# Task 6.1: Analysis of data gaps

Ifremer

Sylvie  Van Iseghem, Dominique Obaton (Ifremer), Alessandra Giorgetti, Elena Partescano (OGS), Lennert Schepers, Laura Marquez (VLIZ), Peter Thijsse, Dick Schaap, Claudia Coman (MARIS), Örjan Bäck, Mikael Hedblom, Karin Wesslander (SMHI), Antonio Reppucci (MOI), Sissy Iona (HCMR), Nathalie Tonné, Oonagh McMeel, Conor Delaney (SSBE), Stéphane PESANT (EMBL-EBI)

# About this document

| Title | D6.1, Analysis of Data Gaps |
|---|---|
| Work Package | WP6 Data Management and Open Science Practices |
| Lead Partner | Ifremer, French Research Institute for the Exploitation of the Sea |
| Lead Author (Org) | Sylvie Van Iseghem (Ifremer) |
| Contributing Author(s) | Sylvie VanIseghem, Dominique Obaton (Ifremer), Alessandra Giorgetti, Elena Partescano (OGS), Lennert Schepers, Laura Marquez (VLIZ), Peter Thijsse, Dick Schaap, Claudia Coman (MARIS), Örjan Bäck, Mikael Hedblom, Karin Wesslander (SMHI), Antonio Reppucci (MOI), Sissy Iona (HCMR), Nathalie Tonné, Oonagh McMeel, Conor Delaney (SSBE), Stéphane PESANT (EMBL-EBI) |
| Reviewers | Ifremer, OGS, VLIZ, MARIS, SMHI, MOI, SSBE |
| Due Date | 31.07.2024, M5 |
| Submission Date | 25.07.2024 |
| Version | **1.1** |

## Acronyms and abbreviations

ADCP          Acoustic Doppler Current Profiler
BODC          British Oceanographic Data Centre
EMODnet       European Marine Observation and Data Network
EU            European Union
EOV           Essential Ocean Variable
CC-BY         Creative Commons Attribution
DNA           Deoxyribonucleic acid - *molecule* that carries genetic information
NODC          National Oceanographic Data Center
RI            Research Infrastructure
RNA           Ribonucleic Acid
TA (or TNA)   TransNational Access

## Figures

## Tables

# Table of Contents

# 1. Aquarius Background

The AQUARIUS project will provide a comprehensive suite of integrated research infrastructures appropriate to address significant challenges for the long-term sustainability of our unique oceans, seas and freshwater ecosystems.

Diverse Research Infrastructures (RI) will be combined to facilitate the work of researchers and key stakeholders focused on challenges and opportunities for both marine and freshwater systems; 57 research infrastructure services will be made available to include research vessels, mobile marine observation platforms, fixed marine facilities, experimental research facilities, river and basin supersites, aircraft, drones, satellite services and sophisticated data infrastructures.

In order for Transnational Access TA projects to be selected, they must convincingly integrate multiple infrastructures and contribute to the core policy objectives of Mission Ocean, that is, to protect and restore marine and freshwater ecosystems and biodiversity; to prevent and eliminate pollution of our oceans, seas and waters; and to ensure a sustainable, carbon-neutral and circular blue economy.

The proposed TA Calls will be based on a thematic and regional approach, aligning with the Lighthouse Regions (See Figure 1), that is
1. Atlantic/Arctic
2. Black Sea
3. Mediterranean Sea,
4. Baltic and the North Sea Basins

*along with their associated rivers.*
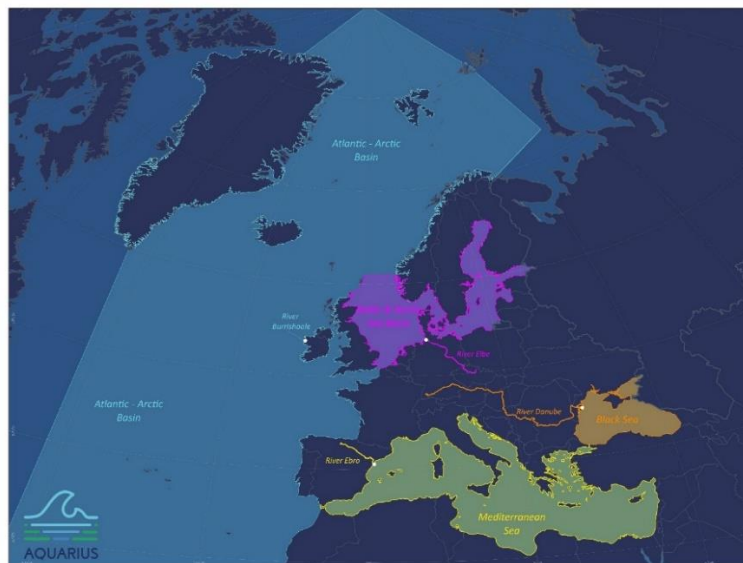


*Figure 1. Lighthouse Regions – Aquarius regional approach*

## 1.1. The role of this deliverable

This deliverable is the contribution from Task 6.1 and is part of Aquarius WP6 on Data Management and Open Science Practices towards serving the targets of the Mission 'Restore our Ocean and Waters by 2030', the European Partnership 'A climate neutral, sustainable and productive Blue Economy' and associated initiatives and projects.

The main aims of WP6 are;

- Ensure TA project researchers benefit from Copernicus, GEOSS, EMODnet and the European Digital Twin of the Ocean (DTO) initiatives, while making their TA project results available for these initiatives;
- Plan, develop and operate a federated data management and analytical system for supporting the implementation of an Open Data Strategy; this system will give FAIR access to new data, data products, and scientific knowledge, derived from the TA projects, through a common AQUARIUS Dataflow Dashboard.

This task, T6.1, involves analyzing data gaps and data availability. This analysis will aid in formulating the TA Calls for proposals and selecting proposals for each TA call by identifying the most useful and relevant data to collect for each lighthouse region. It is composed of:

- Analysing knowledge targets of the Mission 'Restore our Ocean and Waters by 2030', the European Partnership 'A climate neutral, sustainable and productive Blue Economy', the 'Marine Strategy Framework Directive' and the 'Green Deal'.
- Reviewing existing and known data availability for European waters in RIs, EMODnet, Copernicus, and beyond, for supporting research into the knowledge targets, identifying data gaps that might be included as targets in the TA project Calls.

In parallel, WP3 of AQUARIUS is designing, developing and managing the Calls for TA. Task 3.1 has identified key research and innovation actions streamlining with the topical European initiatives for marine and water issue.

Task 6.1 is complementary to Task 3.1 and provides an *Analysis on data gaps* for the four Lighthouse regions. The final Call priorities will be derived from the work of these two Tasks.

## 1.2. Contributors to this deliverable

The contributors to this deliverable are the French National Institute for Ocean Science (Ifremer) as the main author of the deliverable and leader of task 6.1; Flanders Marine Institute (VLIZ) as representative of EMODnet Biology, OGS as representative of EMODnet Chemistry, MARIS as representative of EMODnet Bathymetry, MOI as representative of Copernicus Marine, Marine Institute (MI), HCMR, and SMHI, as partners of Aquarius WP6.

## 2. Methodology

The methodology of task 6.1 is summarized in figure 2 and can be divided into four stages;
- Definition of concepts used
- Survey on data needs detailed in section 2.4.
- Compilation of results and analysis of data gaps per Lighthouse region
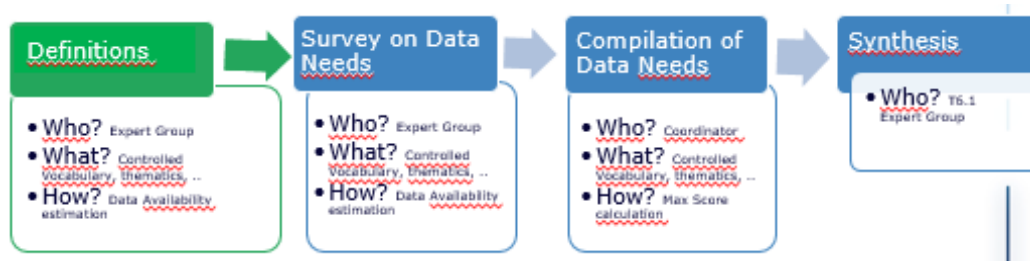- Synthesis Data gaps and key messages for TA call



*Figure 2. AQUARIUS T6.1 Methodology*

## 2.1. Group of Experts

Task 6.1 brings together experts with diverse backgrounds including National Oceanographic Data Centres (NODC) representatives, RI representatives, EMODnet and Copernicus marine representatives.

EMODnet and Copernicus Marine are two major EU initiatives for generating and making available in situ data, products and services. Their products are complementary and together cover history, present and future. For their data provision, EMODnet and Copernicus Marine work together and build upon European Research Infrastructures, SeaDataNet, EuroGOOS and in close interaction with international initiatives on data management led by the Intergovernmental Oceanographic Commission (IOC), the World Meteorological Organization (WMO), the Food and Agricultural Organization (FAO), the Group on Earth Observations (GEO), the International Council for the Exploration of the Sea (ICES) and others. They together contribute to the marine knowledge value chain: from observation to information, knowledge, public benefits or from data originators towards end-users.

The group of experts is selected from this European marine data management community, NODC relevant for lighthouse sea areas and leading representatives of EMODnet thematic groups, while each member is also part of the AQUARIUS consortium.

The expert group varies from one lighthouse to another; EMODnet representatives remain unchanged but the RI/NODC representatives differ. Here is the panel composition for each lighthouse region;

1. Atlantic/Arctic: NODC (Ifremer, MI) + EMODnet thematic representatives
2. Black Sea: NODC (NIMRD) + EMODnet thematic representatives
3. Mediterranean Sea: NODC (OGS, HCMR)+ EMODnet thematic  representatives
4. Baltic and the North Sea Basins: NODC (VLIZ, SMHI)+ EMODnet thematic  representative

Each T6.1 expert has been involved in AQUARIUS T6.1 online meetings to discuss and define controlled vocabularies, see 2.2, and thematic approach, see section 2.3 and indicator of data availability, see section 2.4.

This approach is compliant with existing ones which have been made on data gaps through EMODnet or other European initiatives see list of references in section 5.

## 2.2. Controlled vocabularies

Controlled vocabularies consist of lists of standardized terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community. The use of controlled vocabularies is an important prerequisite towards *FAIR* principle: Findable, Accessible, Interoperable and Reusable.

AQUARIUS Task 6.2 uses the NERC Vocabulary Server (NVS), which predominantly targets oceanography and its associated domains. Moderated by the BODC, the NVS hosts a list of controlled vocabulary for labelling oceanographic variables.

The SeaDataNet Parameter Disciplines (P08) and SeaDataNet Agreed Parameter Groups (P03) [1] have been used to cover the holistic approach taken by AQUARIUS Project (those

---

[1]http://seadatanet.maris2.nl/v_bodc_vocab_v2/vocab_relations.asp?lib=P08

parameters are named SDN-P03 Parameters in this document). This list has already been used in the project *Mission Atlantic, Review of available data and gaps for regional case studies*, for example, see section 5 for complete references

The use of Essential Ocean Variables (EOVs[2]) has been proposed to complete the analysis made with SDN-P03. Measurements for those variables are instrumental for generating and delivering ocean forecasts and early warnings, climate projections and assessments of measures for protecting ocean health and their impacts.

This list of EOV can also be found on the NVS[3]. The use of EOV especially important for biological aspects which do not easily use SDN-P03 Parameters for the moment.

## 2.3. Thematic approach

To ensure a complete gap analysis, four themes have been distinguished in Task 6.1;
1. Biology and Ecosystems
2. Physics, geology and Atmosphere
3. Biogeochemistry **(with cross discipline and environment)**
4. Human Activities

Table below presents the parameters studies in Task 6.1 (SDN-P03 and EOV's) for the four themes.

| Biology and Ecosystem | |
|---|---|
| *SeaDataNet Agreed Parameter Groups (P03)* | *Essential Ocean Variables* |
| B070 - Biota abundance, biomass and diversity | *Emerging EOV* "Microbe biomass & diversity" |
| B070 - Biota abundance, biomass and diversity - deep sea (< 200m depth) | EOV - Macroalgal canopy cover and composition |
| B015 - Birds, mammals and reptiles | EOV - Seagrass cover and composition |
| B060 - Disease, damage and mortality | EOV - Phytoplankton biomass and diversity |
| B020 - Fish | EOV - Zooplankton biomass and diversity |
| B055 - Macroalgae and seagrass | EOV - Fish abundance and distribution |
| B030 - Phytoplankton and microphytobenthos | EOV - Hard coral cover and composition |
| G055 - Rock and sediment biota | EOV - Marine turtles, birds, mammals abundance and distribution (marine turtles) |
| B045 - Zooplankton | EOV - Marine turtles, birds, mammals abundance and distribution (mammals) |
| B005 - Bacteria and viruses | EOV - Marine turtles, birds, mammals abundance and distribution (birds) |
| **BioGeochemistry** *with cross discipline and environment* | |
| *SeaDataNet Agreed Parameter Groups (P03)* | *Essential Ocean Variables* |
| C015 - Dissolved gases | EOV - Oxygen |
| C040 - Nutrients | EOV - Nutrients |
| C010 - Carbonate system | EOV - Inorganic carbon |
| C005 - Carbon Nitrogen and phosphorus | EOV - Transient tracers |
| O005 - Fluxes | EOV - Particulate matter |
| O010 - Rate measurements | EOV - Nitrous oxide |
| B050 - Habitat | EOV - Stable carbon isotopes |
| H001 - Pollution | EOV - Dissolved organic carbon |
| **Physics, Geology, Atmosphere** | |
| *SeaDataNet Agreed Parameter Groups (P03)* | *Essential Ocean Variables* |
| G005 - Gravity, magnetics and bathymetry | EOV - Sea state |
| G040 - Rock and sediment physical properties | EOV - Ocean surface stress |
| G060 - Sedimentation and erosion processes | EOV - Sea ice |
| M010 - Meteorology | EOV - Sea surface height |
| D005 - Acoustic | EOV - Sea surface temperature |
| D020 - Currents | EOV - Subsurface temperature |
| D025 - Water column temperature and salinity | EOV - Surface currents |
| D034 - Wave | EOV - Subsurface currents |
| | EOV - Sea surface salinity |
| | EOV - Subsurface salinity |
| | EOV - Ocean surface heat flux |
| | EOV - Ocean bottom pressure |
| **Human Activities** | |
| Aquaculture | Fisheries |
| H006 - Aquaculture | H004 - Fisheries |
| H007 - Area Management | H011 - Hydrocarbon extraction |
| H008 - Cables | H012 - Mining |
| H002 - Construction and structures | H015 - Tourism |
| H009 - Cultural Heritage | H016 - Transport |
| H010 - Energy | |

*Table 1. Parameters studies in Task 6.1: SDN-P03 and EOV's*

---

[2]https://goosocean.org/what-we-do/framework/essential-ocean-variables/
[3] http://seadatanet.maris2.nl/v_bodc_vocab_v2/vocab_relations.asp?lib=A05

## 2.4. Data Availability estimation

Data needs were recorded by experts using a shared Excel file presenting, for each Lighthouse region, all the parameters described in section 2.3 (line dimension).

Each expert had a dedicated column (either EMODnet thematic representative or RI/NODC representative) to fill. The score for each controlled variable is the following;
- ✓ 0 if the Data Need is not significant or if the Expert cannot identify a need for this variable (White colour)
- ✓ 1 if the Data Need is important (light green colour)
- ✓ 2 if the Data Need is very important (light green colour)

Among the criteria for the data to be available, Task 6.1 wishes to highlight the importance of two prerequisite technical criteria (Metadata and Reusability of the data).

### Metadata
According to the International Organization for Standardization (ISO): "Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often called data about data." Data without metadata is meaningless. A variety of questions arise about data such as Where did it come from? Who is the originator? What are the restrictions to reuse it? Metadata can answer these questions and many more.

When metadata follows a standard format, such as *ISO 19115 - ISO 19139*[4] for geographic information datasets or *Oceansites*[5] for long-term, high-frequency observations at fixed locations, this enables interoperability, increases the quality of data and facilitates greater use of data.

In parallel, the use of common vocabularies at the metadata description as well as the data formats will contribute to the AQUARIUS data interoperability and enhance their integration with the EU marine data infrastructures and services.

### Data reusability
A key success factor for research activities in general and especially in AQUARIUS project is the degree of openness of the data.

Many EU countries offer local or national portals to provide access to open data for public research. However, datasets recorded in catalogues and their reuse conditions are extremely diverse. Therefore, it is necessary to analyse the license and the data accessibility provided in the metadata.

The license is an essential information for data reusability, and it must be verified to ensure the data is open for reuse. Most of the marine data are shared under Common Creative Licenses, offering seven types of licenses with main provisions and conditions applied to reuse available data (cf. Table 2).

---

[4]https://www.iso.org/standard/26020.html

[5]https://goosocean.org/who-we-are/observations-coordination-group/global-ocean-observing-networks/oceansites/

| Common Creative (CC) licenses | Public domain | Attribution | Share-alike (SA) | Non-commercial (NC) | No derivatives (ND) |
|---|---|---|---|---|---|
| CC0 | ■ | | | | |
| CC BY | | ■ | | | |
| CC BY SA | | ■ | ■ | | |
| CC BY NC | | ■ | | ■ | |
| CC BY ND | | ■ | | | ■ |
| CC BY NC SA | | ■ | ■ | ■ | |
| CC BY NC ND | | ■ | | ■ | ■ |

*Table 2. Types of licenses and conditions applied to reuse available data*

The dissemination of marine data is governed by several international conventions and European directives, which aim to ensure broad access to environmental observation data but also specify some restrictions due to data sensitivity (protection of the economic resources of countries, protection of sensitive sites, and protection of individual data).

Most data providers offer access to the data through local or national catalogues, which provide a download link or a web service. In other cases, data access is restricted and a request is mandatory.

There is an international trend to move to CC-BY (4.0) as default license as that is most open and easy to deploy. It is also promoted and used as part of EMODnet, Copernicus Marine (CC-BY equivalent is used), and several data management networks and infrastructures such as SeaDataNet.

Besides license and data accessibility, provenance information is another critical element for ensuring data reusability.

AQUARIUS within its data management activities in WP5 and WP6 will ensure that the required information on how the data was created will be available.

## 2.5. Compilation of data needs for AQUARIUS project

From the shared Excel file, results have been compiled by lighthouse region with the "Maximum strategy": for Each Controlled parameter (each line), the maximum score of the expert have been kept.

Data needs are presented in section 3 first globally and then by Lighthouse Regions; references to scientific expertise have been added and used to establish the final compilation of data needs.

## 3. Data Gaps analysis

The present analysis will help the selection of proposals submitted to the different TA calls by providing an assessment, for each lighthouse region, of the current gaps in data supporting Essential Ocean Variables (EOVs).

At the European level, some gaps have been identified across all lighthouse regions and are described below for each AQUARIUS theme; Biology and Ecosystems; Physics, geology and Atmosphere and Biogeochemistry.

For Biology and Ecosystems theme, different needs have been highlighted;

- High frequency observations of phytoplankton biomass and diversity including harmful algal blooms species[6]: "*phytoplankton biomass and diversity have been identified as Essential Ocean Variables by the UN IOC-Global Ocean Observing System.*

- Genomics data (i.e. data derived from DNA and RNA material) collected in the aquatic environment are very abundant but its ingestion by data infrastructures for marine biodiversity, such as EMODnet, is still limited. The European Bioinformatics Institute (EBI), part of the European Molecular Biology Laboratory (EMBL), provides a comprehensive open record of the world's genomics data, including raw sequencing data, sequence assembly information, taxonomic and functional annotations, and high granularity provenance metadata. Its content is measured in millions of marine samples, hundreds of thousands of sequence libraries and petabytes of storage. EBI coordinates the ELIXIR European infrastructure for life science, and acts as both a Core Data Resource, and a Deposition Database. As a founding member of the celebrated International Nucleotide Sequence Database Collaboration (INSDC), EBI drives international standards and best practice in its domain. EBI currently contributes data via APIs to the European Blue-Cloud and the Global Biodiversity Information Facility (GBIF) and works in close collaboration with VLIZ who operates WoRMS and EMODnet Biology. Standards used by the biodiversity and genomics communities were developed separately, which created challenges for the integration and publication of sequences and DNA-derived data into marine biodiversity data platforms, such as EMODnet. Hence, the creation of quality-controlled genomic Darwin Core occurrence tables was identified as a need. The capacity to ingest, store and share DNA-derived datasets is a priority in the current and future EMODnet Biology phases, in close collaboration with EMBL-EBI.

- Other specific and innovative data types transformed to Darwin Core format and quality controlled, such as imaging data, biologging data, passive acoustics data and citizen science data have been highlighted in current EU projects [7].

- Transversal data needs on the deep sea (waters below 200 m) with an ecological focus with different scientific studies associated[8][9] .The importance of Vulnerable Marine Ecosystems (VMEs) data is emphasized for use in annual ICES advisory processes and the development of new methods to further our understanding of deep-sea ecosystems, and further suggests novel management tools to ensure human activities do not adversely affect them.

- Less represented taxonomic groups in the EMODnet Biology database (< 1% of the total number of occurrences in the EMODnet Biology - EurOBIS database):

    o Bacteria (Bacteria and Archaea). The number of Bacteria occurrences in the EMODnet Biology database is approximately 190 000, less than 0.50% of the total database occurrences. Due to the complex and diverse nature of microbial communities, DNA methods and sequencing technologies are the most effective techniques to capture bacterial biodiversity. Current developments for DNA-derived data integration into EMODnet Biology will

---

[6]http://www.e-pages.dk/ku/1516/
[7]DTO-Bioflow:
[8]Global Observing Needs in the Deep Ocean
[9]Working Group on Deep-water Ecology (WGDEC)

facilitate the ingestion of these occurrences and improve the availability of microbial data in marine biodiversity data initiatives in the future.
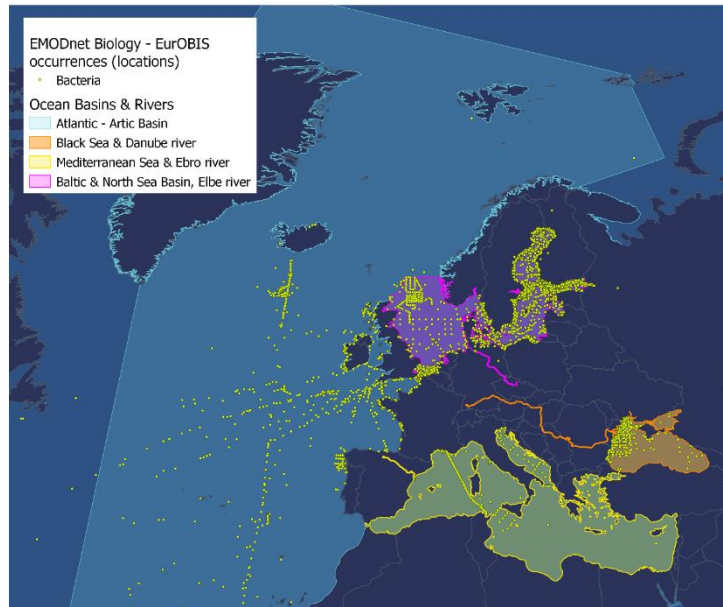


*Figure 3. EMODnet Biology less represented EOV taxonomic groups: Bacteria (Bacteria and Archaea) location of occurrence records.*

o Seagrasses (monocotyledon plants from Hydrocharitaceae, Zosteraceae, Cymodoceaceae, Ruppiaceae and Posidoniaceae families, figure 3). The number of seagrasses occurrences in EMODnet Biology – EurOBIS database is limited to less than 45 000, representing 0.11% of the database. EMODnet Seabed Habitat has done efforts to delineate the known distribution of this group in European waters in this data product: https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/meta data/39746d9c-4220-425c-bc26-7cb3056c36a5.
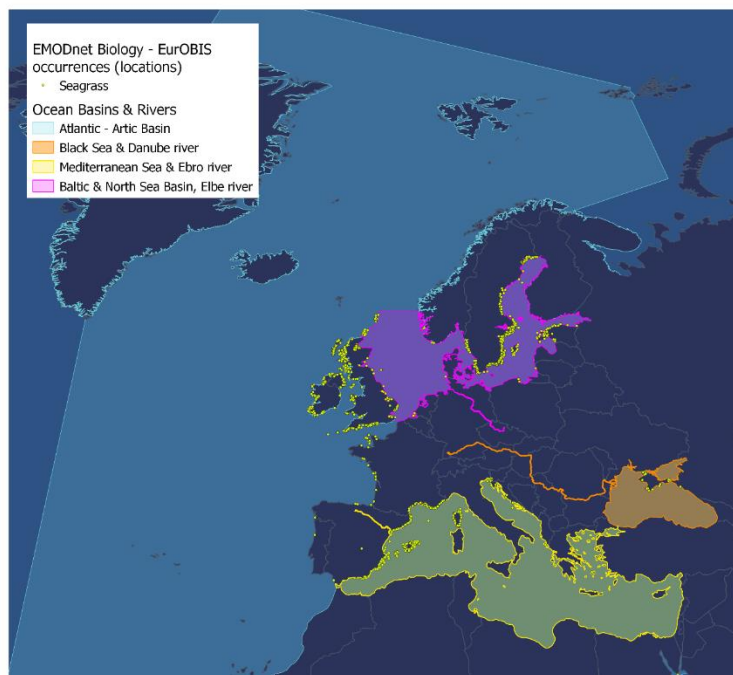
*Figure 4. EMODnet Biology less represented EOV taxonomic groups: Seagrasses (Hydrocharitaceae, Zosteraceae, Cymodoceaceae, Ruppiaceae and Posidoniaceae families) locations of occurrence records.*

- o Hard coral (Order Scleractinia, figure 4). Hard corals are less abundant and diverse in European regions compared to tropical regions. Hence the number of hard coral occurrences in the EMODnet Biology – EurOBIS database is low, around 50 000 occurrences, 0.13% of the total database (see annex 1).
- o Annex 1 presents the species occurrences distribution as geospatial grids for three more EOV taxonomic groups (mammals, fishes and birds) which present data gaps in various Aquarius basins, particularly the Artic Sea and the                                       Black                                       Sea.

For the chemistry, cross discipline and environment topic, a general need is on following parameters;
- Carbonate system: variable pH, pCO2, TA (Total Alkalinity) - SDN-P03-C010
- Carbon, nitrogen and phosphorus - SDN - P03-C005:

Those parameters are linked with the Sustainable Development Goal 14.3.1 Indicator (SDG 14.3.1) for 'average marine acidity (pH) measured at agreed suite of representative sampling stations', to support the United Nations 2030 Agenda.

Looking at the chemical groups of EMODnet Chemistry, the above parameters are included in the acidity and organic matter groups for the carbon component. Apart from the chemical pollution data, these parameters have a very non-uniform spatial resolution with very large and uncovered areas, which does not allow the creation of distribution maps.



*Figure 6. EMODnet Chemistry less represented parameter groups in the field of eutrophication and ocean acidification. Organic matter includes chemical species (primarily forms of carbon and nitrogen) resulting from sewage outfalls and aquaculture while Acidity includes the parameters that quantify the carbonate system in seawater (pH, alkalinity, TCO2 and pCO2).*

Global Data Needs in Physics, Geology and Atmosphere are about Bathymetry and Acoustics Data.

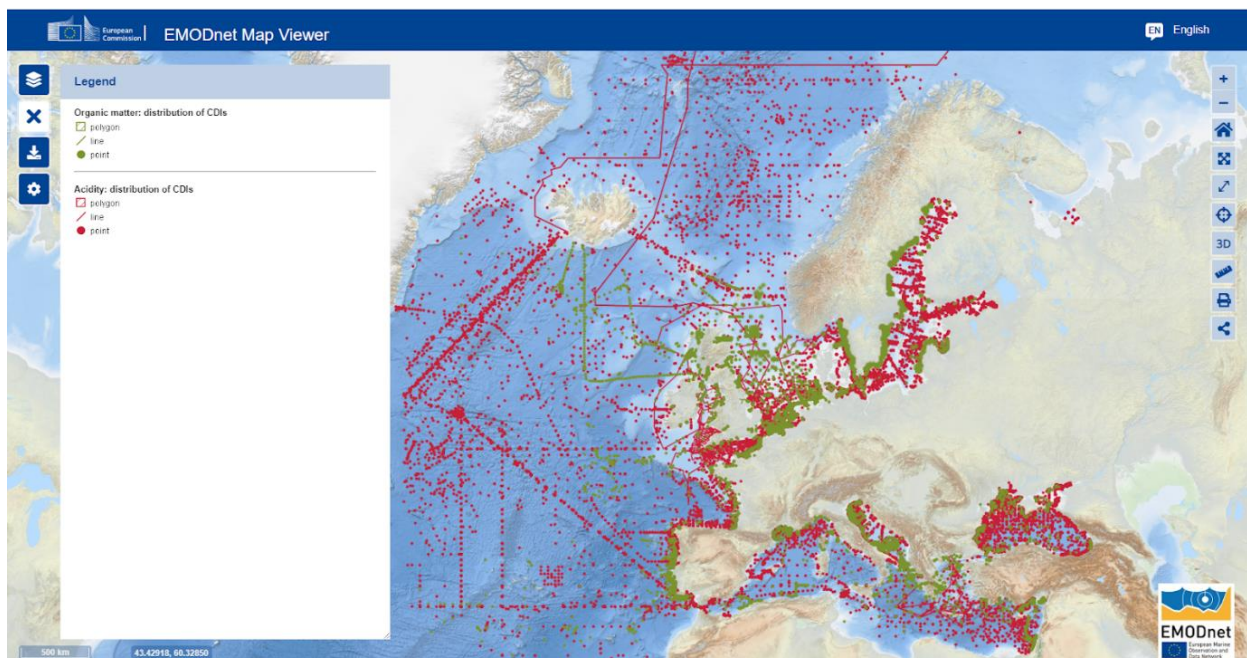For the main physical parameters: temperature, salinity, waves and sea level, the coverage of data is usually good. Of course, it needs to be maintained and can be enhanced especially for wave data that are very few in South Hemisphere. Each area has reported that currents data are missing, most regions report it has a very important gap.

But other parameters are more critical as ice data, river data (at the limit of tidal influence) and sounds data (acoustics data in water-mass and sea floor). Although more common, atmospheric data over the sea, as wind and humidity, are also of great interest.

Objective of EMODnet physics is to harvest all available data in European sea as soon they are free and open. This is done directly from the providers or through various databases or data integrators such as the Copernicus marine service.

Today, the main data need is about sound data (acoustics data in water-mass and sea floor, in particular pollution from echosounders and other sources) from any area and it appears that data is scarce and difficult to find.

Concerning Bathymetry, EMODnet Bathymetry is the leading initiative in Europe for bringing together bathymetric survey data sets and compiling a harmonised Digital Terrain Model.

The EMODnet Bathymetry consortium is active since 2008 and is developing the EMODnet Digital Terrain Model (DTM) for the European seas, which is published each 2 years, each time extending coverage, improving quality and precision, and expanding functionalities for viewing, using, and downloading.

The EMODnet DTM has a uniform grid resolution of $115 * 115 \text{ m}^2$ and this way is a unique product on the market, which is very popular for multiple user communities.

As each grid cell in the resulting EMODnet DTM contains a reference to the prevailing basis dataset that has been used for determining the depth parameters of the specific grid area, this facilitates to generate and publish a 'Source Reference Map', allowing to retrieve metadata.
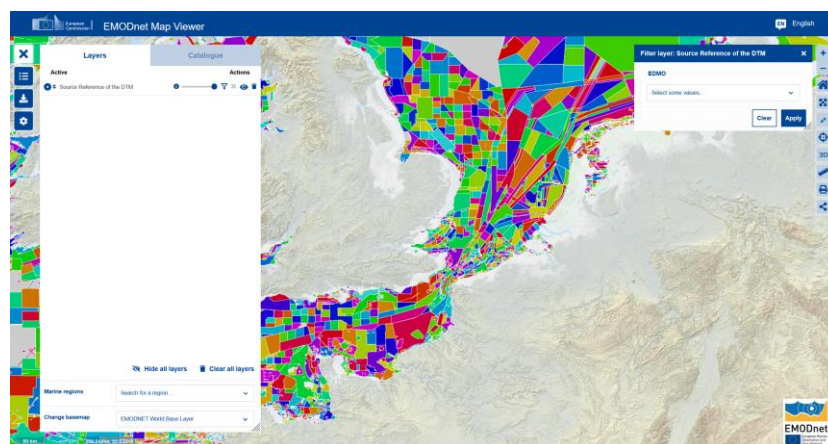


*Figure 7: Source Reference Map to retrieve metadata on used input data sets and option for filtering on data provider*

In addition, EMODnet Bathymetry agreed on a classification of qualitative aspects of the source data, which provides the so-called **Quality Index**. Those aspects rely on the precision reached by the individual components of the system, characterized by the horizontal indicator (QI_Horizontal) and the vertical indicator (QI_Vertical). Associated with these elements are the age of the survey indicator (QI_Age) and the purpose of the survey indicator (QI_purpose). These four fields, supported by controlled classes, have been added to the CDI schema and are maintained by data providers. Moreover, the quality information from the input data sets is used to generate **Quality Index** map layers, included in the Central Map Viewer service, associated with the latest EMODnet DTM.

*The DTMs are produced from survey data sets (derived from single beam, multi beam, LIDAR, and other observation techniques, including Satellite Derived Bathymetry) and aggregated data sets (Composite DTMs) that are referenced with metadata adopting the SeaDataNet Catalogue services. Bathymetric survey data sets are gathered and populated by national hydrographic services, marine research institutes, and companies in the SeaDataNet CDI Data Discovery & Access service.*

*Currently, this amounts to more than 42.000 datasets, received from 63 data providers. Further gathering is on-going and contributes each cycle to improvements of the EMODnet DTM. There are also cooperations and exchanges with regional and international bathymetry networks, such as NSDB (North Sea), BSDB (Baltic Sea), IBCAO (Arctic), GEBCO, Seabed 2030, and IHO, which ensure interoperability, quality assurance, and promotion. Moreover, the EMODnet DTM is integrated in the Global DTM (GEBCO) as European contribution to the SeaBed2030 initiative, while GEBCO data is inserted in the EMODnet DTM for areas where no survey data sets are available.*
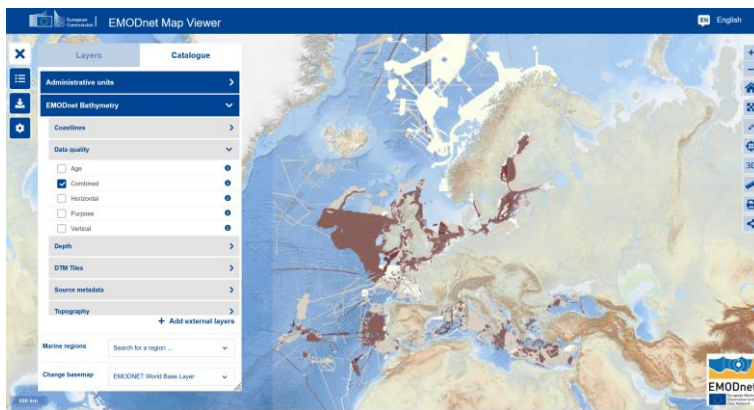


*Figure 8: Quality Index Map for age of input data sets. Most dark brown are youngest. Users can also switch on maps for horizontal and vertical resolution, instrumentation used, and overall combined quality indicators. Darkest areas have highest quality indication.*

.

From the current Quality Index map it can be concluded that Bathymetry represents an important gap in all 4 sea basins. The Baltic seems the best as it has quite a dark brown colour, but zooming in one will find it still has large areas with less quality.

*Through its activities and gathering of bathymetric survey data for the European seas and major part of the North West Atlantic Ocean, using the SeaDataNet CDI Data Discovery & Access Service infrastructure, EMODnet Bathymetry maintains an excellent overview of available bathymetry data sets. The SeaDataNet catalogue makes use of the ISO19115-19139 metadata standard, which includes many metadata elements that are completed using SeaDataNet vocabularies and directories. The overview can also be used to get an indication of areas which are not yet covered with survey data and otherwise the quality of surveys in specific areas, which are covered.*

To conclude, it is recommended that bathymetry surveying is included in any TA proposals that concern organising scientific cruises with research vessels.

The data management should then ensure that the collected and processed survey data become available for EMODnet Bathymetry, enriching the basis for generating the EMODnet DTM.

## 3.1. Atlantic-Arctic basin

Both Atlantic and Arctic Basins have been defined in deliverable 3.1; the first is within the Atlantic Ocean and the second covers the areas at the northernmost part of our planet.

In terms of specific data needs, we can note;
- For Biology and Ecosystem theme: 4 Very Important gaps (dark green): Biota, abundance, biomass and diversity, Zooplanction, Bacteria and Virus and lots of important gaps with many EOV identified

- For Physic thematic: 4 Very Important gaps (dark green) in Acoustics, Current, Sea Ice and Ocean Bottom pressure

- For Chemistry thematic: 3 Very Important gaps (dark green) in Dissolved gases, Carbonate system and Habitat

| Biology and Ecosystem | | | | |
|---|---|---|---|---|
| *SeaDataNet Agreed Parameter Groups (P03)* | *Gaps* | *Essential Ocean Variables* | *Gaps* | |
| B070 - Biota abundance, biomass and diversity | 2 | *Emerging EOV "Microbe biomass & diversity"* | 2 | |
| B070 - Biota abundance, biomass and diversity - deep sea (< 200m depth) | 2 | EOV - Macroalgal canopy cover and composition | 1 | |
| B015 - Birds, mammals and reptiles | 1 | EOV - Seagrass cover and composition | 1 | |
| B060 - Disease, damage and mortality | | EOV - Phytoplankton biomass and diversity | 1 | |
| B020 - Fish | 1 | EOV - Zooplankton biomass and diversity | 1 | |
| B055 - Macroalgae and seagrass | 1 | EOV - Fish abundance and distribution | 1 | |
| B030 - Phytoplankton and microphytobenthos | 1 | EOV - Hard coral cover and composition | | |
| G055- Rock and sediment biota | | EOV - Marine turtles, birds, mammals abundance and distribution (marine turtles) | | |
| B045 - Zooplankton | 2 | EOV - Marine turtles, birds, mammals abundance and distribution (mammals) | | |
| B005 - Bacteria and viruses | 2 | EOV - Marine turtles, birds, mammals abundance and distribution (birds) | 1 | |
| **Biogeochemistry** *with cross discipline and environment* | | | | |
| *SeaDataNet Agreed Parameter Groups (P03)* | *Gaps* | *Essential Ocean Variables* | *Gaps* | |
| C015 - Dissolved gases | 2 | EOV - Oxygen | | |
| C040 - Nutrients | 1 | EOV - Nutrients | | |
| C010 - Carbonate system | 2 | EOV - Inorganic carbon | | |
| C005 - Carbon Nitrogen and phosphorus | | EOV - Transient tracers | | |
| O005 - Fluxes | | EOV - Particulate matter | | |
| O010 - Rate measurements | | EOV - Nitrous oxide | | |
| B050 - Habitat | 2 | EOV - Stable carbon isotopes | | |
| H001 - Pollution | 1 | EOV - Dissolved organic carbon | | |
| **Physics, Geology, Atmosphere** | | | | |
| *SeaDataNet Agreed Parameter Groups (P03)* | *Gaps* | *Essential Ocean Variables* | *Gaps* | |
| M010 - Meteorology | | EOV - Sea state | 1 | |
| G005 - Gravity, magnetics and bathymetry | 1 | EOV - Ocean surface stress | | |
| G040 - Rock and sediment physical properties | | EOV - Sea ice | 2 | |
| G060 - Sedimentation and erosion processes | 1 | EOV - Sea surface height | | |
| D005 - Acoustic | 2 | EOV - Sea surface temperature | | |
| D020 - Currents | 2 | EOV - Subsurface temperature | | |
| D025 - Water column temperature and salinity | 1 | EOV - Surface currents | 1 | |
| D034 - Wave | | EOV - Subsurface currents | 1 | |
| | | EOV - Sea surface salinity | | |
| | | EOV - Subsurface salinity | | |
| | | EOV - Ocean surface heat flux | | |
| | | EOV - Ocean bottom pressure | 2 | |
| **Human Activities** | | | | |
| Aquaculture | Data Gaps | Fisheries | | |
| Area Management | | Hydrocarbon extraction | | |
| Cables | | Mining | | |
| Construction and structures | | Tourism | | |
| Cultural Heritage | | Transport | | |
| Energy | | | | |

Table 3: Data gaps for Atlantic-Arctic basin

## 3.2. Black Sea and Danube River system

In terms of Data gaps; Black Sea and Danube River system are particular and represent the lighthouse region with the highest need in terms of data. This can be explained with its geographical situation, covering both EU and non-EU countries, with diverse political systems and different socio-economic backgrounds.

Different projects have reported[10][11] gaps identification in this area which are complementary from the T6.1 analysis on this lighthouse.

Very important data needs are
- For Biology and Ecosystem : Bacteria and Virus, Seagrass Cover and composition, Fish Abundance and distribution, Mammals and Birds abundance distribution

- For Physics with acoustic and currents parameters and EOV: Sea Ice and Ocean Bottom pressure

- For Chemistry: Carbonate System and Pollution

| **Biology and Ecosystem** | | | |
| --- | --- | --- | --- |
| *SeaDataNet Agreed Parameter Groups (P03)* | *Gaps* | *Essential Ocean Variables* | *Gaps* |
| B070 - Biota abundance, biomass and diversity | | *Emerging EOV "Microbe biomass & diversity"* | 2 |
| B070 - Biota abundance, biomass and diversity - deep sea (< 200m depth) | | EOV - Macroalgal canopy cover and composition | |
| B015 - Birds, mammals and reptiles | 2 | EOV - Seagrass cover and composition | 2 |
| B060 - Disease, damage and mortality | | EOV - Phytoplankton biomass and diversity | |
| B020 - Fish | 2 | EOV - Zooplankton biomass and diversity | |
| B055 - Macroalgae and seagrass | | EOV - Fish abundance and distribution | 2 |
| B030 - Phytoplankton and microphytobenthos | | EOV - Hard coral cover and composition | |
| G055 - Rock and sediment biota | | EOV - Marine turtles, birds, mammals abundance and distribution (marine turtles) | |
| B045 - Zooplankton | | EOV - Marine turtles, birds, mammals abundance and distribution (mammals) | 2 |
| B005 - Bacteria and viruses | 2 | EOV - Marine turtles, birds, mammals abundance and distribution (birds) | 2 |
| **Biogeochemistry** *with cross discipline and environment* | | | |
| *SeaDataNet Agreed Parameter Groups (P03)* | *Gaps* | *Essential Ocean Variables* | *Gaps* |
| C015 - Dissolved gases | | EOV - Oxygen | |
| C040 - Nutrients | | EOV - Nutrients | |
| C010 - Carbonate system | 1 | EOV - Inorganic carbon | |
| C005 - Carbon Nitrogen and phosphorus | | EOV - Transient tracers | |
| O005 - Fluxes | | EOV - Particulate matter | |
| O010 - Rate measurements | | EOV - Nitrous oxide | |
| B050 - Habitat | | EOV - Stable carbon isotopes | |
| H001 - Pollution | 1 | EOV - Dissolved organic carbon | |
| **Physics, Geology, Atmosphere** | | | |
| *SeaDataNet Agreed Parameter Groups (P03)* | *Gaps* | *Essential Ocean Variables* | *Gaps* |
| M010 - Meteorology | 1 | EOV - Sea state | 1 |
| G005 - Gravity, magnetics and bathymetry | 1 | EOV - Ocean surface stress | |
| G040 - Rock and sediment physical properties | | EOV - Sea ice | 2 |
| G060 - Sedimentation and erosion processes | | EOV - Sea surface height | 1 |
| D005 - Acoustic | 2 | EOV - Sea surface temperature | 1 |
| D020 - Currents | 2 | EOV - Subsurface temperature | 1 |
| D025 - Water column temperature and salinity | 1 | EOV - Surface currents | 1 |
| D034 - Wave | 1 | EOV - Subsurface currents | 1 |
| | | EOV - Sea surface salinity | 1 |
| | | EOV - Subsurface salinity | 1 |
| | | EOV - Ocean surface heat flux | |
| | | EOV - Ocean bottom pressure | 2 |
| **Human Activities** | | | |
| Aquaculture | Data Gaps | Fisheries | |
| Area Management | | Hydrocarbon extraction | |
| Cables | | Mining | |
| Construction and structures | | Tourism | |
| Cultural Heritage | | Transport | |
| Energy | | | |

Table 4: Data gaps for Black Sea and Danube River system

[10]ANEMONE Deliverable 1.1, 2021. "Overview of monitoring programs, gaps identification, and research
[11]Emblas Project

## 3.3.  Mediterranean Sea

Through T6.1 experts, very important Data gaps in the Mediterranean Sea have been recorded on Bacteria and virus, on Seagrass Cover and composition, Fish Abundance and distribution, Birds abundance distribution, Acoustics and Ocean pressure.

Scientific studies emphasize the need for more coordinated approaches on a) microplastics marine litter sampling, management, formats (especially on biota), and b) on monitoring activities with a stable sampling frequency for plankton time series. Regarding data needs, the following have been identified:
-   Chemical data: on pH, pCO2, DIC and TA (Total Alkalinity), which are related with the ocean acidification - a major stressor on marine ecosystems.

-   Biological data: on important habitats such as posidonia, Coralligenous habitats, Maerl habitats as well as plankton time series[12]:

-   Physics data: on direct deep current measurements (until 500-600 m) since drifters or satellites cover only the surface up to 50-60 m. An approcah that has already proven successful in the Atlantic between USA and Europe traverses is leveraging the existing network of ships of opportunity or volunteer observing ships by equipping these vessels with vessel-mounted Acoustic Doppler Current Profilers (ADCPs). It would be valuable if AQUARIUS could address this data gap challenge and manage to collect deep current data, b) on acoustic data to better understand the underwater noise levels effect[13] to the distribution of the species habitats.

| Biology and Ecosystem | | | | |
|---|---|---|---|---|
| *SeaDataNet Agreed Parameter Groups (P03)* | Gaps | *Essential Ocean Variables* | | Gaps |
| B070 - Biota abundance, biomass and diversity | 1 | *Emerging EOV "Microbe biomass & diversity"* | | 2 |
| B070 - Biota abundance, biomass and diversity - deep sea (< 200m depth) | 1 | EOV - Macroalgal canopy cover and composition | | |
| B015 - Birds, mammals and reptiles | 2 | EOV - Seagrass cover and composition | | 2 |
| B060 - Disease, damage and mortality | | EOV - Phytoplankton biomass and diversity | | 1 |
| B020 - Fish | 2 | EOV - Zooplankton biomass and diversity | | 1 |
| B055 - Macroalgae and seagrass | | EOV - Fish abundance and distribution | | 2 |
| B030 - Phytoplankton and microphytobenthos | 1 | EOV - Hard coral cover and composition | | |
| G055 - Rock and sediment biota | | EOV - Marine turtles, birds, mammals abundance and distribution (marine turtles) | | 1 |
| B045 - Zooplankton | 1 | EOV - Marine turtles, birds, mammals abundance and distribution (mammals) | | 1 |
| B005 - Bacteria and viruses | 2 | EOV - Marine turtles, birds, mammals abundance and distribution (birds) | | 2 |
| **Biogeochemistry** with cross discipline and environment | | | | |
| *SeaDataNet Agreed Parameter Groups (P03)* | Gaps | *Essential Ocean Variables* | | Gaps |
| C015 - Dissolved gases | | EOV - Oxygen | | |
| C040 - Nutrients | 1 | EOV - Nutrients | | |
| C010 - Carbonate system | 2 | EOV - Inorganic carbon | | |
| C005  Carbon, nitrogen and phosphorus | 2 | EOV - Transient tracers | | |
| O005 - Fluxes | | EOV - Particulate matter | | |
| O010 - Rate measurements | | EOV - Nitrous oxide | | |
| B050 - Habitat | | EOV - Stable carbon isotopes | | |
| H001 - Pollution | 1 | EOV - Dissolved organic carbon | | |
| **Physics, Geology, Atmosphere** | | | | |
| *SeaDataNet Agreed Parameter Groups (P03)* | Gaps | *Essential Ocean Variables* | | Gaps |
| M010 - Meteorology | | EOV - Sea state | | |
| G005 - Gravity, magnetics and bathymetry | 1 | EOV - Ocean surface stress | | |
| G040 - Rock and sediment physical properties | | EOV - Sea ice | | |
| G060 - Sedimentation and erosion processes | | EOV - Sea surface height | | |
| D005 - Acoustic | 2 | EOV - Sea surface temperature | | |
| D020 - Currents | 1 | EOV - Subsurface temperature | | |
| D025 - Water column temperature and salinity | | EOV - Surface currents | | 1 |
| D034 - Wave | | EOV - Subsurface currents | | 1 |
| | | EOV - Sea surface salinity | | |
| | | EOV - Subsurface salinity | | |
| | | EOV - Ocean surface heat flux | | |
| | | EOV - Ocean bottom pressure | | 2 |
| **Human Activities** | | | | |
| Aquaculture | Gaps | Fisheries | | |
| Area Management | | Hydrocarbon extraction | | |
| Cables | | Mining | | |
| Construction and structures | | Tourism | | |
| Cultural Heritage | | Transport | | |
| Energy | | | | |

Table 5: Data gaps for Mediterranean Sea

[12] Monitoring and modelling marine zooplankton in a changing climate
[13] Addressing underwater noise in Europe

## 3.4. Baltic & North Sea

For Baltic & North Sea lighthouse, very important data need have been noted on Bacteria and virus, biota abundance, biomass and diversity, and phytoplankton and microphytobenthos.

One scientific Article presents[14] gaps in current Baltic Sea environmental monitoring saying that Baltic Sea monitoring is insufficient with respect to the regulatory requirements and that Biodiversity monitoring is the category containing most gaps.

| Biology and Ecosystem | | | |
|---|---|---|---|
| SeaDataNet Agreed Parameter Groups (P03) | Gaps | Essential Ocean Variables | Gaps |
| B070 - Biota abundance, biomass and diversity | 2 | Emerging EOV "Microbe biomass & diversity" | 2 |
| B070 - Biota abundance, biomass and diversity - deep sea (< 200m depth) | 1 | EOV - Macroalgal canopy cover and composition | |
| B015 - Birds, mammals and reptiles | 1 | EOV - Seagrass cover and composition | 1 |
| B060 - Disease, damage and mortality | 0 | EOV - Phytoplankton biomass and diversity | 1 |
| B020 - Fish | 1 | EOV - Zooplankton biomass and diversity | |
| B055 - Macroalgae and seagrass | 1 | EOV - Fish abundance and distribution | 1 |
| B030 - Phytoplankton and microphytobenthos | 2 | EOV - Hard coral cover and composition | |
| G055 - Rock and sediment biota | 0 | EOV - Marine turtles, birds, mammals abundance and distribution (marine turtles) | |
| B045 - Zooplankton | 0 | EOV - Marine turtles, birds, mammals abundance and distribution (mammals) | 1 |
| B005 - Bacteria and viruses | 2 | EOV - Marine turtles, birds, mammals abundance and distribution (birds) | |
| Biogeochemistry with cross discipline and environment | | | |
| SeaDataNet Agreed Parameter Groups (P03) | Gaps | Essential Ocean Variables | Gaps |
| C015 - Dissolved gases | 2 | EOV - Oxygen | |
| C040 - Nutrients | 2 | EOV - Nutrients | 2 |
| C010 - Carbonate system | 1 | EOV - Inorganic carbon | |
| C005 - Carbon, nitrogen and phosphorus | | EOV - Transient tracers | |
| O005 - Fluxes | 2 | EOV - Particulate matter | |
| O010 - Rate measurements | | EOV - Nitrous oxide | |
| B050 - Habitat | 2 | EOV - Stable carbon isotopes | |
| H001 - Pollution | 2 | EOV - Dissolved organic carbon | |
| Physics, Geology, Atmosphere | | | |
| SeaDataNet Agreed Parameter Groups (P03) | Gaps | Essential Ocean Variables | Gaps |
| M010 - Meteorology | 1 | EOV - Sea state | 1 |
| G005 - Gravity, magnetics and bathymetry | 1 | EOV - Ocean surface stress | |
| G040 - Rock and sediment physical properties | 0 | EOV - Sea ice | 2 |
| G060 - Sedimentation and erosion processes | 0 | EOV - Sea surface height | |
| D005 - Acoustic | 1 | EOV - Sea surface temperature | |
| D020 - Currents | 2 | EOV - Subsurface temperature | |
| D025 - Water column temperature and salinity | 2 | EOV - Surface currents | 1 |
| D034 - Wave | 1 | EOV - Subsurface currents | 1 |
| | 1 | EOV - Sea surface salinity | |
| | | EOV - Subsurface salinity | |
| | | EOV - Ocean surface heat flux | |
| | | EOV - Ocean bottom pressure | 2 |
| Human Activities | | | |
| Aquaculture | Gaps | Fisheries | |
| H006 - Aquaculture | | H004 - Fisheries | 1 |
| H007 - Area Management | 1 | H011 - Hydrocarbon extraction | |
| H008 - Cables | 1 | H012 - Mining | 1 |
| H002 - Construction and structures | 1 | H015 - Tourism | 1 |
| H009 - Cultural Heritage | | H016 - Transport | 2 |
| H010 - Energy | 1 | | |

Table 6: Data gaps for Baltic & North Sea

---

[14] Gaps in current Baltic Sea environmental monitoring – Science versus management perspectives

# 4. Aquarius Data gaps: Key messages for TA calls

For each Lighthouse region, task 6.1 has listed a comprehensive number of data gaps within the four thematics: Biology and Ecosystems, Physics, geology and Atmosphere, Chemistry (with cross discipline and environment) and Human Activities.

As a data gaps analysis summary, it can be said that;
- Biology and Ecosystem is a very important topic with lots of data gaps identified including Genomics data, High frequency observations of phytoplankton and microbes biomass and diversity, Transversal data needs on the deep sea (waters below 200 m), etc.
- Chemistry theme presents also data gaps and in particular for carbonate system, carbon, nitrogen and phosphorus linked to the average marine acidity.
- Data gaps on Physics thematic are mainly about Bathymetry and Acoustics Data. It is recommended that bathymetry surveying is included in any TA proposals that concern organising scientific cruises with research vessels.

Specific data needs have been identified on each Lighthouse;
- Atlantic/Arctic: data needs have been identified for
  o Biology and Ecosystem theme, Biota, abundance, biomass and diversity, Zooplanction, Bacteria, Deep Sea ecosystem.
  o Physic thematic: Acoustics, Current, Sea Ice and Ocean Bottom pressure.
  o For Chemistry thematic: Dissolved gases, Carbonate system and Habitat.

- Black Sea: is the lighthouse with the highest data need, in particular for
  o Biology and Ecosystem: Bacteria and Virus (i.e. microbes), Seagrass Cover and composition, Fish Abundance and distribution, Mammals and Birds abundance distribution.
  o Physics: Acoustic and currents parameters and EOV: Sea Ice and Ocean Bottom pressure.
  o Chemistry: Carbonate System and Pollution.

- Mediterranean Sea: data gaps have been recorded in
  o Biology and Ecosystem: Bacteria and virus (i.e. microbes), on Seagrass Cover and composition, Fish Abundance and distribution, Birds abundance distribution. On important habitats such as posidonia, Coralligenous habitats, Maerl habitats as well as plankton time series.
  o Physics: Acoustics and Ocean pressure. On direct deep current measurements (until 500-600m) since drifters or satellites cover only the surface up to 50-60m.
  o Chemical data: on pH, pCO2, DIC and TA (Total Alkalinity), which are related with the ocean acidification -a major stressor on marine ecosystems.
  o *Scientific studies emphasize the need of more coordinated approaches on microplastics marine litter sampling, management, formats (especially on biota), and on monitoring activities with a stable sampling frequency for plankton time series.*

- Baltic and the North Sea Basins: main gaps have been highlighted in
  o Biology and Ecosystem: Bacteria and virus (i.e. microbes), Biota abundance, biomass and diversity, and Phytoplankton and microphytobenthos.
  o Chemical data: important gaps for habitat, pollution, dissolved gases and nutrient.
  o Human activities with main gaps on transports.
  o *Scientific study highlight that Baltic Sea monitoring is insufficient with respect to the regulatory requirements and that Biodiversity monitoring is the category containing most gaps.*

Those results come from T6.1 methodology described in this deliverable and based on Task 6.1 Experts group, partners of the project. This presents what is the most useful and relevant data to collect for each lighthouse region and will help the projects selection in each TA call.

The data management should then ensure that the collected and processed survey data become available for EMODnet Bathymetry, enriching the basis for generating the EMODnet DTM.

Task 6.1 insists on the fact that data collection itself is not sufficient to solve data needs challenges. The selected TA projects have to include a Data Management Plan for the data to be preserved, retained and made accessible for analysis and application by current and future users.

The guarantee of advanced data management, interoperability, as well as the connection of digital services to the European Open Science Cloud is crucial to better serve research, operational and commercial users. That will be the challenges of AQUARIUS WP6, Task 6.2 and Task 6.3.

## 5. References

Creative Commons, 2017. Licenses. Available at: https://creativecommons.org/licenses

EMODnet Sea-Basin Checkpoints added value, 2023

EuroSea, D3.17 Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans

JERICO-DS D.2.2 – WP2 - "Gap analysis for JERICO-RI technologies"

Copernicus Marine Service requirements for the evolution of the Copernicus In Situ Component: https://marine.copernicus.eu/sites/default/files/inline-files/CMEMS-requirements-In_Situ_03_2021_VF.pdf

Mission Atlantic, Review of available data and gaps for regional case studies, 2021

Ocean FAIR Data Services, 2019, https://www.frontiersin.org/journals/marine-science/

# 6. Annex 1 EMODnet Biology occurrences

Annex 1 presents the EMODnet Biology - EurOBIS data for: 1) locations of hard coral (Scleractinia) occurrences, and the species occurrence data as geospatial grids (30x30 minute) for 3 EOV taxonomic groups which present data gaps in various Aquarius basins (particularly the Artic Sea and the Black Sea): 2) mammals, 3) fishes and 4) birds.

In the geospatial grids, a quantile distribution was used for creating the colour scale, to denote the heterogeneity of records' distribution across the basins. The white and pink areas present few occurrences, the first three and four classes, has less than 10 records per grid cell.