species Inventories for Europe (DAISIE database) .................................................13

       Wish list of participants ........................................................................................................................................14

2. Proposed research topics ........................................................................................................................................15

3. Breakout groups .........................................................................................................................................................19

   Group I: Using presence data as a proxy for calculation biodiversity indices, absence data and probabilities of occurrences ..................................................................................................................................................19

   Group II: Predator-prey relationships: trends in space and time ..............................................................................22

   Group III: The effect of invasive species on the European marine biodiversity ....................................................25

   Group IV: Biogeographic classifications ....................................................................................................................28

4. Conclusions .................................................................................................................................................................30

   References .................................................................................................................................................................32

   Appendix: Participants ................................................................................................................................................34
1. Introduction

The Flanders Marine Institute (VLIZ) organized a workshop on biological data products in Oostende, Belgium over 25th-26th February, 2010. This workshop was organized within the framework of the European Marine Observation and Data Network, EMODnet, launched by the Maritime Policy of the European Commission. User groups from the scientific community, European marine policy, regional sea commissions and coastal and marine practitioners identified a number of biological data products relevant for science, policy and practice (Claus et al., 2010). In different user discussion groups, four sets of marine biological data products were identified as priority products:

- biodiversity indices maps
- species distribution maps and trends
- species sensitivity and vulnerability maps
- species attributes (functional groups, HAB’s, invasive, red list species,...)

Based on these observations, EMODnet Biology organised from 25th to 27th of October 2011 in Heraklion, Crete a 3-day Biological data analysis workshop to test a number of hypotheses related to the distribution and patterns of European marine biodiversity, by creating some of the above-mentioned data products. During this scientific workshop, a selected group of marine scientists explored the EurOBIS data (> 14,000,000 distribution records of which 40% have abundance data) as a baseline, but also additional marine biological datasets not yet included in EurOBIS were used. These additional data could be used to increase the data coverage and strengthen the data analysis. The data analysis workshop had following objectives:

- to create biological data products identified during the data product workshop
- by creating these data products
  - test a number of scientific hypotheses related to biodiversity indices and the distribution and patterns of European marine biodiversity
  - identify and integrate new relevant biological data holdings for EMODnet

During this scientific exercise, the fitness for purpose of the biological data was evaluated. In addition to the biological data, the other EMODnet projects provided data layers (physico-chemical data, geological data, habitat and seabed maps) that could be used to strengthen the analysis and provide the necessary abiotic background data and information.
2. Data availability for the workshop

2.1. The European Ocean Biogeographic Information System (EurOBIS)

To give the participants an idea of the content of EurOBIS, a general overview of the available information was given, both in terms of the used field names and the content of those fields. At the time of the workshop, 304 datasets were available within EurOBIS, representing some 14.7 million distribution records. Users of the data have to keep in mind that the data are very heterogeneous, which can cause a bias when combining and comparing data (Vandepitte et al., 2011).

Data come from different sources: the majority is the result of research activities (57%) – typically short term, less than 5 years – directly followed by monitoring data (27%; long term activity, more than 5 years). In addition, EurOBIS also harbours data from – amongst others - museum collections or more general distribution records derived from literature. Apart from the origin, there is also a great variety in content: researchers focus on different (species) groups, leading to a wide variety of available species data. Over 40% of the datasets available in EurOBIS represent distribution information on fish, whereas less than 2% of the datasets contain information on marine mammals or reptiles. Other major groups that can be distinguished are phytoplankton (4%), zooplankton (9%), benthos (10%) and macro-algae & plants (4%). For some datasets, the distinction between planktonic and benthic species is harder to make: these have been grouped together and represent about 21% of the available distribution records.

Fig 1: Datasets available in EurOBIS split up according to their origin and datasets in EurOBIS split up according to their functional groups/ taxonomic content
Table 1: Number of datasets, taxa, species and observation records in EurOBIS per species group

* The number of datasets listed represents the datasets of which the majority of the data belong to a specific group. When a species of this group is also present in another dataset (with another focus), the total number of datasets containing data on relevant species is mentioned in brackets.

** Taxon names are not yet unambiguously assigned to either benthos or plankton, as for this exercise one should also take into account the life stage of a species. A species attributes workshop is planned, where this issue will be addressed and dealt with in the future. The number of records for these groups is estimated from the number of records of the datasets assigned with the keywords “benthos” or “plankton”.

Content-wise, a number of graphs were presented, indicating the amount of records that could be used for certain purposes. If someone for example wanted to perform analyses that involve the abundance of species at a location, they should realize that only about 50% of the records in EurOBIS are available for this kind of analysis. For species distribution related analyses, 80% of the records qualify. Graphs were presented showing the available data and precision in terms of taxonomy, time and geography.
EurOBIS has a lot of data available that can be used in different kinds of analyses. However – as is always the case in large integrated databases – one has to keep in mind that there are some limitations to the data use. As some analyses for example require the use of abundance or biomass data, a lot of records in EurOBIS cannot be used simply because of the absence of such data. Sample size and sampling methodology – documented at metadata level – are also necessary to check whether certain data are comparable and needed to be able to standardize the selected data to an arbitrary sample size. Regarding taxonomic precision, not all records contain species information. Also, not all species known to be present in the European marine waters are already documented within EurOBIS. There are also some geographical gaps in the data, most notably in the Mediterranean Sea. The gap in the Mediterranean is currently being addressed through the organisation of a mini-data grant program to mobilize relevant marine data from the area. From a temporal point of view, seasonal analyses can only be performed using the records that have detailed temporal information available (day-month-year). There is also a known gap in the temporal scope of the EurOBIS database: data from the pre-1950s are hard to obtain, as they are mostly only available in paper format and it takes a large effort to digitize these data and prepare them for submission. On the other hand, researchers are very protective on recently collected data.

Fig 3: Taxonomic precision: indication of how many records contain specified taxonomic information, Temporal precision: indication of the level of detail on sampling time for each record, Geographical precision: the level of precision for the given coordinates and for the amount of records which are documented.
(post-2005), as they first want time to analyse these data and publish their results before sharing them with other scientists. As mentioned before, the very diverse origin and content of the contributing data-sets might give a certain bias in the sampling intensity and methods that are documented within EurOBIS.

2.2. QC procedures of EurOBIS

An automated routine has been developed to check the completeness of an EurOBIS record and to detect possible (import) errors. Each record undergoes 22 different quality checks, in which each check is a question to be answered by ‘yes’ or ‘no’. The results of these QC steps are stored in an additional field in the database, enabling users to filter data based on whether they meet the defined quality and completeness standards. Records not meeting the QC checks are sent back to the provider(s) to ask for clarification or additional information. A distinction is made between basic quality controls and additional quality controls. The basic QCs are seen as the ‘cut-off’ for a record and indicate whether it is fit for publication on the portal or not.

Basic quality control steps of EurOBIS (EurOBIS cut-off):

1. Are the fields defined as ‘required’ completed?
2. Is the given taxon name documented in WoRMS?
3. Is the taxon level/rank lower than ‘family’?
4. Are latitude and longitude different from zero (0)?
5. Is latitude between -90 and +90 and is longitude between -180 and +180 (in WGS84)?
6. Are sampling locations situated in sea or along coastline (20 km buffer; to implement)?
7. Is at least the year of data collection completed?

<table>
<thead>
<tr>
<th># records where basic QC steps are positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC_1</td>
</tr>
<tr>
<td>------------</td>
</tr>
<tr>
<td>QC_1</td>
</tr>
<tr>
<td>100</td>
</tr>
</tbody>
</table>

Fig 4: Results basic QC steps of EurOBIS. QC 6 is not implemented yet so no data available
Extra quality control steps of EurOBIS:

8. Is the taxon a non-marine taxon documented in WoRMS?
9. Are the sampling locations within the expected geographical area?
10. Is the given date (year, month, day) possible?
11. When a start and end date are given, does start date precede end date?
12. If time is given, is this number different from 0 and smaller than 24 and is time zone completed?
13. Does ‘Basis of record’ follow the proposed OBIS-codes?
14. Does ‘Sex’ follow the proposed OBIS-codes?
15. Is the ‘Observed individual count’ empty or >0?
16. Is the ‘Observed weight’ empty or >0?
17. If the ‘Observed individual count’ is >0, is the ‘sample size’ documented?
18. If depth values are given, is the minimum depth < maximum depth?
19. If a depth value is given, is this a possible value compared to depth map?
20. If a depth value is given, is this a possible value compared to the depth range of that taxon?
21. Dataset outlier: is the sampling location situated within 4 times the standard deviation from the centroid sampling location of the dataset?
22. Species outlier: is the observation located within 4 times the standard deviations from the centroid of the observations of that species?

The extra QC steps related to the correctness of the depth values, the land-sea distinction of the geographical positions and the outlier analysis are still under development. A regular execution of these QC procedures will help in mapping the quality of the records available within EurOBIS. Additionally, it will also be able to show the progress made in upgrading the quality of the available records through contact with the original providers.

2.3. European Seabirds at Sea Database (ESAS)

The European Seabirds at Sea Database was established in the early 1980s. ESAS is managed by an informal organisation, mostly supported by 1 to 2 teams per involved country. The database holds the results of both ship-based and aerial seabird surveys in the Northwest European waters, using a common format. Data collection is driven by projects, not by regular surveys and all data are collected using standard methods (Camphuyzen & Garthe, 2004).

The ESAS database contains about 3 million records of seabird and marine mammal observations since the early 1980s. Data is collected both from airplane and ship platforms. The bird density is estimated by
following a transect method: where the area = Transect width * length. Data is collected by institutes from Germany, Belgium, Netherlands, Denmark, UK and Portugal. The ESAS database has a strong focus on the North Sea, but also contains data from other European marine waters.

![ESAS Database v5 (Oct 2011): Observation effort second quarter (Apr-Jun)](image)

2.4. SCANS – Small Cetaceans in the European Atlantic and North Sea datasets

The Small Cetacean Abundance in the North Sea and Adjacent waters (SCANS) survey was conducted in the North Sea and related areas. To estimate absolute abundance, a combination of shipboard and aerial surveys were carried out in 1994 and 2005. Both visual and acoustic survey methods are being used to detect cetaceans on the shipboard surveys. The SCANS I data are already available in EurOBIS. The SCANS II data were at the moment of the workshop being integrated into EurOBIS.

![Harbour porpoise estimated density surface (animals per km²) in 1994 (left) and 2005 (right). Source: SCANS II report (available at http://www.vliz.be/imis/imis.php?module=ref&refid=210295).](image)
The workshop participants could get insight in the graphs and maps of the SCANS II report, comparing the observations made during the first SCANS (1994) with the second SCANS (2005). These maps have proven to be very useful in the (re-)formulation of the proposed topic 6 (see further).

2.5. Other available data resources

Seabed substrate data

The EMODnet Geology project created a freely available seabed substrate map of the North Sea, the Baltic Sea and the Irish Sea at a 1:1 million scale. This shapefile -created by EMODnet Geology- provides information on the seabed substrate for the defined regions and is based on a combination of different maps. The smallest cartographic unit on this map with a scale of 1:1 million corresponds with areas of about 4 km². Harmonization of national substrate categories into one classification scheme is essential for interoperability. Therefore, the sea-bed surface-sediment maps that were not originally in the Folk classification system were reclassified. 4 substrate classes on the basis of the modified Folk triangle (mud to sandy mud; sand to muddy sand; coarse sediment; mixed sediment) were used. This map is very useful if one seeks to analyze for example the benthic communities, at European sea basin scale.

![Image of seabed substrate map](image)

*Figure 7: Sea-bed substrate map of the EMODnet-Geology study area, visualized through the Biology Portal*

Seabed habitat maps

The EMODnet Pilot Portal for broadscale modeled seabed habitats (EU SeaMap) has made available predicted (modeled) broadscale seabed habitat maps for over 2 million km² in the Celtic Sea, North Sea, Baltic Sea and Western Mediterranean Sea. The data layers that were needed to create this model include the seabed substrate, seabed energy, biological zones and seabed salinity. These layers were then
combined to produce EUNIS habitat maps for the defined region. All these seabed habitat layers can also be consulted at the EMODnet Biology Portal, again making it possible to investigate links between the presence of species and the seabed habitat.

Fig 8: Overview of different data layers necessary to come to a predictive seabed habitat map and the EUNIS levels.

Figure 9: Sea-bed habitat maps of the North Sea, Baltic and Western Med., visualized through the Biology Portal

Geographical boundaries

Different geographical boundaries are available for geographic and biogeographic analysis: the European Exclusive Economic Zones (EEZ), the IHO sea areas and the ICES Ecoregions, the Longhurst Biogeographical Provinces, Large Marine Ecosystems of the World (LME), Marine Ecoregions of the World (MEOW) and FAO Fishing Areas.
FAO distribution maps

In cooperation with FAO, a number of their ‘aquatic species distribution maps of interest to fisheries’ have been made available as layers through the EMODnet Biology Portal. The main source of information of FAO to create these species distribution maps was their habitat descriptions and geographical distributions derived from official FAO Catalogues of Species. The main FAO purpose was to help in the identification of areas of higher biodiversity or species associated with vulnerable marine habitats. Together with the available species information in EMODnet/EurOBIS, these FAO maps can be validated or complemented and possible gaps can be identified, giving an added value to both systems.
World Ocean Database

Abiotic information such as depth, temperature, salinity, primary productivity can be retrieved from the World Ocean Database (WOD), managed by the National Oceanographic Data Centre (NODC) of the National Oceanic and Atmospheric Agency (NOAA). The World Ocean Atlas WOA consists of objectively analyzed global grids, at standard of temperature, salinity, dissolved oxygen, AOU, percent oxygen saturation and different nutrients. All these data are freely available (http://www.nodc.noaa.gov/OC5/WOA09/pr_woa09.html).

Human impacts

Halpern et al. created a global map representing the impact of humans on marine ecosystems (Science 319, Feb 2008). This map is an ecosystem-specific, multiscale spatial model and synthesizes 17 global datasets dealing with anthropogenic drivers of ecological change for 20 marine ecosystems. This map is available as a raster file (www.nceas.ucsb.edu/globalmarine).

![A Global Map of Human Impacts to Marine Ecosystems (Halpern et al., 2008) zooming at the North Sea](image)

Delivering Alien Invasive Species Inventories for Europe (DAISIE database)

The DAISIE database gives an inventory of invasive species that threaten European terrestrial, freshwater and marine environments. It was compiled under the DAISIE project, funded by the European Commission under the Sixth Framework Programme (Contract Number: SSPI-CT-2003-511202). It contains information of more than 2500 invasions of marine species in European regional seas.
Wish list of participants

There seems to be a need to have a map available, showing the (most important) shipping lanes in the Greater North Sea, as there might be a link between these shipping lanes and the appearance and/or abundance of certain species, more specifically for marine mammals and birds. Such a map is only available from the European Space Agency and can only be downloaded as a .gif or .jpg file (http://www.esa.int/esaEO/SEMBDI0OWUF_index_1.html). A GIS-layer would be highly appreciated by different users.

Fig 13: European shipping routes from ESA website (http://www.esa.int/esaEO/SEMBDI0OWUF_index_1.html)
2. Proposed research topics

A set of eight research topics and relevant hypotheses was selected in preparation of and was presented to the workshop attendees. The selection of topics was made taking by into account the following criteria: (1) the results of the analysis should generate data products of biodiversity indices and species distribution maps, (2) the topics should lead to the identification and integration of new biological data holdings not yet available in the system and (3) the hypothesis should be tested using the EurOBIS integrated database, therefore facilitating the usage of the integrated database by the scientific community.

The proposed topics were grouped into two main categories, corresponding to two important sets of biological data products that were identified during the EMODnet data product workshop in February 2010 (Claus et al., 2010).

- **Category I**: Can we identify diversity patterns, hotspots (using diversity indices) and macro ecological trends in European Marine Waters based on EurOBIS data?

- **Category II**: Can we observe (predicted) species distribution patterns and trends of some selected higher taxa based on available data (a.o. Common guillemot, Fulmar, Harbour porpoise, Harbour seal)?

**Topic 1**: Identify latitudinal gradients of biodiversity and hotspot of European biodiversity.

Diversity of benthic communities or of component groups has been shown to decline with increasing latitude in the northern hemisphere. Other regional studies, however, have failed to detect this pattern or even showed regionally slightly the opposite trends (Renaud et al., 2009). To illustrate patterns of species diversity, diversity indexes should be calculated on integrated data holdings. In any case large databases are of utmost importance to answer questions across regional spatial scales.

- **Available data**: > 14 million occurrence records; > 7 million abundance records; environmental data: bathymetry, seabed substrate, temperature, salinity, primary production.

- **Possible methods**: Data selection, aggregation, calculation of indices, interpolation and validation.

- **Possible output products**: Grid map of diversity index; Full coverage map of diversity index.
**Topic 2:** Can we verify or document the existing systems for biogeographic and/or managerial zonation of the European Seas (e.g. LEOW, LME, IHO, ICES Ecoregions...) by using the observation biological datasets of EurOBIS.

The Marine Strategy Framework Directive states that marine regions and their subregions have, according to Art. 3(2), been identified by taking into account hydrological, oceanographic and biogeographic features. Use of such factors should help support the ecosystem-based approach to management of human activities that is required under Art. 1(3). The Directive pursuant to Art. 3(1) provides a definition of marine waters. It does not provide any specific geographic boundaries or coordinates delimiting boundaries between marine regions or their subregions.

- **Available data:** Shapefiles of different marine classification: e.g. MEOW, IHO, ICES, LME...
- **Possible methods:** Compare patterns of alpha, beta, gamma diversity; and endemicty against area for different classifications. Comparing classification strength of biogeographic classifications based on species composition similarities.
- **Possible output products:** Documentation of the systems proposed for the zonation of the European Seas for biogeographic and managerial purposes useful for MSFD implementation.

**Topic 3:** Presence data can be used as proxy in the calculation of biodiversity (indices). While abundance data is generally a component of most diversity indices, the use of the number of records for a particular species as a proxy for its abundance often forms the common practice on databases with missing abundance information (Ardron et al., 2009). Scientific evidence for this surrogacy could significantly increase biodiversity research on existing integrated databases.

- **Available data:** > 14 million occurrence records; > 7 million abundance records
- **Possible methods:** Correlation between index values calculated on occurrence data vs those calculated on abundance data
- **Possible output products:** Map of diversity index based on occurrences; Map of diversity index based on abundances; Initial step that allows for wide range of mapping possibilities based on occurrence data
**Topic 4:** Based on the EuroBIS database, analyse and detect macro-ecological patterns and relationships across European marine waters. Given that human activity can disrupt even macro-ecological relationships, it would be valuable to include measures of European anthropogenic activity in macro-ecological analyses (Webb et al., 2009).

- **Available data:** EurOBIS data; Map on Human impact on oceans (fig 15)
- **Possible methods:** Occupancy/density relationship vs human impacts
- **Possible output products:** Maps of biologically disturbed areas of the European Waters

**Topic 5:** A great deal of our current scientific debate goes on the homogenization of biota caused by the native species loss and their successive replacement by the exotics (many losers, few winners), but there is not much scientific evidence at the regional/continental scale. Can the available data provide us with the scientific evidence required to conclude on the issue of whether species inventories in different regions have undergone homogenization (lower biodiversity than in the past) through time? This could be likely explored by focusing on benthic inventories (as they emerge from the EurOBIS datasets) and examining temporal trajectories in biodiversity indices (also by taking into account major human influences and predicting major homogenization in areas heavily disturbed).

Do the changes in biodiversity caused by the presence of exotic benthic species vary with latitude? There is a paper by Sax 2001 on this subject, but it is focused on terrestrial/freshwater species. Therefore, it might be worth assessing whether there is a latitudinal pattern in the distribution of introduced species in the marine realm. There is strong theory that would let us predict a smaller number of introductions towards higher latitudes.

- **Available data:** EurOBIS benthic data, DAISIE European invasive checklist
- **Possible methods:** Calculate taxonomic and diversity indices with and without invasive species taxa
- **Possible output products:** maps of invasive vs non invasive EurOBIS distributions, changes of invasive distributions in time.

**Topic 6:** Can we explain distribution shifts of higher organisms (eg harbour porpoise) being triggered by local reductions or shifts in principal prey availability (eg sand eel) based on prey–predator data (Camphuysen C.J. 2004).
Topic 7: Can we observe (changes in) temporal trends in population size and population condition of higher organisms based on available data? Northern Fulmars are clearly more abundant in summer than in winter. The distribution of this species is closely linked to the Central North Sea Water while the Continental Coast Water is usually avoided (Garthe, S., 1997).

Topic 8: How could we validate model-based, European-scale predictions of some higher taxa, based on available data (Kashner et al. 2006, Guisan & Zimmerman 2000).
3. Breakout groups

Four breakout groups were formed. In each group one or more proposed topics were selected for further analysis. While selecting the topics, specific attention was given to take into consideration

- the added value and the unique characteristics of the EurOBIS data
- the relevance for different users and stakeholders
- the possible creation of useful data products that can be visualised on the EMODnet Portal

Group I: Using presence data as a proxy for calculation biodiversity indices, absence data and probabilities of occurrences.

Peter Herman, Edward Vanden Berghe, Stéphane Pesant, Sarah Faulwetter, Evangelos Pavilis, Christos Arvanitidis, Francisco Hernandez

The EurOBIS database contains over 15,000,000 distribution records of marine species in European waters. However absence data are at this moment not stored in the database and abundance data are not always available. However, abundance data or the number of individuals of a given species over a given sample size, can be critical when calculating different biodiversity indices. In order to make best use of the available data, this working group analyzed if and how presence data can be used to calculate biodiversity indices. This can be done either by making use of indicators that only require presences and no absences, by restricting the database to those records that contain abundances, or by using presence/absence data to estimate (relative) abundance. Absences are not explicitly recorded in the databases, and there are many reasons why a species can be unrecorded in the data base of a particular area. The species can be truly absent, it can be present but not recorded because it has not been looked for, or present and not recorded due to random chance in sampling (especially for rare species). This working group investigated the possibility to estimate the probability of occurrence of a species based on presence data only.

1. Using presence data for calculating biodiversity indices

The basic idea of a diversity index is to obtain a quantitative estimate of biological variability. It is important to distinguish ‘richness’ from ‘heterogeneity indices of diversity’ which include both richness and the relative abundance of species. There are numerous such diversity indices. A suitable index that is relatively insensitive to observation bias and therefore valuable to assess diversity patterns on integrated databases is the Hurlbert’s index (ES50), calculated as the number of species in a random subsample of the available data. The formula of the ES50 is:
where \( N \) is the total number of individuals in a sample, “\( i \)” is the species index and \( S \) is the total number of species recorded in the sample.

While abundance data is normally an essential component of the ES50 value, we could use the frequency of the species (relative number of samples where the species is present) as a proxy for its abundance and calculate the ES50 value per gridcell (Ardron et al., 2009) on the basis of these proxy values. However, this method does not take into account the specificities of the individual contributing datasets. Some datasets contain many distribution data on a limited number of species only. The beam trawl monitoring surveys, with millions of records but only few species recorded, are an example. Obviously, when a large number of records from such a data set is available in a particular grid cell, this will influence the recorded frequency of the species: species that have not been part of the species set in the beam trawl monitoring will have a low frequency overall, even if they were present in all samples that explicitly looked for them. Reversely, even species that were only moderately frequent in the beam trawl monitoring will appear to be much more dominant than very frequent species in poorly represented data sets. Some correction for this bias is obviously needed if one is to use frequency as a proxy for abundance. Therefore we propose to calculate the weighted ES50 value, providing a weighted value over the individual contributing datasets. A first approximation that could correct for this bias is based on separate calculations of ES50 per data set. Within each of the \( k \) datasets, an ES50 is calculated for gridcell \( j \) based on the relative frequency of the species within their own dataset. Per dataset we also calculate an overall ES50 (noted ES50i,Set) over the entire domain where the dataset is found, using again frequency as a proxy for abundance. We then estimate the ES50 for grid cell \( j \) as:

\[
ES50_j = \frac{\sum_{i=1}^{k} ES50_{i,j}}{ES50_{i,\text{Set}}}
\]

where \( ES50_{i,\text{Set}} \) is the \( i \)th ES50 value of dataset \( i \) over its entire domain and \( ES50_{i,j} \) is the \( i \)th ES50 value of the “\( i \)th” dataset in grid cell \( j \).

The method validation still has to be done. A possibility would be through simulation in an artificial dataset. The weighted ES50 European data grid including the different EurOBIS datasets can be calculated as a data product to estimate the diversity of European marine waters, based on the EurOBIS data. This index thus weights each dataset equally in the final derivation of the ES50j. This can give undue weight to relatively small or unreliable datasets, for which some care should be exerted.
2. Using presence data to generate absence data

Species distribution models can be constructed from estimates of the environmental niche they occupy. For example, Aquamaps calculates environmental envelopes of marine species with respect to depth, salinity, temperature, primary productivity, and its association with sea ice or coastal areas and then determines the relative suitability of specific geographic areas for a given species (Kashner et al., 2007). Most modern algorithms only require presence data as input for niche modeling. A tacit assumption is, usually, that the dataset is extensive enough to have looked for the species outside the environmental range where presences have been recorded, and that efforts have not been distributed too unequally over space. Model results are presented in the form of some probability of occurrence of the species in geographic space or alternatively in niche space. In order to validate these results, some form of absence data are needed. It should be known at least where the species has been looked for but not recorded. While EurOBIS is the largest database in Europe of marine species observations, real absence data are very scarce. Absence data could be produced from presence data by means of mathematical techniques in some cases, better known as pseudo-absence data (Van Der Wale et al., 2009). We investigated if and how the EurOBIS database could be used to generate the so-called pseudo-absence data for marine species.

We propose to generate pseudo-absence data within a specific dataset. The absence data for a species that occurs in a dataset will then correspond to all the locations or sampling stations from that dataset where the species was not reported. For a specific species, occurring in EurOBIS, we can then combine the absence data from the different datasets where the species occurs to create the pseudo-absence map of the species based on the presence data from EurOBIS. Following SQL selection query on the EurOBIS database returns the geographic coordinates of pseudo-absences of species with APHIAID=148776, taking into account only those datasets where species with APHIAID=148776 occur:

```
SELECT round(a."Latitude",0) as lat,round(a."Longitude",0) as lon FROM eurobis a WHERE a.imis_dasid IN (SELECT distinct imis_dasid FROM eurobis_geo WHERE "AphialID"=148776 ) AND a."AphialID" <> 148776 GROUP BY round("Latitude",0),round("Longitude",0), imis_dasid
```

where imis_dasid = the unique dataset identifier, APHIAID = the unique taxa identifier and Latitude, Longitude = the geographic coordinates of the EurOBIS observations.

This methodology still needs to be validated. There is definitely bias since it focuses on the species included in the EurOBIS database and not on the entire inventory of a certain area/sector. Also, a certain degree of bias comes into the picture because of the specific sampling gear, either quantitative or qualitative. The dataset containing both presence and absence data from the Skagerrak experiment (SKAGEX) could be an appropriate dataset for this purpose. An on-the-fly SQL view in the dataportal will be able to visualize the pseudo-absences for some species.
Group II:  Predator-prey relationships: trends in space and time

Eric Stienen, Stefan Garthe, Eunice Pinn, Leen Vandepitte, Bart Vanhoorne & Frederica Camisa

The group investigated trends of mammals and seabirds in space and time separately. For the purpose of the analyses, a grid of ¼ min (horizontal) and ¼ minute (vertical) of the Greater North Sea has been created.

Partim mammals

The original idea was to link the distribution of harbour porpoises (*Phocoena phocoena*) in the Greater North Sea to the shifts in principle prey availability (whiting, herring, sand eel and sprat). The data to be used for this analysis would be the harbour porpoise data from the ESAS database, in combination with distribution data of the different prey species from EurOBIS. Relevant environmental data – salinity, SST and transparency – would be obtained from the World Ocean Database (WOD).

A seasonal approach would be considered: the data would be aggregated for the months June through August, as winter data are rather scarce and the geographical coverage varied during that time of the year. The seasonal (summer) data would also be aggregated over several years (blocks of 2 - 3 – 5 – 7 years). Similar aggregations would be made for the defined prey species: whiting, herring, sand eel (*Ammodytidae*), gobies (*Gobiidae*) and sprat (*Sprattus sprattus*). After careful consideration, a slightly different approach was defined and the analyses will be subdivided into two parts. Firstly, the recent publication of the harbour porpoise distribution map based on SCANS II gave rise to the idea to try to reconstruct the distribution of the harbour porpoise in the Greater North Sea between the first and the second SCANS survey – based on a combination of the ESAS and EurOBIS data – to see if the shift which is apparent from the SCANS surveys was a one-time shift over 11 years or if multiple shifts in distribution have appeared during this time period. Once this is clear, it can be investigated if this shift(s) are related to prey availability or not. The seasonal approach remains, but data will be considered per year or in a combination of several years (2-3-5). Secondly, once these maps have been developed, it will be investigated whether these shifts can be linked with a shift in prey availability. The same methodologies for data aggregation will be used for the prey species as for the harbour porpoise data. Possible expected results might be that there is a potential shift in prey, followed by the spatial shift in predator, as the predator will follow his prey. By following the prey however, it might be possible that the predator discovers new areas with an increased variety in available prey species and thus not follows the original prey shift completely. If these expected links would not show or not be satisfactory, a link with available environmental data can be made, to investigate whether the geographical shift in predator abundance has to do with prevailing environmental conditions.
Fig 17: Maps based on preliminary analyses of the available ESAS harbour porpoise data. Left: summer data (June-August) of 1995; right: summer data of 2003. Green indicates low estimated density per grid cell, red indicates high estimated densities per grid cell.

When creating the maps based on the SCANS, ESAS and EurOBIS data, the ‘absence’ has to be taken into account: the map should be able to show the difference between a grid that was visited, but where nothing was observed and a grid that was not visited. Possible data products from this exercise could be seasonal or yearly density/surface plots of the harbour porpoise, showing the shifts in their distribution since the early 1980s. Additional maps will be released in spring 2012 in the UK on the distribution of common dolphin, white-beaked dolphin, bottlenose dolphin, minke whale and Risso’s dolphin. It can then be negotiated whether these maps can become available through EMODnet as a data product. The data of the second SCANS survey will also be made available through EurOBIS/EMODnet during 2012.

Partim seabirds

Also here, the link between a predator species and its prey species would be examined. Data would also be seasonally aggregated, grouping all data from June until August. The temporal trends in population size and density would then be compared with the distribution patterns and density of the relevant prey species and linked with environmental variables. Three bird species were selected, and a list of relevant prey species and environmental variables was compiled.

<table>
<thead>
<tr>
<th>Prey</th>
<th>Fulmar glacialis Northern fulmar</th>
<th>Uria aalge Common guillemot</th>
<th>Melanitta nigra Common scoter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calanus finmarchicus</td>
<td>- Sprat</td>
<td>Small bivalves such as Spisula or Ensis</td>
</tr>
<tr>
<td>Environmental</td>
<td>- Salinity</td>
<td>- Sand eel</td>
<td>- Current speed</td>
</tr>
<tr>
<td>variables</td>
<td>- Transparency (Secchi)</td>
<td>- Herring</td>
<td>- Water velocity</td>
</tr>
<tr>
<td></td>
<td>- Water temperature</td>
<td></td>
<td>- Water depth</td>
</tr>
<tr>
<td></td>
<td>- Sea level height?</td>
<td></td>
<td>- Shipping lanes</td>
</tr>
</tbody>
</table>
The main focus during the workshop was on the northern fulmar. Available distribution data from the ESAS database were grouped in 10-year time periods to check the geographical distribution. Data were then combined with either environmental data (e.g. salinity from the World Ocean Database, WOD – see map below) or the corresponding prey species. It was decided not to look into the patterns of the common scoter, as the necessary data are not (readily) available. For the prey bivalves for example, information on length is necessary as the common scoter specifically feeds on small individuals. Size-related information is however not captured in EurOBIS. The relevance of shipping lanes should also not be underestimated and such a map is not available in a GIS-format.

![Fig 18: Northern fulmar distribution in June – August and aggregated for the year 1990-1999 in combination with sea surface salinity for the same period (retrieved from WOD)](image)

For the northern fulmar, information on the presence of its prey species – *Calanus finmarchicus* – is available from the Continuous Plankton Recorder dataset in EurOBIS. The absence of the prey species was reconstructed based on the available location information of the CPR within EurOBIS. All transects were plotted on the northern fulmar distribution map, indicating the presence and absence of *C. finmarchicus* in a station. Based on this map, five ‘boxes’ were identified where both northern fulmar and the prey species had a good coverage. These five boxes will be used in further more detailed analyses, where it will be investigated if there is a significant link between the distribution of the northern fulmar and its prey *Calanus finmarchicus*. As there is a documented shift in the appearance of *C. finmarchicus* towards *C. helgolandicus*, it should be examined if the observations of both species should be taken into account in relation to the geographical distribution and abundances of the northern fulmar. It is noted that a link with environmental variables might be necessary to fully explain certain trends.
Fig 19: Left: observations of northern fulmar combined with the CPR-transects. Right: identification of 5 boxes for trend analyses, where there is a good overlap between the CPR transects and the available northern fulmar distributions.

Fig 20: Graphs showing the probability of occurrence of Calanus helgolandicus and C. finmarchicus in the identified boxes, defined as the number of Calanus sp. per number of total Calanus observation during summer for a specific area. Data have been broken down into 5-year periods (based on the available CPR data in EurOBIS).
There is a lot of discussion on the homogenization (biodiversity smothering) of biota caused by the native species replacements by the exotics (many losers and few winners), but so far there is not much scientific evidence. Therefore, we could explore the available data and include a temporal component that may allow us to analyze whether taxon inventories in different regions have undergone homogenization through time. This could be likely explored by focusing on benthic species lists and examining temporal trajectories in biodiversity indices, taking into account major human influences and perhaps predicting major homogenization in areas heavily disturbed. We can also look if the results of the presence of exotic benthic species vary with latitude. Sax et al. (2001) focused on terrestrial/freshwater species; it might be interesting assessing whether there is a latitudinal pattern in the distribution of introduced species in the marine realm. There is strong theory that would let us predict a smaller number of introductions towards higher latitudes.

Therefore, based on the EurOBIS presence data we could explore following hypotheses:

- $H_0$: At pan European level, taxonomic distinctness does not change with the inclusion of the invasive species in the corresponding species lists
- $H_0$: Invasive together with non-invasive species will follow the same latitudinal trends as the latter category

The working group focused on the first hypothesis and determined a methodological approach to test the hypothesis.

- The records in EurOBIS should be flagged as invasive or not, rather than flagging species. This could be done using the Alien Invasive Species Inventories for Europe (DAISIE DB).
- Use habitat as covariate in analysis; by flagging the species as benthic or pelagic. This information is partly available in the World Register of Marine Species.
- Calculate taxonomic distinctness and other biodiversity indices, including and excluding the invasive species.
- Perform both temporal (times series) and spatial analysis
- Look for additional “quick and easy” macroecological analyses (e.g. Rank-occupancy curves)
The group started with a data selection on the EurOBIS database. In a first phase, they looked only at the benthic distribution data of EurOBIS. Therefore, from the entire EurOBIS database (distributions between Lat: 90/26-; Long -45/70), only those taxa which are flagged as ‘benthic taxa’ in WoRMS were selected. This resulted in 8,543 unique benthic taxa occurring in EurOBIS. From that list, we selected the species occurring in the DAISIE database, which resulted in 169 species, considered as invasive.

![Fig 21.: Geographic distribution of species present in the Alien Invasive Species Inventories for Europe DB](image)

We then used this data to check the taxonomic distinctness. The taxonomic distinctness is a biodiversity index which is based on species relatedness and calculates both the average 'distance' as well as the variation in distance between all pairs of species in a sample. The distance is defined as the path length through the phylogenetic tree, provided by the classification tree of the World Register of Marine Species, connecting these species. Preliminary results indicate that there are indeed differences in taxonomic distinctness, when including information of invasive species at the pan-European scale. When the invasive species are not included in the total species lists then the “sample” of their species list is left outside the expected distribution, as refers to the variation in taxonomic distinctness. On the contrary, when this species list of the invasive species is included in the total pan-European one, then, the “sample” with the invasive species is included within the expected distribution of the values of the variation in taxonomic distinctness index.

A strong assumption we made was the consideration that a species was invasive for Europe, when it was present in the DAISIE database. By testing the same hypothesis on a smaller scale, for example testing the hypothesis by regional sea (DAISIE flags species as invasive in the Atlantic Ocean, Azov Sea, Baltic Sea, Barentz Sea, Baltic Sea, Caspian Sea, Mediterranean Sea, North Sea), by grid or even a by sample we would increase the precision of the results.
Group IV: Biogeographic classifications

Mark Costello, Klaas Deneudt, Thanos Dailianis, Christos Arvanitidis

The biological EMODnet portal provides access to the largest biogeography data collection that has ever been available for European marine waters. This creates a unique opportunity to investigate whether, based on the available biological observation data, it is possible to distinguish the most biologically meaningful biogeographic marine classifications.

Marine Ecosystems of the World (MEOW), Large Marine Ecosystems (LME), Longhurst provinces and ICES Eco-regions are some of the classifications that are currently in use to delineate marine areas. Most of these classifications have been proposed based on geographic and physical characteristics (Longhurst), geo-political (LME), or by expert judgment (MEOW) rather than based on biological data. Providing indications on the most optimal marine classifications could inform policy and support ecosystem-based management. For example in the current implementation of the Marine Strategy Framework Directive a classification, where it is proposed to use a classification based on the ICES eco-regions.

Fig 22.: Three commonly used marine region classifications: ICES Eco-regions, Marine Ecosystems of the World (MEOW), Longhurst Biogeographical Provinces

For the validation of the biogeographic classifications the workgroup considered two separate approaches. The first approach looked into obtaining a classification based on similarity in species composition. A second approach assesses the classification strength of the biogeographic classifications by looking at the endemcity (as % of all species that only occur in a region) against surface area of the regions for each of the classifications.

Fig 23: Preliminary map of c-square clustering based on species composition
Obtaining a classification based on similarity in species composition involved performing a cluster analysis on the data for 1 degree latitude-longitude grid cells (c-squares). Grid cells falling within the same clusters distinguished at a certain level of similarity are given the same color code on a map. The resulting map is simplified by re-coloring those grid cells that are isolated by surrounding differently colored cells. As a final result the map gives a data driven classification of the area under consideration. The initial results suggest some bias by the spatial distribution of particular datasets for which comparable datasets do not occur in other areas. This exercise is to be repeated at different levels of similarity and further validation of the results is needed before the outcome can be served as a reliable product. A parallel approach was to use the SIMPROF test of statistical significance in PRIMER-E to cluster grid cells that were significantly different from others. Unfortunately the computational power available was not sufficient to perform this analysis.

Following the second approach three existing classifications are validated based on the endemicity of their regions. For this analysis the data was extracted from the data system and observations were linked to so-called pseudo stations (unique locations). Spatially joining the pseudo stations with each of the regions in the three classifications, the percentage endemic species per region could be calculated.

The preliminary results did not show one of the three classifications scoring significantly higher to endemicity than the others, but these results will be further discussed and a planned re-analysis on specific subsets of the data should produce more reliable results.
4. Conclusions

During the EMODnet Biological data analysis workshop we assessed the feasibility of developing biological data products as demonstration cases for EMODnet Biology. These data products were identified during a previous EMODnet data product workshop. We focused on the creation of data products on species distribution maps and trends and on biodiversity indices. A set of eight research topics, likely to result in data products, was presented at the workshop. Four topics were analyzed in more detail. While selecting the topics, specific attention was given to take into consideration the added value and the unique characteristics of the EurOBIS data, the relevance for different users and stakeholders and the possible creation of useful data products that can be visualised on the EMODnet Portal. The four working groups focused on:

- Using presence data as a proxy for calculation biodiversity indices, absence data and probabilities of occurrences
- Predator-prey relationships: trends in space and time
- The effect of invasive species on the European marine biodiversity
- Biogeographic classifications

The EurOBIS database proved to be a very valuable data resource to calculate value added data products. The presence data could be used to calculate other derived biological parameters, useful for system analysis like the calculation of absence data or diversity indices. The variety of different contributing datasets gave rise to new possible correlations, for example between predator (mammals or seabirds) and prey (fish, zooplankton) datasets. As a very valuable side-effect of the workshop it should be noted that a significant amount of new data sources were identified and mobilized into the system through the organization of this workshop. The workshop generated a number of concrete results with immediate potential for use and visualization on the EMODnet data portal. For some of the workshop outcomes further research is needed before results are usable or publishable on the portal. This is to be expected as the workshop generated ideas and methods, and considerably more time would be needed to conduct the analyses to demonstrate the potential and policy value of the data. This exercise also triggered new scientific questions. Some of them will be analyzed in more detail by the workshop attendees.

The invited scientists and experts contributed in many ways to the further development of the EMODnet pilot portal:

- providing feedback on the fitness for purpose and suggestions on further developments
- validation and scientific substantiation of data product generation
- stimulate data usage by scientific community and thereby generating new science
The workshop confirmed the involvement of the scientific community to be invaluable for future EMODnet activities. Finally the workshop also confirmed some of the findings of the earlier data products workshop, like the importance of biological information on functional and structural species attributes and the need for an EMODnet workshop initiating a structural capture and compilation of this information.
References

Ardron Jeff, Daniel Dunn, Colleen Corrigan, Kristina Gjerde, Patrick Halpin, Jake Rice, Edward Vanden Berghe, Marjo Vierros (2009). Defining ecologically or biologically significant areas in the open oceans and deep seas: Analysis, tools, resources and illustrations. A background document for the CBD expert workshop on scientific and technical guidance on the use of biogeographic classification systems and identification of marine areas beyond national jurisdiction in need of protection.


Halpern, Benjamin S.; Walbridge, Shaun; Selkoe, Kimberly A.; Kappel, Carrie V.; Micheli, Fiorenza; D’Agrosa, Caterina; Bruno, John F.; Casey, Kenneth S.; Ebert, Colin; Fox, Helen E.; Fujita, Rod; Heinemann, Dennis; Lenihan, Hunter S.; Madin, Elizabeth M. P.; Perry, Matthew T.; Selig, Elizabeth R.; Spalding, Mark; Steneck, Robert; Watson, Reg (2008). A Global Map of Human Impact on Marine Ecosystems. Science, Volume 319, Issue 5865, pp. 948-952


32

Appendix: Participants

Arvanitidis Christos, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email: arvanitidis@her.hcmr.gr

Camisa Frederica, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email: camfamily@alice.it

Claus Simon, Vlaams Instituut voor de Zee, VLIZ – InnovOcean site, Wandelaarkaai 7, 8400 Oostende, Belgium, Email: simon.claus@vliz.be

Chatzinikolaou Eva, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email:

Costello Mark J., University of Auckland; Leigh Marine Laboratory, Box 349, Warkworth, Northland 0941, New Zealand, Email: m.costello@auckland.ac.nz

Deneudt Klaas, Vlaams Instituut voor de Zee, VLIZ – InnovOcean site, Wandelaarkaai 7, 8400 Oostende, Belgium Email: klaas.deneudt@vliz.be

Dailianis Thanos, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email: thanosd@hcmr.gr

Faulwetter Sarah, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email: sarifa@hcmr.gr

Garthe Stefan, Christian-Albrechts-University Kiel; Research and Technology Centre Hafentörn 1, 25761 Büsum, Germany Email: garthe@ftz-west.uni-kiel.de

Herman Peter, Royal Netherlands Institute for Sea Research; Centre for Estuarine and Marine Ecology; Department of Spatial Ecology, NIOZ-KNAW, Postbus 140, 4400 AC Yerseke,
Netherlands,
Email: peter.herman@nioz.nl

Hernandez Francisco, Vlaams Instituut voor de Zee, VLIZ – InnovOcean site, Wandelaarkai 7, 8400 Oostende, Belgium, Email: francisco.hernandez@vliz.be

Pafilis Evangelos, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email:

Pavloudi Christina, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email: cpavloud@hcmr.gr

Pesant Stéphane, Centre National de la Recherche Scientifique; Laboratoire d’Océanographie de Villefranche, UMR 7093, BP 28, 06234 Villefranche-sur-Mer Cédex, France, Email: pesant@obs-vlfr.fr

Pinn Eunice, Joint Nature Conservation Committee, Dunnet House, 7 Thistle Place, Aberdeen, AB10 1UZ Email: Eunice.Pinn@jncc.gov.uk

Stienen Eric, Vlaamse Overheid; Beleidsdomein Leefmilieu, Natuur en Energie; Instituut voor Natuur- en Bosonderzoek, Hoofdzetel, Kliniekstraat 25, 1070 Brussel, Belgium, Email: eric.stienen@inbo.be

Vandepitte Leen, Vlaams Instituut voor de Zee, VLIZ – InnovOcean site, Wandelaarkai 7, 8400 Oostende, Belgium, Email: leen.vandepitte@vliz.be

Vanhoorne Bart, Vlaams Instituut voor de Zee, VLIZ – InnovOcean site, Wandelaarkai 7, 8400 Oostende, Belgium, Email: bart.vanhoorne@vliz.be

Vasileiasou Aikaterini, Hellenic Centre for Marine Research; Institute of Marine Biology and Genetics; Biodiversity & Ecosystem Management Department, Thalassocosmos, Former US base at Gournes, P.O.Box 2214, Heraklion 71003 (Crete), Greece, Email: kvasileiadou@hcmr.gr

Waumans Filip, Vlaams Instituut voor de Zee, VLIZ – InnovOcean site, Wandelaarkai 7, 8400 Oostende, Belgium, Email: filip.waumans@vliz.be

Webb Tom, Department of Animal and Plant Sciences, The University of Sheffield, Western Bank, Sheffield S10 2TN Email: UK t.j.webb@sheffield.ac.uk