https://doi.org/10.5194/sp-2024-23
Preprint. Discussion started: 27 September 2024

Merging and serving Ocean Observations: a description of Marine Data Aggregators

Antonio Novellino[1], Pierre-Yves Le Traon[2], Andy Moore[3]

[1] ETT SpA, Genoa, Italy
[2] Mercator Ocean International, Toulouse, France
[3]Physical & Biological Sciences Division, Ocean Sciences Department Institute of Marine Sciences, Institute of Marine Sciences, University of California Santa Cruz, Santa Cruz, California, USA

Correspondence to: Antonio Novellino (antonio.novellino@grupposcai.it)

Abstract. Observations are a fundamental element in ocean predictions: they are crucial not only for monitoring the ocean state, but also for improving the forecasting systems and validating the model outputs. In this framework, it is essential to adequately access, manage, and integrate such information in the ocean value chain. Data providers are in charge of collecting, processing and analyzing these observations, delivering comprehensive datasets that can be used for informed decision making and by forecasters to improve ocean models. In this paper, several examples of data services are discussed – ranging from the Copernicus Marine In-Situ Thematic Assembly Center to European Marine Observation and Data network (EMODnet) to SeaDataNet – recognized as key players in the framework of monitoring and management of the marine resource. The paper offers an outlook on future directions in ocean data integration, particularly on the opportunities offered by the standardization of protocols for data dissemination and the role of cost-effective and citizen-based data collection.

1 Introduction

The importance of ocean observation in met-ocean forecasting is emphasized, as it provides crucial data for understanding oceanic behavior and coastal areas. The integration of parameters like temperature, salinity, currents, and atmospheric conditions enhances model accuracy, crucial for effective management of human impacts and resource exploitation. The complex ocean data collection framework involves numerous in situ platforms, remote sensors, and types of data, necessitating the provision of multidisciplinary, aggregated datasets (Belbéoch et al., 2022).

Marine data integrators play a pivotal role in managing, integrating, and advancing the understanding of marine environments. They collect, process, and analyze diverse data types to create comprehensive datasets, contributing to informed decision-making in areas such as fisheries management, offshore energy development, and marine conservation. Additionally, these integrators support the development of technologies for monitoring the marine environment, continually refining data collection processes to enhance accuracy.

Over the past three decades, progress in marine data management has been marked by the establishment of international programs and networks, such as the International Oceanographic Data and Information Exchange (IODE) and the IOC Ocean

https://doi.org/10.5194/sp-2024-23
Preprint. Discussion started: 27 September 2024

Data and Information System (ODIS). These initiatives, including the World Ocean Database, involve collaborative efforts globally, led by organizations like the Intergovernmental Oceanographic Commission (IOC) and the International Council for the Exploration of the Sea (ICES).

Under the Global Ocean Observing System (GOOS) framework, coordinated efforts by OceanOPS and GOOS Regional Alliances (GRAs) contribute to the provision of ocean observing information (Moltmann et al., 2019). GRAs integrate national monitoring needs into a regional system, facilitating data assembly and exchange (Corredor, 2018). Data Assembly Centers (DACs) and Global DACs (GDACs) play a key role in this process by receiving, quality controlling, and assembling data from various sources. They act as the primary point of access according to a common data format.

In this intricate and dispersed framework, integration services play a crucial role by harmonizing metadata, applying standardized data quality checks, and facilitating the integration of diverse datasets and models. The use of unified controlled vocabularies, common data models, and transport formats ensures the seamless integration of real-time and delayed mode observations into numerical models. The following section describes the European marine data integration landscape which is characterized by three prominent initiatives: the Copernicus Marine Service, the European Marine Observation and Data network and the SeaDataNet network of the National Oceanographic Data Centers associated with the International Oceanographic Commission (IOC).

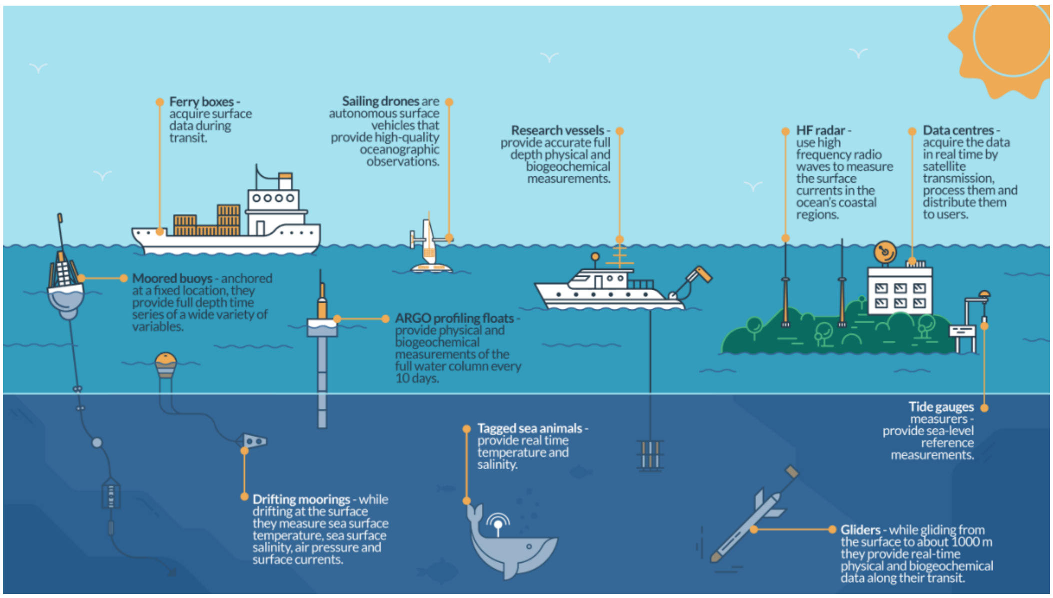

Figure 1: In-situ platforms for ocean data collection (from https://marine.copernicus.eu/explainers/operational-oceanography/monitoring-forecasting/in-situ)

2 European Marine Data Integrators

To exemplify the importance of data integrators, a few relevant examples from Europe are presented.

Copernicus Marine Service INS TAC

https://doi.org/10.5194/sp-2024-23
Preprint. Discussion started: 27 September 2024

Within this programme, the Copernicus INS TAC is a distributed service integrating data from different sources for operational needs in oceanography. The Copernicus INS TAC integrates and quality controls in a homogeneous manner in situ data from outside CMEMS data providers in order to fit the needs of internal and external users. It provides access to integrated datasets of core parameters for initialization, forcing assimilation and validation of ocean numerical models, which are used for forecasting, analysis and re-analysis of ocean physical and biogeochemical conditions. Since the primary objective of CMEMS is to forecast ocean state, the initial focus has been on observations from autonomous observatories at sea (e.g. floats, buoys, gliders, ferryboxes, drifters, and ships of opportunity). The second objective is to set up a system for re-analysis purposes that requires products integrated over the past 25 to 60 years. The Copernicus INS TAC comprises a global in-situ center and 6 regional in-situ centers, one for each EuroGOOS Regional Ocean Observing System (ROOS). The INS TAC has been designed to fulfill the Copernicus Marine Core Service needs and the EuroGOOS ROOS needs. The focus is on parameters that are presently necessary for Copernicus Monitoring and Forecasting Centers, namely temperature, salinity, sea level, current, waves, chlorophyll / fluorescence, oxygen and nutrients. Additional atmospheric parameters (such as wind, air temperature, air pressure, etc.) are added by some ROOSes to these regional in-situ portals to fulfill additional downstream applications needs.

For NRT (Near-Real-Time) and DM (Delayed Mode) products, the Copernicus INS TAC (In-Situ Thematic Assembly Centre) is connected to the OceanOPS networks and each Regional Ocean Observing System (ROOS) of EuroGOOS. In the case of DM products, it is also connected to the SeaDataNet Network, which comprises National Oceanographic Data Centers (NODCs). The Copernicus INS TAC integrates data from various observation programs, including Argo, Everyone's Gliding Observatories (EGO), Data Buoy Cooperation Panel (DBCP), OceanSITES, and ship data obtained via NODCs, leveraging the AtlantOS observations. Whenever possible, the Copernicus INS TAC adheres to the standards developed within the SeaDataNet framework.

EMODnet

The European Marine Observation and Data Network (EMODnet) is the EU infrastructure for in situ marine data. The goal of EMODnet is to provide access to a wide range of standardized and harmonized marine data, making it easier for researchers, policymakers, and the public to access and use marine information. EMODnet focuses on various thematic areas, including bathymetry, geology, physics, chemistry, biology, and human activities in the marine environment (Shepherd, 2018). By pooling and harmonizing data from various sources, EMODnet aims to create a comprehensive and easily accessible marine data infrastructure that supports a wide range of marine and maritime activities.

EMODnet Physics (https://emodnet.ec.europa.eu/en/physics, Figure 4.1-3) is the domain-specific project (Míguez et al., 2019) that provides in situ ocean physics data and data products built with common standards, free of charge, and without restrictions. These services encompass a wide range of parameters, including temperature, salinity, current profiles, sea level trends, wave height and period, wind speed and direction, water turbidity (light attenuation), underwater noise, river flow, and sea-ice coverage.

https://doi.org/10.5194/sp-2024-23
Preprint. Discussion started: 27 September 2024

EMODnet Physics offers an array of in situ data collections (time-series, profiles, and datasets) obtained from various platforms (such as tide gauges, river stations, floats, buoys, gliders, drifters, and ship-based observations). EMODnet Physics does not operate platforms; instead, it integrates and federates key data infrastructures and programs. For example, it is synchronized with CMEMS INS TAC and includes supplementary in situ data from PANGAEA, the International Council for the Exploration of the Sea (ICES), the European Multidisciplinary Seafloor and Water Column Observatory (EMSO), and other Global Ocean Observation System networks (https://goosocean.org/). The data and data products are accompanied by metadata, offering users comprehensive information regarding the provenance, content, location, time, data sources, and quality check procedures.

It supports human-based data discovery (https://https://emodnet.ec.europa.eu/geoviewer/) and machine-to-machine interoperability (https://data-erddap.emodnet-physics.eu) and contributes to enhancing our understanding of the physical aspects of the marine environment. EMODnet Physics supports various applications, including scientific research, coastal management, maritime operations, and policymaking

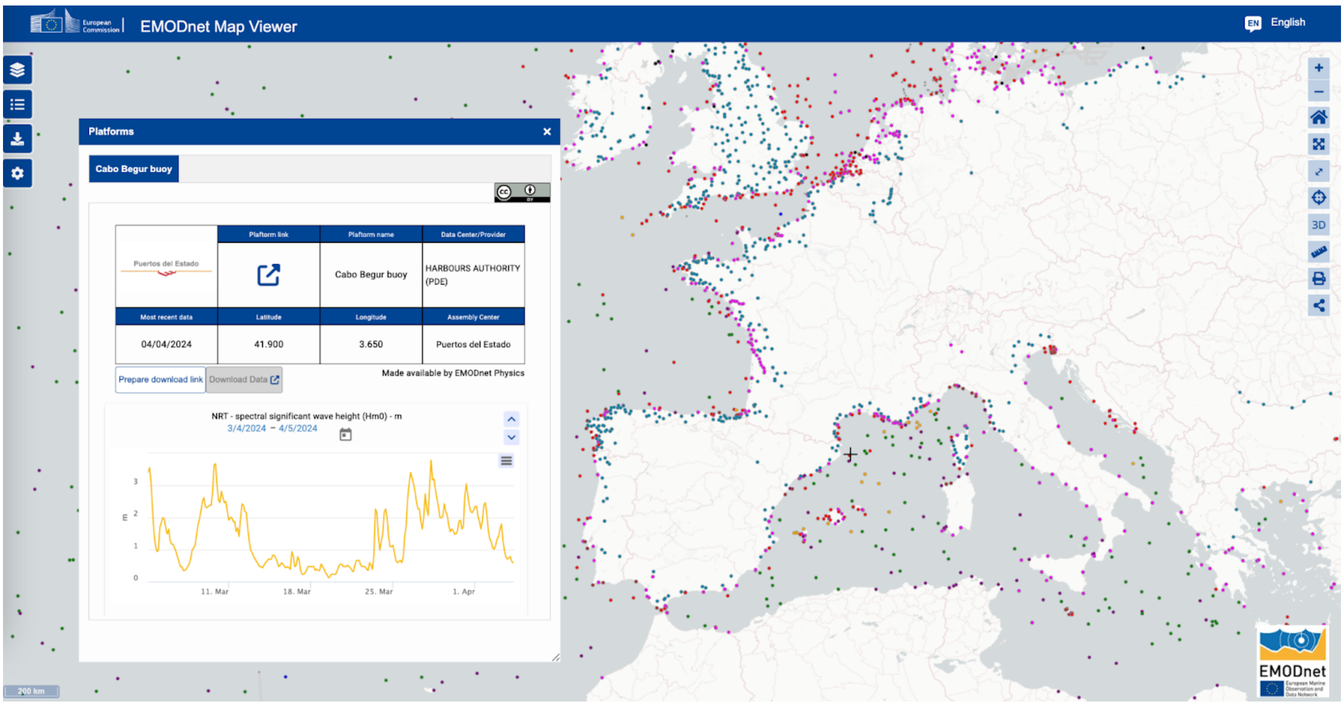

Figure 2: In-situ data discovery in EMODnet Physics. Waves Height chart.

SeaDataNet

SeaDataNet (http://www.seadatanet.org) is a Pan-European network of professional marine data centers providing data and metadata standards for the marine community, and on-line access to their data holdings of standardized quality (Schaap and Lowry, 2010). Founding partners are National Oceanographic Data Centres (NODCs), major marine research institutes, UNESCO-IOC, ICES, and EC-JRC. Over three decades, SeaDataNet has expanded its network of data centers and

https://doi.org/10.5194/sp-2024-23
Preprint. Discussion started: 27 September 2024

infrastructure in a long series of EU projects, mostly funded through EU DG RTD. SDN operates an infrastructure for managing, indexing and providing access to ocean and marine environmental data sets and data products (e.g. physical, chemical, geological, and biological properties) and for safeguarding the long term archival and stewardship of these data sets. Data are derived from many different sensors installed on research vessels, satellites and in-situ platforms that are part of various ocean and marine observing systems and research programs. A core SeaDataNet service is the Common Data Index (CDI) data discovery and access service which provides harmonized discovery and access to a large volume of marine and ocean data sets, both from research and monitoring organizations, which increasingly are major input for developing added-value services and products that serve users from government, research and industry. Currently, more than 110 data centers are connected to the CDI service from 34 countries around European seas, giving access to more than 2.5 million data sets, originating from more than 650 organizations in Europe, covering physical, geological, chemical, biological and geophysical data, and acquired in European waters and global oceans. Data resources are quality controlled and managed at distributed data centers that are interconnected by the SeaDataNet CDI system and accessible for users through an integrated portal. The data centers are mostly National Oceanographic Data Centres (NODCs) which are part of major marine research institutes that are developing and operating national marine data networks, and international organizations such as IOC/IODE and ICES. The data sets managed come from various originators, instruments, platforms and periods. This imposes strong requirements towards ensuring quality, elimination of duplicate data and overall coherence of the integrated data set. This is achieved in SeaDataNet by establishing and maintaining accurate metadata directories and data access services, as well as common standards like vocabularies, metadata formats, data exchange formats, quality control methods and quality flags.

3 Single Source Integrators

Besides these key European multi-parameter ocean data integrators, there are a number of initiatives that focus on single platforms or specific ocean variables. These initiatives concentrate on specific aspects of the marine environment, targeting a particular platform or variable for data collection and integration. Examples include projects that solely focus on buoys or floats for collecting oceanic data, or initiatives that specifically address parameters such as sea surface temperature, ocean currents, or marine biodiversity. By specializing in a single platform or variable, they can provide detailed and focused data products and services that cater to specific user needs and applications, as well as provide a simplified source for specific forecasting systems. The following Table 1 summarizes the most used ones.

Table 1: List of single data type sources for modeling community.

Data Type	Description	Source
Upper Ocean T&S	Global Temperature and Salinity Profile Programme (GTSPP)	https://www.ncei.noaa.gov/products/global-temperature-and-salinity-profile-programme

https://doi.org/10.5194/sp-2024-23
Preprint. Discussion started: 27 September 2024

Surface underway T&S	Global Ocean Surface Underway Data (GOSUD)	https://www.gosud.org/
Underway surface ocean and marine MET data (SAMOS)	Shipboard Automated Meteorological and Oceanographic System (SAMOS)	https://samos.coaps.fsu.edu/html/
Argo profiling float data	Argo Global Data Assembly Centres (GDACs)	https://argo.ucsd.edu/data/
Drifters	Global Drifter Programme (GDP) Drifter Data Assembly Centre (DAC)	https://www.aoml.noaa.gov/phod/gdp/data.php
Meteorological Moored Buoys	National Meteorological and Hydrological Services (NMHSs) operating the buoys	https://www.ndbc.noaa.gov/
Tsunameters	Tsunami Information Centre (ITIC)	http://itic.ioc-unesco.org/index.php
Tsunameters	Tsunamy Alert Device (JRC)	https://webcritech.jrc.ec.europa.eu/TAD_server/Home
Deepwater reference stations	OceanSITES stations	http://www.oceansites.org/
Surface marine observational records from ships, buoys, and other platform types	International Comprehensive Ocean-Atmosphere Data Set (ICOADS)	http://icoads.noaa.gov/
Tide gauges	Global Sea Level Observing System (GLOSS)	http://www.gloss-sealevel.org/data/
Tide gauges	IOC Sea Level Station Monitoring Facility	http://www.ioc-sealevelmonitoring.org/list.php
Tide gauges	The UHSLC - GLOSS Fast-Delivery Center	https://uhslc.soest.hawaii.edu/data/?fd
Gliders	OceanGliders (formerly EGO)	http://www.ego-network.org/dokuwiki/doku.php https://erddap.ifremer.fr/erddap/tabledap/OceanGlidersGDACTrajectories.html
HR SST data from satellites	Group for High Resolution Sea Surface Temperature (GHRSST) Regional Data Assembly Centres	https://www.ghrsst.org/
Bottle data	International Council for the Exploration of the Sea (ICES)	https://data.ices.dk/view-map

MEOP-CTD database	Over 600,000 vertical profiles (since 2004) of Temperature and Salinity collected by sea mammals	https://www.meop.net/database/meop-databases/
CLIVAR and Carbon Hydrographic Data Office	CCHDO (CLIVAR and Carbon Hydrographic Data Office) hosts vessel-based CTD and hydrographic data from GO-SHIP, WOCE, CLIVAR and other repeat hydrography programs.	https://cchdo.ucsd.edu/

4 Ways forward in ocean data integration

In advancing ocean data integration, several key strategies can propel our understanding of marine ecosystems and facilitate more informed decision-making. Shared data repositories and standardized data formats can streamline the integration process, ensuring compatibility and accessibility. Harnessing the power of emerging technologies, such as artificial intelligence and machine learning, offers opportunities to analyze vast datasets swiftly and extract meaningful insights. Implementing autonomous sensors and advanced monitoring systems enhances real-time data collection, providing a more comprehensive and dynamic picture of oceanic conditions. To follow the evolution of ocean general metocean models in terms of spatial resolution, which, in the future, will reach the kilometric scale at the global level, there is a clear need for more sensors deployed at the global, regional, and local scale. In this framework, the inclusion of cost-effective and citizen-based data collection is also a key forward-looking element. Timeliness is also an important parameter to be improved to ensure that data are available at each model run, particularly crucial for coastal applications where ocean dynamics evolve rapidly. Nevertheless, data usability/consumability strongly depends on the data policy license, and there is an increasing push for adopting the Common Creative framework and, in particular, the CC-BY license, where the only limitation is that credit must be given to the creator. Integrating these strategies collectively will not only advance ocean data integration but also contribute to the ongoing evolution of ocean general metocean models and foster a more comprehensive and accessible understanding of the marine environment.

References

Belbéoch, M., Jiang, L., Kramp, M., Krieger, M., Lizé, A., Rusciano, E., Turpin, V.: International coordination of the in situ met-ocean observing networks. 9th EuroGOOS International conference, Shom; Ifremer; EuroGOOS AISBL, May 2021, Brest, France. Available at https://hal.science/hal-03328358v1/file/EuroGOOS2021_extended_abstract_Belbeoch.pdf, 2022 (last access: 27/07/2024).

https://doi.org/10.5194/sp-2024-23
Preprint. Discussion started: 27 September 2024

Corredor, J.E.: Coastal Ocean Observing - Platforms, Sensors and Systems. Springer Cham, XIV, 159. https://doi.org/10.1007/978-3-319-78352-9, 2018.

Martín Míguez, B., Novellino, A., Vinci, M., Claus, S., Calewaert, J.B., Vallius, H., Schmitt, T., Pititto, A., Giorgetti, A., Askew, N., Iona, S., Schaap, D., Pinardi, N., Harpham, Q., Kater, B., Populus, J., She, J., Palazov, A.V., McMeel, O., Oset, P., Lear, D., Manzella, G.M.R., Gorringe, P., Simoncelli, S., Larkin, K., Holdsworth, N., Arvanitidis, C.D., Molina, J.M.E., Chaves Montero, M., Herman, P.M.J., Hernandez, F.: The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe. Frontiers in Marine Science, 6. https://doi.org/10.3389/fmars.2019.00313, 2019.

Moltmann, T., Turton, J., Zhang, H.-M., Nolan, G., Gouldman, C., Griesbauer, L., Willis, Z., Piniella, Á.M., Barrell, S., Andersson, E., Gallage, C., Charpentier, E., Belbeoch, M., Poli, P., Rea, A., Burger, E.F., Legler, D.M., Lumpkin, R., Meinig, C., O'Brien, K., Saha, K., Sutton, A., Zhang, D., and Zhang, Y.: A Global Ocean Observing System (GOOS), Delivered Through Enhanced Collaboration Across Regions, Communities, and New Technologies. Frontiers in Marine Science, 6, 291. https://doi.org/10.3389/fmars.2019.00291, 2019.

Schaap, D.M.A., and Lowry, R.K.: SeaDataNet – Pan-European infrastructure for marine and ocean data management: unified access to distributed data sets. International Journal of Digital Earth, 3(1), 50-69. https://doi.org/10.1080/17538941003660974, 2010.

Shepherd, I.: European efforts to make marine data more accessible. Ethics in Science and Environmental Politics, 18, 75-81. https://doi.org/10.3354/esep00181, 2018.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Data and/or code availability

This can also be included at a later stage, so no problem to define it for the first submission.

Authors contribution

This can also be included at a later stage, so no problem to define it for the first submission.

Acknowledgements

This can also be included at a later stage, so no problem to define it for the first submission.