STATE OF THE
PLANET

Merging and serving ocean observations: a description of marine data aggregators

Antonio Novellino[1]**, Pierre-Yves Le Traon**[2]**, and 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: Antonio Novellino (antonio.novellino@dedagroup.it)

Abstract. Observations are a fundamental component of ocean predictions: they are critical not only for monitoring the state of the ocean but also for improving forecasting systems and validating model outputs. In this context, it is essential to effectively access, manage, and integrate such information into the ocean value chain. Data providers play a pivotal role in collecting, processing, and analysing these observations, delivering comprehensive data sets that support informed decision-making and enable forecasters to enhance ocean models. This paper discusses several examples of data services, including the Copernicus Marine In Situ Thematic Assembly Centre (Copernicus Marine INS TAC), the European Marine Observation and Data Network (EMODnet), and SeaDataNet, all of which are recognized as key players in the monitoring and management of marine resources. Additionally, the paper provides an outlook on future directions for ocean data integration, emphasizing the opportunities offered by the standardization of data dissemination protocols and the role of cost-effective, citizen-based data collection.
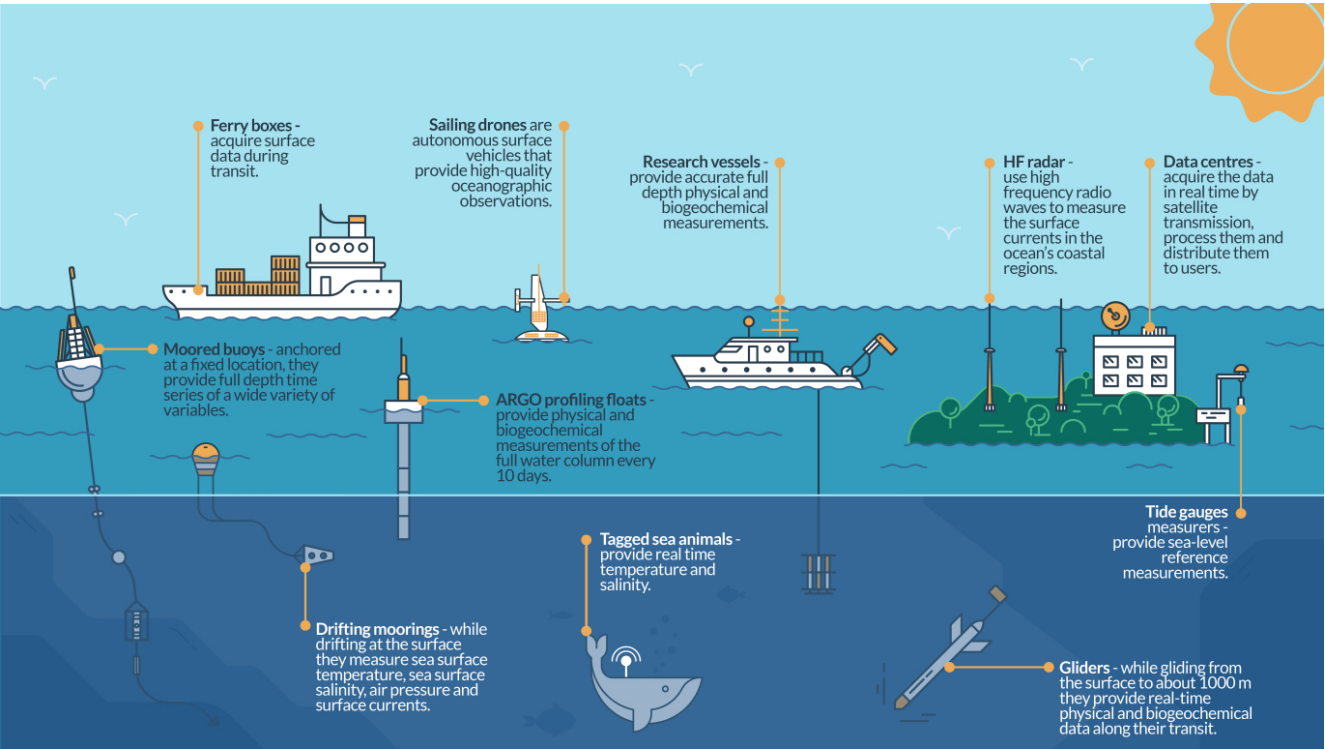
1 Introduction

The importance of ocean observation in metocean forecasting is emphasized, as it provides crucial data for understanding oceanic behaviour and coastal areas. The integration of parameters like temperature, salinity, currents, and atmospheric conditions enhances model accuracy, crucial for the effective management of human impacts and resource exploitation. The complex ocean data collection framework involves numerous in situ platforms (Fig. 1), remote sensors, and types of data, necessitating the provision of multidisciplinary, aggregated data sets (Belbéoch et al., 2022).

Marine data aggregators, also referred to as integrators, play a pivotal role in managing, integrating, and advancing the understanding of marine environments. They collect, process, and analyse diverse data types to create comprehensive data sets, contributing to informed decision-making in areas such as fisheries management, offshore energy development, and marine conservation (see e.g. Novellino et al., 2025, in
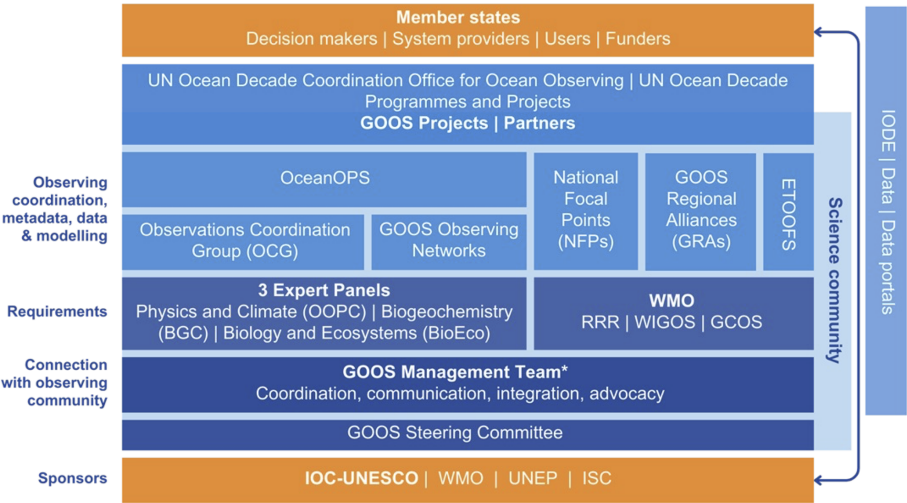
this report). Additionally, these aggregators support the development of technologies for monitoring the marine environment, continually refining data collection processes to enhance accuracy.

Over the past 3 decades, progress in marine data management has been marked by the establishment of international programmes and networks, such as the International Oceanographic Data and Information Exchange (IODE), the Global Ocean Observing System (GOOS), and the Ocean Data Information System (ODIS). These initiatives, including the World Ocean Database, involve collaborative efforts globally, led by organizations such as the Intergovernmental Oceanographic Commission (IOC), the World Meteorological Organization (WMO), the United Nations Environment Programme (UNEP), and the International Council for the Exploration of the Sea (ICES).

Under the GOOS framework (Fig. 2), the Observations Coordination Group (OCG), supported by OceanOPS (the GOOS in situ ocean observations programme support cen-

Figure 1. In situ platforms for ocean data collection (from https://marine.copernicus.eu/explainers/operational-oceanography/monitoring-forecasting/in-situ, last access: 19 March 2025).

Figure 2. GOOS framework (from https://goosocean.org/, last access: 19 March 2025).

tre) and GOOS Regional Alliances (GRAs), coordinates the GOOS observing networks to provide 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 centres (DACs) and global DACs (GDACs) play a critical role in this process by receiving, quality-controlling, and assembling data from various sources. They act as primary access points for this information, adhering to a common data format (netCDF).

Despite these efforts, GOOS networks and data represent only a subset of the overall ocean data framework. While progress has been made in modernizing the WMO data exchange system – transitioning from the Global Telecommunication System (GTS) to WIS 2.0 – by leveraging new web technologies and existing DAC/GDAC infrastructures,

full data integration between OCG networks and national/regional initiatives has yet to be achieved.

In this intricate and dispersed framework, integration services play a crucial role in harmonizing metadata, applying standardized data quality checks, and facilitating the integration of diverse data sets and models. GOOS networks, guided by the OCG data strategy (O'Brien et al., 2024), are establishing global data nodes that progressively enhance overall data delivery while maintaining "GOOS quality" within the broader ocean data lake. Furthermore, the adoption of unified controlled vocabularies, common data models, and standardized transport formats ensures the seamless integration of real-time, near-real-time (NRT), and delayed-mode (DM) observations into numerical models.

At the international level, various marine data integrators exist, and Table 1 lists the most active. Some lead the way in adopting new standards and tools, while others take the approach of following them. Europe, along with the US and Australia, is at the forefront of introducing new tools and standards. The following section outlines the European marine data integration landscape, which is shaped by three key initiatives: the Copernicus Marine Service (specifically, the In Situ Thematic Assembly Centre); the European Marine Observation and Data Network (with a focus on physics); and the SeaDataNet network of national oceanographic data centres (NODCs), affiliated with the International Oceanographic Commission.

2 European marine data integrators

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

2.1 Copernicus Marine In Situ Thematic Assembly Centre (Copernicus Marine INS TAC)

Within this programme, the Copernicus Marine INS TAC is a distributed service integrating data from different sources for operational needs in oceanography. The Copernicus Marine INS TAC integrates and quality-controls in a homogeneous manner in situ data from data providers in order to fit the needs of internal and external users. It provides access to integrated data sets of core parameters for initialization of, assimilation into, 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 Copernicus Marine is to forecast ocean state, the initial focus has been on observations from autonomous observatories at sea (e.g. floats, buoys, gliders, FerryBox systems, 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 Marine INS TAC comprises a global in situ centre and six regional in situ centres: one for each EuroGOOS Regional Operational Oceanographic System (ROOS). The INS TAC was designed to fulfil the Copernicus Marine Service and EuroGOOS ROOS needs. The focus is on essential ocean variables (EOVs) that are presently necessary for Copernicus monitoring and forecasting centres, namely temperature, salinity, sea level, current, waves, chlorophyll/fluorescence, oxygen, and nutrients. Additional atmospheric parameters (such as wind, air temperature, and air pressure) are added by some ROOSs to these regional in situ portals to fulfil additional downstream applications needs.

For near-real-time and delayed-mode products, the Copernicus Marine In Situ Thematic Assembly Centre is connected to the GOOS global networks and each Regional Operational Oceanographic System (ROOS) of EuroGOOS. In the case of DM products, it is also connected to the SeaDataNet Network, which comprises national oceanographic data centres (NODCs). The Copernicus Marine INS TAC integrates data from various observation programmes, including Argo, OceanGliders, the Data Buoy Cooperation Panel (DBCP), OceanSITES, and ship data obtained via NODCs, leveraging the GOOS network observations. Whenever possible, the Copernicus Marine INS TAC adheres to the standards developed within the SeaDataNet framework.

2.2 European Marine Observation and Data Network (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 (Schaap et al., 2022).

EMODnet Physics (https://emodnet.ec.europa.eu/en/physics, last access: 19 March 2025; Fig. 3) is the domain-specific project (Martín 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.

EMODnet Physics offers an array of in situ data collections (time series, profiles, and data sets) 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

Table 1. International marine data integrators (alphabetical order).

Region	Name	Description	Link
Africa	ODINAFRICA (Ocean Data and Information Network for Africa)	Marine data collection, capacity building, and information sharing across Africa.	http://odinafrica.org/ (last access: 19 March 2025)
Arctic Ocean	SAON (Sustaining Arctic Observing Networks)	Enhances and coordinates monitoring efforts in the Arctic region.	https://www.arcticobserving.org/ (last access: 19 March 2025)
Australia	IMOS (Integrated Marine Observing System)	Australia's national ocean observing system, providing open-access marine data.	https://portal.aodn.org.au/ (last access: 19 March 2025)
India	INCOIS (Indian National Centre for Ocean Information Services)	Provides oceanographic data, forecasts, and early warnings.	https://incois.gov.in/ (last access: 19 March 2025)
Japan	JAMSTEC (Japan Agency for Marine-Earth Science and Technology)	Conducts deep-sea research, ocean observations, and marine forecasting.	https://www.jamstec.go.jp/e/ (last access: 26 April 2025)
South America	REMARCO (Research Network of Marine-Coastal Stressors in Latin America and the Caribbean)	Latin American and Caribbean initiative for coastal and marine environmental monitoring.	http://remarco.org/ (last access: 19 March 2025)
Southern Ocean	SOOS (Southern Ocean Observing System)	A network for sustained ocean observations in the Southern Ocean.	https://www.soos.aq/ (last access: 19 March 2025)
United Kingdom	UK-IMON (UK Integrated Marine Observing Network)	The UK's coordinated ocean and coastal observing system, supporting climate monitoring, marine safety, and biodiversity research.	https://noc.ac.uk/ (last access: 19 March 2025)
United States (IOOS)	IOOS (Integrated Ocean Observing System)	US national ocean observation network providing real-time and long-term data.	https://ioos.noaa.gov/ (last access: 19 March 2025)

integrates and federates key data infrastructures and programmes. For example, it is synchronized with Copernicus Marine INS TAC and includes supplementary in situ data from PANGAEA (https://www.pangaea.de/, last access: 19 March 2025), the International Council for the Exploration of the Sea (https://www.ices.dk/data/data-portals/Pages/ocean.aspx, last access: 19 March 2025), the European Multidisciplinary Seafloor and water column Observatory (EMSO) (https://emso.eu/, last access: 19 March 2025), the SeaDataNet network of national oceanographic data centres (NODCs), and other Global Ocean Observing 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://emodnet.ec.europa.eu/geoviewer/, last access: 19 March 2025) and machine-to-machine interoperability (https://data-erddap.emodnet-physics.eu/erddap/, last access: 19 March 2025) 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.

2.3 SeaDataNet

SeaDataNet (http://www.seadatanet.org, last access: 19 March 2025) is a Pan-European network of professional marine data centres providing data and metadata standards for the marine community and online access to their data holdings of standardized quality (Schaap and Lowry, 2010). Founding partners are the national oceanographic data centres, major marine research institutes, UNESCO-IOC, ICES, and the European Commission Joint Research Centre (EC JRC). Over 3 decades, SeaDataNet has expanded its network of data centres and infrastructure in a long series of EU projects, mostly funded through EU DG RTD. SeaDataNet 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
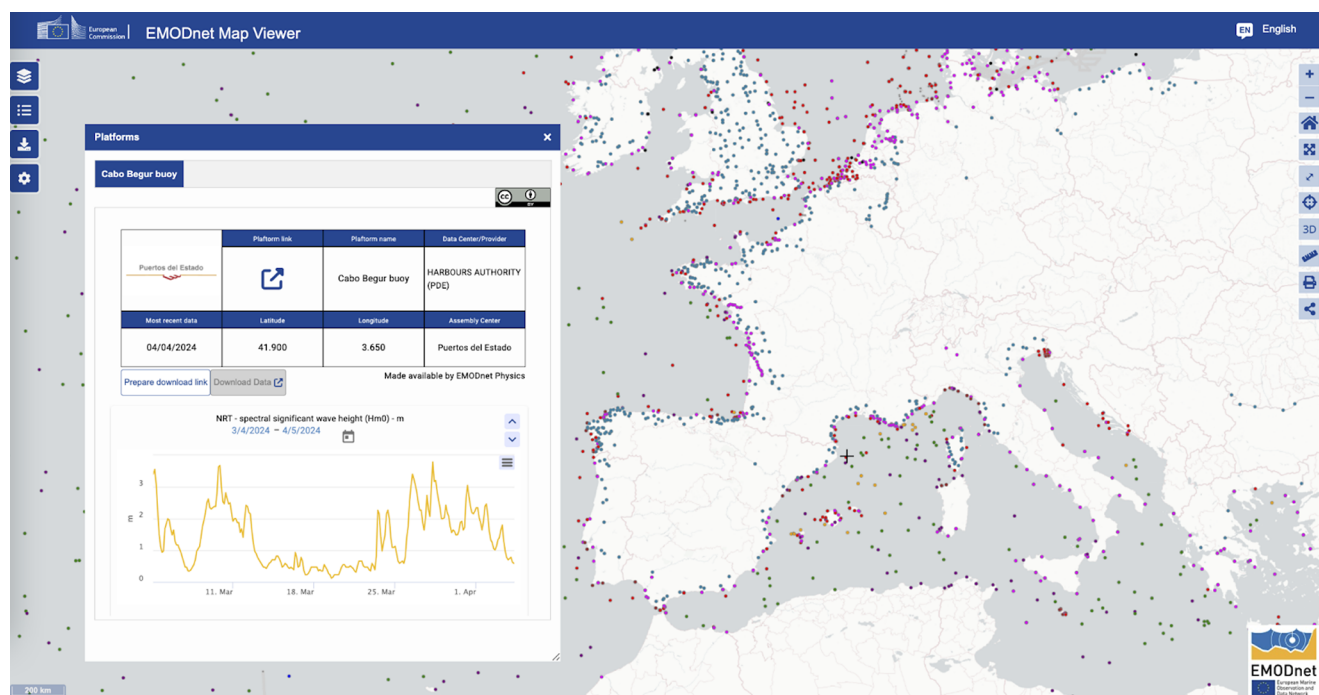
Figure 3. In situ data discovery in EMODnet Physics. Wave height chart.

of various ocean and marine observing systems and research programmes. 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. Currently, more than 110 data centres 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. 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. SeaDataNet data resources are quality-controlled and are major input for developing added-value services and products that serve users from government, research, and industry (Simoncelli et al., 2022).

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 on 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 and provide a simplified source for specific forecasting systems. The following Table 2 summarizes the most used ones.

4 Ways forward in ocean data integration

In advancing ocean data integration, several key strategies can push 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 and, more generically, fair data (Wilkinson et al., 2016). Harnessing the power of emerging technologies, such as artificial intelligence and machine learning, offers opportunities to analyse vast data sets 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

Table 2. List of single-data-type sources for the modelling community.

Data type	Description	Source	Last access
Upper-ocean T&S	Global Temperature and Salinity Profile Programme (GTSPP)	https://www.ncei.noaa.gov/products/global-temperature-and-salinity-profile-programme	September 2024
Surface underway T&S	Global Ocean Surface Underway Data (GOSUD)	https://www.gosud.org/	March 2025
Underway surface ocean and marine MET data (SAMOS)	Shipboard Automated Meteorological and Oceanographic System (SAMOS)	https://samos.coaps.fsu.edu/html/	February 2025
Argo profiling float data	Argo global data assembly centres (GDACs)	https://argo.ucsd.edu/data/	March 2025
Drifters	Global Drifter Program (GDP) Drifter Data Assembly Centre (DAC)	https://www.aoml.noaa.gov/phod/gdp/data.php	March 2025*
Meteorological moored buoys	National Meteorological and Hydrological Services (NMHSs) operating the buoys	https://www.ndbc.noaa.gov/	March 2025
Tsunameters	International Tsunami Information Centre (ITIC)	https://www.tsunami.gov/	March 2025
Tsunameters	Tsunami Alert Device (JRC)	https://webcritech.jrc.ec.europa.eu/TAD_server/Home	March 2025
Deepwater reference stations	OceanSITES stations	https://dods.ndbc.noaa.gov/oceansites/	March 2025
Surface marine observational records from ships, buoys, and other platform types	International Comprehensive Ocean-Atmosphere Data Set (ICOADS)	http://icoads.noaa.gov/	January 2025
Tide gauges	Global Sea Level Observing System (GLOSS)	http://www.gloss-sealevel.org/data/	March 2025
Tide gauges	IOC Sea Level Station Monitoring Facility	http://www.ioc-sealevelmonitoring.org/list.php	March 2025
Tide gauges	The UHSLC GLOSS Fast-Delivery Centre	https://uhslc.soest.hawaii.edu/data/?fd	March 2025
Gliders	OceanGliders (formerly EGO)	http://www.ego-network.org/dokuwiki/doku.php, https://erddap.ifremer.fr/erddap/tabledap/OceanGlidersGDACTrajectories.html	March 2025
HR SST data from satellites	Group for High-Resolution Sea Surface Temperature (GHRSST) regional data assembly centres	https://www.ghrsst.org/	March 2025
Bottle data	International Council for the Exploration of the Sea (ICES)	https://data.ices.dk/view-map	December 2023
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/	February 2024
CLIVAR and Carbon Hydrographic Data Office (CCHDO)	CCHDO hosts vessel-based CTD and hydrographic data from GO-SHIP, WOCE, CLIVAR, and other repeat hydrography programmes.	https://cchdo.ucsd.edu/	February 2024

* drifter_6hour_qc.

citizen-based data collection is also a key forward-looking step, and long-term initiatives, such EMODnet, may play a crucial role in setting up the data flow capacities for emerging networks not organized under GOOS networks.

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 licence, and there is an increasing push for adopting the Creative Commons framework and, in particular, the CC-BY licence, 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 general metocean models, including digital twins of the oceans, 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., and Turpin, V.: International coordination of the in situ met-ocean observing networks, 9th EuroGOOS International conference, Shom, Ifremer, EuroGOOS AISBL, 3–5 May 2021, Brest, France, https://hal.science/hal-03328358v1/file/EuroGOOS2021_extended_abstract_Belbeoch.pdf (last access: 27 July 2024), 2022.

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., and Hernandez, F.: The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe, Front. Mar. Sci., 6, 313, 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, Front. Mar. Sci., 6, 291, https://doi.org/10.3389/fmars.2019.00291, 2019.

Novellino, A., Arnaud, A., Schiller, A., and Wan, L.: End User Applications for Ocean Forecasting: present status description, in: Ocean prediction: present status and state of the art (OPSR), edited by: Álvarez Fanjul, E., Ciliberti, S. A., Pearlman, J., Wilmer-Becker, K., and Behera, S., Copernicus Publications, State Planet, 5-opsr, 25, https://doi.org/10.5194/sp-5-opsr-25-2025, 2025.

O'Brien, K., Heslop, E., and Belbeoch, M.: Observations Coordination Group (OCG) Data Implementation Strategy (2024), GOOS Report No. 296, https://goosocean.org/document/33970 (last access: 19 March 2025), 2024.

Schaap, D. M. A. and Lowry, R. K.: SeaDataNet – Pan-European infrastructure for marine and ocean data management: unified access to distributed data sets, Int. J. Digit. Earth, 3, 50–69, https://doi.org/10.1080/17538941003660974, 2010.

Schaap, D. M. A., Novellino, A., Fichaut, M., and Manzella, G. M. R.: Chapter Three – Data management infrastructures and their practices in Europe, in: Ocean Science Data, Elsevier, 131–193, https://doi.org/10.1016/B978-0-12-823427-3.00007-4, 2022.

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.

Simoncelli, S., Manzella, G. M. R., Storto, A., Pisano, A., Lipizer, M., Barth, A., Myroshnychenko, V., Boyer, T., Troupin, C., Coatanoan, C., Pititto, A., Schlitzer, R., Schaap, D. M. A., and Diggs, S.: Chapter Four – A collaborative framework among data producers, managers, and users, in: Ocean Science Data,

Elsevier, 197–280, https://doi.org/10.1016/B978-0-12-823427-3.00001-3, 2022.

Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., Bonino da Silva Santos, L., Bourne, P., Bouwman, J., Brookes, A., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C., Finkers, R., Gonzalez-Beltran, A., Gray, A., Groth, P., Goble, C., Grethe, J., Heringa, J., 't Hoen, P., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S., Martone, M., Mons, A., Packer, A., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3, 160018, https://doi.org/10.1038/sdata.2016.18, 2016.