



EMODnet



European Marine
Observation and
Data Network

EMODnet Thematic Lot n° V – Biology

EASME/EMFF/2017/1.3.1.2/02/SI2.789013

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

D3.8. General report on data entry

Disclaimer

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the EASME or of the European Commission. Neither the EASME, nor the European Commission, guarantee the accuracy of the data included in this study. Neither the EASME, the European Commission nor any person acting on the EASME's or on the European Commission's behalf may be held responsible for the use which may be made of the information.

Document info

Title	D3.8. General report on data entry
WP title	WP3: Data archaeology and rescue
Task	Task 1: a common method of access to data held in repositories
Authors	Vasilis Gerovasileiou (HCMR), Dimitra Mavraki (HCMR), Georgia Sarafidou (HCMR), Laura Boicenco (NIMRD), Adrian Filimon (NIMRD), Camelia Dumitrache (NIMRD), Menashè Eliezer (OGS), Marina Lipizer (OGS), Ruben Perez (VLIZ), Leen Vandepitte (VLIZ), Joana Beja (VLIZ),
Dissemination level	Public
Submission date	07/04/2021
Deliverable due date	18/04/2021

Contents

1	Generic datasets presentation	4
	Geographical coverage	5
	Temporal coverage	6
	Taxonomic coverage	7
	Challenges of historical datasets digitization	7
2	Extensive datasets presentation	9
	List and description of datasets	15
3	Abbreviations	24
4	References	25

General report on data entry

1 Generic datasets presentation

This report compiles all the information regarding the datasets processed and submitted to the EMODnet Biology portal within the Work Package 3 (WP3), during the phase III of the EMODnet Biology project. Information such as occurrence records, species number and temporal and spatial coverage, are presented, among others (Table 1). Datasets are also categorized according to the Essential Ocean Variables (EOVs), enhancing their interoperability and management overall (Lear *et al.* 2020). The EOVs are a series of variables that aim to consistency and cost-effective marine monitoring, within common standards for data collection around the world.

The datasets are organized in DarwinCore Archive (DwC-A) standard before submitted, where all the relevant information is included in three tables linked with each other. The Occurrences table, that contains information on the taxonomy of a taxon and its presence in a specific place (occurrence record); the Events table, that gives information on the sampling events, the coordinates or the depth, among others; and the extended Measurements or Facts (eMoF) table that contains information about other biotic or abiotic measurements or facts, such as biomass, temperature etc.

Overall 27 new datasets were processed and submitted. In addition, one already published dataset (#4) which was updated. A total of 46,041 occurrence records have been collated, covering a temporal period of almost 120 years; from 1898 to 2017 (Table 1). The area studied concerned mostly the Mediterranean Sea, but datasets from other marine regions such as the Eastern Atlantic Ocean, the Black Sea and the Suez Canal were also included (Figure 2). In total, 1,526 species and 1,767 taxa were reported.

As far as taxonomy was concerned, there was a wide coverage observed, a great part of which consisted of benthic invertebrates with phytoplankton, zooplankton, fish and macroalgae being also reported. The aforementioned groups correspond to five out of the ten EOVs followed in EMODnet, thus addressing different needs of the project (e.g. to support policy decisions and potential future development of data products):

- EOV2: Phytoplankton biomass and diversity (3 datasets),
- EOV3: Zooplankton biomass and diversity (5 datasets),
- EOV4: Fish abundance and distribution (2 datasets),
- EOV6: Benthic invertebrate abundance and distribution (22 datasets),
- EOV10: Macroalgal canopy cover and composition (1 dataset).

Table 1. General metadata of the 28 EMODnet Biology phase III datasets.

Occurrence records	Events	Extended Measurements or Facts (eMoF)	Species number	Taxa number	Temporal coverage (years)
46,041	5,706	47,919	1,526	1,767	1898 - 2017

At this point, it is worth mentioning the data flow as it is shown in Figure 1. EMODnet Biology uses the EurOBIS data system to harmonize and centralize biodiversity data on marine species collected by European institutions. The data that flows to EurOBIS is being mapped to the Darwin Core Terms as it is said before. The purpose of

these terms is to facilitate data sharing by providing a well-defined standard core vocabulary in a flexible framework to minimize the barriers to adoption to ensure interoperability and to maximize reusability, according to FAIR principles.

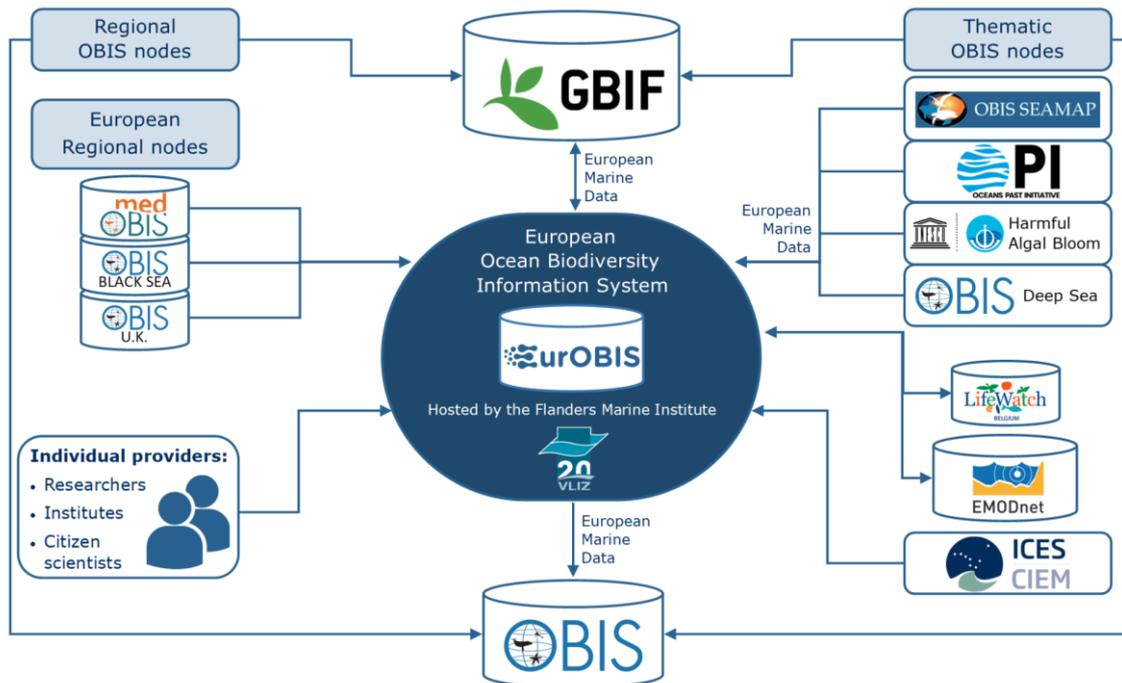


Figure 1: Representation of [data flow](#) used within EMODnet Biology.

EurOBIS aims to centralise biogeographic data on marine species collected by European institutions and receives its data through different pathways (http://www.eurobis.org/data_flow):

- Individual providers via e.g. email or the EurOBIS IPT instance;
- OBIS European regional nodes - [OBIS Black Sea](#), [MedOBIS](#) and [OBIS UK](#);
- OBIS thematic nodes - [OBIS Seamap](#), [Oceans Past Initiative \(OPI\)](#), [OBIS Deep Sea](#), [Harmful Algal Blooms \(HAB\) OBIS](#);
- Other networks - e.g. [ICES](#), [EMODnet Biology](#), [LifeWatch](#)

Geographical coverage

The main focus of WP3 is the Mediterranean and Black Seas. Other regional European Seas and adjacent marine areas are also covered, where appropriate, depending on the historical data sources. Therefore, the Mediterranean Sea is the region where data were mostly collected from (Figure 2). In addition, three datasets covered the Black Sea, the Eastern Atlantic Ocean and the Suez Canal (Red Sea), respectively.

The occurrence points were evenly distributed among the Mediterranean Sea, but the highest values of occurrence records are observed in the eastern part as it is shown in Figure 2. The following maps highlight the effort put into discovering, prioritizing and digitizing datasets within WP3.

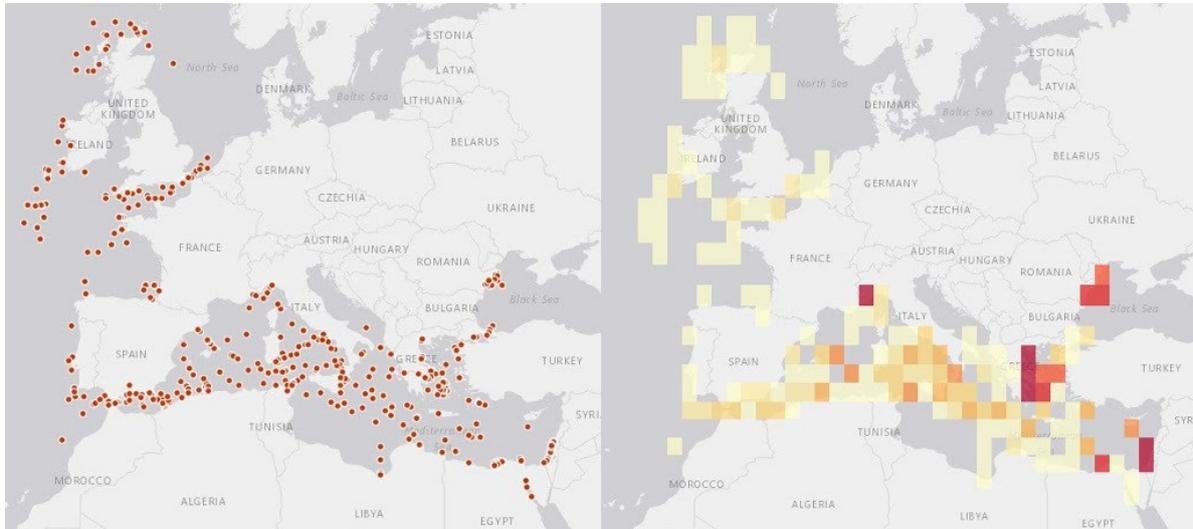


Figure 2: Distribution maps of the occurrence records for all 28 datasets: with toggle points (left) and gradient colours (red colour indicates the highest values, right).

Temporal coverage

The time period covered by all 28 datasets is 1898-2017. In Figure 3, the number of occurrence records per year for all 28 datasets is depicted with bar plots. As the graph shows, a relatively small number of historical occurrence records from the beginning of the 20th century was reported, while the number of records significantly increased after the 80's. Most occurrence records belong to recent datasets; those of the last 20 years. This graph underlines both the effort put in recent years but also the need of further prioritization and digitization of historical datasets from the Mediterranean Sea. Filling temporal and spatial gaps of biodiversity knowledge for this area is of utmost importance since it enhances the understanding of the species distribution changes, shedding light to the climate change causes (McClenachan *et al.* 2012; Costello *et al.* 2013; Mouquet *et al.* 2015).

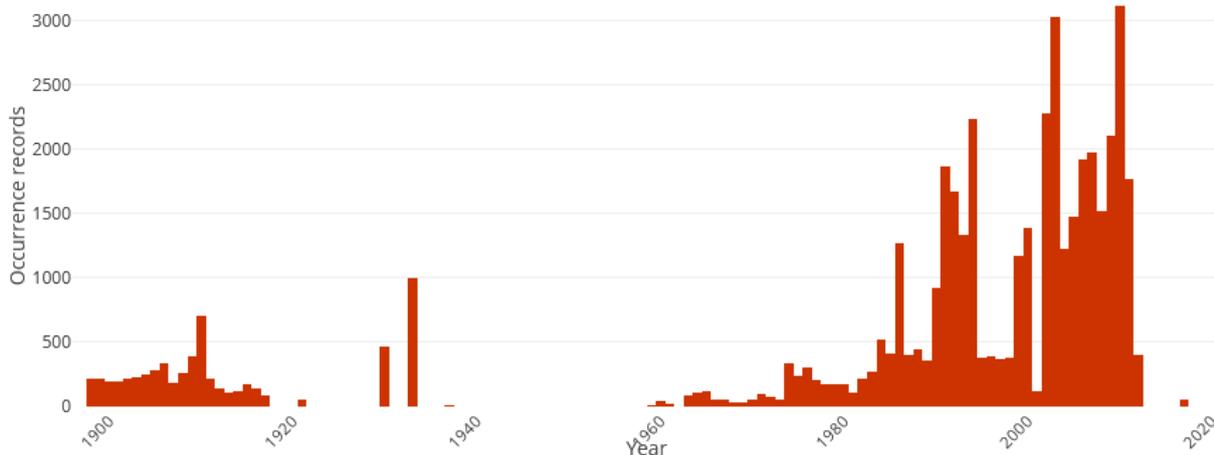


Figure 3: Bar plot depicting the occurrence records per year for all 28 digitised datasets.

Taxonomic coverage

As previously mentioned, the datasets included in the renewal stage of EMODnet Biology Phase III (April 2019 to April 2021) cover a wide variety of taxa, more specifically from 24 phyla: Annelida, Arthropoda, Brachiopoda, Bryozoa, Cercozoa, Chaetognatha, Chordata, Ciliophora, Cnidaria, Ctenophora, Echinodermata, Euglenozoa, Foraminifera, Haptophyta, Hemichordata, Mollusca, Myzozoa, Nematoda, Nemertea, Ochrophyta, Porifera, Radiozoa, Rhodophyta, Sipuncula.

The top 10 species, with the highest number of occurrence records are presented in Table 2. An occurrence record is defined as the evidence of the occurrence of a species (or other taxa) at a particular place on a specified date. The scientific name is one of the core pieces of information required for any record and ideally, the name supplied is at species level or below.

Table 2: The top ten species with the highest number of occurrence records from all 28 datasets published within WP3. Their higher taxonomy is also included.

Scientific Name	Occurrence Records	Phylum	Class
<i>Liriope tetraphylla</i>	1,028	Cnidaria	Hydrozoa
<i>Thalia democratica</i>	974	Chordata	Thaliacea
<i>Centropages typicus</i>	778	Arthropoda	Hexanauplia
<i>Lensia subtilis</i>	777	Cnidaria	Hydrozoa
<i>Muggiaea kochii</i>	639	Cnidaria	Hydrozoa
<i>Abylopsis tetragona</i>	595	Cnidaria	Hydrozoa
<i>Parasagitta setosa</i>	572	Chaetognatha	Sagittoidea
<i>Pagellus erythrinus</i>	568	Chordata	Actinopterygii
<i>Salpa fusiformis</i>	554	Chordata	Thaliacea
<i>Muggiaea atlantica</i>	536	Cnidaria	Hydrozoa

Challenges of historical datasets digitization

The WP3 of EMODnet Biology has been in charge of prioritizing, digitizing and submitting historical and rescue biodiversity datasets mainly from the Eastern Mediterranean and Black Seas. Historical data are defined as data that have been collected at least 70 years ago (before 1950), while rescue data are those data that have been published from the 1960s to 2000s, which currently only exist on paper or digital text files. More recent datasets are usually processed within WP2 but since some of them concerned the Mediterranean and Black seas, it was convenient to keep them in WP3 and publish them through the MedOBIS node. In this context, only datasets that have been successfully submitted to the MedOBIS/VLIZ IPT and subsequently harvested by the EMODnet



Biology portal are presented; yet, there are datasets, for which curation is ongoing and are due to be published during the next phase of EMODnet Biology.

“Data curation is the act of discovering a data source(s) of interest, cleaning and transforming the new data, semantically integrating it with other local data sources, and deduplicating the resulting composite” (Stonebraker *et al.* 2013). Biodiversity data follow a certain curation process; the same applies to historical data: data curators initially identify and prioritize the available literature sources, then they digitize the selected documents with standardized procedures and equipment, they extract data and metadata with structured knowledge identifiers in order to become computer readable and manageable and finally, they publish these data using controlled vocabularies (Sarafidou *et al.* 2020).

There are a number of reasons that complicate the curation of historical data (Faulwetter *et al.* 2016). To begin with, there is a major difficulty in locating the original data sources if they are not already digitized. Integral part of the digitization of historical data is the transformation of scanned documents to text through the process of Optical Character Recognition (OCR). This is a crucial step as all the subsequent steps rely on its results, so the poor quality of the historical documents affects the whole process. Apart from being hard to locate, since they are often hidden in floppy disks or scattered sheets of paper and forgotten in office drawers and institute libraries, when discovered, they are also hard to understand.

The frequent absence of the data provider -since they are often retired or deceased- further complicates the process since clarifications on the dataset are unavailable. Some other challenges -that have also been encountered during the curation of many of the abovementioned datasets- are the following:

- the lack of essential metadata, such as coordinates, date or sampling methods
- the lack of data and metadata standardization
- the unclear or self-contradictory data
- the use of ambiguous symbols and verbatim instead of numerical data
- the typographic errors or misspellings
- the use of languages other than English and
- the taxonomic inconsistencies or absences from the WoRMS database

Therefore, a detailed comprehension of the dataset is needed by the curator in order for it to be properly digitized. Moreover, personal contact with taxonomic specialists is often of great assistance. Apart from the human capabilities in resolving the mentioned challenges, it is worth mentioning some digital tools that facilitate the curator’s work, such as the [WoRMS taxon match tool](#) for validating the taxonomy or the [Marine Gazetteer](#) for finding the accurate coordinates. In this context, a fully featured data curation workflow has been developed recently; [Deco](#) - bioDivErsity data Curation wORkflow (Paragkamian *et al.* submitted) with the aim of assisting data curation needs and thus accelerating the whole process.



2 Extensive datasets presentation

This section presents a concentrated metadata table for all the 28 datasets (Table 3) and a brief description for each available dataset with the corresponding links to the EMODnet portal.

Table 3: Metadata for the 28 datasets published within WP3. Occurrence, Event, eMoF and Species numbers are given for each dataset, as well as temporal, geographical and taxonomic coverage, along with the corresponding EOVs.

Dataset	Occurrences	Events (core)	eMoF	Species number	Temporal coverage (year)	Geographical coverage	Taxonomic coverage	EOV
1	49	12	49	2	2017	MED, Greece, Crete	Rhynchonellida	EOV6
2	883	599	-	22	1905-1930	MED, Mediterranean	Polychaeta	EOV6
3	3	3	-	2	1937	ISW, Egypt, Suez Gulf MED, Egypt, Nile Delta	Cladocera	EOV3
4	898	16	898	275	1986	MED, Greece, Aegean	Actinopterygii Annelida Bryozoa Echinodermata Mollusca Nematoda Porifera	EOV4, EOV6
5	421	48	-	66	2004	MED, Greece, Aegean	Mollusca	EOV6
6	59	16	-	21	1908-1910	AE, East Atlantic MED, Mediterranean	Corallinales Rhodophyta	EOV6
7	646	341	-	5	1904-1914	MED, Mediterranean	Actinopterygii	EOV10
8	7,099	1,273	7,099	18	1966-1999	MED, France, Provence-Côte d'Azur MED, North-western Mediterranean	Arthropoda Chaetognatha Cnidaria Ctenophora	EOV4
9	22	12	32	5	1933	MED, Egypt, Alexandria	Cumacea Leptostraca	EOV3, EOV6

Dataset	Occurrences	Events (core)	eMoF	Species number	Temporal coverage (year)	Geographical coverage	Taxonomic coverage	EOV
							Stomatopoda	
10	4,149	212	4,149	-	2004-2010	MED, France, Provence-Côte d'Azur MED, North-western Mediterranean	Larvacea Branciopoda Gastropoda Hexanauplia Hydrozoa Malacostraca Ostracoda Sagittoidea Scyphozoa Tentaculata Thaliacea	EOV6
11	5,629	1,278	5,621	16	1959-2010	MED, France, Provence-Côte d'Azur	Annelida Arthropoda Chaetognatha Chordata Ciliophora Cnidaria Ctenophora Echinodermata Foraminifera Mollusca Sarcomastigophora	EOV3, EOV6

Dataset	Occurrences	Events (core)	eMoF	Species number	Temporal coverage (year)	Geographical coverage	Taxonomic coverage	EOV
12	3,606	212	-	66	1898-1917	MED, France, Provence-Côte d'Azur MED, North-western Mediterranean	Amoebida Annelida Arthropoda Cercozoa Chordata Cnidaria Ctenophora Mollusca Nemertea Radiozoa	EOV3, EOV6
13	32	24	-	13	1933	MED, Egypt, Alexandria	Hydrozoa	EOV3, EOV6
14	303	100	24	117	1933	MED, Egypt, Alexandria	Polychaeta	EOV6
15	8	3	-	2	1933	MED, Egypt, Nile Delta	Ochrophyta	EOV6
16	38	31	-	23	1933	MED, Egypt, Alexandria	Ascidiacea	EOV2
17	165	81	-	33	1933	MED, Egypt, Alexandria	Echinodermata	EOV6
18	861	63	1,548	204	1999	MED, Mediterranean	Bacillariophyceae Dinophyceae Haptophyta	EOV6
19	133	46	212	42	1933	MED, Egypt, Alexandria	Amphipoda	EOV2
20	8	7	-	1	1933	MED, Egypt, Alexandria	Anopla Enopla	EOV6
21	94	51	112	26	1933	MED, Egypt, Alexandria	Isopoda Tanaidacea	EOV6

Dataset	Occurrences	Events (core)	eMoF	Species number	Temporal coverage (year)	Geographical coverage	Taxonomic coverage	EOV
22	10	1	-	9	1933	MED, Egypt, Alexandria	Invertebrata	EOV6
23	5,115	224	7,268	366	2002-2003	MED, Aegean	Bacillariophyceae Dinophyceae Prymnesiophyceae	EOV6
24	4	4	-	4	1933	MED, Egypt, Alexandria	Arachnida	EOV2
25	122	58	-	46	1933	MED, Eastern Mediterranean	Porifera	EOV6
26	60	57	-	10	1933	MED, Egypt, Alexandria	Brachiopoda Enteropneusta Leptocardii Phoronida Sipuncula	EOV6
27	13,941	694	13,941	238	1990-1994	MED, Eastern Mediterranean	Annelida Arthropoda Articulata Chordata Cnidaria Mollusca	EOV6
28	1683	240	6966	76	2000-2010	MED, Black Sea	Annelida Cnidaria Crustacea Echinodermata Mollusca Nemertea Phoronida	EOV6

Dataset	Occurrences	Events (core)	eMoF	Species number	Temporal coverage (year)	Geographical coverage	Taxonomic coverage	EOV
							Platyhelminthes	

List and description of datasets

1. Brachiopoda in a *Posidonia oceanica* meadow in Plakias, SW Crete, Greece, 2017

The dataset contains occurrence records of 963 living individuals belonging to two species of brachiopods in a shallow-water *Posidonia* meadow in Crete, sampled in two seasons in 2017, from 5 to 20 m depth. Organisms on the leaves were collected with a net mounted on a metal frame with a 40×20 cm opening. The rhizomes were sampled with air-lift suction sampling. The sampler consisted of a PVC tube with a length and diameter of 100 cm and 8 cm, respectively. A SCUBA cylinder supplied air and was fitted at 10 cm above the mouth of the tube. At its other end, it was attached to a removable 0.5-mm mesh nylon bag that could be closed and replaced underwater. Sampling on the rhizomes was carried out on 1-m² square areas after defoliation in order to enhance collecting efficacy. The study was carried out in Plakias, Crete, Greece. The method steps included: Net on *Posidonia* leaves: 60 strokes per replicate (L6 samples) and 20 strokes per replicate (L2 samples). Suction sampling on *Posidonia* rhizomes: 1 square meter per replicate.

2. Report on the Danish Oceanographical expeditions 1908-1910 to the Mediterranean and adjacent seas - Pelagic Polychaetes

The present historical paper deals with the pelagic Polychaetes except the Tomopterids collected on the cruises of the "Thor", 1908-1910 in the Mediterranean and adjacent waters. The tables included in this report present also the scientific results from other research vessels such as "Dana" (years 1921 and 1930) and "S/S Pagan" (1911). The majority of the catches were made with "Petersen's Young-fish Trawl" mostly with a diameter of 200 cm and 330 cm and stramin nets with diameters of 100, 150 and 200 cm: the fishing parts of these consist of stramin, a coarse canvas with about 20 threads to 3 cm. A few samples containing Polychaeta were taken with other apparatus, viz. 5 samples with ring-trawl 130 cm, 200 cm and 300 cm in diameter; 2 samples with open conical silk-net 100 cm in diameter; one with Nansen's closing net, 50 cm in diameter and finally one with rectangular dredge opening 27 x 117 cm.

3. F.M.Ghazzawi (1938) Plankton of the Egyptian waters. Two Cladocera from the plankton. Notes and Memories of the Hydrobiology and Fisheries Directorate of Egypt, Notes and Memories No 31

This paper presents the identification of two species, namely *Podon polyphemoides* and *Penilia avirostris* by F.M.Ghazzawi, through plankton catches. It includes microphotographs and drawings of the above-mentioned species. As far as the study extent was concerned, plankton catches were collected from the Suez Canal and the Eastern and Western Harbour of Alexandria (Egypt). Information about the sample collection was sourced from Adolf Steuer's preliminary report: The Fishery Grounds Near Alexandria. Notes and Memories No 8.

4. Kyklades-data of the Central Aegean Sea

In the framework of the project "Oceanographic studies in open Seas", a survey of the benthic fauna of the Central Aegean Sea (Greece) was performed in June 1986. The hand operated Smith-McIntyre sampler of 0.1 m² was used in 16 stations and the samples were sieved through a 1 mm mesh size net. During the quality control

the abundance was transformed to individuals per square meter. The dataset includes 275 species, with Polychaeta being the most abundant class (Fig. 4) and specifically the species *Notomastus latericeus* being the one with the most records.

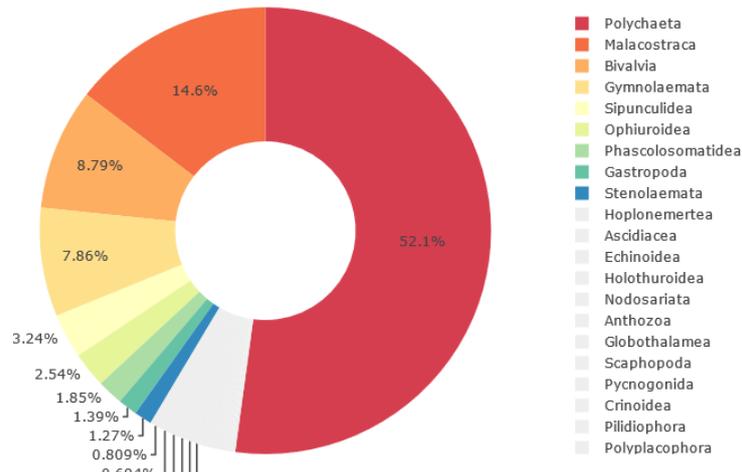


Figure 4: Class composition of dataset 4.

5. Spatial and seasonal variability of the molluscan macrofauna at the marine-lagoonal environmental gradient at Kalloni saltworks (Lesvos Island, NE Aegean Sea, Greece)

This focuses on the seasonal community structure of the macrobenthic molluscs at the marine-lagoonal environmental gradient that takes place at Kalloni solar saltworks (Lesvos Island, NE Aegean Sea, Greece). The molluscan community was studied by means of four seasonal samplings carried out in 2004 at four sites located along the marine-lagoonal environmental gradient in the study area. Four sampling sites, spread over a total distance of approximately 1 km, were chosen along the course of seawater flow from Kalloni Gulf into the saltworks and termed "gulf", "channel", "pond1" and "pond2", respectively. Sampling was repeated as four seasonal samplings in 2004 (in February – "Winter", May – "Spring", September – "Summer" and November – "Autumn"). Samples for the macrobenthic molluscs analyses (three replicates, 0.1 m² each, tens of meters apart) were collected with a Hydrobios Van Veen grab sampler and were subsequently sieved through a 0.5 mm mesh size net. The samples were fixed with 4% formaldehyde and stained with Rose Bengal.

6. Report on the Danish Oceanographical Expeditions 1908-10 to the Mediterranean and adjacent seas-Calcareous Algae

This dataset is based on the paper entitled "Calcareous Algae. Report on the Danish Oceanographical Expeditions 1908-10 to the Mediterranean and adjacent seas" published by Lemoine (1915). It comprises of 59 records of Rhodophyta (Corallinales and Peyssonneliales) collected during the 1908-9 and 1910 expeditions of Thor vessel from a wide geographical (from the Aegean Sea to the eastern Atlantic Ocean) and bathymetric



range (3-98 m). Sampling gear used for the collection of rhodophytes is not clarified in the respective paper by Lemoine (1915). A thorough examination of the different types of sampling gear used in the 16 stations, where calcareous algae were sampled from, suggests that they were collected using dredges (e.g. D1: dredge with rectangular opening of 27 x 117 cm; D2: dredge with triangular opening of 45 x 45cm; and hand-dredge of 18 x 14 cm) and also possibly trawls (e.g. Y200: young-fish trawl, 200 cm in diameter at opening and/or Monaco trawl, 56 x 170 cm at opening). In the introductory paper by Schmidt (1912) it is mentioned in page 12 that "When at anchor the motorboat of the "Thor" was often used for fishing in shallow water, for example, with the same dredges as used from the ship and with small hand-dredges intended for the collection of marine algae" supporting that the hand-dredges were also used in certain cases. However, given that different types of sampling gear were deployed in all stations, it is quite probable that samples were mostly collected on an opportunistic basis with various sampling methods.

7. Report on the Danish Oceanographical Expeditions 1908-1910 to the Mediterranean and adjacent seas-Engraulidae-Clupeidae

Fish larvae of the families Engraulidae and Clupeidae were sampled mainly by trawling with Young fish trawl and Petersen trawl, depending on the expedition. Both methods are most suitable for pelagic fish larvae rather than coastal ones, as reported by Schmidt (1912). The dataset proceeds from the report by Fage (1920), merging data from Thor winter and summer expeditions to the Mediterranean and adjacent seas (1908-1910), with data from Thor expeditions to the N-W Atlantic (1905 and 1906). Some data from the Red Sea are also present, although this area was not the target of the expedition. The dataset was digitized manually from scanned documents. Original names regarding taxonomy and sampling location were kept as a reference. At the same time taxon names were cross-checked against the World Register of Marine Species (WoRMS); while the original sampling coordinates were plotted using Google Earth in order to double check that they are located in the sea. All data were checked for inconsistencies, both for the range of values, but also compared to other datasets from the same series, collected at the same station. Any problematic records were recorded in a remarks field in the data.

8. Temporal evolution of zooplankton by Juday-Bogorov net in the Northwestern Mediterranean Sea. Villefranche-sur-mer, 1966-1999

The Observatoire de Villefranche sur Mer supports the daily sampling of zooplankton since 1966 at a time series station located at the entrance of the Bay of Villefranche sur Mer. Samples were collected by a vertical tow from bottom to surface (75-0 m), using a Juday-Bogorov net with 380 µm of mesh, opening of 0.50 m and length of 1.80 m. The core data table contains 1,569 records. Two extension data tables also exist. An extension record supplies extra information about a core record.

9. The fishery ground near Alexandria. XVI. Cumacea, Stomatopoda, Leptostraca by Adolf Steuer (1938). Notes and Memoirs No 26





This is a historical dataset that was published in 1938 by A. Steuer and concerns of the faunistic report on the Cumacea, Stomatopoda and Leptostraca collected during the floristical and faunistic survey of himself on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 15/5/1933 to 13/11/1933 and contains occurrence data of Cumacea, Stomatopoda and Leptostraca from the coasts of Alexandria, Egypt. Besides occurrence data, this dataset contains individual counts, sex, life stage and body length. Twelve (12) stations were sampled, 2 of them were planktonic stations. *

10. Temporal evolution of zooplankton by WP2 net in the Northwestern Mediterranean Sea, Villefranche-sur-mer 2004-2010

The Observatoire de Villefranche sur Mer supports the daily sampling of zooplankton since 1966 at a time series station located at the entrance of the Bay of Villefranche sur Mer. Samples were collected by a vertical tow from bottom to surface (75-0 m), using a WP2 net: 220 µm of mesh and opening of 0.25 m². Zooplankton abundance was counted by specialists at OOV under microscopes. The sampling site is located at the mouth of the Bay over a bathymetry of 80 m and is thus open to the Ligurian Sea. Data was entered manually to an Excel template. After entry into the Excel template data was transferred to DwC format. The list of species after applying taxonomic quality control, contains the presently valid names according to the World Register of Marine Species (WoRMS) and the AphiaID is matched against every single species name. Spatial quality control was also performed and the coordinates are provided.

11. Temporal evolution of zooplankton by Regent net in the Northwestern Mediterranean Sea, Villefranche-sur-mer, 1959-2010

The Observatoire de Villefranche sur Mer supports the daily sampling of zooplankton since 1966 at a time series station located at the entrance of the Bay of Villefranche sur Mer. Samples were collected by a horizontal tow at 5 layers, using a Regent net. The sampling site is located at the mouth of the Bay over a bathymetry of 80 m and is thus open to the Ligurian Sea. Zooplankton abundance was counted by specialists at OOV under microscopes. The data in this sampling event resource has been published as a Darwin Core Archive (DwC-A), which is a standardized format for sharing biodiversity data as a set of one or more data tables. The core data table contains 1,517 records. 2 extension data tables also exist. An extension record supplies extra information about a core record. In this case also data was entered manually to an Excel template. After entry into the Excel template data was transferred to DwC format. The list of species after applying taxonomic quality control, contains the presently valid names according to the World Register of Marine Species (WoRMS) and the AphiaID is matched against every single species name. Spatial quality control was also performed and the coordinates are provided.

12. Temporal evolution of zooplankton, surface observations, in the Northwestern Mediterranean Sea, Villefranche-sur-mer 1898-1917

The Observatoire de Villefranche sur Mer supports the regular observation of large zooplankton from 1898-1917 at a time series station located at the entrance of the Bay of Villefranche sur Mer. The data in this sampling event resource has been published as a Darwin Core Archive (DwC-A), which is a standardized format for sharing biodiversity data as a set of one or more data tables. The core data table contains 214 records. Extension data table also exists. An extension record supplies extra information about a core record. Again, data was entered manually to an Excel template. After entry into the Excel template data was transferred to DwC format. The list





of species after applying taxonomic quality control, contains the presently valid names according to the World Register of Marine Species (WoRMS) and the AphiaID is matched against every single species name. Spatial quality control was also performed and the coordinates are provided.

13. Billards A. 1936. The fishery grounds near Alexandria. VI Hydroidea. Notes et memoires No 13

This is a historical dataset that was published in 1936 by A. Billards and concerns the faunistic report on the Hydroidea collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 2/9/1933 to 12/11/1933 and contains occurrence data of Hydroidea in the coasts of Alexandria, Egypt. Hydroidea were found in about 20 sampling stations near Alexandria, Egypt, eastern Mediterranean. *

14. Fauvel P. 1937. The fishery grounds near Alexandria. XI. Annelida Polychaeta. Notes et memoires No 19

This is a historical dataset that was published in 1937 by P. Fauvel and concerns the faunistic report on the Polychaetes collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 3/9/1933 to 18/11/1933 and contains occurrence data of polychaetes in the coasts of Alexandria, Egypt. Almost 150 stations were sampled in the coasts near Alexandria, with 303 occurrence records and 117 species recorded, *Nephtys hombergii* being the one with the most records. *

15. H. Bachmann (1936) The fishery grounds near Alexandria. XIV. Phytoplankton from the Nile. Fisheries Research Directorate of Egypt, Notes and Memoirs No 22

Egyptian plankton samples for revision as to its phytoplankton content. First samples are dated in 1933 owned by Prof. Steuer, while the other samples are from the Nile area and collected by Dr. Hans Amberg, in 1904. The samples that come from the large district of the Nile prove that the *Melosira* in Prof. Steuer's material represents a real phytoplankton component of Nile water and not merely a passive swimming organism. Apart from the samples of 1933, there are samples of Nile owned by H. Bachmann and collected by Dr Hans Amberg, in 1904. The equipment used is a very small net. The samples are from the following stations: (1) Assuan, 1904/02/09; (2) Above Edfu, 1904/02/08; (3) Dendera, 1904/02/04; (4) Assiut, 1904/02/02; (5) Above Fashn, 1904/01/30; (6) Cairo, 1904/04/08 and (7) Bedreshein, 1904/01/29. The samples 1, 2 and 3 contain the genus *Melosira* and were very frequent. The samples 4, 6 and 7 contain *Melosira granulata* var. *augustissima* and the typical *Melosira crucipunctata*. These samples showed hardly any other phytoplankton. One may mention: *Pediastrum clathratum*, with varieties, *Surirella splendida*, *Eudorina elegans*.

16. Harant H. 1939. The fishery grounds near Alexandria. Ascidiacea (Cartes 1-5). Notes et Memoirs No 28

This is a historical dataset that was published in 1939 by H. Harant and concerns of the faunistic report on the Ascidiacea collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria





mainly with the vessel "El Hoot". This dataset contains occurrence data of Ascidiacea in the coasts of Alexandria, Egypt. Temporal coverage for this dataset is not described in the historical paper but only in the preliminary report of the expedition. Thus, it is assumed that the sample collection was done in the autumn of 1933 (beginning from the end of August to end of November). Ascidiacea were found in about 30 sampling stations near Alexandria, Egypt, eastern Mediterranean. *

17. [Mortensen Th. And Steuer Ad. 1937. The fishery grounds near Alexandria. XIII Echinoderma. Notes and Memoirs No 21. Hydrobiology and Fisheries Directorate, Egypt](#)

This is a historical dataset that was published in 1937 by Th. Mortensen and Ad. Steuer and concerns of the faunistic report on the Echinodermata collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". In this historical paper, there is no information about exact sampling temporal coverage. The dataset contains occurrence data of Echinodermata in the coasts of Alexandria, Egypt. It also contains information about individual counts, sex and life stage. Echinodermata were found in about 80 sampling stations in Alexandria, Egypt, Eastern Mediterranean. *

18. [Phytoplankton in longitudinal east-west transect of the Mediterranean Sea, 1999](#)

The scope of this study was to explore the diversification of the phytoplankton community structure and related ecological determinants along a large scale (3,188 km) longitudinal transect, covering the region from eastern Mediterranean to Western Basin. Sampling was performed during the period June – July of 1999. Six stations were located in the Eastern Basin, one station in the Sicily strait and two stations in the Western Basin. A total of 216 species were identified (48 diatoms, 112 dinoflagellates and 56 coccolithophores). Samples were collected from routine hydrocasts using a CTD rosette sampler from 1, 10, 20, 50, 75, 100 and 120 m. Temperature and salinity were recorded using a Seabird Electronics SBE-911+ CTD profiler. Samples for chlorophyll-a determination (2 L from each sampling depth) were filtered on board through 0.2 mm polycarbonate Millipore filters and kept frozen until analysis in the laboratory with a TURNER 00-AU-10 fluorometer. Samples for the determination of nutrients were collected in 100 ml polyethylene bottles and kept at -20 °C, until their analysis in the laboratory with a Bran+Luebbe II autoanalyser. Phosphate was measured with a Perkin Elmer Lambda 2S UV/VIS Spectrometer. Stations were plotted on Google earth for a visual check. Taxonomic quality control was done using the online taxon match tool of WoRMS.

19. [Schellenberg A. 1936. The fishery grounds near Alexandria. Amphipoda Benthonica. Notes and memoirs no 18](#)

This is a historical dataset that was published in 1936 by Schellenberg A., and concerns of the faunistic report on the Amphipoda collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 29/8/1933 to 18/11/1933 and contains occurrence data of Amphipoda in the coasts of Alexandria, Egypt. It also contains information about





individual counts, sex, life stage and body length. Amphipoda were found in about 45 benthic stations near Alexandria, Egypt, Eastern Mediterranean. *

20. The fishery Ground near Alexandria. XX. Nemertini by Herman Friedrich (1940). Notes and Memoirs No 38

This is a historical dataset that was published in 1940 by Herman Friedrich and concerns the collection of 9 tubes of Nemerteans collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 6/9/1933 to 12/11/1933. The material was unsatisfactory, so they were identified mostly to the class level; namely 75% Pilidiophora and 25% Hoplonemertea. Only one species, *Wijnhoffella alexandrinensis* was identified. *

21. The fishery ground near Alexandria. XXI. Tanaidacea and Isopoda by H.J. Larwood (1940). Notes and Memoirs No 35

This is a historical dataset that was published in 1940 by H.J. Larwood and concerns the faunistic report on the Tanaids and Isopods collected during the floristic and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 29/8/1933 to 15/11/1933 and contains occurrence data of Tanaids and Isopods in the coasts of Alexandria, Egypt. It also contains information about individual counts, sex, life stage and body size. Tanaidacea and Isopoda species were found in 49 sampling stations near Alexandria, Egypt, eastern Mediterranean. *

22. Vatova A., 1935. The fishery grounds near Alexandria, II. A bottom sample taken at Alexandria, Notes and Memoirs No 9. Fisheries Research Directorate, Egypt

This is a historical report, published in 1935 by A. Vatova. It contains specific information about a bottom sample that was taken once in the eastern harbour of Alexandria, near the Marine Laboratory during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This historical report contains information about a bottom sample that was taken with a Petersen grab of 0.2 square meter surface. During A. Steuer's expedition in the coasts of Alexandria, this sampler was used only once on 2 September 1933. According to A. Steuer's preliminary report difficulties in the manipulation of this sampler prevented further investigations with this gear. *

23. Phytoplankton of coastal waters in Aegean Sea, Eastern Mediterranean, 2002-2003

This dataset contains abundance data of phytoplankton species (diatoms, dinoflagellates and coccolithophores) collected in four eutrophic gulfs of the Aegean Sea, Eastern Mediterranean Sea, during the years 2002–2003 and data on environmental parameters. The samples were collected from 1 m depth with a Hydro-bios sampler equipped with an inverted thermometer. For species identification and enumeration samples were fixed with Lugol's solution (without acetic acid), settled in 50-ml sedimentation chambers. Samples for nutrient analysis were filtered (0.2 mm Millipore filter), dispensed into 100 mL polyethylene bottles and kept continuously under





deep freeze (-20 °C) until their analysis in the laboratory. Samples (0.5–1 L) for chlorophyll-a analysis were collected from each sampling station, filtered immediately through Whatman GF/F filters and kept deep frozen till the day of analysis in the laboratory by a TURNER AU-10 fluorometer. Phytoplankton was studied in four gulfs of Aegean Sea (Saronikos Gulf, Evoikos Gulf, Pagasitikos Gulf and Thermaikos Gulf) during the years 2002-2003. Stations were plotted on Google earth and errors were corrected. Taxonomic quality control was done using the online taxon match tool of WoRMS.

24. Dr.Karl Viets. 1935. The fishery grounds near Alexandria. IV. -Some Marine Mites From Alexandria. Notes and Memoirs No 11. Fisheries Research Directorate, Egypt

This is a historical dataset that was published in 1935 by Dr Karl Viets and concerns some marine mites collected from Egyptian littoral waters near Alexandria. They collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 9/9/1933 to 13/11/1933 and 4 species of Arachnida are identified in it. *

25. M. Burton (1936). The Fishery grounds near Alexandria IX – Sponges. Notes and Memoirs No.17. Department of Zoology, British Museum. Fisheries Research Directorate. Ministry of Commerce and Industry, Egypt

This historical collection consists of 150 specimens and it is of particular interest since it is the first to be taken from the waters of the Eastern Mediterranean during summer-autumn of 1933. 46 Porifera species with 122 occurrence records were identified with 95.9% of them belonging to Demospongiae and 4.1% to Calcarea. The main point of importance brought out by the present investigations is that there are definite signs of an immigration of sponges from the Red Sea and Indian Ocean. *

26. Steuer Ad., 1939. The fishery grounds near Alexandria. XVIII Sipunculoidea, Phoronidea, Brachiopoda, Enteropneusta and Acrania. Notes and Memoirs No. 30. Fouad I Institute of Hydrobiology and Fisheries, Egypt

This is a historical dataset that was published in 1939 by A. Steuer and concerns of the faunistic report on the Sipunculoidea, Phoronidea, Brachiopoda, Enteropneusta and Acrania collected during the floristical and faunistic survey of Adolf Steuer on the coasts near Alexandria mainly with the vessel "El Hoot". This dataset covers the time span of 1/4/1933 to 17/11/1933 and contains occurrence data of Sipunculoidea, Phoronidea, Brachiopoda, Enteropneusta and Acrania in the coasts of Alexandria, Egypt. Sipunculoidea, Phoronidea, Brachiopoda, Enteropneusta and Acrania were found in 52 benthic and 5 planktonic sampling stations in the coasts near Alexandria, Egypt, eastern Mediterranean. *

27. Fish and marine invertebrates from the Israeli Eastern Mediterranean Sea

Fish and marine invertebrates caught with trawl surveys (fish and shrimp trawls) from the Mediterranean continental shelf and upper slope of Israel from 1990 to 2012 (1990-1994, 2000, and 2008-2012). Surveys took





place by on-board trawlers (fish and shrimp trawls) fishing the Israeli continental shelf and upper slope. Data were extracted from hauls replicating commercial trawl hauls, that is, the skippers were asked to perform their own routes within a given depth, location or time and they are therefore hybrids between fisheries dependent and independent surveys. Depth range was between 8-180.5 meters. Historical and modern catch data of Israeli trawlers were used to evaluate trends in abundance, biomass and diversity of Levantine fishes and marine invertebrates. Historical data were collected from trawl hauls from 1990 to 1994. Current data were collected in trawl hauls conducted in 2000 and between 2008 and 2011, covering the same fishing fields and depth range on the continental shelf and upper slope.

28. Macrozoobenthos collected in the northern part of the Romanian littoral (Danube mouths) between 2000-2010

The macrozoobenthos dataset includes a number of 243 samples collected in the Northern part of the Romanian littoral between Sulina and Chituc from 10 m to 57 m depths. Van Veen grab with a sampling area of 0.05 m² was employed as a macrozoobenthos sampler. One replicate was collected at every station. Sampling was done in spring, summer and autumn months. The samples have been analyzed qualitative (number of species) and quantitative (densities and biomasses). The macroinvertebrates have been identified at species level. Dataset contains abundance data (ind/m²) and biomass (g/m²). The macroinvertebrates have been identified at species level.

*Datasets 9, 13, 14, 16, 17, 19, 20, 21, 24 and 26 are part of a large faunistic [expedition](#) that took place in the coasts near Alexandria. The samples were collected by Adolf Steuer. The results of this expedition were published in separate volumes for each taxonomic group. Thus, sampling methods were described in a preliminary report by A. Steuer. Species identification and description were done by the corresponding author of each dataset. The HCMR team was responsible for the digitization of these datasets.





3 Abbreviations

DwC-A: Darwin Core Archive is the standard for packaging and publishing biodiversity data using Darwin Core terms. It is the preferred format for publishing data in OBIS and GBIF. The format is described in the Darwin Core text guide. A Darwin Core Archive contains a number of text files, including data tables formatted as CSV http://www.eurobis.org/data_formats.

eMoF: The expansion of the OBIS schema to OBIS-ENV was accomplished by the use of two new core types, Event and extended Measurements or Facts (eMoF). The Event core allows for the association of measurements with nested events, whereas the eMoF allows for the association of the occurrence records with other biotic or abiotic measurements and facts. The terms in the eMoF core can be populated with free text annotation and with controlled vocabularies.

EOVs: Essential Ocean Variables. The Global Ocean Observing System (GOOS) aims to promote common standards for data collection around the world. As part of this it has identified a series of variables that it hopes will lead to consistency and cost-effective marine monitoring, globally; these are known as 'Essential Ocean Variables' (EOVs).

FAIR Principles: It is a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. The term FAIR was launched at a Lorentz workshop in 2014, the resulting [FAIR principles](#) were published in 2016.

IPT: [Integrated Publishing Toolkit](#). The Integrated Publishing Toolkit is a free open source software tool written in Java that is used to publish and share biodiversity datasets through the GBIF network.

OCR: Optical Character Recognition. It is the electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document or a photo of a document

WP: Work Package of EMODnet Biology phase III





4 References

- Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE (2013) Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution* 28: 454–461. <https://doi.org/10.1016/j.tree.2013.05.002>
- Faulwetter S, Pafilis E, Fanini L, Bailly N, Agosti D, Arvanitidis C, Boicenco L, Capatano T, Claus S, Dekeyzer S, Georgiev T, Legaki A, Mavraki D, Oulas A, Papastefanou G, Penev L, Sautter G, Schigel D, Senderov V, Teaca A, Tsompanou M (2016) EMODnet Workshop on mechanisms and guidelines to mobilise historical data into biogeographic databases. *Research Ideas and Outcomes* 2: e9774. <https://doi.org/10.3897/rio.2.e9774>
- Lear D, Herman P, Van Hoey G, Schepers L, Tonné N, Lipizer M, Muller-Karger FE, Appeltans W, Kissling WD, Holdsworth N, Edwards M, Pecceu E, Nygård H, Canonico G, Birchenough S, Graham G, Deneudt K, Claus S, Oset P (2020) Supporting the essential - Recommendations for the development of accessible and interoperable marine biological data products. *Marine Policy* 117: 103958. <https://doi.org/10.1016/j.marpol.2020.103958>
- McClenachan L, Ferretti F, Baum JK (2012) From archives to conservation: why historical data are needed to set baselines for marine animals and ecosystems. *Conservation Letters* 5: 349–359. <https://doi.org/10.1111/j.1755-263X.2012.00253.x>
- Mouquet N, Lagadeuc Y, Devictor V, Doyen L, Duputié A, Eveillard D, Faure D, Garnier E, Gimenez O, Huneman P, Jabot F, Jarne P, Joly D, Julliard R, Kéfi S, Kergoat GJ, Lavorel S, Gall LL, Meslin L, Morand S, Morin X, Morlon H, Pinay G, Pradel R, Schurr FM, Thuiller W, Loreau M (2015) REVIEW: Predictive ecology in a changing world. *Journal of Applied Ecology* 52: 1293–1310. <https://doi.org/10.1111/1365-2664.12482>
- Paragkamian S, Sarafidou G, Mavraki D, Pavludi C, Beja J, Eliezer M, Lipizer M, Boicenco L, Zafeiropoulos H, Arvanitidis C, Pafilis E, Gerovasileiou V (submitted) Automating the curation process of historical biodiversity data using image processing and text mining technologies: new workflow and comparison of curation tools.
- Sarafidou G, Paragkamian S, Gerovasileiou V, Pafilis E, Mavraki D, Pavludi C, Arvanitidis C, Beja J, Eliezer M, Lipizer M, Boicenco L (2020) D3.7 Scientific document on the design of the workflow of text mining technologies in data archaeology. EMODnet Biology project. 29 pp.

