Science of the Total Environment 668 (2019) 531–546

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

An identification and a prioritisation of geographic and temporal data gaps of Mediterranean marine databases

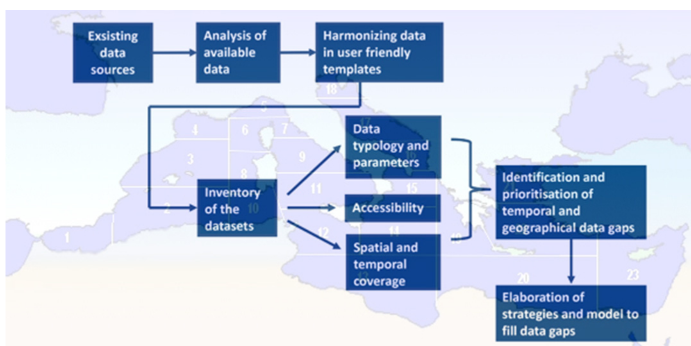
Davide Astiaso Garcia [a,*], Marina Amori [b], Franco Giovanardi [b], Giuseppe Piras [c], Daniele Groppi [c], Fabrizio Cumo [c], Livio de Santoli [a]

[a] Department of Astronautics, Electrical and Energy Engineering (DIAEE), Sapienza University of Rome, Via Eudossiana, 18, 00184 Rome, Italy
[b] ISPRA, Italy
[c] Interdepartmental Centre for Landscape, Building, Conservation, Environment (CITERA), Sapienza University of Rome, Via A. Gramsci, 53, 00197 Rome, Italy

HIGHLIGHTS

- The research aims to find temporal - geographical data gaps of the Mediterranean.
- Platforms, satellite and remote sensing multidisciplinary data have been analysed.
- A method to prioritise temporal and geographic data gaps has been developed.
- An inventory of existing data source has been developed for the whole basin.
- Data gaps priorities are useful for the planning of marine management strategies.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 October 2018
Received in revised form 26 February 2019
Accepted 26 February 2019
Available online 04 March 2019

Editor: Damia Barcelo

Keywords:
Geographical and temporal data gaps
Mediterranean marine database
Data gaps prioritisation

ABSTRACT

Getting an overall view of primary data available from existing Earth Observation Systems and networks databases for the Mediterranean Sea, the main objective of this paper is to identify temporal and geographic data gaps and to elaborate a new method for providing a prioritisation of missing data useful for end-users that have to pinpoint strategies and models to fill these gaps. Existing data sources have been identified from the analysis of the main projects and information systems available. A new method to perform the data gap analysis has been developed and applied to the whole Mediterranean basin as case study area, identifying and prioritise geographical and temporal data gaps considering and integrating the biological, geological, chemical and physical branches of the total environment. The obtained results highlighted both the main geographical data gaps subdividing the whole Mediterranean Sea into 23 sub-basins and the temporal data gaps considering data gathered since 1990. Particular attention has been directed to the suitability of data in terms of completeness, accessibility and aggregation, since data and information are often aggregated and could not be used for research needs. The elaborated inventory of existing data source includes a database of 477 data rows originated from 122 data platforms analysed, able to specify for each dataset the related data typologies and its accessibility. The obtained results indicate that 76% of the data comes from ongoing platforms, while the remaining 25% are related to platforms with non-operational monitoring systems. Since the large amount of analysed records includes data gathered in inhomogeneous ways, the prioritisation values obtained for each identified data gap

* Corresponding author.
E-mail addresses: davide.astiasogarcia@uniroma1.it (D. Astiaso Garcia), marina.amori@isprambiente.it (M. Amori), fabrizio.cumo@uniroma1.it (F. Cumo), livio.desantoli@uniroma1.it (L. de Santoli).

https://doi.org/10.1016/j.scitotenv.2019.02.417

simplify the data comparison and analysis. Lastly, the data gaps inventory contains geographic and temporal information for any missing parameter at the whole basin scale, as well as the spatial resolution of each available data.

1. Introduction

The main anthropic pressures on the marine environment derive directly from the increase of the human population and its consumptions (Bugnot et al., 2019). Together with the widespread urbanization of the shorelines (Mazaris and Germond, 2018), tourism (Huang et al., 2016), agriculture (Xu et al., 2016), fishing (Fu et al., 2018), aquaculture (Lauria et al., 2018), industrial activities (Carlucci et al., 2017) and maritime transport (Rodríguez-Rodríguez et al., 2015) are the main factors that determine the overall impact on the marine environment, causing degradation of natural habitats and producing consequent decrease and loss of biodiversity (Marignani et al., 2017). These effects are mainly due to the following causes: a) nutrient pollution (nitrogen compounds and phosphorus), which causes eutrophication of coastal waters (Garnier et al., 2018), often accompanied by blooms of toxic algae and phenomena of anoxia (absence of dissolved oxygen) (Howarth, 2008); b) pollution from Hazardous and Noxious Substances (HNS) (Astiaso Garcia et al., 2013), such as trace elements (heavy metals) (Moura et al., 2018), and organohalogenated compounds, such as DDT and PCBs (polychlorinated biphenyls mainly used as additives for pesticides), which accumulate in the marine food chain and contaminate tissues of organisms (Jepson and Law, 2016); c) oil and hydrocarbon pollution (Al Shami et al., 2017; Gugliermetti et al., 2007); d) pollution from pathogenic microorganisms (Curren and Leong, 2019); e) impoverishment of species caused by over-fishing (Hinz et al., 2013); f) Disturbance and damage caused to the aquatic fauna by the presence of an increasing number of various vessels, from watercraft to super-tankers (Harik et al., 2017); g) marine litter, and especially the micro-plastics and their degradation products (Hardesty et al., 2017); h) increased levels of underwater noise with both acute and chronic effects on marine life (Tasker, 2016).

Regarding the European regulatory framework, the Water Framework Directive - Directive 2000/60/EC of the European Parliament and the Council (2000) established a framework for Community action in the field of water policy. The main goal of the Directive is to qualitative and quantitative improve the water status and ensuring its sustainable use, based on the long-term protection of available water resources. Furthermore, the Marine Strategy Framework Directive (MSFD) - Directive 2008/56/EC of the European Parliament and the Council (2008) established a framework for community action in the field of marine environmental policy. MSFD acts as an important and innovative tool for sea protection since it is the first binding regulatory framework for EU Member States which considers the marine environment in a systemic perspective. Both legislative documents require systematic environmental monitoring data to assess the status of aquatic and marine ecosystems, following the classification system introduced. Environmental monitoring is the periodic and systematic detection of chemical, physical and biological parameters, through specific tools, in order to assess the current status or to identify the trends of complex systems. Furthermore, the INSPIRE Directive - Directive 2007/2/EC of the European Parliament (2007) established an infrastructure for spatial information in the European Community, providing guidelines for the dissemination of data and metadata and addressing 34 spatial data themes (among them atmospheric conditions, agricultural and aquaculture facilities, environmental monitoring facilities and oceanographic geographical features).

However, data and information reported in the European Directives are generally aggregated and could not be used for research needs that are rapidly expanding by satellite observations and more precise numerical modelling in two-way nested resolutions (Lecacheux et al., 2018; Declerck et al., 2016). Consequentially, the situation in the various European countries is still very complex, as the existing information systems are fragmented.

The main functions of monitoring systems are, in brief: i) the comparison between the expected and the actual environmental effects, considering the monitoring as a tool for qualifying the environmental report (Gharaibeh, 2017); ii) the verification of compliance with the environmental conditions imposed by the current regulatory framework; iii) the verification of the compliance of the programme implementation with the measures envisaged to avoid, reduce or mitigate the negative environmental effects.

Therefore, the need to organize marine data collected in accessible databases has become more and more essential in the latest years, including monitoring operations, in-situ and satellite surveys.

Indeed, many scientific researches are based on the gathering and analysis of measurement data. Data set are becoming more and more important and can be considered as a primary intellectual output of a given research useful for the implementation of future researches and subsequent studies (Dekker, 2006).

On the international scene, it appears of great importance that the marine sector is equipped with an integrated environmental platform, operational and usable, that obeys the international standards required by the various projects. Use of common vocabularies in all meta-databases and data formats is an important prerequisite towards consistency and interoperability. Common vocabularies consist of lists of standardised categories' terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community. Using standardised terms of reference solves the problem of ambiguities associated with data mark-up and enables records to be easily interpreted by other researchers and end users.

In this context, this research has been carried out within the H2020 research project ODYSSEA (*Operating a network of integrated observatory systems in the Mediterranean Sea*) whose main goal is to develop, operate and demonstrate an interoperable and cost-effective platform that fully integrates networks of observing and forecasting systems across the Mediterranean basin, addressing both the open sea and the coastal zone. In order to protect, conserve and manage the marine environment, ODYSSEA generally aims to collect, integrate and make interoperable the numerous existing systems and platforms managing marine environmental data for the Mediterranean Sea.

Bearing in mind the above mentioned ODYSSEA goal, the particular aims of this research, that represent the first essential pillar for achieving the overall project objectives, are: i) to analyse the reliability and usefulness of the existing monitoring systems at the Mediterranean basin level, organizing them into a structured inventory and identifying the critical issues and shortcomings considering and integrating biological, geological, chemical and physical properties of the Mediterranean Sea; ii) to elaborate and apply a new method for identifying geographical and temporal marine data gaps and for providing a prioritisation of missing data.

Therefore, this paper illustrates the methodological approach used both for the preparation of the data inventory, reviewing all the existing databases, and for the elaboration of a new procedure for prioritise data gaps, based on precise and adequate statistical method.

Summarizing the state of the art of this research, several studies dealt with marine monitoring, describing innovative techniques and technologies (Chapman, 2015), sampling methods, and analysing the data gathered for monitoring different themes and parameter, such as water quality (Karydis and Kitsiou, 2013), marine pollutants (Cumo et al., 2008), fish community health (Nicholson and Jennings, 2004), and many other biodiversity indicators (Birk et al., 2012). Elliott (2013) argued that that marine monitoring should foresee modelling and management systems in order to optimize its effectiveness.

Recently, Caro et al. (2018) conducted a review of existing spatial databases that can be used to map coastal and marine ecosystem services in Europe starting from the open source databases available in institutional websites and project websites. They classified the analysed databases into five categories (biodiversity, environment, administrative, social and economic) providing useful information about selected variables together with statistical summaries of their contents. However, they did not analyse geographical coverage neither pinpointed the main data gaps of these databases.

Anyway, papers analysing marine database pinpointing data gaps are fewer in numbers and are often focused on particular themes, as the ones above mentioned. Moreover, already existing gap analysis, such as ecological one firstly defined by Scott et al. (1993), are mainly aimed to compare current situation of available data with the best potential condition pinpointing where or when knowledge is lacking (Brown and Dodd, 1998).

Turrell (2018) analysed marine two monitoring themes in the northeast Atlantic improving monitoring implementation by identifying gaps and overlaps.

Dimarchopoulou et al. (2017) assessed the level of knowledge of numerous biological characteristics of the Mediterranean marine fishes by carrying out a gap analysis based on existing datasets. Indeed, Mediterranean is currently one of the seas where main biological information is missing (Stergiou and Tsikliras, 2006).

Within the H2020 ONION (Operational Network of Individual Observation Nodes) project, Lancheros et al. (2018) estimated the spatial resolution, revisit time, precision and temporal continuity data gaps of the Copernicus system for polar regions, underlining the importance of this kind of researches for satisfying end user needs in any area of the planet.

In this existing framework, the novelty factors of this research are to analyse and integrate all the main available data on the Mediterranean Basin, considering almost all the data typologies, and consequentially to develop a new and user-friendly methodology to identify and to prioritise the main geographical and temporal data gaps. Indeed, a similar research that analyses the whole Mediterranean Basin as case study area, including also not European coastal and marine waters and considering both main platforms, satellite data and remote sensing analysis, has never been carried out. Consequentially the research does not include particular case studies at smaller scale for filling temporal and spatial gaps since it inevitably involves new site measurements, remote sensing or numerical model outputs that are not the focus of this paper.

Indeed, the main research aim is to elaborate a new methodology for assessing and prioritising geographical and temporal data gaps for the whole Mediterranean basin starting from an analysis of the existing data sources.

Obviously, the obtained results could not be considered as an exhaustive overview of the whole available Mediterranean data, since the research analyses existing databases accessible to stakeholders, researches and to the scientific community in general and it cannot take into account the confidential and not accessible databases of the national ministries or other institutions of the Mediterranean basin countries. Anyway, research outcomes provide for the first time a useful and interesting framework of the priorities that end users, decision makers, local and national authorities should take into account for improving their marine monitoring systems for prioritise conservation actions in the Mediterranean basin. Furthermore, the elaborated methodology for prioritising data gaps should be replicate to carry out similar researched in different basins, oceans or smaller seas. Indeed, as stated by Carstensen (2014), the need to understand the human pressures is increasing and requires integrated and multidisciplinary environmental mapping and monitoring (Nilssen et al., 2015).

2. Methods

The research methods include a data gathering methodology for developing an inventory of existing data sources and a new approach for data gaps identification and prioritisation.

The starting point was the organization of the data sources on a Mediterranean scale, by an analysis of multidisciplinary existing input datasets. The general framework has been provided by the "Methodology to assess and communicate the economic benefits of consensus-based standards", developed by the International Organization for Standardization (ISO) (2010). Indeed, the definition of key vocabulary terminology was essential, and it was based on ISO standard definitions as detailed here below.

- Data Typology: Refers to the prevailing attributes of similar nature related to the marine environment and provided by measurements and observations.
- Data Parameters: i) a variable derived from an observation network or from in-situ measurements; ii) a numerical model output simulating a particular environmental process; iii) a geographical representation of an object on a map (i.e. a layer such as a protected area, a coastline or fishing fleet distribution) by a set of vectors (polygon, curve, point); iv) a raster (a spatial data model that defines space as an array of equally sized cells, such as a grid or an image).
- Data: Re-interpretable representation of information in a formalized manner suitable for communication, interpretation or processing (ISO, 2014).
- Dataset: An identifiable collection of data (ISO, 2014). It can be a time series, a lithological description of a marine sample, a gridded dataset such as a DTM, a hydrodynamic model output, a GIS dataset or a feature layer of a GIS dataset, a data base or a table of values in a publication. A data set can be constituted of several files (e.g., the set of seismic data files recorded along the same line).
- Input Dataset: The collection of existing data to be imported in the data gaps analysis. Assessment criteria: The criteria are focused on two questions: "what" and "how" data is made available from the information platform.
- Data adequacy: Adequacy indicates "data sufficient to satisfy a requirement or meet a need". Using this definition, "adequacy" relates to meeting both requirements, as well as needs, and it is normally applied within the framework of an ISO 9001-based Quality Management System.

2.1. Data gathering methodology for the inventory of existing data source

The Mediterranean Sea is the largest semi-enclosed sea on Earth which lies between Europe, Asia and Africa. It covers about 2.5 million km^2 or 0.82% of Earth's surface and is connected to the Atlantic Ocean, surrounded by the Mediterranean Basin and almost completely enclosed by land.

The Mediterranean Sea is, in oceanographic terms, a mostly enclosed system that has limited exchange of water with outer oceans and has water circulation caused mainly by salinity and temperature differences, attributed to the limited freshwater fluxes and enhanced evaporation, rather than by winds (Llasses et al., 2018).

Considering heat and salinity balances, one of the most typical property of this basin is that it is a concentration sea since the surplus of evaporation over precipitation and river runoff is balanced by a net incoming water flux from the Atlantic Ocean (Criado-Aldeanueva et al., 2014). Moreover, the Mediterranean Sea has a complex orography and almost all its sub-basins are evaporative, mainly to the air–sea interactions that regulate the ocean variability (Cusinato et al., 2018).

As regard its environmental value, the Mediterranean basin is characterized by a significant variety of ecosystems, including oligotrophic and fully eutrophic areas, and consequentially it includes a high biodiversity rate, corresponding to the 8% of the known marine species (Jordà et al., 2017).

The main regional subdivision of the Mediterranean Sea divides it into: the Levant Sea, the Aegean Sea, the Adriatic Sea, the Ligurian Sea, the Tyrrhenian Sea, the Ionian Sea and the western basin with the Balearic and Alboran Seas. A more detailed partition is shown in Fig. 1.

Within this research, ODYSSEA partners identified existing data sources from their own country, analysed Earth Observation facilities and networks (platforms with online models, satellite data, in-situ systems, citizen scientist networks) containing Mediterranean status databases maintained by agencies, public authorities, research institutions and universities.

Subsequently, the authors implemented the collected information with data provided in research journals (both open access or not) and in international databases, analysing the researches mentioned in the introduction paragraph and in other specific studies such as Bonsignore et al., 2018, Marbà et al., 2014, Zaccaroni et al., 2018.

It was therefore necessary to design a structured format in which each row (record) represents a data source and each column represents a data attribute. The records in the table contain a fixed number of fields and all records have the same format. This method has considerable advantages:

- It is easy to use;
- It significantly narrows the margin of error in the insertion and mapping of data;
- It allows data gap analysis to be easily performed;
- It allows the rigorous design, which will facilitate all the subsequent work of the platform.

The flowchart in Fig. 2 describes the methodological process necessary to build up the inventory and achieve the prefixed objectives.

The information and data gathered under these procedures were managed and dealt using the MS Excel software, to provide second-level information for subsequent uses of the data gaps analysis.

Common vocabularies were set-up and implemented by SeaDataNet Project (European Commission, 2007). SeaDataNet is a distributed Marine Data Infrastructure for the management of large and diverse sets of data which are derived from in situ surveys of the seas and oceans. Professional data centres, active in data collection, constitute a Pan-European network providing on-line integrated databases of standardised quality. The on-line access to in-situ data, meta-data and products is provided through a unique portal interconnecting the interoperable node platforms composed of the SeaDataNet data centres. The SeaDataNet metadata services provide overviews of marine research organisations in Europe and their engagement in marine research projects, managing large datasets, and data acquisition by research vessels and monitoring programmes for the European seas and global oceans. For communication and cataloguing purposes, the SeaDataNet Common Vocabulary, that identifies monitoring groups and categories of characteristics by a code, giving a definition for each code at different levels of aggregation, has been used.

The vocabulary terms utilized are:

- P02 SeaDataNet Parameter Discovery Vocabulary - Terms describing fine-grained related groups of measurement phenomena designed to be used in dataset discovery interfaces;

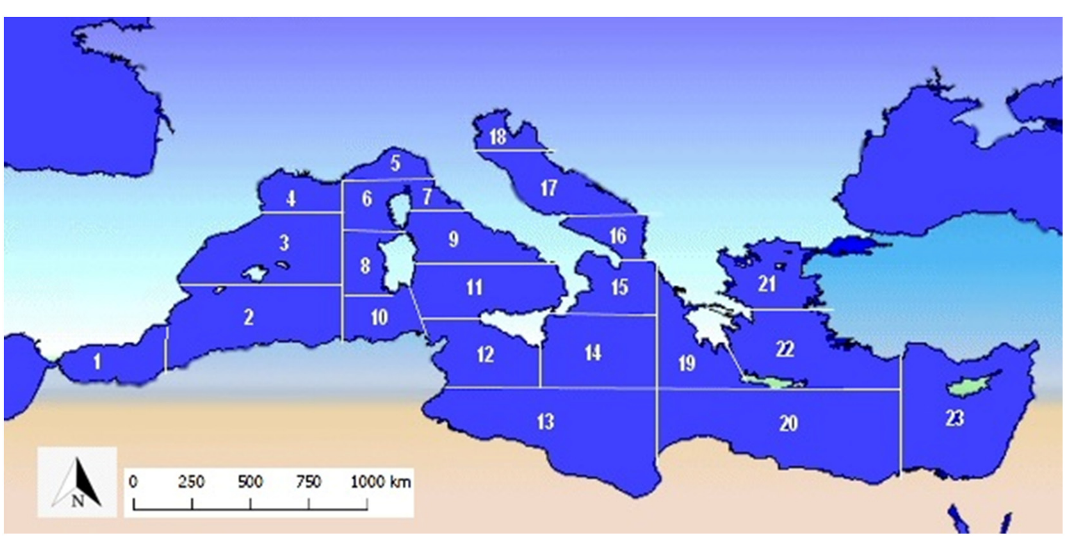


Fig. 1. Mediterranean Sea's regional subdivision in seas and smaller basins 1) Alboran sea; 2) Southern Balearic Sea; 3) Northern Balearic Sea; 4) Gulf of Lion; 5) Ligurian sea; 6) Corsica sea; 7) Northern Tyrrhenian Sea; 8) Sardinia sea; 9) Central Tyrrhenian Sea; 10) Channel of Sardinia; 11) Southern Tyrrhenian Sea; 12) Strait of Sicily; 13) Libyan Sea; 14) Southern Ionian Sea; 15) Northern Ionian Sea; 16) Northern Adriatic Sea; 17) Central Adriatic Sea; 18) Southern Adriatic Sea; 19) Eastern Ionian sea; 20) Sea of Crete; 21) Northern Aegean; 22) Southern Aegean; 23) Levant sea.

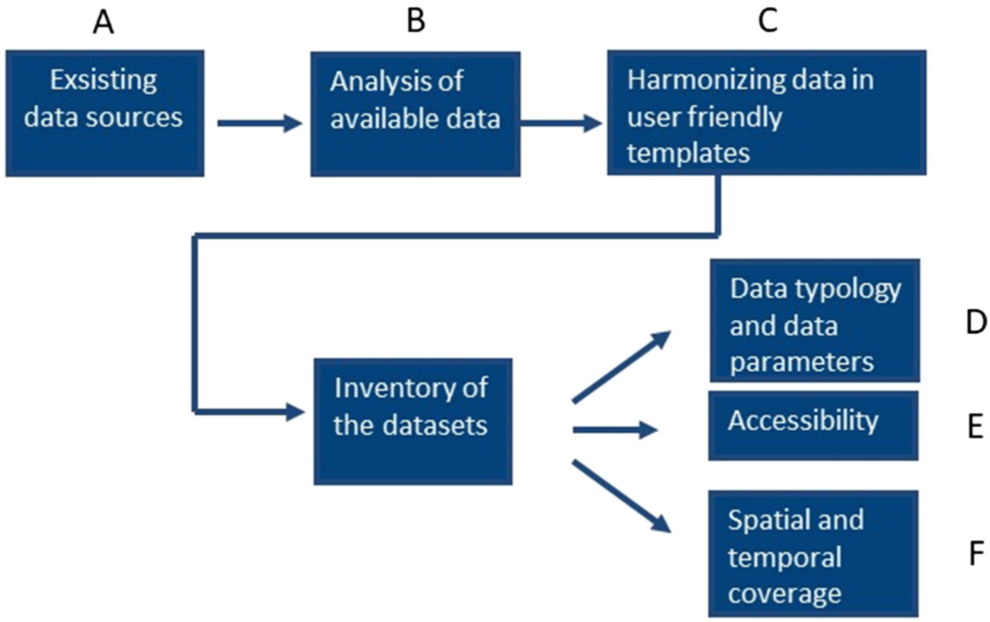


Fig. 2. Methodological processes for elaborating the datasets inventory (A, B and C) and analysed data features (D, E, F).

- P03 SeaDataNet Agreed Parameter Groups - Terms agreed within the EU SeaDataNet community to describe coarse-grained groupings of related measurement phenomena;
- P04 Global Change Master Directory Science Keywords V5 - Concatenated category, topic, term and variable separated by '>'. From Olsen et al. (2006) NASA/GCMD Earth Science Keywords Version 5.3.3.

Chemical, biological, ecological, physical, meteorological and human pressure datasets have been collected considering both the Mediterranean open sea and the coastal zone. All the collected data were included in an explicit inventory of existing data sources specifying for each dataset its data typology and accessibility (open or non-accessible). The main typologies were relative to Earth Observation facilities and networks (platforms with online models, satellite data, in-situ systems, citizen scientist networks) containing Mediterranean status databases maintained by agencies, public authorities, research institutions and universities.

Overall, the meta-database (*Master_Data*) includes 122 platforms related to 26 groups (*Data Typology*) (SDN: P03 GROUP) and 101 categories of characteristics (*Data Parameter*) (SDN:P02 GROUP and SDN: P04GROUP), i.e. monitoring environmental and human activity information. These descriptors identify potentially usable information for data gap analysis implementation. The elaborated excel template for data collection is divided into two sections: Section I describes the Standard Template structure; Section II includes the Data Dictionary related to the fields to be filled in the Standard Template. The Data Dictionary contains the organized list of all data concerning the system and defines them by means of data description of preliminary information entities.

Specific default lists of fields containing the options to be entered by an automatic selection using a dropdown menu have been elaborated to facilitate data entry and to significantly reduce the possibility of errors when compiling these data.

The default lists are shown here below for each considered field.

- Data Provider: citizen scientist networks; collecting data from other providers; integrate and disseminate knowledge and expertise; platform.
- Platform Typology: online models; in situ system; geographical; monitoring systems and cruises; remote sensing data; numerical models; Corine land cover; GIS (Geographic Information System). The term "Geographical Platform" refers to a platform with an application programming interface (API) provided by Google Maps, that allows the overlay of geo-referenced data with a personalized Google map. It should not be confused with a GIS (Geographic Information System), although the acronym GIS is often used to mean science or studies on geographical information. The differences between "Online models" and "Numerical models" are not substantial; both platforms provide various types of data which is produced by simulation models. The separation between the two types of platforms was maintained, to respect the original information provided by the respective websites and, in the case of the "numerical models" platforms, to highlight the numerical nature of the outputs that usually refer to projects and services granted by the EC, aimed above all at ocean monitoring and forecasting such as myOcean projects in the frame of the CMEMS Copernicus Programmes (Masina et al., 2017; Buongiorno Nardellia et al., 2013).
- Platform Access: open; non-accessible.
- Platform Start Year/Platform Finish Year: from before 1990 to 2017 or ongoing.
- Geographical Coverage: the formal list of the 23 regional sub-basins was used, correcting the different names used in the various inventoried platforms: Whole basin; Northern Alboran Sea; Alboran Island; Southern Alboran Sea; Algeria; Balearic Island; Northern Spain; Gulf of Lion; Corsica; Ligurian Sea and North Tyrrhenian Sea; Southern and Central Tyrrhenian Sea; Western Sardinia; Eastern Sardinia; Northern Tunisia; Gulf of Hammamet; Gulf of Gabes; Malta; Southern Sicily; Northern Adriatic; Southern Adriatic Sea; Western Ionian Sea; Eastern Ionian Sea; Southern Ionian Sea; Aegean Sea; Crete; Northern Levant Sea; Cyprus; Southern Levant Sea; Eastern Levant Sea; Marmara Sea; Black Sea; Azov Sea; Coastal water EU member states; Italy; Southern European Seas; Greece.
- Spatial Resolution: 1 arcsec; 3 arcsec; 7.5 arcsec; 30 arcsec; 1 arcminute; 0.004166°; 0.0027°; 0.003°; 0.06°; 0.1°; 0.12°; 0.125°; 0.2°; 0.25°; 0.4°; 0.5°; 1°; 1 km; 4 km; 25 km; 12 m × 12 m; 500 k, 1 M, 1 .5M based on region; 1:1000000 scale (1 M); 0.06°, 1.0 km, 4.0 km; 7 km, 14 km; 1 km, 4 km; 4 km, 9 km; 0.01°, 0.02°; 0.01°, 0.04°; 5 m; 2.5 km; 7.5 km; 30 min spatial cells.
- Temporal Resolution: real time based on station; hourly, daily, monthly mean; daily, monthly mean; monthly mean; 9-monthly mean; 3-hourly mean; 6-hourly mean; 12-hourly mean; 48-hourly

mean; monthly mean or daily instantaneous based on date and dataset (satellite); weekly, monthly mean; annual mean.
- Data Format: vector point; vector polygon; vector shapefile; grib; netcdf; asci; pdf.

Table 1 shows the different SDN codes for Data Typology. For each Data Typology a list of ad-hoc parameters has been considered in the appropriate drop-down menu. The considered parameters for each data typology are reported in Annex 1 and have been selected starting from the SeaDataNet vocabularies, with particular focus on the following vocabularies: SDN: P02 GROUP, SDN: P03 GROUP, SDN: P04 GROUP.

2.2. A new approach for data gaps identification and prioritisation

An overall Mediterranean data gaps inventory has been developed starting from the comprehensive information available, the *Master_Data* sheet, in order to identify and assess data gaps and lack of information in existing platforms and systems of data collection. It was also recommended to utilize experience from other relevant EU projects (e.g. PERSEUS project). In particular, the authors made reference to the experience already acquired from other relevant EU projects and the procedures already applied and developed for data gap analysis in the MSFD assessment elements of the Southern European Sea countries, as described by Laroche et al. (2013). In the development of this work, in addition to space/time possible data gaps, another element of interest has been considered: the "data suitability", in terms of relevance, importance and representativeness of the available data contained in the *Master_Data* sheet.

An elaborated *Whole_Master_Data* sheet represents all of the individual templates referred to specific *Data Typology* merged into a unique data matrix (e.g., the baseline dataset).

The various results obtained by data sorting enabled the graphical representation of the entire set of information contained in the *Master_Data*. The quality of this information, in terms of frequency of occurrence of the *Data Typology* and *Data Parameters*, referred alternatively to *Platform_Name* and to *Geographical_Coverage*. Data quality has also been evaluated, by means of appropriate explanatory diagrams.

Data gap analysis starts from a *Preliminary_Matrix*, which assigns to each *Data Typology* (rows), the number of *Data Parameters* related to each regional Sea (columns) in terms of occurrences, taking into account also all-inclusive categories, as the *whole basin* and *coastal water EU member states* that collect data without distinguishing between individual Mediterranean sub-basins. Each cell of the preliminary matrix was then marked with a score obtained through the normalization of occurrences with respect to the sum of occurrences per row. Accordingly, authors assigned to this sum the meaning of total available amount of data per *Data Typology*, as it is provided by the *Whole_Master_Data* sheet.

The *Parameters_Gap_Matrix* was evaluated for each score provided by the preliminary matrix. *Parameters_Gap_Matrix* represents the final output of the data gap analysis. Following this approach, a new procedure has been developed to assign an objective value to quantify data availability, evaluating the maximum achievable value for a given *Data Typology* as the sum of occurrences of *Data Parameters* per each Sea registered in the master data sheet, and attributing the data gap as the difference needed to reach the maximum value (100%) of information attainable. Consequentially, the elaborated matrix assigns to each Data Typology (DT), the number of Data Parameters (DP) related to each regional Sea (columns) in terms of occurrences, taking into account also all-inclusive categories, as the whole basin and coastal waters of EU member states that collect data without distinguishing between individual Mediterranean sub-basins. Geographical data gaps values have been therefore assessed according to Eq. (1):

$$DT_{a,1} = 1-\left(\frac{DP_{a,1}}{\sum_{i=1}^{i=16} DP_{a,i}}\right) \quad (1)$$

where $DT_{a,1}$ is the geographical data gaps value for Data Typology *a* in Mediterranean sub basin *1*, $DP_{a,1}$ is the number of data parameters include in Data Typology *a* available in the sub basin *1*, and $DP_{a,i}$ is the number of data parameters include in Data Typology *a* available in the *i-th* sub basin.

For each *Data Typology* (rows), different data gap values (within 0 and 1) were then obtained and classified on the following data gap classes: "maximum" for score gaps: 1; "high" for score gaps between 0,8 and 1; "medium" for score gaps between 0,5 and 0,8; "minimum" for score gaps lower than 0,5.

3. Inventory of existing data source

All the obtained results of the mapping of existing platforms models and tools have been reported in a database representing the inventory of existing data sources (Annex 2).

This inventory contains 477 data rows originating from 122 data platforms analysed. In terms of data accessibility, 34 data sets are not accessible while 403 are open. However, it was not possible to check the remaining 40 datasets originating from remote sensing platform typology.

Analysing the elaborated inventory reported in Annex 2, here below a summary of the gathered biological, geological, chemical and physical

Table 1
Data Typology from SDN: P03 GROUP.

Data Typology	Data Typology code	Data Typology	Data Typology code
Meteorology	M010	Rock and sediment sedimentology	GSED
Biota abundance biomass and diversity	B070	Terrestrial including bathymetry and under-sea features	T001
Birds mammals and reptiles	B015	Positioning references and data management	Z005
Fish	B020	Habitat	B050
Anthropogenic contamination	H001	Macroalgae and seagrass	B055
Construction and structures	H002	Pigments	B035
Fisheries	H004	Dissolved gases	C015
Human activity	H005	Carbon nitrogen and phosphorus	C005
Currents	D030	Sedimentation and erosion processes	G060
Sea level	D032	Rock and sediment physical properties	G040
Water column temperature and salinity	D025	Optical properties	D015
Waves	D034	Suspended particulate material	G015
Rock and sediment lithology and mineralogy	G045	Earth science oceans marine volcanism	VOLC

properties of the Mediterranean Sea, integrating multidisciplinary data available for this basin.

Biota and biomass characteristic available in the Mediterranean include the following data: fauna abundance per unit area of the bed, bacteria taxonomic abundance in water bodies and in sediment, biodiversity indices, bacteria generic abundance in water bodies and in sediment, phytoplankton taxonomic biomass in water bodies, biological detritus in the water column suspended particulate material, zooplankton wet weight biomass, microzooplankton taxonomic abundance in water bodies, zoobenthos generic abundance, zoobenthos taxonomy-related abundance per unit area of the littoral zone and non taxonomy-related wet weight biomass per unit area of the bed, zooplankton dry weight biomass per unit volume of the water column, microzooplankton abundance in water bodies, phytoplankton biomass in the water bodies, phytoplankton taxonomic abundance in water bodies, plankton abundance per unit volume of the water column, shellfish abundance and biomass in water bodies, shellfish morphology, age and physiology, virus abundance in water bodies, water quality bioindicators, zooplankton taxonomy-related abundance per unit volume of the water column, zoobenthos taxonomic abundance and taxonomy-related counts, macroalgae generic abundance in water bodies, macroalgae and seagrass taxonomy-related counts, chlorophyll pigment and phaeopigment concentrations in the water column.

Fish data includes abundance in water bodies, reproduction data, fish and shellfish catch statistics, morphology, age and physiology data, fish taxonomy-related abundance and ash-free dry weight biomass per unit area of the bed, fish taxonomy-related counts and fish biomass in water bodies.

Other vertebrates' properties include behaviour, reproduction and abundance data of birds, cetacean and reptiles.

Chemical and physical properties gathered and available for the Mediterranean water column are: salinity, temperature, electrical conductivity, turbulence, density, structure and stability, alkalinity, acidity and pH, concentration of dissolved organic matter, primary production and dissolved metal concentrations. Furthermore, regarding carbon, nitrogen and phosphorus components, the following data have been gathered: nutrient fluxes between the bed and the water column, particulate total and organic nitrogen concentration in the water column, phosphate, nitrate and nitrite concentration parameters in the water column, dissolved total or organic phosphorus concentration in the water column, ammonium concentration parameters in the water column, carbon concentrations in sediment, particulate total and organic carbon concentrations in the water column, dissolved organic carbon concentration in the water column, nitrogen concentrations in sediment and in suspended particulate material, phosphorus concentrations in suspended particulate material, silicate concentration parameters in the water column, total dissolved inorganic carbon (TCO2) concentration in the water column, dissolved inorganic nitrogen concentration in the water column, dissolved total and organic nitrogen concentrations in the water column, total particulate and organic phosphorus concentrations in the water column, dissolved oxygen parameters in the water column. Regarding the optical properties data, transmittance and attenuance of the water column, ocean colour and earth-leaving visible waveband spectral radiation, light extinction and diffusion coefficients, optical backscatter, visible waveband radiance and irradiance measurements in the water column have been considered.

Meteorology properties gathered include air pressure, air temperature and density, atmospheric humidity, as well as wind speed and direction data.

As regard waves and currents data, spectral wave data parameters, wave direction, height and period statistics, horizontal velocity of the water column, river flow and discharge, transport in the water column, vertical velocity of the water column as well as wind stress and shear have been considered.

Considering geological properties, available data are: lithology, mineralogical composition, sedimentary structure, depositional environment, sediment accumulation rate, rock grain size, redox potential in sediment, concentration of organic matter in sediments, coastal geomorphology, sediment grain size parameters.

Lastly, as regard anthropogenic contamination, the following parameters have been considered: litter abundance and type, pollution events, concentration of polycyclic aromatic hydrocarbons (PAHs) in biota, concentration of metal and organic contaminants in biota, concentration of polychlorobiphenyls (PCBs) in biota, bioassay and contaminant biological impact, industrial discharges, acoustic noise in the water column, concentration of polycyclic aromatic hydrocarbons (PAHs) in the water column, pesticide concentrations in biota, concentration of polychlorobiphenyls (PCBs) in the water column, pesticide concentrations in sediment and in water bodies, concentration of polycyclic aromatic hydrocarbons (PAHs) and other organic contaminants in sediment samples, concentration of other organic contaminants in sediment samples and in the water column, radioactivity in the water column.

The obtained results indicate that 76% of the data comes from ongoing platforms, while the remaining 25% are related to platforms with non-operational monitoring systems in 2016 or even prior.

Fig. 3 summarizes the variability in platform typology, pinpointing that almost half of the data originates from monitoring systems and cruises or in-situ systems.

Spatial and temporal resolutions are not specified in most of the datasets of the Mediterranean Sea. More specifically, spatial resolution is reported only at the 16% of the datasets and temporal resolution at the 38% of the total datasets. On the other hand, data format information is reported in the 78% of the data, with details summarized in Fig. 4.

A preliminary assessment in terms of data geographical coverage indicated that 43% of the total data are available for the whole Mediterranean Basin while the remaining 57% is available only for some specific sub-seas. An analysis of the datasets collected for each data typology is given in Annex 3.

4. Data gaps analysis and prioritisation results

The following elaborations starting from the contents of the Master_Data sheet collecting data and information provided by the existing platforms on the Mediterranean, as they were registered in the preliminary data survey phase (Annex 1). The objective is to provide a complete framework on the state of knowledge at the basin level, without entering into details on the individual platforms (for which reference is made directly to the master dataset). The aim of summarizing and highlighting the various types of data available is to identify all possible shortcomings (gaps) relative to the degree of spatial and temporal coverage.

The census made it possible to record data relating to 122 surveyed platforms, with 475 records of data corresponding to a relevant number of data typologies, which are listed in Table 1, in which the respective SeaDataNet code is also reported.

As for the individual parameters, they are listed in Annex 4, where, in addition to the Parameters Code, the occurrence of each parameter in the master dataset is also provided. Annex 4 and the subsequent graphical representation of the results (Fig. 5) show that the most frequent and represented parameters refer to the hydrographic data (temperature and salinity along the water column and sea level) and meteorological (wind speed and direction), with high percentages of occurrence, >8%. Information on ecology of marine habitats is also well represented with frequencies around 5%. It is worth noting that an “emerging pollutant” such as marine litter, occupies the first places in the ranking (>4%). The last places in the ranking (<0.5% of the cases), refer to parameters mainly related to the taxonomic aspects, to the presence of contaminants in the biota and sediments, and in general, to all

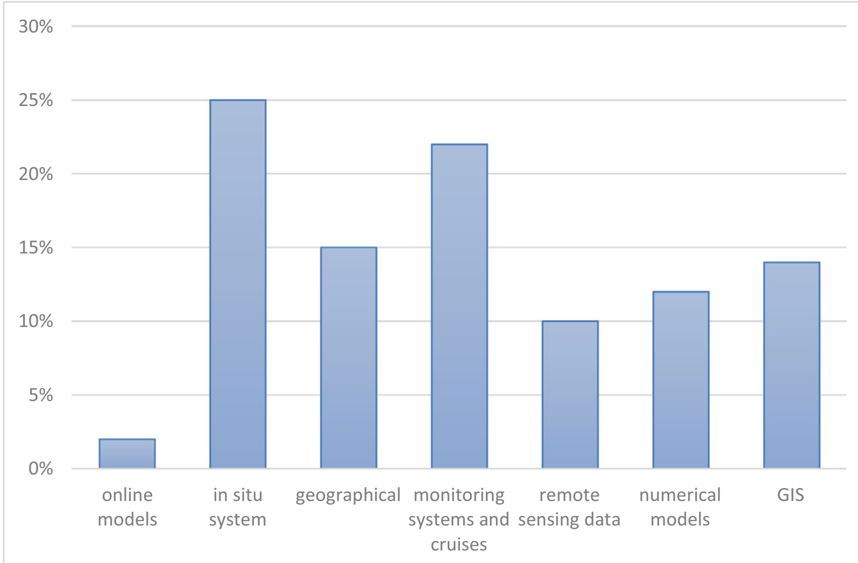


Fig. 3. Variability in platform typologies.

those measures that require complex instruments and procedures for their determination.

In Fig. 6, the Data Typologies are characterized based on the number of platforms providing Data Parameters referred to these typologies.

Habitats and "biota abundance biomass and diversity" results as the most frequent Data Typologies reported (29 and 26 platforms, respectively), with a number of Parameters data referring to these two types of 2 out of 2 and 12 out of 25 for habitat and biota abundance, respectively. Currents, Meteorolgy, Human Activity, Bathymetry and Sea level are also well represented (No. of Platforms >10). The last places in the ranking refer to those types related to the geology (mineralogical and physical nature of rocks and sediments) and to the earth's science (volcanism and thermal vents).

A separate discussion deserves the type D025 (water column temperature and salinity). The limited number of platforms providing data of this type should not mislead. These platforms are characterized by the maximum geographical coverage (with more sub-basins or to the whole basin extension).

In fact, Fig. 7 reports on the number of times (% of occurrence) in which individual seas appear in the Whole_master_Data. The disparity between the "Whole basin" and "Coastal water EU member states" categories becomes immediately evident, if compared to the occurrence of individual seas. On the other hand, this difference arises from the decision taken from the beginning, to consider the geographical coverage referred to several sub-basins or extended to the entire Mediterranean basin as a unique category of its own.

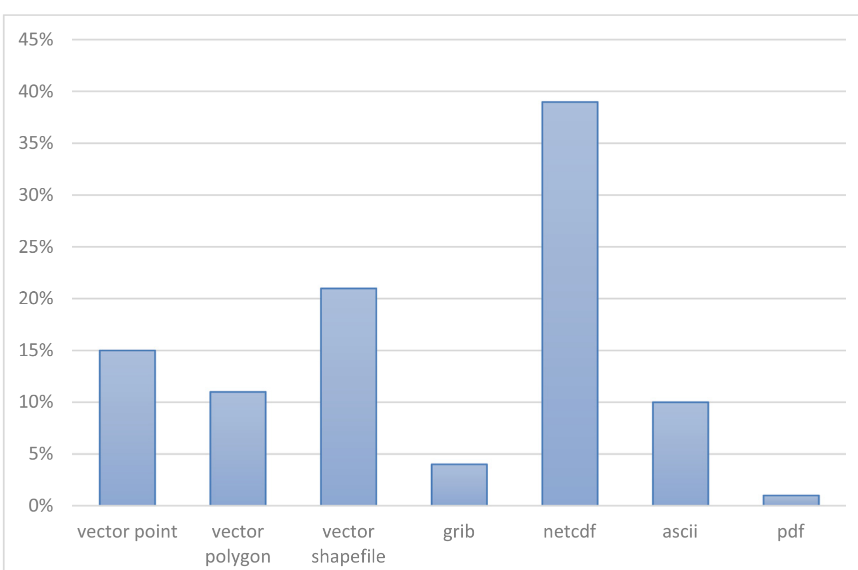


Fig. 4. Variability in data format typology.

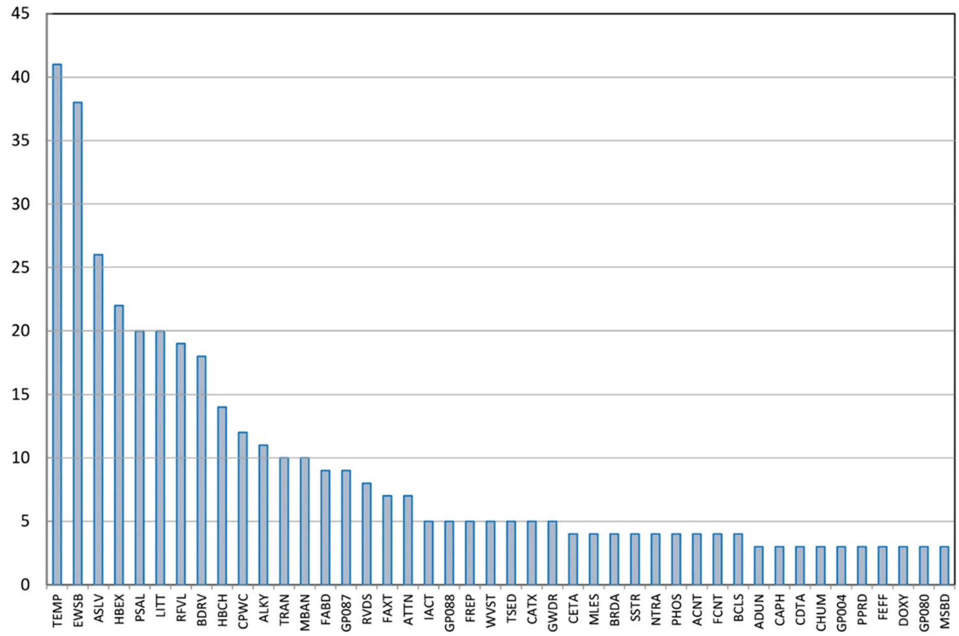


Fig. 5. Parameters list- occurrence in the database (parameters with No. of times <3 are not represented). Parameter codes are the following: TEMP = temperature of the water column; EWSB = wind speed and direction; HBEX = habitat extent; PSAL = salinity of the water column; LITT = litter abundance and type; RFVL = horizontal velocity of the water column (currents); BDRV = biodiversity indices; HBCH = habitat characterisation; CPWC = chlorophyll pigment concentration in the water column; ALKY = alkalinity, acidity and pH of the water column; TRAN = transport activity; MBAN = bathymetry, elevation and undersea features; FABD = fauna abundance per unit area of the bed; GP087 = fishery characterisation; RVDS = river flow and discharge; FAXT = fish abundance in water bodies; ATTN = transmittance and attenuance of the water column; IACT = industrial activity; GP088 = bird behaviour; FREP = fish reproduction; WVST = wave height and period statistic; TSED = concentration of suspended particulate material in the water column; CATX = phytoplankton taxonomic biomass in water bodies; GWDR = wave direction; CETA = cetacean abundance; MLES = marine environment leisure usage; BRDA = bird taxonomy-related counts; SSTR = sedimentary structure; NTRA = nitrate concentration parameters in the water column; PHOS = phosphate concentration parameters in the water column; ACNT = macroalgae and seagrass taxonomy-related counts; FCNT = fish taxonomy-related counts; ADUN = administrative units; CAPH = air pressure; CDTA = air temperature and density; CHUM = atmospheric humidity; GP004 = bird reproduction; PPRD = primary production in the water column; FEFF = fishing effort; DOXY = dissolved oxygen parameters in the water column; GP080 = fishing by-catch; MSBD = zooplankton dry weight biomass per unit volume of the water column.

For coastal water of EU member states, the only inventories used are: 1) The European environment information and observation network (Eionet; link: www.eionet.europa.eu); 2) The EMODnet Data Ingestion portal (link: http://www.emodnet.eu/). The temporal coverage starts from year 2000 and refers to monitoring programmes mostly still in progress. In general, the access

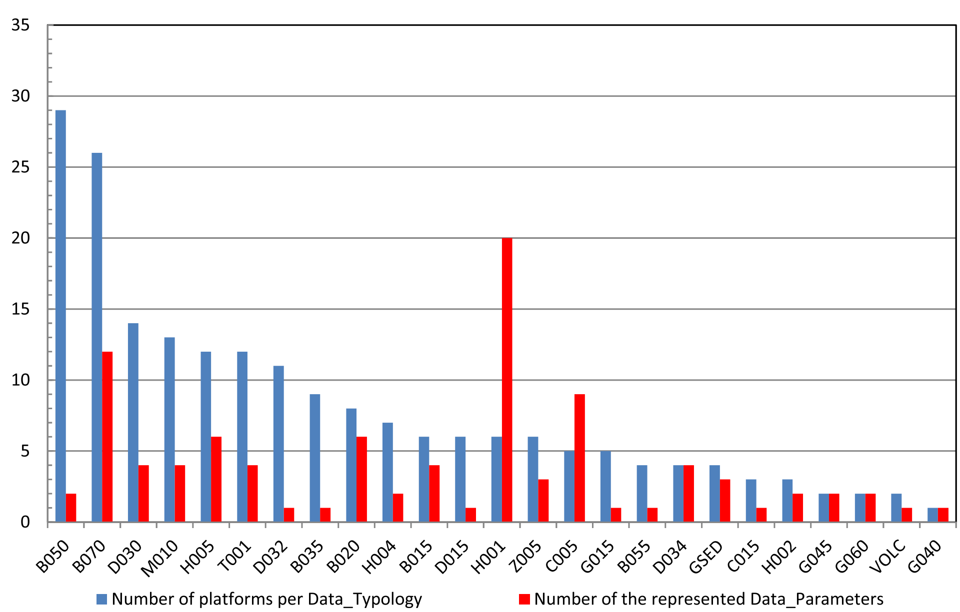


Fig. 6. Number of platforms per Data Typology and corresponding number of the Data Parameter reported for each typology. Data Typology list: B050 = habitat; B070 = biota abundance biomass and diversity; D030 = currents; M010 = meteorology; H005 = human activity; T001 = terrestrial including bathymetry and under-sea features; D032 = sea-level; B035 = pigments; B020 = fish; H004 = fisheries; B015 = birds mammals and reptiles; D015 = optical properties; H001 = anthropogenic contamination; Z005 = positioning references and data management; C005 = carbon nitrogen and phosphorus; G015 = suspended particulate material; B055 = macroalgae and seagrass; D034 = waves; GSED = rock and sediment sedimentology; C015 = dissolved gases; H002 = construction and structures; G045 = rock and sediment lithology and mineralogy; G060 = sedimentation and erosion processes; VOLC = earth science oceans marine volcanism; G040 = rock and sediment physical properties.

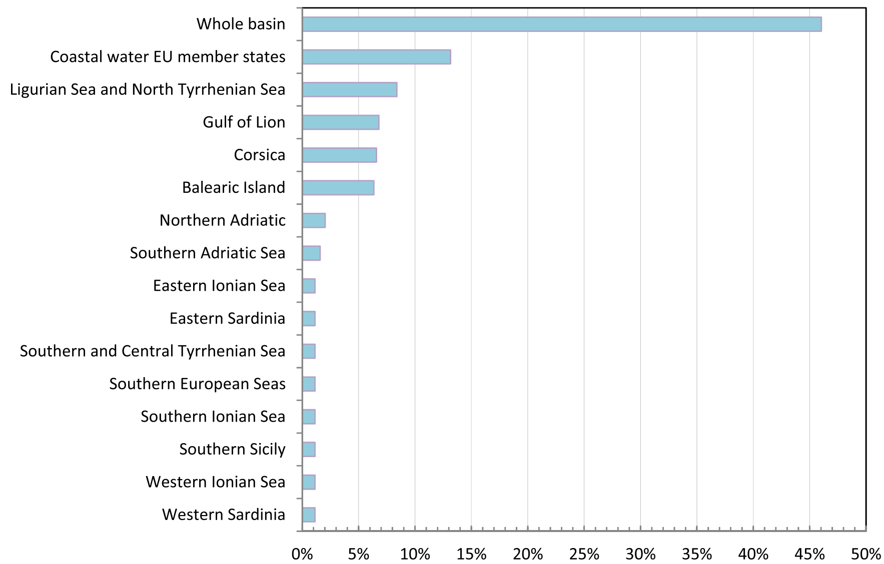


Fig. 7. Geographical coverage: No. of occurrences of the monitored sea in the Whole_Master_Data sheet.

to the platforms is open. As for the temporal resolution, the data are in general available as monthly means, but in the case of Eionet platform, the original data are directly accessible at the EEA web site.

Information on the spatial resolution is not reported always in detail (Table 3). The spatial scales may vary depending on the EU member countries and should respond to the requirements of the European Directives (mainly WFD and MSFD) that prescribe the sea monitoring. In particular, as detailed in Table 3, spatial resolution is indicated as: 1, 3, 7.5 or 30 arcsec, 1 arcminute, between 0.004166 and 1°, or between 1 and 25 km. The available typologies of data for this category, are shown in the Table 2.

The whole basin is represented in a considerable number of platforms. Spatial and temporal resolutions, together with the monitoring period of each platform are summarized in Table 3. Furthermore, all the additional information for each data source is listed in Annex 2.

Table 2
Data typologies represented in the EU source platforms and number of the monitored parameters, per each typology.

Data Typology	Data Typology code P03	N. of available parameters
Pigments	B035	2
Habitat	B050	3
Macroalgae and seagrass	B055	1
Biota abundance biomass and diversity	B070	7
Carbon nitrogen and phosphorus	C005	9
Dissolved gases	C015	1
Water column temperature and salinity	D025	6
Currents	D030	2
Sea level	D032	1
Suspended particulate material	G015	1
Rock and sediment physical properties	G040	1
Sedimentation and erosion processes	G060	1
Rock and sediment sedimentology	GSED	1
Anthropogenic contamination	H001	18
Construction and structures	H002	1
Human activity	H005	1
Terrestrial including bathymetry and under-sea features	T001	2

About 30% of the "Whole basin" platforms provide data starting from year 2000 and 66% of the monitoring programmes referred to these platforms, are still in progress. The access to the platforms with a few exceptions is in general open. Spatial and temporal resolutions of the available data are very varied, depending on the type of parameter considered.

As can be seen from the next diagram (Fig. 8), almost all types of data are represented. The sequence of data types by order of importance is very similar to the trend already reported in Fig. 6, relating to all remaining platforms. In this case, however, the type D025 (water column temperature and salinity), is placed in the leading positions, as expected.

As discussed above, the procedure adopted to perform the data gap analysis could not start from a punctual examination of data related to the geographical coverage, since the large amount of records in the Whole_master_Data refers to the entire Mediterranean basin or in any case to most sub-basins and includes data gathered in inhomogeneous ways. The same apply to the temporal coverage. It is a fact that the examined data were provided in a very inhomogeneous manner leading to difficult and time-consuming adaptation of these data. Nevertheless, the temporal coverage provided by most platforms can be considered satisfactory. The individual situations referred to parameter of interests could however be accessed directly from the Whole_Master_Data.

Consequently, data analysis focused on the assessment of the degree or lack of representativeness of the various data types, evaluating this representativeness to the different sub-basins that appear in the dataset.

The first step of the analysis was to order the dataset by sub-basin and by type. From this ranking, a preliminary matrix was obtained, as reported in Table 4. This matrix has two entries: the first (rows) is related to typologies, the second (columns) is relative to the sub-basins, including the "whole basin" category and the "Coastal water EU member states" category. References to sub-basins inadequately represented by the data, with a negligible occurrence in the whole dataset, are not taken into account.

Each cell of this preliminary matrix contains the number of data_parameters available and related to the corresponding data_typology (row) referring to the individual sub-basins (column). This

Table 3
Monitoring period and spatial-temporal resolutions of the analysed platforms.

N.	Platform name	Start year	Finish year	Spatial resolution	Temporal resolution
1	ACCOBAMS	2001	Ongoing		
2	AEMET	Ante 1990–2006	Ongoing	0.06–0.2 deg	3-Hourly mean, daily, monthly mean
3	Agence des aires marines protégées	2006	Ongoing		
4	AQUAMAPS for Marine Species	2010	Ongoing	0.5 deg	Hourly, daily, monthly mean
5	ARPEGE	1994	Ongoing	7,5 km	9-Monthly mean
6	AVISO CLS	1993	Ongoing	1 arcminute, 0.25 deg, 6–7 km, 1/60°, 1/16°	Daily, monthly mean
8	BirdLife-001	2005	Ongoing		
9	Boussole	2000	Ongoing		Hourly, daily, monthly mean
10	CBD-001	2008	Ongoing		
11	CERSAT IFREMER MEDSPIRATION	2004	Ongoing	0,01 deg	Daily, monthly mean
12	ChEssBase-002	2010	2010		
13	CISL Research Data Archive	Ante 1990	Ongoing		Real time based on station
14	CoastColour	2002	2012	5 m	Hourly, daily, monthly mean
15	COCONET	2011	2016		
16	CoL-001 - Catalogue Of Life	2001	Ongoing		Monthly mean, annual mean
17	CMEMS	1993, 1997, 2014	Ongoing	1–25 km, 0.1–0.12 deg	Daily, weekly, monthly mean
18	Corine Land Cover (EEA)	1990	1990		Hourly, daily, monthly mean
19	Data Collection (DCR-DCF) for the Common Fisheries	2000	Ongoing		
20	DATA.SHOM.FR	Ante 1990	Ongoing		
21	DIRM MéDITERRANéE	2010	Ongoing		
22	ECMWF	Ante 1990	Ongoing		Hourly, daily, monthly mean
23	EEA - EIONET	2000	Ongoing		Monthly mean
24	EMODNET	2009	Ongoing		
25	EoL-001 - Enciclopedia of Life	2007	Ongoing		
26	EUMETSAT	1999–2016	Ongoing	0.3–50 km, 0.05 deg	Real time based on station, 12-hourly mean
27	FAO Fish and Aquaculture	Ante 1990	Ongoing		
28	FishBase	Ante 1990	Ongoing		
29	GBIF	1999	Ongoing		
30	GEBCO	Ante 1990	Ongoing		
31	Global Fisheries Catch dataset	Ante 1990	2014	30 min spatial cells	Annual mean
32	GRDC	Ante 1990	Ongoing		Daily, monthly mean
33	GridA-001	Ante 1990	Ongoing		
34	GROMS	2001	Ongoing		
35	HIRLAM, WRF	Ante 1990	Ongoing	0.1 deg	12-Hourly mean
36	HyMeX	2007	Ongoing		
37	IBAT	Ante 1990	Ongoing		
38	Instituto Español de Oceanografia (IEO)				Hourly, daily, monthly mean
39	International Maritime Organization IMO-001	2016	2016		
40	IntRid-001	Ante 1990	2011		
41	IUCN - World Database on Protected Areas (WDPA)	Ante 1990–2013	Ongoing		
42	IWC-001 - Ship Strike Database	Ante 1990	Ongoing		
43	MAPAMED	2008	Ongoing		
44	MarBEF	2004	Ongoing		
45	Marina Platform	2010	2014		
46	Mediseh-001	Ante 1990	2014		
47	MEDISEH-MAREA	1994	2012	Vector data	Monthly mean
48	MEDITS surveys	1994	2016		Real time based on station
49	MONGOOS - AFS	2013	Ongoing		Hourly, daily, monthly mean
50	MYOCEAN	Ante 1990–2007	Ongoing	0.06 deg, 1–25 km	Real time based on station, hourly, daily, monthly mean
51	NASA - Oceancolor	Ante 1990	Ongoing		
52	NCEAS - Global Health Index	2012	2012		
53	NOAA-001	2013	2013		
54	OBIS	1999	Ongoing		
55	OBPG/NASA's Goddard Space Flight Center	2002–2011	Ongoing	1 deg, 4–9 km	Daily, monthly mean
56	Oceana 2011	2001	Ongoing		Monthly mean
57	OTN	2015	Ongoing		
58	PANGAEA		Ongoing		
59	PO.DAAC	1996	2003	0.5 deg	Monthly mean
60	PSMSL	Ante 1990	2013		
61	Puertos del Estado (REDEXT)	1996			Real time based on station, Hourly, daily, monthly mean
62	RAC-SPA	Ante 1990	Ongoing		
63	RivDIS	Ante 1990	1991		Monthly mean
64	RMN (Italy)	2008	Ongoing		
65	SeaDataNet	2006	Ongoing		Hourly, daily, monthly mean
66	SEDNET - Sediment Network	2002	Ongoing		
67	SESAME	2006	2010		
68	SIH - Système d'Informations Halieutiques	2002–2009	Ongoing		
69	SKIRON	1992	Ongoing		
70	SLBase-001 - SeaLifeBase	2008	Ongoing		

(continued on next page)

Table 3 (*continued*)

N.	Platform name	Start year	Finish year	Spatial resolution	Temporal resolution
71	SWOT-001	1993	1999		
74	TNC-002	Ante 1990	Ongoing		
75	UBC-003 - Global Estuary Database	2003	2003		
78	VLIZ-006 - World Porifera Database	2002	Ongoing		
81	WaDNR-001	2001	2013		
82	WCMC	Ante 1990–2015	1993–ongoing	6 raster	
83	WoRMS	2000	Ongoing		
84	ZSL-001	Ante 1990	2012		

preliminary matrix can be considered as an indicator of the "overall information" available, as provided by the Whole_Master_Data sheet, and likewise the individual rows of the matrix summarize all information related to each Data_Typology.

Parameter's Gap matrix (Table 5) represents the final output of the data gap analysis. It has been obtained by scaling each cell value of the Preliminary_Matrix by the total amount of available information per each Data_Typology, i.e., by dividing the content of each cell by the sum of the values per row. The score so obtained represents the punctual value (or the goodness) of the information provided by crossing Data_Typology with the corresponding seas. By making the complement to 1 of each of these scores, we will obtain a measure of the data gap that must be filled for those sub-basins in relation to the corresponding Data_Typology.

In the matrix, the numerical values of data gaps have been highlighted with appropriate colours, to favour the overall presentation of results obtained through this procedure.

5. Conclusions

The main goals of this research were the identification of geographical and temporal data gaps in the Mediterranean basin and the elaboration of an innovative methodology to prioritise these gaps.

Analysing the reliability and usefulness of the existing data sources, it was possible to identify their critical issues and shortcomings that are mainly related to data aggregation and accessibility as well as low data frequency and quality in terms of spatial and temporal resolutions.

The assignment of ad-hoc scores to each data gap, identified through the described approach, allowed the implementation of an explicit data gap analysis, which led to the final product of this research: a prioritisation of existing geographical and temporal data gaps in the whole Mediterranean basin. In particular, as regard the elaboration and application of the new developed method for identifying geographical and temporal marine data gaps and for providing a prioritisation of missing data, since the large amount of analysed records includes data gathered in heterogeneous ways, the prioritisation values obtained for each identified data gap simplify the data comparison and analysis. Consequentially it could be useful for end-user's specific needs including environmental quality, ecological health and human health monitoring activities due to the multidisciplinary analysis of data parameters that integrates in the elaborated inventory the biological, geological, chemical and physical branches of the total environment.

Lastly, the implementation of all the existing information in a single dataset may be considered as a starting point for further processing and for detailed analysis for fill data gaps with in situ monitoring surveys ad smaller scale.

Acknowledgements

ODYSSEA project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727277. A special thanks to all the ODYSSEA partners who provided data filling in the template shared by Sapienza University team.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2019.02.417.

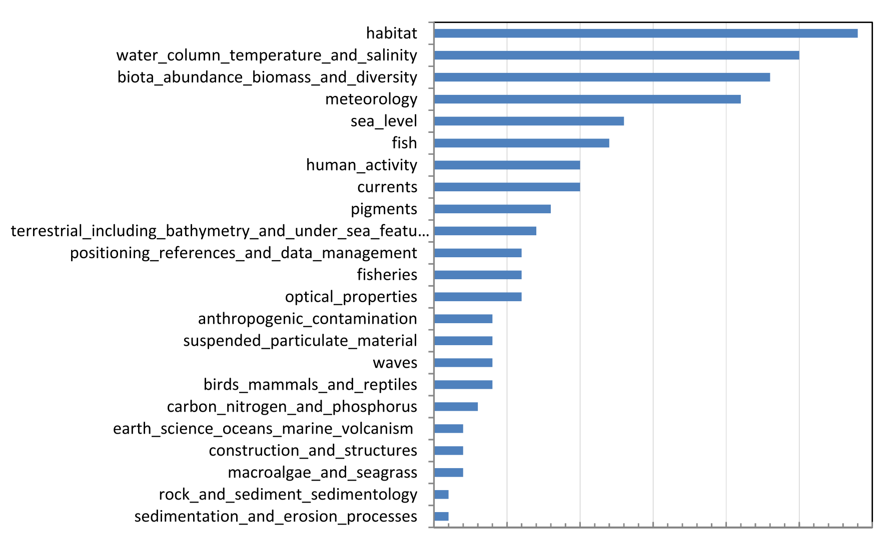


Fig. 8. Data typologies represented in the "Whole basin" source platforms and number of the monitored parameters, per each typology.

Table 4
Preliminary Matrix containing the number of available data_parameters for each data_typology (row) and sub-basin (column).

		Whole basin	Western Sardinia	Western Ionian Sea	Southern Sicily	Southern Ionian Sea	Southern European Seas	Southern and Central Tyrrhenian Sea	Southern Adriatic Sea	Northern Adriatic	Ligurian Sea and North Tyrrhenian Sea	Gulf of Lion	Eastern Sardinia	Eastern Ionian Sea	Corsica	Coastal water EU member states	Balearic Island
B015	Birds mammals and reptiles	4	0	0	0	0	0	0	0	0	4	4	0	0	4	0	0
B020	Fish	12	0	0	0	0	0	0	0	0	3	4	0	0	4	0	1
B035	Pigments	8	0	0	0	0	0	0	0	0	1	0	0	0	0	2	1
B050	Habitat	29	0	0	0	0	0	0	1	2	0	0	0	0	0	3	1
B055	Macroalgae and seagrass	2	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0
B070	Biota abundance biomass and diversity	23	0	0	0	0	0	0	0	0	3	3	0	0	3	7	1
C005	Carbon nitrogen and phosphorus	3	0	0	0	0	3	0	0	0	0	0	0	0	0	9	4
C015	Dissolved gases	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
D015	Optical properties	6	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
D025	Water column temperature and salinity	25	2	2	2	2	1	2	2	2	5	0	2	2	0	6	2
D030	Currents	10	0	0	0	0	1	0	0	1	1	1	0	0	1	2	1
D032	Sea level	13	1	1	1	1	0	1	1	1	1	0	1	1	0	1	2
D034	Waves	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
G015	Suspended particulate material	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
G040	Rock and sediment physical properties	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
G045	Rock and sediment lithology and mineralogy	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
G060	Sedimentation and erosion processes	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
GSED	Rock and sediment sedimentology	1	0	0	0	0	0	0	0	0	1	2	0	0	0	1	0
H001	Anthropogenic contamination	4	0	0	0	0	0	0	0	0	8	2	0	0	8	18	0
H002	Construction and structures	2	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
H004	Fisheries	6	0	0	0	0	0	0	0	0	2	2	0	0	2	0	0
H005	Human activity	10	0	0	0	0	0	0	0	0	4	7	0	0	5	1	0
M010	Meteorology	21	2	2	2	2	0	2	2	2	2	1	2	2	1	0	10
T001	Terrestrial including bathymetry and under-sea features	7	0	0	0	0	0	0	0	0	1	2	0	0	1	2	1
VOLC	Earth science oceans marine volcanism	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Z005	Positioning references and data management	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 5

Data Gaps Matrix. data gap classes: “maximum” for score gaps: 1; “high” for score gaps between 0,8 and 1; “medium” for score gaps between 0,5 and 0,8; “minimum” for score gaps lower than 0,5.

		Whole basin	Western Sardinia	Western Ionian Sea	Southern Sicily	Southern Ionian Sea	Southern European Seas	Southern and Central Tyrrhenian Sea	Southern Adriatic Sea	Northern Adriatic	Ligurian Sea and North Tyrrhenian Sea	Gulf of Lion	Eastern Sardinia	Eastern Ionian Sea	Corsica	Coastal water EU member states	Balearic Island
B015	birds mammals and reptiles	0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	0.75	1.00	1.00	0.75	1.00	1.00
B020	fish	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.83	1.00	1.00	0.83	1.00	0.96
B035	pigments	0.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00	1.00	1.00	0.83	0.92
B050	habitat	0.19	1.00	1.00	1.00	1.00	1.00	1.00	0.97	0.94	1.00	1.00	1.00	1.00	1.00	0.92	0.97
B055	macroalgae and seagrass	0.60	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.80	1.00	1.00	1.00	1.00	1.00	0.80	1.00
B070	biota abundance biomass and diversity	0.43	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.93	1.00	1.00	0.93	0.83	0.98
C005	carbon nitrogen and phosphorus	0.84	1.00	1.00	1.00	1.00	0.84	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.53	0.79
C015	dissolved gases	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	0.50
D015	optical properties	0.14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.86	1.00	1.00	1.00	1.00	1.00	1.00
D025	water column temperature and salinity	0.56	0.96	0.96	0.96	0.96	0.98	0.96	0.96	0.96	0.91	1.00	0.96	0.96	1.00	0.89	0.96
D030	currents	0.44	1.00	1.00	1.00	1.00	0.94	1.00	1.00	0.94	0.94	0.94	1.00	1.00	0.94	0.89	0.94
D032	sea level	0.50	0.96	0.96	0.96	0.96	1.00	0.96	0.96	0.96	0.96	1.00	0.96	0.96	1.00	0.96	0.92
D034	waves	0.33	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.67
G015	suspended particulate material	0.20	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	1.00
G040	rock and sediment physical properties	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00
G045	rock and sediment lithology and mineralogy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	1.00	0.50
G060	sedimentation and erosion processes	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.50	1.00
GSED	rock and sediment sedimentology	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.60	1.00	1.00	1.00	0.80	1.00
H001	anthropogenic contamination	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.80	0.95	1.00	1.00	0.80	0.55	1.00
H002	construction and structures	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.75	1.00	1.00	1.00	0.75	1.00
H004	fisheries	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.83	0.83	1.00	1.00	0.83	1.00	1.00
H005	human activity	0.63	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.85	0.74	1.00	1.00	0.81	0.96	1.00
M010	meteorology	0.60	0.96	0.96	0.96	0.96	1.00	0.96	0.96	0.96	0.96	0.98	0.96	0.96	0.98	1.00	0.81
T001	terrestrial including bathymetry and under sea features	0.50	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.86	1.00	1.00	0.93	0.86	0.93
VOLC	Earth science oceans marine volcanism	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Z005	positioning references and data management	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

	maximum	gap = 1
	high	gap = 0,8 - 1
	medium	gap = 0,5 - 0,8
	minimum	gap = < 0,5

References

Al Shami, A., Harik, G., Alameddine, I., Bruschi, D., Astiaso Garcia, D., El-Fadel, M., 2017. Risk assessment of oil spills along the Mediterranean coast: a sensitivity analysis of the choice of hazard quantification. Sci. Total Environ. 574, 234–245. https://doi.org/10.1016/j.scitotenv.2016.09.064.

Astiaso Garcia, D., Cumo, F., Gugliermetti, F., Rosa, F., 2013. Hazardous and noxious substances (HNS) risk assessment along the Italian coastline. Chem. Eng. Trans. 32, 115–120. https://doi.org/10.3303/CET1332020.

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecol. Indic. 18, 31–41.

Bonsignore, M., Salvagio Manta, D., Al-Tayeb Sharif, E.A., D'Agostino, F., Traina, A., Quinci, E.M., ... Sprovieri, M., 2018. Marine pollution in the Libyan coastal area: environmental and risk assessment. Mar. Pollut. Bull. 128, 340–352.

Brown, F.W., Dodd, N.G., 1998. Utilizing organizational culture gap analysis to determine human resource development needs. Leadership Org Dev J 19 (7), 374–385.

Bugnot, A.B., Hose, G.C., Walsh, C.J., Floerl, O., French, K., Dafforn, K.A., ... Hahs, A.K., 2019. Urban impacts across realms: making the case for inter-realm monitoring and management. Sci. Total Environ. 648, 711–719. https://doi.org/10.1016/j.scitotenv.2018.08.134.

Buongiorno Nardellia, B., Tronconi, C., Pisano, A., Santoleri, R., 2013. High and ultra-high resolution processing of satellite sea surface temperature data over Southern European seas in the framework of MyOcean project. Remote Sens. Environ. 129, 1–16. https://doi.org/10.1016/j.rse.2012.10.012.

Carlucci, R., Maglietta, R., Buscaino, G., Cipriano, G., Milella, A., Pollazzon, V., ... Fanizza, C. (2017). Review on research studies and monitoring system applied to cetaceans in the gulf of Taranto (northern Ionian sea, central-eastern Mediterranean sea). Paper presented at the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2017, doi:https://doi.org/10.1109/AVSS.2017.8078473.

Caro, C., Pinto, R., Marques, J.C., 2018. Use and usefulness of open source spatial databases for the assessment and management of European coastal and marine ecosystem services. Ecol. Indic. 95, 41–52. https://doi.org/10.1016/j.ecolind.2018.06.070.

Carstensen, J., 2014. Need for monitoring and maintaining sustainable marine ecosystem services. Front. Mar. Sci. 1 (33).

Chapman, P.M., 2015. Future challenges for marine pollution monitoring and assessment. Mar. Pollut. Bull. 95 (1), 1–2.

Criado-Aldeanueva, F., Soto-Navarro, J., Garcıa-Lafuente, J., 2014. Large-scale atmospheric forcing influencing the longterm variability of Mediterranean heat and freshwater budgets: climatic indices. J. Hydrometeorol. 15, 650–663.

Cumo, F., Cinquepalmi, F., Astiaso Garcia, D., 2008. Data gathering guidelines for the mapping of environmental sensitivity to oil spill of the Italian coastlines. WIT Transactions on the Built Environment, 99 119–125 https://doi.org/10.2495/CENV080111.

Curren, E., Leong, S.C.Y., 2019. Profiles of bacterial assemblages from microplastics of tropical coastal environments. Sci. Total Environ. 655, 313–320.

Cusinato, E., Zanchettin, D., Sannino, G., Rubino, A., 2018. Mediterranean thermohaline response to large-scale winter atmospheric forcing in a high-resolution ocean model simulation. Pure Appl. Geophys. 175 (11), 4083–4110. https://doi.org/10.1007/s00024-018-1859-0.

Declerck, A., Ourmières, Y., Molcard, A., 2016. Assessment of the coastal dynamics in a nested zoom and feedback on the boundary current: the north-western Mediterranean sea case. Ocean Dyn. 66 (11), 1529–1542.

Dekker, R., 2006. The importance of having data-sets. Proceedings of the IATUL Conferences, p. 16.

Dimarchopoulou, D., Stergiou, K.I., Tsikliras, A.C., 2017. Gap analysis on the biology of Mediterranean marine fishes. PLoS One 12 (4).

Eionet. The European environment information and observation network. https://www.eionet.europa.eu/

Elliott, M., 2013. The 10-tenets for integrated, successful and sustainable marine management. Mar. Pollut. Bull. 74 (1), 1.

EMODnet Data Ingestion portal, d. http://www.emodnet.eu/welcome-new-emodnet-data-ingestion-portal-turn-your-data-valuable-asset-marine-and-maritime.

European Commission, 2007. SeaDataNet I3 - Pan-European infrastructure for Ocean & Marine Data management. Ifremer pp 15.

European Parliament and of the Council, 2000. Directive 2000/60/EC. Off. J. Eur. Union L327/1–72.

European Parliament and of the Council, 2007. Directive 2007/2/EC. Official Journal of the European Union. L108/1-17.

European Parliament and of the Council, 2008. Directive 2008/56/EC. Official Journal of the European Union. L164/19-40.

Fu, X.-., Zhang, M.-., Liu, Y., Shao, C.-., Hu, Y., Wang, X.-., ... Wang, C., 2018. Protective exploitation of marine bioresources in China. Ocean Coast. Manag. 163, 192–204. https://doi.org/10.1016/j.ocecoaman.2018.06.018.

Garnier, J., Ramarson, A., Billen, G., Théry, S., Thiéry, D., Thieu, V., ... Moatar, F., 2018. Nutrient inputs and hydrology together determine biogeochemical status of the Loire river (France): current situation and possible future scenarios. Sci. Total Environ. 637-638, 609–624.

Gharaibeh, A.A., 2017. Environmental Impact Assessment on Oil Shale Extraction in Central Jordan. FOG - Freiberg Online Geoscience. vol. 50 pp. 1–157.

Gugliermetti, F., Cinquepalmi, F., Astiaso Garcia, D., 2007. The use of environmental sensitivity indices (ESI) maps for the evaluation of oil spill risk in Mediterranean coastlines and coastal waters. WIT Trans. Ecol. Environ. 102, 593–600. https://doi.org/10.2495/SDP070572.

Hardesty, B.D., Harari, J., Isobe, A., Lebreton, L., Maximenko, N., Potemra, J., ... Wilcox, C., 2017. Using numerical model simulations to improve the understanding of microplastic distribution and pathways in the marine environment. Front. Mar. Sci. 4 (MAR). https://doi.org/10.3389/fmars.2017.00030.

Harik, G., Alameddine, I., Maroun, R., Rachid, G., Bruschi, D., Astiaso Garcia, D., El-Fadel, M., 2017. Implications of adopting a biodiversity-based vulnerability index versus a shoreline environmental sensitivity index on management and policy planning along coastal areas. J. Environ. Manag. 187, 187–200. https://doi.org/10.1016/j.jenvman.2016.11.038.

Hinz, H., Murray, L.G., Lambert, G.I., Hiddink, J.G., Kaiser, M.J., 2013. Confidentiality over fishing effort data threatens science and management progress. Fish Fish. 14 (1), 110–117.

Howarth, R.W., 2008. Coastal nitrogen pollution: a review of sources and trends globally and regionally. Harmful Algae 8 (1), 14–20. https://doi.org/10.1016/j.hal.2008.08.015.

Huang, W., Zeng, J.N., Chen, Q.Z., Du, P., Tang, Y.B., Yang, H., 2016. Preliminary research on the zoning method of the marine ecological red line: a case study of Hainan province. Shengtai Xuebao. Acta Ecol. Sin. 36 (1), 268–276. https://doi.org/10.5846/stxb201404030635.

International Organization for Standardization (ISO), 2014. ISO 19115-1:2014 geographic information – metadata. https://www.iso.org/standard/53798.html (accessed 28 August 2018).

International Organization for Standardization (ISO), 2010. Assessing economic benefits of consensus-based standards - the ISO methodology. https://www.iso.org/sites/materials/benefits-of-standards/benefits-detail57da.html?emid=6 (accessed 28 August 2018).

Jepson, P.D., Law, R.J., 2016. Persistent pollutants, persistent threats perspectives: polychlorinated biphenyls remain a major threat to marine apex predators such as orcas. Science 352 (6292), 1388–1389.

Jordà, G., Von Schuckmann, K., Josey, S.A., Caniaux, G., García-Lafuente, J., Sammartino, S., ... Macías, D., 2017. The Mediterranean sea heat and mass budgets: estimates, uncertainties and perspectives. Prog. Oceanogr. 156, 174–208. https://doi.org/10.1016/j.pocean.2017.07.001.

Karydis, M., Kitsiou, D., 2013. Marine water quality monitoring: a review. Mar. Pollut. Bull. 77 (1), 23–36.

Lancheros, E., Camps, A., Park, H., Sicard, P., Mangin, A., Matevosyan, H., Lluch, I., 2018. Gaps analysis and requirements specification for the evolution of copernicus system for polar regions monitoring: addressing the challenges in the horizon 2020–2030. Remote Sens. 10 (7). https://doi.org/10.3390/rs10071098.

Laroche, S., Andral, B., Pantazi, M., Vasilopoulou Vassiliki, C., Gonzalez-Fernandez, D., Hanke, G., Cadiou, J.F., Secrieru, D., Begun, T., Gomoiu, M.T., 2013. A tool for gap analysis of the MSFD assessment elements of the Southern European seas countries, carried out in the framework of the Perseus FP7 project. Rapp. Comm. int. Mer Médit. 40.

Lauria, V., Das, I., Hazra, S., Cazcarro, I., Arto, I., Kay, S., ... Fernandes, J.A., 2018. Importance of fisheries for food security across three climate change vulnerable deltas. Sci. Total Environ. 640-641, 1566–1577.

Lecacheux, S., Bonnardot, F., Rousseau, M., Paris, F., Pedreros, R., Lerma, A.N., ... Barbary, D., 2018. Probabilistic forecast of coastal waves for flood warning applications at Reunion island (Indian ocean). J. Coast. Res. 85, 776–780.

Llasses, J., Jordà, G., Gomis, D., Adloff, F., Macías, D., Harzallah, A., ... Sannino, G., 2018. Heat and salt redistribution within the Mediterranean sea in the med-CORDEX model ensemble. Clim. Dyn. 51 (3), 1119–1143. https://doi.org/10.1007/s00382-016-3242-0.

Marbà, N., Díaz-Almela, E., Duarte, C.M., 2014. Mediterranean seagrass (Posidonia oceanica) loss between 1842 and 2009. Biol. Conserv. 176, 183–190.

Marignani, M., Bruschi, D., Astiaso Garcia, D., Frondoni, R., Carli, E., Pinna, M.S., ... Blasi, C., 2017. Identification and prioritization of areas with high environmental risk in Mediterranean coastal areas: a flexible approach. Sci. Total Environ. 590–591, 566–578. https://doi.org/10.1016/j.scitotenv.2017.02.221.

Masina, S., Storto, A., Ferry, N., Valdivieso, M., Haines, K., Balmaseda, M., ... Parent, L., 2017. An ensemble of eddy-permitting global ocean reanalyses from the MyOcean project. Clim. Dyn. 49, 813–841. https://doi.org/10.1007/s00382-015-2728-5.

Mazaris, A.D., Germond, B., 2018. Bridging the gap between climate change and maritime security: towards a comprehensive framework for planning. Sci. Total Environ. 635, 1076–1080. https://doi.org/10.1016/j.scitotenv.2018.04.136.

Moura, J.F., Tavares, D.C., Lemos, L.S., Acevedo-Trejos, E., Saint'Pierre, T.D., Siciliano, S., Merico, A., 2018. Interspecific variation of essential and non-essential trace elements in sympatric seabirds. Environ. Pollut. 242, 470–479. https://doi.org/10.1016/j.envpol.2018.06.092.

Nicholson, M.D., Jennings, S., 2004. Testing candidate indicators to support ecosystembased management: the power of monitoring surveys to detect temporal trends in fish community metrics. ICES Journal of Marine Science: Journal du Conseil 61 (1), 35–42.

Nilssen, I., Ødegård, Ø., Sørensen, A.J., Johnsen, G., Moline, M.A., Berge, J., 2015. Integrated environmental mapping and monitoring, a methodological approach to optimise knowledge gathering and sampling strategy. Mar. Pollut. Bull. 96 (1), 374–383.

Olsen L.M. et al., 2006. NASA/GCMD Earth Science Keywords Version 5.3.3. Natural Environment Research Council.

Rodríguez-Rodríguez, D., Sánchez-Espinosa, A., Schröder, C., Abdul Malak, D., Rodríguez, J., 2015. Cumulative pressures and low protection: a concerning blend for Mediterranean MPAs. Mar. Pollut. Bull. 101 (1), 288–295. https://doi.org/10.1016/j.marpolbul.2015.09.039.

Scott, J.M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., et al., 1993. Gap analysis: a geographic approach to protection of biological diversity. Wildl. Monogr. 123, 3–41.

Stergiou, K.I., Tsikliras, A.C., 2006. Underrepresentation of regional ecological research output by bibliometric indices. Ethics Sci Environ Polit–17. Pilling GM.

Tasker, M.L., 2016. How might we assess and manage the effects of underwater noise on populations of marine animals? Adv. Exp. Med. Biol. 875, 1139–1144. https://doi.org/10.1007/978-1-4939-2981-8_141.

Turrell, W.R., 2018. Improving the implementation of marine monitoring in the Northeast Atlantic. Mar. Pollut. Bull. 128, 527–538. https://doi.org/10.1016/j.marpolbul.2018.01.067.

Xu, H.G., Ding, H., Ouyang, Z.Y., Zhang, W.G., Cui, P., Xu, W.H., ... Lei, J.C., 2016. Assessing China's progress toward the 2020 global biodiversity targets. Shengtai Xuebao. Acta Ecol. Sin. 36 (13), 3847–3858. https://doi.org/10.5846/stxb201411142254.

Zaccaroni, A., Andreini, R., Franzellitti, S., Barceló, D., Eljarrat, E., 2018. Halogenated flame retardants in stranded sperm whales (Physeter macrocephalus) from the Mediterranean sea. Sci. Total Environ. 635, 892–900.