ISSN 2039-6651
anno 2024 numero 80

MISCELLANEA INGV

Proceedings Volume

International Conference on Marine Data and Information Systems

IMDIS 2024 - Bergen (Norway), 27-29 May 2024

80

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

REGISTRAZIONE AL TRIBUNALE DI ROMA N.174 | 2014, 23 LUGLIO

Sede: Via di Vigna Murata, 605 | Roma

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

ISSN 2039-6651
anno 2024 numero 80

MISCELLANEA INGV

Proceedings Volume
International Conference on Marine Data and Information Systems

IMDIS 2024 - Bergen (Norway), 27-29 May 2024

Editors: Simona Simoncelli[1], Marine Vernet[2], Christine Coatanoan[2]

[1]INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy
[2]IFREMER | France

Acceped 20 November 2023 | *Accettato 20 novembre 2023*

How to cite | *Come citare* S. Simoncelli, M. Vernet, C. Coatanoan (Eds), (2024). International Conference on Marine Data and Information Systems - Proceedings Volume, IMDIS 2024 Bergen (Norway) 27-29 May 2024, Misc. INGV, 80: 1-398. https://doi.org/10.13127/misc/80

Cover Picture by MAO YUQING - Unsplash

80

Organised by

Organising Committee

Helge Sagen - IMR, Norway
Marte Louise Strømme - IMR, Norway
Veslemøy Villanger - IMR, Norway
Christine Coatanoan - IFREMER, France
Vanessa Tosello - IFREMER, France
Marine Vernet - IFREMER, France

Steering Committee

Alexander Barth – ULiège, Belgium
Taco de Bruin - NIOZ, The Netherlands
Christine Coatanoan - IFREMER, France
Jean-Noël Druon - JRC, EU
Michèle Fichaut – IFREMER, France
Lotta Fyrberg - SMHI, Sweden
Alessandra Giorgetti – OGS, Italy
Mark Hebden – BODC, United Kingdom
Neil Holdsworth – ICES, Denmark
Anastasia Iona – HCMR, Greece
Sebastian Mieruch – AWI, Germany
Volodymyr Myroshnychenko – METU-IMS, Turkey
Leda Pecci – ENEA, Italy
Lesley Rickards – BODC, United Kingdom
Helge Sagen – IMR, Norway
Dick Schaap – MARIS, The Netherlands
Serge Scory – RBINS-BMDC, Belgium
Simona Simoncelli – INGV, Italy
Vanessa Tosello - IFREMER, France
Charles Troupin – ULiège, Belgium
Lennert Tyberghein – VLIZ, Belgium

Acknowledgments. The Editors would like to thank Barbara Angioni for her valuable support and meticulous work carried out for the publication.

INDEX

Introduction

The IMDIS cycle of conferences has the aim of providing an overview of the existing information systems to serve different users in ocean science. It also shows the progresses on development of efficient: infrastructures for managing large and diverse data sets, standards, interoperable information systems, services and tools for education.
The Conference will present different systems for on-line access to data, meta-data and products, communication standards and adapted technology to ensure platforms interoperability. Sessions will focus on infrastructures, technologies and services for different users: environmental authorities, research, schools, universities, etc.

Keywords IMDIS; Marine data; Environmental

SeaDataNet – panEuropean infrastructure for marine and ocean data management – an update

Dick Schaap[1] (dick@maris.nl), Michele Fichaut[2] (m.fichaut@free.fr),
Steven Piel[2] (steven.piel@ifremer.fr), Marine Vernet[2] (marine.vernet@ifremer.fr),
Peter Thijsse[1] (peter@maris.nl), Serge Scory[3] (serge.scory@naturalsciences.be),
Alessandra Giorgetti[4] (agiorgetti@inogs.it), Anastasia Iona[5] (sissy@hnodc.hcmr.gr),
Hong Minh Le[3] (hongminh.le@naturalsciences.be), Lesley Rickards[6] (ljr@noc.ac.uk)

[1]MARIS (The Netherlands)
[2]Ifremer (France)
[3]RBINS (Belgium)
[4]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[5]HCMR (Greece)
[6]National Oceanography Centre -NOC, British Oceanographic Data Centre - BODC (UK)

Introduction
SeaDataNet is a major pan-European infrastructure for managing, indexing and providing access to marine data sets and data products, acquired by European organisations from research cruises and other observational activities in European coastal marine waters, regional seas and the global ocean. Founding partners are National Oceanographic Data Centres (NODCs), major marine research institutes, UNESCO-IOC, and ICES. The SeaDataNet network was initiated in the nineties and over time its network of data centres and infrastructure has expanded, during a successive series of dedicated EU RTD EU projects, such as SeaSearch, SeaDataNet, SeaDataNet 2, and SeaDataClod, and by engaging as core data management infrastructure and network in leading European initiatives such as the European Marine Observation and Data network (EMODnet), Copernicus Marine Service (CMS), and the European Open Science Cloud (EOSC) and closely cooperating with ocean observing communities such as EuroGOOS and Euro-Argo. These facilitated an ongoing development and evolution of the SeaDataNet technical infrastructure, standards, tools, and services, while managing and further expanding a large network of connected data centres and data providers.

SeaDataNet Data Management services
SeaDataNet develops, governs and promotes common standards, vocabularies, software tools, and services for marine data management, which are freely available from its portal and widely adopted and used by many communities and projects. The SeaDataNet network of data centres maintains and publishes a series of European directory services. These give a wealth of data and information, such as overviews of marine organisations in Europe (EDMO), their engagement in marine research projects (EDMERP), managing large datasets (EDMED), data acquisition by research vessels (CSR) and monitoring programmes (EDIOS) for the European seas and global oceans. A core SeaDataNet service is the Common Data Index (CDI) data discovery and access service which provides online unified discovery and access to vast resources of data sets, managed by 115+ connected SeaDataNet data centres from 34 countries around European seas, both from research and monitoring organisations. Currently, it gives access to more than 2.9 Million data sets, originating from 950+ organisations in Europe, covering physical, geological, chemical, biological and geophysical data, and acquired in European waters and global oceans.

SeaDataNet CDI Data Discovery and Access Service
The CDI service infrastructure consists of several components. Offline software packages

(MIKADO, NEMO, OCTOPUS) facilitate manual and machine generation of metadata entries in SeaDataNet standard profiles, based upon ISO19115-19139, and converting data sets into SeaDataNet standard data formats (ODV ASCII and NetCDF), supported by an ever increasing set of controlled vocabularies. Moreover, the separate online directories (EDMO, EDMED, EDMERP, CSR, and EDIOS) are maintained with input from NODCs for activities in their countries. Together, these contribute to marking up the metadata profiles in a semantically controlled way. Overall, this results in rich and highly FAIR metadata and data sets, useful for various applications. Other online CDI service components aim at populating the CDI metadata and harmonised related data sets into a central online CDI catalogue and a cloud based data cache. Once imported, validated, and accepted, the CDI metadata and data sets can be discovered and retrieved by users through an online shopping mechanism with dashboard for users and data centres for overseeing requests and download transactions.

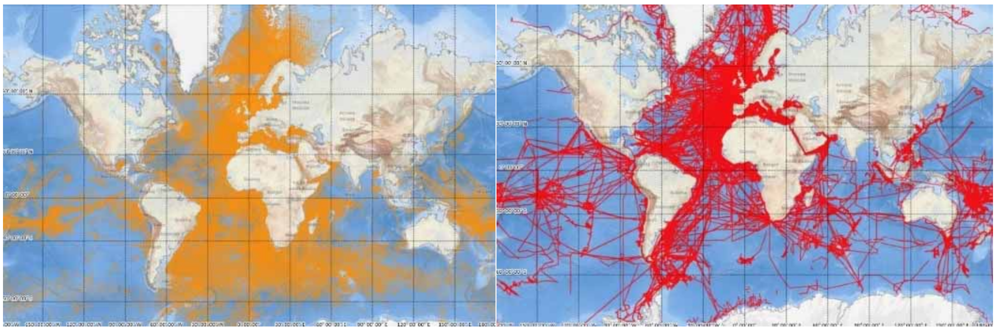


Figure 1 Coverage of data sets for point and track observations in the SeaDataNet CDI data discovery and access service.

The FAIRness of the SeaDataNet services is further amplified by running SPARQL endpoints for machine-to-machine services, using Linked Data principles, and a Swagger API for the CDI service. Use of common vocabularies in all metadatabases and data formats is an important prerequisite towards consistency, interoperability, and FAIRness.

Major pillar under EMODnet

SeaDataNet is providing core services in EMODnet Chemistry, Bathymetry, Physics, and Ingestion for bringing together and harmonizing large amounts of marine data sets from hundreds of data originators (currently >950), which are then used by EMODnet groups for generating thematic data products. Such products are for example: a Digital Terrain Model for all European seas and NW Atlantic Ocean (Bathymetry), and European aggregated and validated data collections for eutrophication, contamination, and marine litter (Chemistry). These products are very popular and find their way to many users from government, research, industry, and public. Users also pose further requirements, which leads to further optimising of the data flows from originators to data centres to SeaDataNet and to EMODnet thematics. Moreover, it implicates regular maintenance, upgrading and expanding of standards, tools, and services, also because of adding new data types, such as e.g. marine litter or vessel-mounted ADCP. Further uptake of standards and services is promoted and established, resulting in more data population, and expansion of the networks of data centres and data originators. Currently, SeaDataNet not only counts NODCs, but also many other data centres from marine research institutes, geological surveys, national hydrographic services, environmental management agencies, and companies.

Further innovation of services

Innovative developments are taking place for adopting new technologies and responding to new challenges. The European Open Science Cloud (EOSC) gives a suitable framework, promoting FAIR data services and interoperability. SeaDataNet is well engaged in several EOSC projects, such as ENVRI-FAIR, EOSC-FUTURE, Blue-Cloud pilot, and currently, in the Blue-Cloud 2026 project. The Blue-Cloud pilot deployed a versatile cyber platform with smart federation of multidisciplinary data repositories, analytical tools, and computing facilities in support of exploring and demonstrating the potential of cloud based open science for ocean sustainability. Blue-Cloud 2026 project aims for a further evolution into a Federated European Ecosystem to deliver FAIR & Open data and analytical services, instrumental for deepening research of oceans, coastal & inland waters. SeaDataNet benefits from further innovating and expanding its services. Using its large knowledge base, network, and experience, it is also a major driver towards other blue data infrastructures stimulating uptake and refinement of SeaDataNet standards, and engagement in open science and virtual research. This contributes to optimizing the overall FAIRness in the European marine data landscape. This is highly important, also in the perspective of the Digital Twins of the Oceans (DTO) initiative as part of the EU Green Deal and DestineE.

SESSION TECH ORAL PRESENTATIONS

The Semantic Analyser

Alexandra Kokkinaki[1] (alexk@noc.ac.uk), Gwenaëlle Moncoiffé[1] (gmon@noc.ac.uk),
David Habgood[2] (david@kurrawong.ai), Enrico Boldrini[3] (enrico.boldrini@cnr.it),
Nicholas Car[2] (nick@kurrawong.ai)

[1]British Oceanographic Data Centre - BODC (UK)
[2]KurrawongAI (AU)
[3]CNR (Italy)

Introduction

The EU FAIR-EASE project aims to improve Earth system, environment, and biodiversity observation and modelling through integrated services. A key component is the Data Discovery and Access service (DDAS), ensuring easy access to diverse distributed datasets. To overcome semantic differences, FAIR-EASE will integrate a semantic brokerage service into DDAS, requiring a comprehensive understanding of datasets, metadata, and employed semantics for effective mapping and alignment.
Extracting the semantic artefacts (terms originating from ontologies, controlled vocabularies, thesauri) from the above-mentioned datasets and metadata records proved to be a laborious and time-consuming task for the following reasons:

- The datasets span a wide array of domains, resulting in semantic artefacts originating from numerous, diverse sources.
- Inconsistent or lack of referencing of semantic artefacts:
  - representation as strings: terms within these datasets are typically represented as mere strings
  - some annotations are composed of two or more distinct semantic artefacts concatenated within a single string
- Typos, wrong URIs.
- Missing information.

In response to these challenges, we have undertaken the development of the Semantic Analyser (SA). The SA is capable of extracting semantic content from data files encoded in various formats (focusing on netCDF in a first instance), and from metadata records like XML ISO19115. Upon completion of the analysis, the SA provides a list of terms that have been successfully extracted, along with information about the matched semantic artefacts.

Architecture

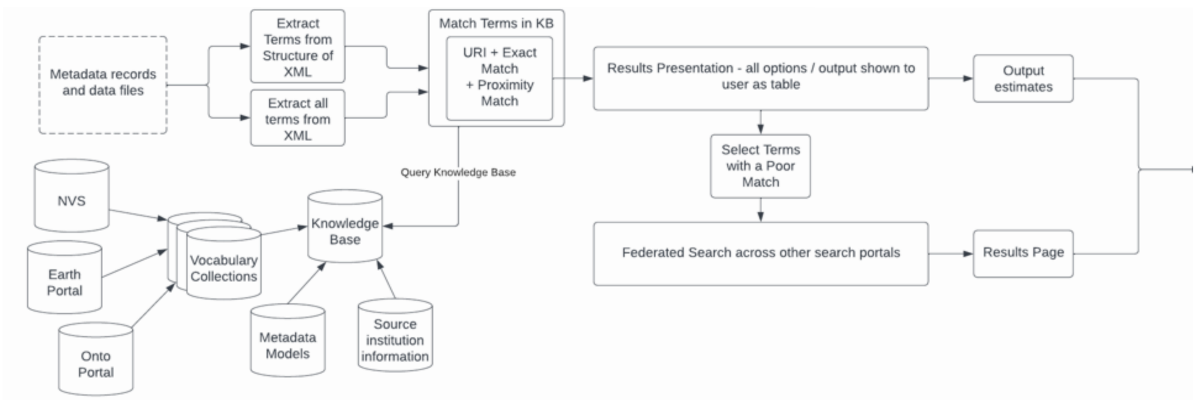


Figure 1 The architecture of the Semantic Analyser.

The input

The SA (Figure 1) is fed with a list of known data providers and research infrastructures that offer datasets. Currently input is limited to data providers that are relevant to the FAIR-EASE project. When one or more providers are selected, the associated datasets are listed. These datasets have been standardised to adhere to the ISO19115 standard, thanks to the geoDAB broker. The broker connects to the data providers web services (e.g., CSW, OpenAPI, OAI-PMH, …) and collects metadata records heterogeneously encoded, which are subsequently harmonised to meet the ISO19115 metadata model and encoded according to its XML schema. In this project, harmonisation has been applied to various aspects, including keywords, instruments, platforms, variables, and organisations.

The Knowledge Base (KB)

The KB, structured as an indexed triple store using RDF format, hosts various ontologies and SKOS vocabularies that were selected to cover the relevant domains and support interoperability between domains. Each semantic resource has been characterised in the KB as being relevant to either variable descriptions, instruments or platforms, or any combination of these. The KB will be enriched with FAIR-EASE stakeholders' contributions of domain-specific semantic resources. A strategy is being developed to keep the semantic resources updated. Additionally, the analyser performs federated searches across external semantic repositories (currently BioPortal and EarthPortal). This enables us to exploit these rich sources of terminologies without the need to bring them into the KB.

The Semantic Analyser (SA)

The SA currently analyses XML metadata records and netCDF files, offering users the choice to employ various analysis methods. Full URL extraction is used to extract all URLs from the select files and compare them against KB URLs based on exact matches only. Structured XML extraction is only applied to XML metadata records and focuses on four predefined metadata elements: keywords, instruments, platforms and variables using exact, wildcard and proximity matching techniques. Data file analysis will be employed on specific file formats starting with various flavours of netCDF files. For each technique, the SA matches text and identifiers with relevant entries in the KB. The SA can refine its searches to only match metadata elements coming from specific categories.

Conclusions and next steps

The SA is a tool that enables semantic landscaping and alignment of diverse data and metadata sources and provides insights as to which semantics each data provider is using, in order to establish semantic brokerage through mappings.
These insights will also allow us to:

- Provide an overview of semantic artefact usage practices across the network.
- Identify gaps, commonalities, and areas for alignment to enhance interoperability of tools and data access strategies.
- Issue recommendations for more efficient data flows, and achieving the "I" (Interoperable) aspect of FAIR (Findable, Accessible, Interoperable, Reusable) data principles.

Next steps include the addition of CSV files as input to the semantic analyser and the provision of a machine-readable output.

Federating discovery metadata to the Ocean InfoHub in a world of federated networks

Justin Buck[1] (justin.buck@noc.ac.uk), Charlotte Dempster[1] (charlotte.dempster@noc.ac.uk),
Thomas Gardner[1] (thogar@noc.ac.uk), Alexandra Kokkinaki[1] (alexk@noc.ac.uk),
Colm Walsh[1] (colm.walsh@noc.ac.uk), Paul Weerheim[2] (paul@maris.nl),
Jeff McKenna[3] (jmckenna@gatewaygeomatics.com)

[1]National Oceanography Centre - NOC (UK)
[2]MARIS (The Netherlands)
[3]Ocean InfoHub (Canada)

Introduction

The Ocean InfoHub (OIH) and underpinning Ocean Data and Information System (ODIS) have become key infrastructure to broker data and metadata through to enable visibility of research data within the UN Decade of Ocean Science for Sustainable Development. The integration of metadata and data from national data centres such as the British Oceanographic Data Centre (BODC) into ODIS needs to be achieved interoperably and as part of aligned activity with existing regional or discipline-focused data brokering services, whilst also leveraging the enhanced metadata exposure possible for end users in the Ocean InfoHub.

BODC are a key partner in the EU SeaDataNet (SDN)[1] and the UK Marine Environmental Data and Information Network (MEDIN) and are working on enhancing metadata and data exposure to ODIS while also leveraging and enhancing SDN and MEDIN systems to enable efficient exposure of holdings to ODIS and thus the Ocean InfoHub.

This paper presents the results of metadata modelling and a hybrid model, combining existing brokering services with additional BODC contextual metadata, of exposure of holdings to ODIS.

Metadata modelling

The ODIS system is designed to allow enhanced metadata beyond those of SDN and MEDIN with exemplar metadata records created by Nico Lange and the US NSF funded research coordination network for marine ecological timeseries in 2023. This detailed metadata model will enable significantly enhanced data discoverability and utility in the biological and ecological research communities. The resulting model using concepts from the Schema.org[2] markup language is shown in Figure 1.

Resulting hybrid model

The hybrid model for the SDN use case developed with linked metadata being served by both BODC and its brokering partners is shown in Figure 2. This model enables BODC to expose metadata enhanced beyond the existing SDN metadata models to ODIS and is readily iterated as SDN metadata models evolve with time. Such a model gives data centres the ability to rapidly meet new user requirements while facilitating the broader evolution of brokering infrastructures.

Next steps

Operational implementaiton of the model is on-going and will need to evolve with time. As data brokers such as SDN and MEDIN infrastructure evolves to enable more detailed metadata these elements will need moving out of the BODC exposure into those of MEDIN and SDN while the

[1] SeaDataNet - SeaDataNet

[2] Schema.org - Schema.org

BODC exposure can continue to be enhanced for more novel data types when required by specific communities. This model facilitates the evolution of ODIS while also feeding results into broader networks, helping the ensure communities develop together.

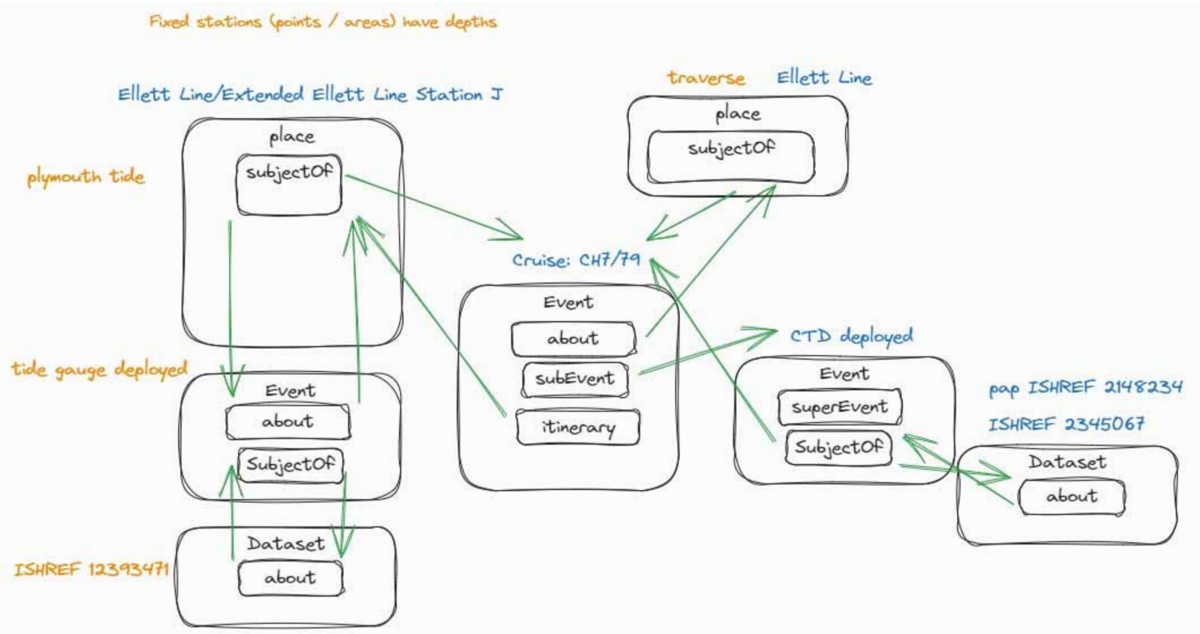


Figure 1 The model for exposure of BODC metadata data to ODIS based on Schema.org concepts the ISHREF concept is a metadata granule in the BODC internal schema and could span one oceanic CTD castor one sampling station within a research cruise, for example.

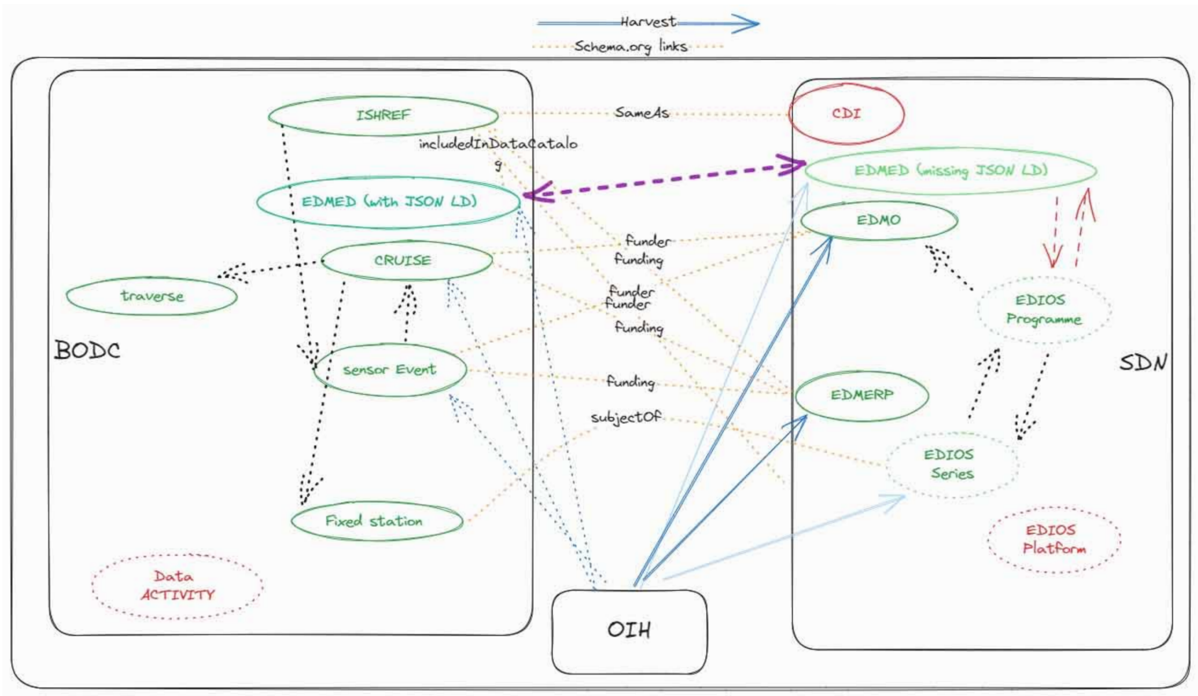


Figure 2 The hybrid model for brokering metadata to ODIS for the BODC SDN use case. The acronyms shown are all SDN concepts. Elements in red are still to be modeled while the elements are fully modeled and agreed by BODC and SDN.

The Ocean InfoHub Project and the Ocean Data and Information System: Developing a Digital Ocean Ecosystem

Lucy Scott[1] (L.Scott@unesco.org), Pier Luigi Buttigieg[2] (pier.buttigieg@awi.de), Peter Pissierssens[1] (p.pissierssens@unesco.org), Carolina Garcia-Valencia[3] (carolina.garcia@invemar.org.co)

[1]UNESCO/IOC Project Office for IODE (Belgium)
[2]Alfred Wegener Institute (Germany)
[3]INVEMAR (Colombia)

The United Nations Educational, Scientific and Cultural Organization's Intergovernmental Oceanographic Commission (UNESCO/IOC) Project Office for International Oceanographic Data and Information Exchange (IODE) has documented over 3100 online repositories of ocean data and information, which shows the highly complex online environment, and challenge of finding the right information from the right source (ODISCat 2023-10; https://catalogue.odis.org).

The Ocean InfoHub Project was designed to address this challenge through the development of the Ocean Data and Information System (ODIS). ODIS is not a new portal or centralised system under the control of a single authority, but a partnership of distributed, independent systems voluntarily sharing (meta)data and information along co-developed and clear conventions in the pursuit of common goals. These conventions are formalised and operationalised in the ODIS Architecture (ODIS-Arch) to allow existing and emerging ocean data and information systems, from any stakeholder, to interoperate with one another. This enables and accelerates more effective development and dissemination of digital technology and sharing of ocean data, information, and knowledge for sustainable development.

The Ocean InfoHub Project now supports a global network of distributed information and data resources related to the ocean. The Project has had a focus on co-design with three pilot regions in particular: Africa, Latin America and the Caribbean (LAC), and the Pacific Small Island Developing States (PSIDs), to meet their unique user community requirements. The three regional nodes facilitate that local, national, and regional digital systems and infrastructures could be interlinked to strengthen science, technology, innovation systems, and policies for the sustainable use and management of marine areas.

The Ocean Data and Information System (ODIS) Architecture links over 32 nodes from 25 partners. This is demonstrated through three regional portals and a global search portal that can now be searched to find Oceans related data and information from multiple sources (https://oceaninfohub.org). The global portal, is an openly accessible online platform and it currently contains over 130,000 content items in 8 content categories: (i) Experts (27,000); (ii) Institutions (13,000); (iii) Documents (42,000); (iv) Training (1,500); (v) Vessels (113); (vi) Projects (3,600); (vii) Datasets (48,000); and (viii) Spatial search (42,000).

The IOC Assembly at its 31st Session (2021) formally established ODIS through Decision A-31/3.4.2., ensuring its long term support as a stand-alone project. Subsequently, IODE-XXVII decided to designate ODIS as one of three IODE Programme Components, which further strengthens its long term support.

Besides its core pilot areas and themes, OIH/ODIS is now supporting additional communities of practice such as those focused on Marine Protected Areas, the GOOS (Global Ocean Observing System) Essential Ocean Variables (EOV) and Areas Beyond National Jurisdiction (ABNJ). Interoperability and strategic alignment deliberations are underway with organisations including the Group on Earth Observations Biodiversity Observation Network (GEO BON), the

Helmholtz Metadata Collaboration (HMC), the Earth Science Information Partners (ESIP), and the Polar Data Discovery Enhancement Research (POLDER) project).
The ocean digital ecosystem concept promoted through OIH/ODIS has also been adopted by the UN Decade of Ocean Science for Sustainable Development and is referred to in the Data & Information Strategy for the UN Ocean Decade. It will furthermore be promoted by the Decade Coordination Office (DCO) for Data Sharing, that was approved for establishment by the IOC Executive Council (2022) and is hosted by the IOC Project Office for IODE, Oostende, Belgium. A Programme called An Ocean Data and Information System supporting the UN Decade of Ocean Science for Sustainable Development (OceanData-2030) has been registered with the UN Decade for Ocean Science for Sustainable Development. Together, ODIS and OD-2030 will contribute to Outcome (6), “An accessible ocean” and Decade challenges 8-10, by provisioning a common digital infrastructure to support discoverability and exchange of data, information and knowledge. Challenge 9 is a particular focus, as ODIS aims to lower the barriers to accessing information and thus improve equitable access.
The overarching goal of the Ocean Data and Information System (ODIS) in the long term, is to provide a sustainable and responsive digital ecosystem where users can discover data, data products, data services, information, information products and services provided by IOC Member States, independent projects, private sector partners, and other partners associated with the UN Decade of Ocean Science for Sustainable Development.

Marine Regions: An interoperable standard for georeferenced marine place names

Salvador J. Fernández Bejarano (salvador.fernandez@vliz.be), Britt Lonneville (britt.lonneville@vliz.be), Lawrence Whatle (lawrence.whatley@vliz.be), Bart Vanhoorne (bart.vanhoorne@vliz.be), Lennert Schepers (lennert.schepers@vliz.be)

VLIZ, Flanders Marine Institute (Belgium)

Marine Regions is a global geographic database containing (i) a gazetteer of marine georeferenced place names and (ii) maritime boundaries. The Marine Regions Gazetteer includes over 60,000 places integrating national, regional as well as global marine gazetteers. In addition, the Marine Regions Gazetteer integrates the maritime boundaries dataset that has currently its 12th release [Flanders Marine Institute, 2023]. This is a digital representation of the administrative areas defined by the United Nations Convention on the Law of the Sea of 1982 such as Exclusive Economic Zones (EEZs) or Areas Beyond National Jurisdiction. The Marine Regions dataset is accessible through a web portal (marineregions.org) and an API. Machine-to-Machine interactions via its API allow integration into larger knowledge networks such as the World Registed of Marine Species (WoRMS), Global Fishing Watch and MarineTraffic. Here, we will discuss the Marine Regions back-end, the new Linked Data Model and two improved access methods based on the linked data model - Linked Data Event Streams and a new R client. Finally, as an endorsed Action of the UN Ocean Decade under the OceanData2030 programme, we explain how we want to expand our gazetteer and impact by involving various existing and new stakeholders.

The Marine Regions back-end: structure and rationale

The Marine Regions gazetteer is composed of a relational database in which geographical entities are characterized by their unique resource identifier (URI), the Marine Regions Geographic Identifier (MRGID). This identifier allows users to unambiguously refer to a Marine Regions entity and is persistent and resolvable at <http://marineregions.org/mrgid/{id}>. Geographical entities are defined both by a geometry and a place type. While the spatial features are represented as different vector data types, for example, point, multipoint, line, or polygon, a place type provides contextual information on the geographic entities, for example, a sea, bay, ridge, sandbank, or oceanic trench. Not only physical place types are considered but also administrative place types, such as countries, EEZs, territorial seas, fishing zones, or marine protected areas. In addition, one or more names can be stored for every gazetteer entity, allowing us to deal with synonyms or names in different languages. The gazetteer also provides a hierarchy between the different entities, based on a parent-child relation (partOf). Other relations to Marine Regions entities can be described in the Marine Regions gazetteer, these can cover topological (adjacentTo, streamsThrough, ...) and non-topological (administrativePartOf, influencedBy, ...) links.

Until recently, these data were accessible through the Marine Regions website, through OGC web services and through a JSON HTTP API. In order to achieve semantic interoperability, we aligned the Web APIs using a common Linked Data model [Lonneville B. et al., 2021].

A new Marine Regions Linked Data model

In this section, we explain the core aspects of the Linked Data Model and elucidate certain encountered challenges. Gazetteer entities are naturally both a description of a place and the location of the place itself. Consequently, each geographical entity becomes both a

<skos:Concept> and a <dct:Location>. The whole Marine Regions dataset is a SKOS concept scheme with the geospatial extent defined as instances of the OGC's SimpleFeatures vocabulary <sf:geometry>. The relationships offered by traditional SKOS predicates such as <skos:narrower> cannot accurately capture the hierarchical nature of the Marine Regions Gazetteer: consider one gazetteer entity that may be spatially contained by another, but only partially overlaps with a third one. To address these relations, we defined <mr:isRelatedTo> along with nine subproperties (e.g., <mr:isPartOf>) and their respective inverses (e.g., <mr:contains>). They are detailed in the Marine Regions ontology at <http://marineregions.org/ns/ontology>.
The documents describing each URI (e.g., <http://marineregions.org/mrgid/5686.ttl>) are currently available through content negotiation in two RDF serializations, Turtle and JSON-LD, together with existing JSON, XML and HTML options. Nevertheless the model presented a performace challenge: How to efficiently handle documents containing large data, including those with detailed geometries or with numerous relationships? Our solution is to provide geometries with an additional link within the document, thereby enabling data consumers to retrieve geometries on demand. In cases where the number of relationships is extensive, the document incorporates the <hydra:next> predicate, directing users to subsequent pages that each contain a maximum of 40 relations.

Retrieving Marine Regions using Linked Data Event Streams
The Marine Regions semantic model provides the basis for replicating the database via Linked Data Event Stream (LDES). It allows data consumers to capture changed entities as immutable objects with unique URIs [Van Lancker et al. 2021]. The pagination strategy is crucial in LDES. Here we opted for fragmenting using <dc:modified> timestamps, allowing to retrieve the full change history from one single feed endpoint at <http://marineregions.org/feed>. This method simplifies the process of keeping up with updates, enabling users to set up services like (Geo)SPARQL endpoints or document stores for full-text searches. The LDES, with its layered architecture, enhances accessibility and flexibility, making it easier for third parties to integrate and utilize the Marine Regions data in various applications.

Bringing the Marine Regions API into an R client: mregions2
As an example of the possibilities of this semantic API we developed mregions2, a client written in the R programming language. mregions2 offers a streamlined interface to access data from Marine Regions in R for researchers, marine scientists, and geospatial analysts seeking marine geographical information. It follows the same rationale behind the API as calls to the Marine Regions gazetteer do not provide geometries or relationships by default. Instead, these can be accessed on demand as the data consumer may need. mregions2 has been peer-reviewed by rOpenSci and will be published on CRAN, ensuring that Marine Regions data is available programatically for R users. Installation instructions are available at:
https://github.com/lifewatch/mregions2

References
Flanders Marine Institute, (2023). *Maritime Boundaries Geodatabase, version 12*. Available online at https://www.marineregions.org/ - https://doi.org/10.14284/628
Lonneville B. et al., (2021). *Publishing the Marine Regions Gazetteer as a Linked Data Event Stream*. S4BioDiv 2021. http://ceur-ws.org/Vol-2969/paper8-s4biodiv.pdf
Van Lancker D. et al., (2021). *Publishing base registries as linked data event streams*, in: Proceedings of the 21th International Conference on Web Engineering.

OGC API - Connected Systems

Christian Autermann (c.autermann@52north.org), Simon Jirka (jirka@52north.org)

52°North GmbH (Germany)

The OGC API family of standards

The Open Geospatial Consortium (OGC) API family of standards represents a suite of specifications designed to enable seamless and interoperable access to geospatial data and services on the web. These standards are developed and maintained by the OGC, a global organization dedicated to advancing the interoperability of geospatial technologies. The OGC API family builds upon the success of earlier OGC standards and embraces a more modern and web-centric approach to geospatial data exchange. At its core, the OGC API family focuses on providing standardized and consistent interfaces for different aspects of geospatial information, ranging from simple feature access to more complex capabilities such as data processing.

Previous web services standardized by the OGC relied for the most part on XML-RPC (eXtensible Markup Language Remote Procedure Call) where XML encoded requests are submitted via HTTP POST to a single endpoint. OGC APIs are designed to be resource-oriented, making use of standard HTTP mechanisms such as content negotiation or using different HTTP methods like GET, POST, PUT and DELETE to interact with geospatial resources. This RESTful architecture enhances simplicity, discoverability, and scalability, making it easier for developers to integrate geospatial data and services into their applications. OGC APIs are formalized using the OpenAPI specification, a programming language-agnostic interface description standard for HTTP APIs that allows both human and machine-readable discovery of service capabilities. In contrast to older, XML-based OGC standards, the preferred content encoding was shifted towards JSON.

One of the notable aspects of the OGC API family is its emphasis on simplicity, flexibility, and scalability. By adopting common web protocols and leveraging RESTful principles, OGC APIs make it easier for developers to consume and produce geospatial data across different platforms and programming languages. This not only promotes interoperability but also facilitates the creation of dynamic and interactive geospatial applications that can seamlessly integrate with diverse data sources.

The OGC API family comprises several individual standards, each addressing specific use cases and functionalities and should be considered building blocks for creating geospatial APIs. Some key members of the OGC API family include OGC API - Features (OAPI-F) replacing the OGC Web Feature Service (WFS), OGC API - Maps (OAPI-M), replacing the OGC Web Mapping Service (WMS), and the OGC API - Coverages (OAPI-C), replacing the OGC Web Coverage Service (WCS). These standards collectively cover a wide range of geospatial data types, from vector features to maps and coverages, providing a comprehensive framework for building interoperable geospatial web applications.

Notably the OGC APIs did not include a true successor to SensorWeb standards such as the OGC Sensor Observation Service (SOS) or OGC Sensor Planning Service (SPS). The popular OGC SensorThings API predates the OGC API initiative and is based on a slightly different architecture. It uses a variant of the Open Data Protocol (OData) and is as such not truly aligned with the OGC API paradigms. Although there are efforts to mend this in upcoming versions of the standards, the focus of the API are Internet of Things (IoT) applications and it lacks ways to convey complex and detailed sensor metadata descriptions that are required in research communities.

OGC API - Connected Systems

The OGC Connected Systems standards working group (SWG) seeks to fill this gap. In the last year the SWG worked on a first draft of a specification that allows the management of descriptions of sensor systems and networks and their data. It builds upon existing and proven conceptual models such as OGC SWE Common Data Model Encoding (SWE Common), the OGC Sensor Model Language (SensorML) and OGC/ISO Observations, measurements and samples (OMS, formerly known as Observation & Measurements, O&M). The SWG worked on creating JSON encodings for SensorML and SWE common, aligning them with new iterations of related standards such as the OGC/W3C Semantic Sensor Network Ontology (SOSA/SSN) and OMS and introduced new concepts that were missing in previous versions of the standard such as a dedicated class to describe deployments of sensors at particular place and time.
This preparatory work was needed for the introduction of the OGC API - Connected Systems, a draft specification aligned with the OGC API family of standards that comes in two parts.
Part 1 is an extension of the OGC API - Features standard that allows the management of static resources. It allows the management of systems, procedures, deployments, sampling features and observed properties. Systems are entities that can produce data and/or receive commands such as sensors and sensor networks, platforms, actuators, processing components, etc. Observing systems produce observations for one or more features of interest. Procedures furnish details about the processes undertaken by a system to fulfill its designated tasks. In the case of hardware equipment, this typically aligns with the device's datasheet. In scenarios involving human sensing or sampling tasks, the procedure delineates the requisite steps or methodology that operators must adhere to. Deployments offer a way to describe where and when a system was deployed, sampling features allow the management of real-world objects that are observed by systems and properties allow the integrated management of derived properties, e.g., more narrower definitions of observed properties, that are referencing properties in a controlled vocabulary. The API allows the implementer and user to choose between different data formats for these entities. This way rich and complex metadata descriptions in SensorML can be used while at the same time a sensor can be represented by a simple GeoJSON feature. All these resources can be grouped logically by the service provider into collections according to any criteria.
While it is supported to link to data in a more suitable API e.g., OAPI-C, Part 2 of the draft specification addresses the management of dynamic data, that is datastreams of observations produced by systems, control channels allowing commands to be sent to systems and the history of systems in the form of historical descriptions and events related to a system. Datastreams enable the retrieval of observations generated by systems in various formats. It allows the flexible grouping of observations, e.g., by sensor network, observed property, measurements campaign or cruise. While the standard format for observations in datastreams is OMS or SWE Common, the draft standard allows other encodings. For example, it is possible to publish observations as Protocol Buffers (protobuf), an efficient binary format, and include the protobuf schema in the datastream description. Control channels allow sending commands to systems to change their state or for example initiate a measurement. API clients can act both as sender or receiver of commands and observations. In previous standards events related to a system were part of the system's description. In cases where hundreds of events related to a system are produced, e.g., the continuous lowering of a sensor into the water and its subsequent recovery, this could lead to a system description that could no longer be handled. System events are now handled as a special datastream of a system. The integration of publish/subscribe patterns via MQTT for data streams, control channels and events is foreseen.

Summary

The OGC API - Connected Systems presented in this contribution seeks to be a successor to established SensorWeb technologies aligned with the OGC API family of standards, it addresses shortcomings and lessons learned of the past decade and allows data managers and scientists a modern way to exchange observational data and metadata. It offers a great degree of freedom in how data is structured and through the flexible choice of data formats it is suitable for embedded systems and metadata-heavy research topics alike.

Workflows for marine metadata and data management

Julien Barde[1] (julien.barde@ird.fr), Emmanuel Blondel[2] (emmanuel.blondel@fao.org),
Bastien Grasset[3] (bastien.grasset@ird.fr)

[1]MARBEC, Univ. Montpellier, CNRS, Ifremer, IRD, Montpellier (France)
[2]UN-FAO, Fisheries and Aquaculture Division (Italy)
[3]MARBEC, Univ. Montpellier, CNRS, Ifremer, IRD, Réunion (France)

Interest of workflows to simplify metadata creation and publication processes

Data managers regularly face the need to manually replicate the same sequence of steps to ensure (meta)data edition, update and publication in different infrastructures. For example fisheries datasets are often time series of spatial catches with biological observations and, as such, can be disseminated in spatial data infrastructures but are also valuable species occurrences for biodiversity research which can ultimately be assigned a DOI in various data repositories (e.g., GBIF, Zenodo, Pangaea, …). Obviously, with the increasing volume of data management tasks, there is no reason to waste time by manually editing the same metadata elements (e.g., Title, abstract, spatial or temporal extent) in different infrastructures complying with different standards. Multiple tasks can indeed be automated through APIs and orchestrated as steps of workflows which can be fed by various sources:

- read and write: meta-data complying with various standards (e.g., OGC, CF conventions, EML);
- transform: metadata standards mapping, data structure harmonization;
- load: metadata and data publication in different tools (e.g., GeoNetwork, GeoServer);
- long term storage and versioning: DOIs assignment on DataCite (e.g., Zenodo data repository).

A method for editing metadata through basic tools

Technology, standards and software regularly evolve over the years and so does the ergonomics of their GUIs. Moreover, the method used to implement rich standards (e.g., ISO 19115) is a key issue to facilitate the metadata editing process which is often too heavy for users. Indeed, once authenticated, metadata producers often face complex forms that are time-consuming to fill. Instead we suggest an alternative method by using basic but robust data formats and tools which can be handled by anybody. Rather than sophisticated GUIs, we use a simple tabular format (e.g., CSV) made of columns named after well known metadata elements. Each line of the table thus describes a dataset and part of the cells values can be automated by browsing the data (e.g., "spatial coverage" column) when others (e.g., "Title" column) have still to be filled manually. Metadata editing is then much easier and efficient as multiple datasets can be described within a single table by using basic desktop (e.g., QGIS, OpenOffice, MS Excel …) or collaborative applications (e.g., nextcloud, google drive …).

A pivot model to comply with differents standards

As explained before, we promote the use of a single spreadsheet to edit metadata of multiple datasets at once. Each column of the table is named after a metadata element and the data structure is built to store metadata from standards which are relevant in the marine domain:

- Dublin Core (DCMI): domain agnostic standard providing a list of key metadata elements.
- OGC: for spatial metadata formats (19115, 19119, 19139, 19110) and protocols (CSW).
- EML and Darwin Core: for biodiversity and ecological domains.
- Datacite: for global and cross domain dissemination and long term access through DOIs.

As these standards overlap, we have built a simple pivot metadata model based on: 1) 15 Dublin Core metadata elements which are used as columns names in our tabular format, 2) additional metadata elements from the other standards which are optional and directly named in the cells of the columns by using simple syntactic rules.

Implementing workflows with R programming language

At first, our main purpose was to implement workflow complying with OGC standards which are widely used in the marine domain, just like R software. However, until recently, only a few programming languages provided expected libraries to handle spatial metadata and data publication within spatial data infrastructures (mainly Java and Python). There was thus a gap to be filled to facilitate metadata management directly with R language. This have been achieved by a set of recent R libraries (geometa, ows4R, geonapi, geosapi, zen4R) with which usual data management steps can be handled : 1) reading and writing metadata, 2) publishing metadata, 3) publishing data, 4) assigning DOIs. Finally, the geoflow R library builds workflows on top of these libraries to automate these steps.

Examples of marine metadata and data management with R workflows

Over the years such R workflows have been improved to implement FAIR data management plans for fisheries data in the context of previous FP7 iMarine, H2020 BlueBridge and BlueCloud, and current HORIZON BlueCloud 2026 european research projects.
The Figure 1 illustrates a workflow which is hosted in a Virtual Research Environment (VRE) provided by the BlueCloud 2026 project and deployed in the “Global Fisheries Atlas virtual laboratory”. The R workflow is entirely driven by metadata and feeds widely used SDI software components which provide FAIR services. Metadata, edited and stored in basic spreadsheets, are published by using standards (e.g., CSW-T) or native software APIs (e.g., GeoNetwork, GeoServer, Zenodo ...). This makes data: 1) findable with discovery metadata (complying with DCMI, ISO/OGC 19115, Datacite), 2) accessible with standard formats (e.g., GML, GeoJSON) and access protocols (e.g., CSW/WMS/WFS/WCS), 3) reusable with usage metadata (e.g., data dictionary complying with ISO 19110 - Feature Catalog).

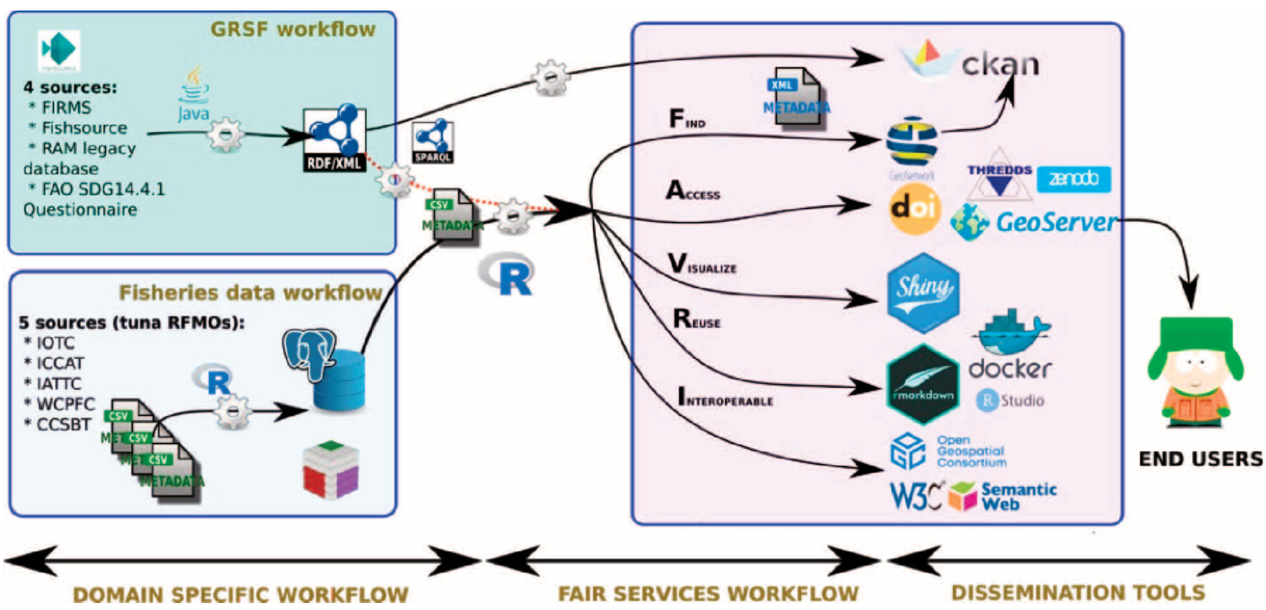


Figure 1 Example of R workflow to provide FAIR services for fisheries datasets.

The FIRMS Global Tuna Atlas workflow orchestrates similar steps: 1) metadata are edited and stored in a Postgres database along with harmonized data complying with CWP standard for fisheries (gridded catches and fishing efforts), 2) metadata are extracted, transformed and published with data in both FAO GeoNetwork and GeoServer as well as in Zenodo to assign DOIs. Similar workflows are used to manage global geospatial and statistical fisheries datasets through the FAO Fisheries & Aquaculture GeoInfo portal, or fisheries regional databases to comply with Data Collection Reference Frameworks (DCRF) such as the WECAFIS in the Caribbean.
Beyond fisheries data, the same method has been successfully tested and validated with other kinds of marine data. In particular, data managed with NetCDF data formats can be directly accessed (local or OPeNDAP access) to extract metadata stored in the header, transform CF metadata into other metadata standards (e.g., OGC 19115 mapping) and publish them in SDI software components.

Using Accreditations to navigate the waves of marine data management

Andrew Conway[1] (andrew.conway@marine.ie), Ramona Carr[1] (ramonda.carr@marine.ie), Sarah Flynn[2] (sarah.flynn@marine.ie), Tara Keena[1] (tara.keena@marine.ie), Rob Thomas[2] (rob.thomas@marine.ie), David Currie[1] (david.currie@marine.ie), Eoin O'Grady[1] (eoin.ogrady@marine.ie)

[1]Marine Institute (Ireland)
[2]Compass Informatics (Ireland)

Background

The Marine Institute (MI) of Ireland is the State agency responsible for marine research, technology development and innovation, providing scientific and technical advice to Government to help inform policy and to support the sustainable development of Ireland's marine resources. As the National Oceanographic Data Centre (NODC) of Ireland, the MI has played a leading role in marine data management and availability for over 30 years, operating a number of data acquisition platforms and programmes which underpin national marine services. In addition to operational activity the MI is a research performing organisation and a marine research funder in Ireland.

To further support data management processes, and to demonstrate the institute's capabilities to provide data and services in compliance with established standards and responsibilities, there has been a strong focus over the past five years on relevant accreditations, including IOC-IODE Quality Management Framework and most recently CoreTrustSeal.

Accreditation as Ireland's National Oceanographic Data Centre

In 2019, the Marine Institute in its remit as Ireland's National Oceanographic Data Centre, received the international accreditation of its Data Management Quality Management Framework by the (UNESCO) International Oceanographic Commission's IODE programme [Leadbetter et al., 2020].

The Data Management Quality Management Framework (DM-QMF) was developed to guide data management activities from collection, storage, quality control, and analysis to data outputs from minting DOIs to legislative reporting. The real-world experience of the Marine Institute in developing a DM-QMF was presented at IMDIS 2021 [Reed and Thomas, 2021].

Since this time, maintaining this accreditation has been a priority, and there has been a focus on the further development of the Institute's DM-QMF in its organisational strategy "Ocean Knowledge that Informs and Inspires" (Marine Institute 2023). In particular, this is identified as a key support for the wider digitalisation of scientific workflows, and the adoption and use of advanced data processes and analytical tools. In March 2023, the Marine Institute has been successful in achieving re-accreditation.

DM-QMF: Data Process Quality Assurance underpinning National Programmes

While being very active in marine data acquisition, including through the national offshore research vessels, the Marine Institute has long acted as a broker for a range of data from partner organisations including research data. These data sets are important contributions to national marine programmes, such as the National Marine Planning Framework, the Ocean Renewable Energy Development Plan and for marine licensing services.

The development of the framework to achieve DM-QMF accreditation have been extensive, with 121 data processes now managed using it. These processes are regularly reviewed and improved to optimise how the processes operate and the resulting access and usability of the data as well as the reliability of the underlying services. The quality assurance of these data processes provides a solid foundation for the evidence-based that underpins national marine decision making.

Next Steps: Managing Data including from 3rd Parties and Research Projects

In extending the remit to discrete datasets and 3rd party data, such as research project data, in addition to data delivered through operational programmes, the Marine Institute explored complimentary accreditations to support data management activities. Some of these datasets are data which has been previously collected and is being repurposed or data which are the result of a discrete scientific activity to answer research questions with a defined time-period and resources.

When managing discrete data, particularly from an external organisation, there is a greater requirement for appropriate documentation to ensure the data are available for reuse. The ability to recover undocumented knowledge and data outputs become increasingly difficult with time. To ensure these discreet data sets are available for future reuse, a specific DM-QMF pack was developed and piloted for data processes resulting from these types of activities.

Following a full analysis of internal digital preservation processes, an application was submitted for CoreTrustSeal accreditation in October 2022. CoreTrustSeal is another framework (aligned to ISO 16363) complimentary to the existing DM-QMF (aligned to ISO 9001) to manage and preserve data over time. All data archival processes follow the Open Archival Information System (OAIS) model.

Conclusion

The benefits of embracing these two Data Management Accreditations are most evident in improving the quality and transparency of internal data processes and how 3rd party data are managed. They increase awareness of and compliance with established standards, building stakeholders confidence and demonstrate that the Marine Institute are following internally recognised best practices. They also offer a benchmark for comparisons to help identify opportunities for improvement.

References

Leadbetter A., Carr R., Flynn S. et al., (2020). *Implementation of a Data Management Quality Management Framework at the Marine Institute*, Ireland. Earth Sci Inform 13, 509–521, https://doi.org/10.1007/s12145-019-00432-w

Reed G., Thomas R., (2021). *Delivering Quality Marine Data and Services: the IODE Quality Management Framework*. IMDIS - Bollettino di Geofisica, Vol.62 – Supplement (2021) pages 220-221.

Streamlining Data Submission: A Guide to Ensuring Data Quality and Compliance with ICES Data Screening Utility (DATSU)

Laura Andreea Petre (laura.andreea.petre@ices.dk), Carlos Pinto (carlos@ices.dk), Anna Osypchuk (anna.osypchuk@ices.dk), Colin Millar (colin.millar@ices.dk)

ICES (Denmark)

Introduction

ICES Data Screening Utility (DATSU) is a powerful tool designed to streamline and enhance the validation process for files before they are uploaded to ICES thematic databases. DATSU is based on an SQL database (DATSU DB), supported by an intuitive user interface (DATSU UI) allowing the Data Centre data managers to work with data formats and SQL-based quality checks. Data submission and validation relies on DATSU web services to verify that files adhere to the specified format, including:

- Ensuring the files conform to the record structure, mandatory records and fields reported.
- Validating the data integrity of each field and the datatype reported.
- Verifying reported field values against associated vocabularies or ranges, when applicable.
- Allowing flexible implementation of data quality checks specific to a given data type.
- Communicating the data screening results by providing the data submitter with a screening session report, which includes errors, warnings and information messages.

History

- DATSU was first developed in the early 2000s and its main components were:
- DATSU DB - an SQL database, containing format definitions and uploaded data.
- DATSU management user interface (DATSU UI) - an ACCESS-based solution used by the data managers to define the format and quality checks.
- DATSU VB6 DLL which used DATSU DB and was registered in IIS on a web server allowing the screening tool to be called from a webpage each time a file was uploaded.

In 2007, the system was extended to enable data managers and programmers to define SQL rules to be applied to a format, and then their results included in the screening report.

In 2016, DATSU was reprogrammed to run as .NET web services, allowing both synchronous and asynchronous runs, improving service management. Asynchronous runs were implemented using a Hangfire service call with a queue to enhance performance. This release also introduced the ability to accept text files, smaller and with predefined column order, or XML files, larger but more readable.

In 2019, an R library (icesDATSU, https://CRAN.R-project.org/package=icesDatsu) that interacts with the web services was implemented. This was done to support experts familiar with R to incorporate the DATSU checks into their data pre-processing workflow, thus simplifying, and improving the data submission process through local checking of files on their own computer prior to submission.

In 2023, the DATSU UI, previously ACCESS-based, was moved to a web application developed using C# .NET and the Blazor framework, providing improved format management. For this new app the focus was on implementing the necessary format management functionality, while providing a more user-friendly and intuitive experience, as well as improved validations when changings the format. One improvement brought in this version is synchronizing the codes used in the DATSU database with the ICES Vocabulary Server automatically using the Vocab Web

API. As part of this revision process, some of the redundant features, like species lists and reports, were left out of the new version.

Data validation process

The data validation process for a new dataset begins with defining a format, which is then saved in the DATSU database. Data managers use the management interface to define the format, records, and fields within those records. Data managers specify the data type and maximum width of each field, and whether it is mandatory or optional. String fields can be associated with vocabularies, which are part of the ICES data governance framework and are defined in ICES Vocabulary Server.

Once the format is defined, data submitters have three options for checking their files:

- They can use the DATSU webpage, where they select the format and upload a file.
- They can directly call the DATSU webservices, preferably the asynchronous option.
- They can install the R library for DATSU and use it to verify that the file conforms to the format and a subset of the SQL checks before proceeding with one of the previous two methods.

Benefits of using DATSU

DATSU (https://datsu.ices.dk/web/index.aspx) exposes format information, data types, and related vocabularies both online and via the web service, making it easily accessible for submitters. The web service allows data submitters and other organizations to run the screening and query DATSU, using various programming languages and scripts, providing a certain freedom for developers and autonomy for data submitters.

DATSU generates a comprehensive report of the file screening process and results, which is accessible through both web pages and services. When displaying quality checks information, DATSU also shows the source code for each check, often in SQL. DATSU R package (available in CRAN) allows a fast start by running a part of DATSU checks locally.

Issues and limitations

The DATSU screening queue can be 'blocked' by large files causing a delay for subsequent uploads.

DATSU is heavily based on shared functionality and definitions for the format which can be both an advantage and a disadvantage.

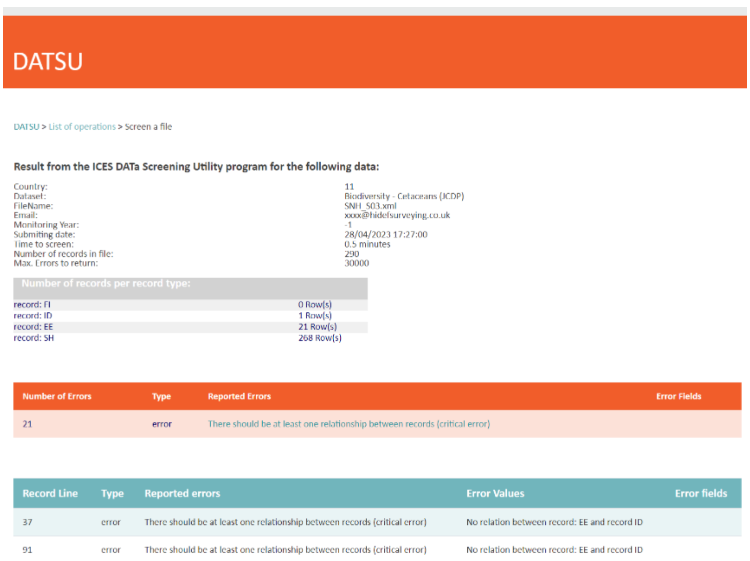


Figure 1 DATSU report of a file screening.

Future developments

- Allow the fields to be reported in another order in the text files if the user reports the headings.
- Accept JSON file format in submissions.
- Add logging for each change made using the DATSU UI by data managers.

Interoperable Management and Provision of Metrological Data in Ocean Sciences

Simon Jirka[1] (jirka@52north.org), Christian Autermann[1] (c.autermann@52north.org), Joaquin Del Rio Fernandez[2] (joaquin.del.rio@upc.edu), Enoc Martínez[2] (enoc.martinez@upc.edu)

[1]52°North GmbH (Germany)
[2]Universitat Politècnica de Catalunya (Spain)

Introduction

The provision of metrological data (e.g., calibration data, information about the precision of instrument) is an important foundation to enable the correct processing and interpretation of sensor measurements. An example is the densification of professional observation networks with low-cost sensing devices. Such additional low-cost sensors can help to close gaps in the geographical coverage of the network. However, at the same time data such low-cost devices would deliver their data with a higher level of uncertainty. To handle such heterogeneities in data quality, it is important that those properties and characteristics of a sensor, which influence the quality of the measured data, are well described. This may include information about calibration processes, quality and precision of sensing devices, applied measurement processes, and any kind of data processing that was conducted.

This paper introduces the work conducted in the European Horizon 2020 project MINKE (Metrology for Integrated Marine Management and Knowledge-Transfer Network) to address this challenge. Based on a comprehensive requirements analysis via interviews and dedicated questionnaires, a review of relevant international standards was carried out. Furthermore, an approach has been developed, how such metrological information can be efficiently shared via large scale research data infrastructures such as the European Open Science Cloud (EOSC).

Relevant international standards

The first step in the development of a management approach for metrological data was a review of relevant international standards. This has led to a set of key findings which are summarised in the following paragraphs.

The OGC Sensor Model Language (SensorML) offers a comprehensive approach to model and encode sensor metadata, including calibration and tests. However, the current XML-based approach is rather complex. Thus, we recommend to especially consider a SensorML JSON encoding which is currently being developed in the standardisation community.

For encoding the measurement data, itself as well as quality information directly associated with the data, the OGC Abstract Specification Topic 20: Observations, measurements and samples (OMS) is recommended. Also, in this case a more lightweight JSON encoding shall be used.

The Web service interface standard for sharing these kinds of information shall be based on modern technologies such as REST and JSON. This includes for example the OGC SensorThings API specification. However, as this standard has only limited support for relevant sensor metadata, we especially recommend to follow the work on the OGC API - Connected Systems, which is currently in development. This standard will offer a dedicated access interface for important sensor metadata which will especially be suited for the requirements to better provide metadata on sensor quality.

To ensure a certain level of semantic interoperability, vocabularies shall be used within SensorML and OMS documents. A good recommendation are the vocabularies provided by the NERC Vocabulary Server of the British Oceanographic Data Centre (e.g., P01 for parameters, P06 for units, W01 - W10 for SensorML elements (history, capabilities, etc.)). However, beyond this, further dedicated vocabularies for metrology will be necessary (e.g., based on the JCGM 100:2008 “Evaluation of measurement data - Guide to the expression of uncertainty in measurement”).

Integration into the European Open Science Cloud

For enabling the publication of sensor quality information to the EOSC, different aspects of data management need to be considered. The basis of the data publication process is in a first step the establishment of an approach for handling sensor quality information on an organisational and project level. During the design of the internal sensor data management of the MINKE project, significant emphasis has been put on using an interoperable approach that relies on international standards for interfaces, data models, and data formats. More specifically, the OGC API - Connected Systems of the Open Geospatial Consortium (OGC) had been chosen as the underlying framework, especially because of its special capabilities to support the management and provision of detailed sensor metadata.

As the OGC API - Connected Systems is based on the REST paradigm, the data publication simply requires an HTTP POST request for uploading a new sensor quality information document to the data management component. Similarly, the download of sensor quality information is realised using HTTP GET requests. The format of the sensor quality information is based on the OGC API - Connected Systems specification as well (in this case a JSON encoding of the OGC Sensor Model Language (SensorML) will be used).

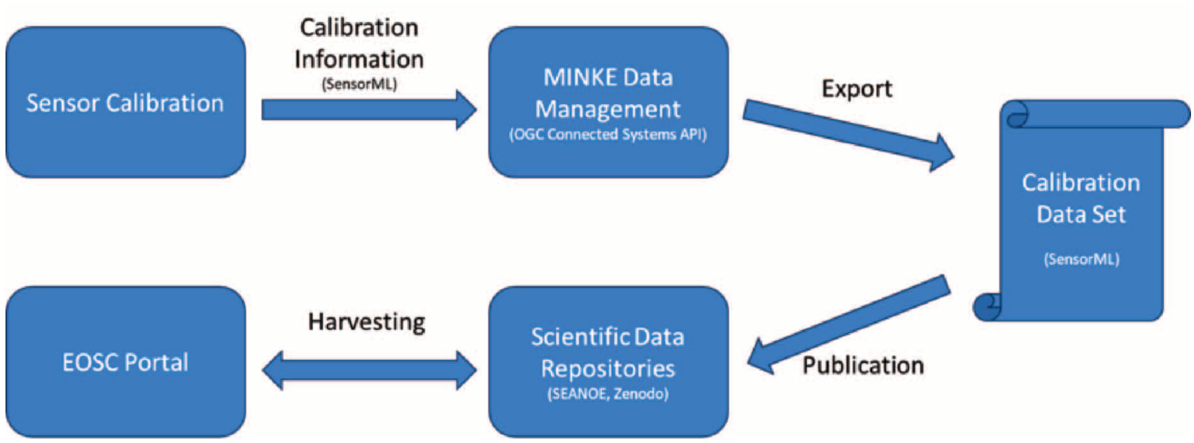


Figure 1 Approach for Publishing Sensor Calibration Information.

After sensor quality information such as a calibration report is available in the MINKE Data Management component (based on the OGC API - Connected Systems) this can serve as a starting point for publishing the data to larger scale research data infrastructures such as the EOSC. The publication workflow mainly includes the export of sensor quality information files via the OGC API - Connected Systems and the publication in commonly used research data repositories such as Zenodo and SEANOE. It is important to note, that a bulk export of sensor quality information is not intended, in order to ensure that the publication of sensor quality information is a consciously executed process. At the same time, the publication to scientific data repositories such as Zenodo and SEANOE will ensure a regular harvesting of the published resources by platforms such as OpenAIRE which will subsequently lead to discoverability of the data via the search interface of the EOSC Portal.

Summary

This contribution has introduced an approach for handling metrological data in ocean sciences based on international standards. After the initial design was completed, the implementation is currently ongoing. The underlying software components that are implemented as part of the European MINKE project will be made available to the community as open-source software. However, even though the presented approach is considered as work in progress, we are confident that the emerging new international standards such as the OGC API - Connected Systems will help to further improve the management of scientific measurement data in ocean sciences.

Functional «Semantic Web» Practice in the Marine Research Domain

Marc Portier[1] (marc.portier@vliz.be), Cedric Decruw[1], (cedric.decruw@vliz.be), Lennert Tyberghein[1] (lennert.tyberghein@vliz.be), Lucy Scott[2] (l.scott@unesco.org), Arno Lambert[3] (a.lambert@unesco.org), Pieter Provoost[3] (p.provoost@unesco.org), Douglas Fils[4] (doug@fils.network), Pier Luigi Buttigieg[5] (pier.buttigieg@awi.de), Bart Vanhoorne[1] (bart.vanhoorne@vliz.be), Simon Claus[6] (simon.claus@vlaanderen.be), Julian Rojas[7] (julianandres.rojasmelendez@ugent.be), Giorgia Lodi[8] (giorgia.lodi@istc.cnr.it), Alessandro Russo[8] (alessandro.russo@istc.cnr.it)

[1]VLIZ, Flanders Marine Institute (Belgium)
[2]IODE (ZA)
[3]IODE (Belgium)
[4]Ronin Institute (USA)
[5]AWI (DE)
[6]Digitaal Vlaanderen (Belgium)
[7]UGent-imec (Belgium)
[8]CNR (Italy)

Introduction

Independently published documents on distributed web servers delivered almost overnight the compelling browsing experience of «reading a single book», the analogous counterpart of «searching a single database» is still only a vague possibility, despite all efforts to publish structured semantic data on the web. The criticism is that the complexity involved is not balanced quickly enough with tangible result; a sentiment that further cripples motivation and slows progress. To combat that narrative, this talk searches for the opposite: easily achievable (low cost) Linked Open Data Publishing (LOD); and showcasing tools that leverage those efforts to ready, and tangible benefits.

Keep it simple

The combination of its central policy to "leave no-one behind", its focus on creating a robust, federated, and scalable backend to "provide robust plumbing that services many fountains", and its continuous global outreach to co-develop its functionality, has resulted in the basic RDF usage patterns from ODIS (Ocean Data and Information System) becoming the de facto underpinning for new projects in the marine research domain. We mostly applaud its technical simplicity and elegance: (i) by focusing on the simplest thing that could possibly work globally: i.e., basic Schema.org usage in JSON-LD (JavaScript Object Notation for Linked Data); and (ii) narrowing the core task to publishing only. This in stark contrast with the frequent technology-driven approach to RDF publishing, which can showcase impressive technologies and capacities, but adds unnecessary components that, more often than not, increase the cost of publishing and sometimes even hinders semantic readability. Explicitly:
You don't need to set up a triple-store or SPARQL (SPARQL Protocol And Query Language) endpoint to do Linked Open Data Publishing.
Just like the web didn't need working search engines upfront before creating linked HTML (HyperText Markup Language) pages.
You don't need to introduce new property paths or highbrow ontologies to share what you have to say, especially when you are aiming for a maximum understanding by an as wide as possible audience. In day-to-day usage, a property-path that navigates the information shape suffices

to get the answer. The optional type assessments are a luxury that ontologists and developers can engage in outside the executed real-time data-access.

Connecting dimensions through common identifiers

Connections between independently published structured data graphs are created by their referencing shared concept URIs (Uniform Resource Identifiers). Indeed, just as cartographic overlays help to discover relations between independently measured values that share a geo-temporal coincidence, the application of knowledge graph technologies allow us to disclose pathways extended down to the level of data points and over many more dimensions, thus allowing both deeper and broader connections to be made.

In the European marine research domain, many such sets of concept URIs exist. Often directly usable in RDF: persistent identifiers that are (mostly) leading to valid triple-exposing representations. Adopting these is yet another (low cost) “no brainer” to create viable LOD records.

The BODC (British Oceanographic Data Center) NVS (NERC Vocabulary Server) vocabulary for “all things measurable”.

The European Directories maintained by SeaDataNet for Marine organisations (EDMO), Environmental Research Projects (EDMERP), and Ocean Observing Systems (EDIOS)

ORCID (Open Researcher and Contributor IDentifier) for individual researchers and ROR (Resource Organisation Registry) for institutes.

The MareGraph project has the ambition to further complete this set with core elements of the taxonomic backbone – (i) WoRMS (World Register of Marine Species) and (ii) the Marine Regions. This work can interoperate with systems like ODIS, augmenting the marine LOD commons, while also innovating around it. Reporting on the work involved here discloses a number of interesting real-life lessons concerning Linked Data publishing.

Showcasing benefits

After making the case for reducing costs we also want to balance even further by introducing a number of equally achievable ways that these efforts can benefit the community. First we will show a useful JavaScript shim script that augments the browser experience when end-users navigate HTML with structured data elements. FAIR-Signposting guidance makes concept URIs discoverable and retrieving their RDF descriptions allows for Wikipedia-inspired previews as well as copy-paste access to the actual concept URIs. The second example will showcase a free and open service component that generates and feeds standard pluggable UI widgets for the selection of concept URIs. These widgets are offered in the ‘web components’ standard for easy integration by developers of data-systems that need to manage these connections. Finally, we present a comprehensive Knowledge Graph Analysis Platform (K-GAP) that brings the potential of available connected Graph Data into the discovering spyglass of the common data scientist.

Conclusion

In short, with a clear focus on the field of the Marine Research domain, we want to drive home the central argument of this quote: “People think RDF is a pain because it is complicated. The truth is even worse. RDF is painfully simplistic, but it allows you to work with real-world data and problems that are horribly complicated. While you can avoid RDF, it is harder to avoid complicated data and complicated computer problems.” – Dan Brickley and Libby Miller (in the foreword of the book “Validating RDF”).

Using data fragments as the foundation for interoperable data access

Marc Portier[1] (marc.portier@vliz.be), Peter Thijsse[2] (peter@maris.nl), Alexandra Kokkinaki[3] (alexk@noc.ac.uk), Tjerk Krijger[2] (tjerk@maris.nl), Gwen Moncoiffé[3] (gmon@noc.ac.uk), Paul Weerheim[2] (paul@maris.nl), Erwan Bodere[4] (erwan.bodere@ifremer.fr)

[1]VLIZ, Flanders Marine Institute (Belgium)
[2]MARIS (The Netherlands)
[3]British Oceanographic Data Centre - BODC (UK)
[4]Ifremer (France)

FAIR-EASE project context

The FAIR-EASE project aims to provide users with a Virtual Research Environment (VRE) for processing and visualizing datasets from diverse Data Infrastructures across environmental subdomains. Blue Data infrastructures developed within the Blue-Cloud and Blue-Cloud 2026 projects will be utilized to streamline access to data in the VRE. However, for non-marine infrastructures in FAIR-EASE, accessing data poses an additional challenge due to varied data access services and metadata models outside marine domain harmonization efforts. Problem areas include metadata catalogues, subsetting services, and different types of aggregations.

Sustainability of the "interoperability" enterprise

Our FAIR-EASE experience shows that as we grow the scope of our integration initiatives, we also re-asses what harmonisation is required. An optimistic view perceives this as fine-tuning on a guaranteed trajectory, assuming the independent domain efforts follow some universal plan. A more critical perspective, however, acknowledges the trail of successively built and now abandoned systems in the rear-view mirror; and rightfully questions the overall cost and long-term sustainability of such an approach.

Confronting these views compels us to leverage local harmonization, embrace diversity, and explore techniques for greater, domain-overreaching harmonization. Regarding sustainability, this further requires separating spaces prone to change from those needing sustained operation—allowing competitive engineering solutions to compete and replace each other without imposing costly changes in the research space. Success is measured by the effective decoupling of introducing new technologies into research practice with no (or minimal) effect to the sustainable operation.

System interfaces with Uniform semantics

In systems engineering, the need to separate concerns between functional spaces equates to defining interfaces. FAIR-EASE engages in this exercise, providing an architectural map with key dividers—the [[Data Provider]] and [[Data Access]] interfaces. The first captures how to (a) provide data (input side) in an open, yet structured and linkable way; while the second describes how to (b) effectively extract from the assembled pool of data those relevant subsets, slices or fragments (output side) to work on inside the VRE. These interfaces aim to operate without assumptions about specific technical solutions.

While aiming for "no" assumptions is idealistically lowered for practical realism, we agree on safe, low-cost assumptions inspired by and reusing commonly recognised success elements from web-architecture—reusing URI addressing, HTTP(s) protocols, and replacing HTML with RDF. So embracing the idea to publish structured data using the basic rules of the Semantic Web.

On that basis, the aim of bridging the demand (data-access) to the distributed, independent, and cross-domain offer (data-provider) in an effective way is forcing us into designing RDF expressions for both the exposed catalogue of datasets and the requested data-frames, proposing a prototype solution for both from existing vocabularies like DCAT, DCTerms, CSVW, and schema.org.

Research questions

The design effort is distinctively modelling the two viewpoints: (i) the data access (demand) point-of-view, being that of the VRE users; and (ii) the data provider (offer) point-of-view, about what is being made available by the data infrastructures. Key questions include:

- What RDF schemas can bridge the gap between different domains to make data (and their descriptions) from behind a wide range of data access services discoverable and accessible?
- What are the most important metadata elements from a researcher point of view? And how should these be encoded so automated assistance can be provided to them.
- What as-is (meta)data are provided by the data infrastructures? And how should they adapt to support sustainable future rehashing.

During this IMDIS session we plan to explain the approach and the latest results of the research and prototyping work within FAIR-EASE.

List of Abbreviations

CSVW (Comma-Separated Values on the Web)
DCTerms (Dublin Core Terms)
DCAT (Data Catalog Vocabulary)
HTML (Hypertext Markup Language)
HTTP(s) (Hypertext Transfer Protocol Secure)
IMDIS (International Conference on Marine Data and Information Systems)
RDF (Resource Description Framework)
URI (Uniform Resource Identifier)
VRE (Virtual Research Environment)

AI-based fish detection and classification at OBSEA underwater observatory

Oriol Prat Bayarri (oriol.prat.bayarri@upc.edu), Pol Baños Castelló (pol.banos.castello@upc.edu), Enoc Martinez (enoc.martinez@upc.edu), Joaquin del Rio (joaquin.del.rio@upc.edu)

Universitat Politècnica de Catalunya (Spain)

Introduction

Nowadays the use of artificial intelligence (AI) is expanding in a multitude of areas of knowledge, helping and accelerating many aspects of work. One of them is the object detection and classification, in this case applied to ecosystem monitoring. The monitoring of fish populations is of great ecological interest, and it is essential to know their status in order to determine what fishing impact is appropriate and sustainable, considering the resilience of these habitats. For many years, biologists have counted single species one by one in pictures in order to obtain several metrics such as abundance, biomass and biodiversity. Today, it is possible to greatly reduce the time spent in this process with computer vision tools such as AI-based object detection. A great example of the comparison between the use of AI and a human has been investigated by Ellen et al. [2020], and they found an improvement in the results when applying these computer vision methods.

There are several models that serve these methodologies through neural networks, one of the most established and stable is YOLO (You Only Look Once). YOLO divides an input image into a grid and each grid cell predicts multiple bounding boxes along with class probabilities. This grid-based approach allows YOLO to achieve impressive detection speeds. In addition, YOLO uses a convolutional neural network architecture to extract relevant features of the input image, enabling accurate object detection.

Data and Methods

This work uses a data-set of images obtained from OBSEA (1505 at the moment), a cabled underwater observatory located 4 km from the coast of Vilanova i la Geltrú (Catalunya, Spain) and at a depth of 20 m. The observatory has several marine sensors, including three video cameras from which the frames are extracted periodically.

In order to train the algorithm, the first step is to label the images, (work carried out with LabelImg software). Then the total number of images is divided into three blocks, one for training, one for testing and the other for validation, normally following this percentages respectively; 70%, 20% and 10%. The first part goes through the different levels of the convolutional neural network in order to determine the characteristics of the images. The second one is used to infer the information contained in images it has never seen before, in order to detect the objects in these new images. And the third part, as the name says, is used to validate the results, to collect metrics that give us information about is the model performance. The most commonly used are accuracy, precision, recall, mean average precision and confusion matrices.

In order to find the best model fit in the OBSEA custom data-set, a series of trainings have been performed using different characteristics in order to get the best results, that is, the best weights. The combinations used three species number groups (5, 12 and 16), five models (*nano*, *small*, *meduim*, *large* and *extra large*), five learning rates (0.01, 0.03, 0.0033, 0.00011, 0.000037), various data augmentation parameters and five random initial seeds.

Key findings

The best results for the 12 species group so far refer to the extra-large model, learning rate of

0.000037 and using data augmentation. In Figure 1 there are two relevant graphs, the metrics/precision(B) and the metrics/recall(B), both with good values indicating a correct prediction measure and sensitivity.
On the other hand, in Figure 1 the confusion matrix provides important information on the detection of each species, and how much it is confused with other species or with the background. The ideal situation would be the diagonal with values of unity for all species, however for some species such as *Mullus surmuletus* the detection is poorer.

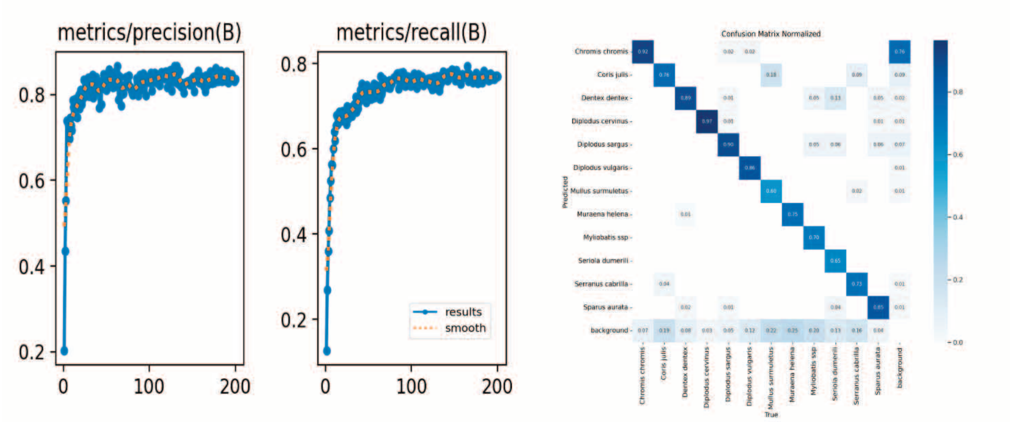


Figure 1 Metrics of a) precision, b) recall and c) confusion matrix normalized of the 12 species considered.

Figure 2 shows some examples of fish detection at OBSEA's environment, with their associated bounding box and detection confidence.
The model is capable of highly accurate detection of a wide diversity of fish. However, some environmental conditions such as water turbidity can affect the detection. The next steps are focused on further improving the model, increasing the number of species and applying real-time detection to video streams.

Figure 2 Examples of fish detections using the extra-large model at OBSEA enviroment (Vilanova i la Geltrú, Catalunya, Spain).

Acknowledgement
This work was developed in the framework of the Research Unit Tecnoterra (ICM-CSIC/UPC) and the following project iMagine, funded from the European Union Horizon Europe Programme – Grant Agreement number 101058625. This work benefited from services and resources provided by EGI with dedicated support from IFCA.

References

Ditria E.M., Lopez-Marcano S., Sievers M., Jinks E.L., Brown C.J., & Connolly R.M., (2020). *Automating the analysis of fish abundance using object detection: optimizing animal ecology with deep learning*. Frontiers in Marine Science, 429.

Jocher G., Chaurasia A., & Qiu J., (2023). *YOLO by Ultralytics (Version 8.0.0)* [Computer software]. https://github.com/ultralytics/ultralytics

AI-based fish detection and classification trials at OBSEA: available November 2023 https://www.youtube.com/watch?v=SEw79Gd6m5o

ADDRESS - An AI powered tool to visualize and analyse real time Ocean data

Sundar Ranganathan (rsundar@niot.res.in), Ramasundaram Subramanian (sundar@niot.res.in), Sachin Kumar Devakumar (dsachin.niot@gov.in), Arulmuthiah M. (arul@niot.res.in), Ramadass G.A. (ramadass@niot.res.in)

National Institute of Ocean Technology (India)

Understanding the oceans and their dynamic behaviour is crucial for scientific research, environmental conservation, and practical applications ranging from weather forecasting to maritime navigation. Moored buoy is one of the primary tools used for in-situ Ocean data collection. The accurate collection and analysis of moored buoy data are critical for understanding ocean dynamics, climate patterns, and ecosystem health. To streamline these tasks and optimize data quality, CORNEA – Centre for Ocean Realtime iNformation viEw and Archives - a State-of-Art IT infrastructure has been built for computational and visual capabilities to bring significant benefits to ocean researchers. The most unique feature of CORNEA is the innovative software “ADDRESS” (ADvanced Data REception and analysiS System). ADDRESS [Venkatesan et al.,2015] is a comprehensive platform that combines data visualization, analysis, quality control (QC), missing data imputation, and database management, with a particular focus on utilizing the AI-ML algorithm for enhancing data quality. The features of ADDRESS tool are discussed in detail.

1. Data Visualization

ADDRESS provides an intuitive and user-friendly interface for visualizing moored buoy data in real-time. Users can also create customized plots and graphs to explore various parameters such as meteorological and subsurface oceanographic parameters like sea temperature, salinity, wave height, and more.

2. Data Mining

ADDRESS is aimed at mining the data in the required format for a predictive analysis, apply quality control routines, plot, and visualise the data in a dashboard on the go. It also reduces manual interaction and encompasses diverse techniques for suggesting conclusions thereby supporting decision making too.

3. Data Analysis

The software offers powerful analytical tools that enable scientists to perform in-depth data analysis. ADDRESS allows users to generate statistical summaries, time series analysis, and anomaly detection to identify trends, patterns, and irregularities within the dataset.

4. Quality Control (QC)

Maintaining data quality is paramount in scientific research. ADDRESS employs advanced QC algorithms to automatically detect and flag erroneous or inconsistent data points (Subramanian et al.,2022). Ocean Best Practices methodology recognized by "International Oceanographic Data and Information Exchange" (IODE) of the Intergovernmental Oceanographic Commission (IOC) - UNESCO, is adopted to achieve quality and consistency in observation.

5. Missing Data Imputation

One of the standout features of ADDRESS is its integration with the Particle swarm optimization (PSO) algorithm for missing data imputation. PSO algorithm optimizes the imputation process by learning from the existing data by iteratively adjusting the positions of particles in the solution space, the PSO algorithm attempts to find an optimal or near-optimal set of imputed values for the missing data in the dataset with high accuracy. This ensures that datasets remain complete and consistent, even in cases of equipment malfunctions or data transmission issues.

6. Database Integration

ADDRESS seamlessly integrates with relational databases, allowing users to store, retrieve, and manage moored buoy data efficiently. This feature simplifies data archiving, retrieval, and sharing among research teams, fostering collaboration and data reuse.

7. Real-time Data Updates

The software continuously receives and updates data from moored buoys in real-time, ensuring that researchers have access to the most recent information and the ADDRESS has been equipped with the provision to navigate to other international buoy operator's web portals such as OMNI-RAMA and so on. This feature is especially valuable for tracking rapidly changing ocean conditions and responding to emerging research needs.

8. Data Dissemination

Further, the buoy data is disseminated to the global community through Global Telecommunication Systems (GTS). The data and metadata from Indian buoy programme are acclaimed by the global scientific community. Data from Indian buoy programme is also published in OceanOPS (formerly JCOMMOPS), the agency that establishes a common platform, coordinates within and amongst ocean observation communities across the globe.

Conclusion

ADDRESS represents a cutting-edge solution and mind mapping of all the met-ocean information to oceanographers, climate scientists, and environmental researchers engaged in the collection and analysis of moored buoy data. By combining data visualization, analysis, quality control, ML-based missing data imputation, and seamless database integration, ADDRESS empowers researchers to extract valuable insights from oceanographic datasets with unparalleled efficiency and accuracy, ADDRESS contributes significantly to the advancement of marine science and our understanding of the oceans.

References:

Subramanian R., Ranganathan S. and Ramasamy V., (2022). *Gateway to Ocean Data Management: Ocean Best Practices*. Information and Visualisation for moored ocean observation network. In OCEANS 2022-Chennai (pp. 1-9). IEEE.

Venkatesan R., Ramasundaram S., Sundar R., Vedachalam N., Lavanya R. and Atmanand M.A., (2015). *Reliability assessment of state-of-the-art real-time data reception and analysis system for the Indian seas*. Marine Technology Society Journal, 49(3), pp.127-134.

AI-based decision tree tools for best practices in ocean applications

Katie Watkins-Brandt[1] (katie.watkins-brandt@noaa.gov), Laura Riihimaki[1] (laura.riihimaki@noaa.gov), Sundar Ranganathan[2] (rsundar@niot.res.in), Virginie van Dongen-Vogels[3] (V.VanDongenVogels@aims.gov.au), Cristian Muñoz Mas[4] (cristian.munoz.mas@hi.no), Jay Pearlman[5] (jay.pearlman@ieee.org), Pauline Simpson[6] (p.simpson@unesco.org)

[1]NOAA (USA)
[2]NIOT (India)
[3]AIMS (Australia)
[4]IMR (Norway)
[5]IEEE (France)
[6]UNESCO-IOC (International)

Introduction

Decision trees have long been used as tools to distill complex processes by visually outlining paths to predictable outcomes. Simply put, it is a decision aid or simple flowchart with logic to reach a decision. The use of trees for prediction and classification dates back to as early as the 3rd century C.E with the development of the Tree of Porphyry developed by the Greek Philosopher Porphyry [Lima, M. et al., 2013]. Today, and since the late 20th century, much of the research on decision trees has focused on the development of algorithms for computationally intense logic and a host of Artificial Intelligence (AI) tasks related to decision-making [de Ville, B., 2013]. Regardless of the application, the design of a decision tree from the simplest form used by humans as visual aids to make decisions, to its most complex form of AI tasking, utilizes the same foundational design principles.

Decision trees to improve discoverability of methods in OBPS

Decision trees are widely used tools across various disciplines in science, applications, and purposes, and as previously noted, vary widely in complexity. In the context of data collection and quality assurance, the focus is to offer guidance on creating an efficient decision tree and to discuss the design of decision tree templates that may integrate or coincide with methodologies and best practices documentation created by the Ocean Best Practices System (OBPS) community. Design components integrate the FAIR Data Principles of Findability, Accessibility, Interoperability, and Reuse of digital assets [Wilkinson, M. et al., 2016]. By integrating FAIR principles into the design, the guidance and templates will be suitable for a wide variety of ocean data applications and ensure decision trees generated from these templates are visible, versatile, and widely adopted by the community. While the focus has mainly been on developing decision trees for distinct applications in ocean sciences, there is a desire within the community for the development of decision tree tools for identifying resources in the OBPS Repository. This effort is directly related to recommendations put forth here to integrate FAIR principles and design criteria that may promote harvestable information for identification of resources in the future, and further allow for data mining of information in decision trees to identify resources for the community.

Areas of application of decision trees

Decision trees are commonly used and applicable to communities across diverse scientific disciplines. They provide tools to understand, operationalize, and implement procedures or practices that influence data quality. The templates provide focus on three common purposes that may limit options and make decisions particularly important, with examples being heavily focused on ocean and atmospheric-related sciences: (1) "Time-based purposes" provide recommendations where availability in terms of time may be limited. Common use cases are field applications targeting the audience of technicians where limited time may be available for the maintenance or management of a particular sensor or system. Decision trees of this nature are meant to focus on maintenance and management practices that improve quality data while considering the time constraints experienced by technicians and their working environment. (2) "Resource-limitation purposes" become useful where resources may be limited; providing alternative workflows based on resource limitation. Common examples are applications where power may be a limiting factor such as buoys, gliders, or floats or where resource limitation is experienced related to supplies, equipment, or access. (3) "Procedural purposes" suggest procedures that are best suited to a particular application or situation. Common examples are applications where data quality may be assessed in the field or a process to determine which practice is best suited to their application or situation based on system, sensor, or working environment (Figure 1).

AI/ML use case for procedural decision trees related to oceanographic data quality control

Quality control (QC) in data management is essential. These checks depend upon the climatic condition of the region, the presumed accuracy of the instrument, and the expected accuracy of the parameter. it is not possible to set rigid QC standards for all the data types, regions, and seasons. Once the raw data is received, the preliminary processing and quality check is implemented in the real-time quality control (RTQC) followed by a detailed quality check conducted on the data in delayed mode quality control (DMQC). The RTQC basic tests are the primary level test to ensure the data delivered are of good quality. In delayed mode, the data undergo correlation test and range test based on long-term climatological data is applied as a double-check measure of the RTQC. With the AI approach, a model on the available data (parameter-wise) is trained to predict new ranges and thresholds which will suggest the decision of quality data.

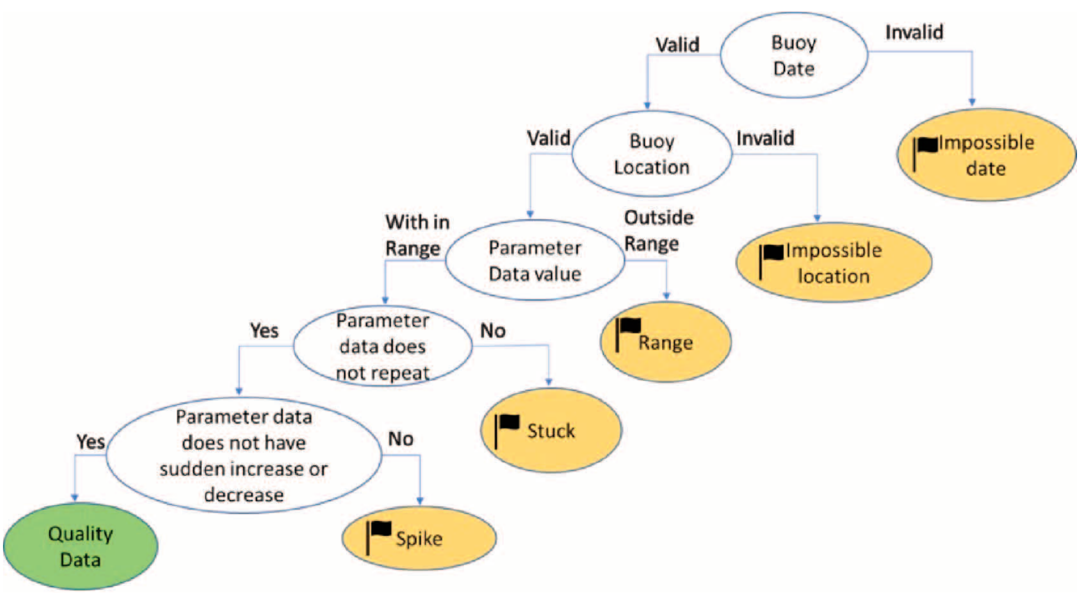


Figure 1 Decision Tree for Moored Buoy quality control procedure. Based on the quality procedures, the decision tree-making system was prepared using AI/ML algorithms with a given range and threshold that suggest the decisions.

Conclusion

An approach for designing and implementing decision trees to be used throughout the entire ocean observing value chain has been established. Decision trees, whether time-based, resourced-limited, or procedure-based might enhance productivity and bring a comprehensive way of guiding (1) users of the OBPS repository to discover methods, and (2) all-around ocean professionals to implement particular methodologies following a subject-matter expert approach. Implementation of decision trees throughout different ocean subject domains may act as a mechanism for establishing convergent processes among dispersed or divergent communities that use similar methods that are adapted to different situational contexts. Finally, AI/ML has proven to play a relevant role in procedural purposes to make better decisions on the quality of data which ensures the data's reliability and relevance.

SESSION TECH POSTERS

MetaTransfer: Harmonizing metadata of heterogeneous data resources in a common database framework

Susanne Feistel (susanne.feistel@io-warnemuende.de), Christiane Hassenrück (christiane.hassenrueck@io-warnemuende.de), Hagen Radtke (hagen.radtke@io-warnemuende.de)

Leibniz Institute for Baltic Sea Research Warnemünde (Germany)

Over the past decade, the increasing variety and number of deployed sensors, observation platforms and analytical devices, has resulted in a significant increase in the volume and complexity of scientific data in all disciplines of marine research. This trend is expected to continue with the development of new observation technologies and modelling approaches, calling for new strategies and approaches for data acquisition, processing, storage, analysis, and re-use. The widespread endeavor to manage such data according to F.A.I.R. principles led to the publication of vast amounts of datasets in a wide spectrum of databases and portals. However, while data publications steadily progress, the quality of the accompanying metadata does not yet keep up (e.g., next generation sequencing data publication, Figure 1). High quality metadata, though, is essential to follow F.A.I.R. in its true intent to support and advance the re-use of published data.

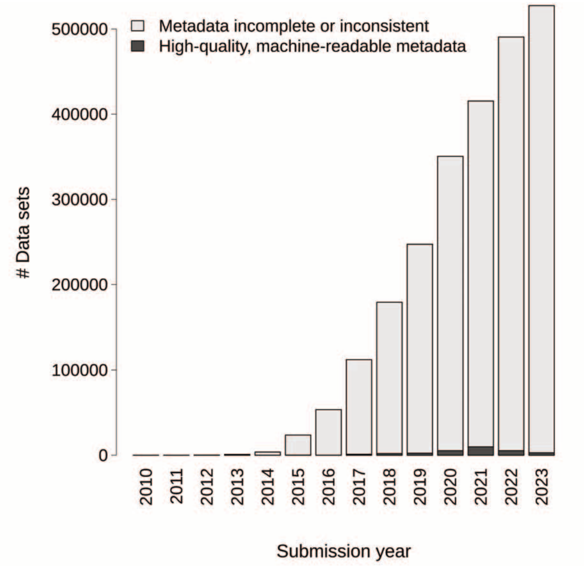


Figure 1 Assessment of metadata quality of publicly available DNA metabarcoding data from environmental samples (data retrieved from the Short Read Archive of the European Nucleotide Archive; date accessed: 03.01.2024; retrieved with: https://git.io-warnemuende.de/hassenru/ENA-metadata/src/branch/master/Scripts_update_2024/status_ena_illumina_amplicon_update2024.R#L1614).

One of the obstacles to archiving high quality data and metadata is the lack of understanding of the added value of re-usable data, as well as the requirements for re-usability, and the often cumbersome and high-effort data archiving procedures and metadata standards.
The current customary way requires scientists to provide metadata for administration and for publication in different locations such as modelling servers, doi servers, nucleotide sequence databases, or metadata catalogues. Meeting the requirements of all the used metadata standards of the different locations is time-intensive, requires expert knowledge, lacks incentivation, and might therefore prove too much effort, since for the scientists in possession of the required meta-information, their provision is only one out of many obligations and not the focus of their work.
Thus two interwoven problems crystalize when publishing data including high-quality metadata:

- Existing (extensive, but often not mandatory) metadata standards are not being used as intended or as comprehensively as necessary.

- Less extensive and easier to fulfil metadata standards may not facilitate data re-use.

A solution needs to be a system that offers easy, one-time entry of metadata while at the same time ensuring that all relevant parameters for data re-use are collected.
This may be achieved in two ways:

- interconnectedness of different data collection systems simplifies metadata submission, since entries can be transferred.
- Implementation of metadata standards beyond only their minimum mandatory fields to improve metadata quality for data re-use.

In the case of the Leibniz Institute for Baltic Sea Research Warnemünde (IOW), several specialized metadata systems for different purposes already exist. These systems contain high-quality, structured metadata:

- a Current Research Information System collecting all project information, involved scientists, publications, events, presentations, and collaborations,
- a Hyrax server for gridded datasets created from ocean model simulations, containing standardized metadata to provide access via the OPeNDAP protocol,
- a GeoNetwork metadata catalogue distributing metadata according to ISO standard for geo-referenced metadata,
- a DOI server to publish datasets and software according to DataCite requirements, and
- a document-oriented database for DNA sequencing data and metadata in compliance with internationally accepted metadata standards (MIxS) and compatible with external data archives to facilitate data submission after project completion.

Our aim is to connect all these systems in a «MetaTransfer Database» (Figure 2), enriching the metadata multifold. Ideally scientists need to input metadata once in an easy-to-use web application and only add partial information in case of using the other systems. This self registration system uses APIs to connect to the concerned databases and retrieve already existing metadata «per mouse click». Shared and specific metadata will be automatically transferred and may be exported in required formats.
This presents one solution tailored to the specific needs and data infrastructure at IOW. It will provide an easier way to collect F.A.I.R. data and metadata and thereby offer incentives to publish better documented and re-usable datasets.

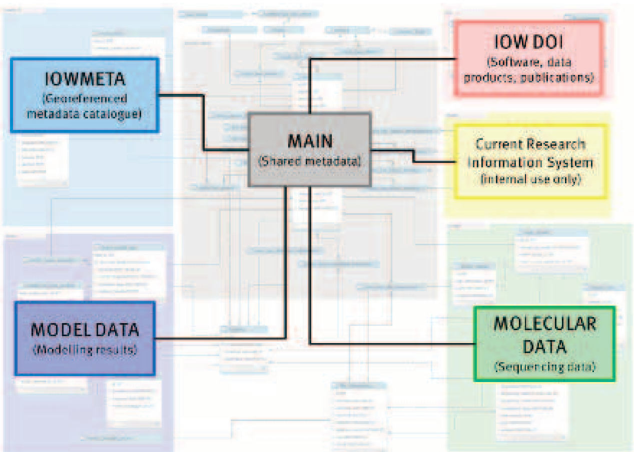


Figure 2 Entity-relationship model of MetaTransfer Database overlaid with simplified interconnected metadata systems.

The Docker Container Service for SeaDataNet Replication Manager: Implementation for ORION RM

Kyriakos Prokopi[1,2] (kyrpro@gmail.com), Elena Zhuk[3] (alenixx@gmail.com),
George Zodiatis[1] (oceanosgeos@gmail.com)

[1]ORION Research (Cyprus)
[2]Roswell Computing (Cyprus)
[3]MHI-RAS (Russia)

RM as a Docker container

Replication Manager (RM) is a Java (Tomcat) web application that runs on Windows, Linux and other Unix platforms. It is managed by IFREMER, a SeaDataNet partner, in conjunction with partners from EUDAT (DKRZ, GRNET, CINECA, CSC, STFC), who handle cloud services, and partner MARIS, who creates the Request Status Manager service and makes sure all three components function properly. The ORION research and development team proposes running RM as a Docker container, a compact software package that encapsulates code and its dependencies to enable fast and dependable operation of the RM service across various computing environments, such as development, testing, and production settings. This technique allows RM to connect to a MySQL DBMS on the host server, another host machine or in a different container, and it optimizes RM to run as several instances on a single cloud or physical system. Significant advantages over the traditional host-level setup are offered by the new proposed setup: (a) deployment on Linux, MacOS X, and Windows; (b) private network, versioned, and persistent storage; (c) simple upgrade of RM dependencies, such as JDK; (d) enhanced security due to isolation of processes, storage, networks, and containers; and (f) dependable upscaling for running multiple RM containers for different data consumers and teams. Combining the suggested configuration with WireGuard VPN to implement improved network access controls is an economical and latency-optimized way to further strengthen overall security.

Setup Overview

- Created with Docker compose.
- Running on Apache Tomcat v9 and JDK v11.
- Ability to connect to MySQL / MariaDB on host / container.
- Persistent storage for DB, data and configuration.
- Non-standard / higher ports preferred.
- Container stats and health can be monitored via Portainer (additional setup).
- Reinforced security and access control via WireGuard VPN (additional setup).

Operational Benefits for RM as a Containerized Service

- Resource optimization - Single cloud server can serve multiple RM services as needed.
- Simplified administration and monitoring via Web UI and CLI.
- Efficient and reliable storage backup with per container persistent volumes.
- Efficient and secure DB access per container.
- Simplified upgrade path using container dependencies.

Future Enhancements

- Automate RM configuration and deployment options.
- Improve multi-level container administration.

- Improve monitoring and event triggers.
- Deploy and publish as a service stack.
- Location-independent orchestration via Kubernetes.
- Improve data and configuration backup automation.

ORION Dockerized RM

The Cyprus ORION RM includes 3666 CDIs and ODVs on hydrological and hydrochemical parameter data, which were also migrated to the Dockerized RM (Figure 1). Currently, data for each station is kept independently in an ODV file (CDI). Furthermore, the host has a MySQL database management system (DBMS) that contains data sourced from many European and global databases, including the SeaDataNet, EMODnet Ingestion, EMODnet Chemistry, Pangea, and Argo portals, as well as data gathered by Cyprus institutions, pertaining to the Eastern Mediterranean Levantine basin. All this data has received the SDN quality control. The ORION RM uses this database [Zhuk and Zodiatis 2019] as a source of data. The ORION RM service can be easily managed and monitored via the Portainer UI (Figure 2).

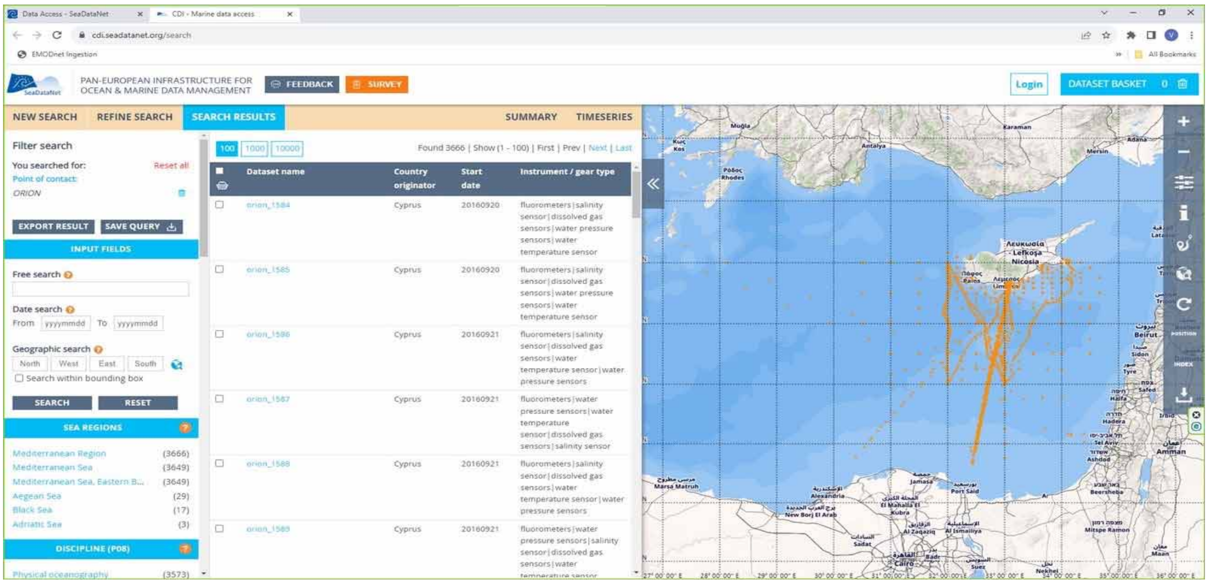


Figure 1 Example of ORION data in the SeaDataNet via the ORION Dockerized RM.

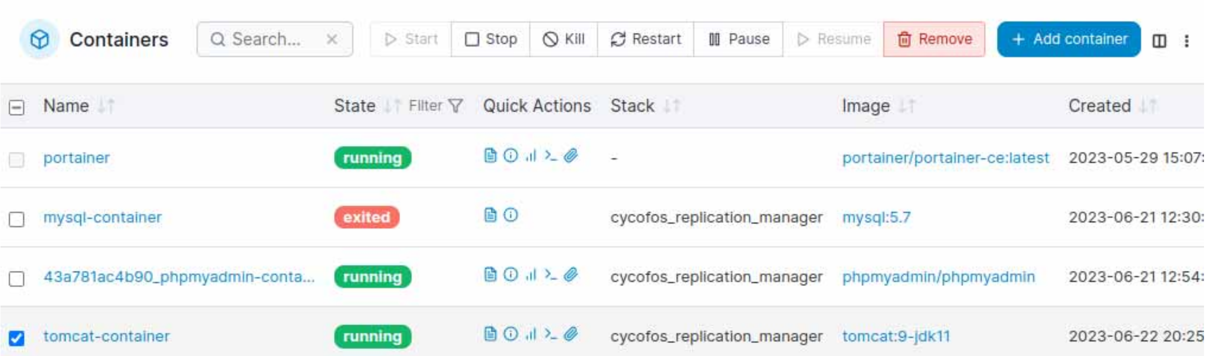


Figure 2 The ORION RM service monitored via the Portainer tool.

Reference

Zhuk E. and Zodiatis G., (2019). *EastMedAtlas: online data access system*, Proc. SPIE. 11174, Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019), 111740P (27 June 2019); https://doi.org/10.1117/12.2533440

New Function of Oceanographic Data Center for Data Publication and Citation

Toru Suzuki (suzuki@mirc.jha.jp)

Marine Information Research Center (Japan)

The National Oceanographic Data Centers (NODCs) and Associate Data Units (ADUs) have the role and responsibility for collection, management and dissemination of oceanographic data observed in IOC Member States and to exchange them among the NODCs and ADUs. Recently major academic associations and scientific journals require to register the data used in author's works to public data repositories for open access with a permanent identifier, such as the Digital Object Identifier (DOI), according to FAIR (Findable, Accessible, Interoperable and Reusable) data principles. It means that NODCs and ADUs should undertake a new and important task as a public data repository. The simplest way is that NODC assigns one DOI to its massive database or product, but oceanographic data are usually obtained by measurements and observations operated by organizations, institutes and national and international projects for ocean sciences using vessels/platforms with many variables for long time and global coverage, then DOI should be assigned to each measurement or observation. Actually, oceanographic data are submitted to the Japan Oceanographic Data Center from one hundred and more Japanese institutes with several hundreds of vessels and platforms, then a large number of DOIs will be required to assign all submitted data individually. In fact, it is not realistic to refer to many DOIs by data users in their scientific works. One solution to facilitate citation, identification and proof of existence for oceanographic data is that NODCs and ADUs should have a function for DOI assignment not only for data management statistically but also according to result for data query conditions, such as time/period, position/area, variable, measurement/sensor, vessel/platform, organization/project or other parameters which are requested by data users. For oceanographic data citation, namely, it should be considered to be allow a DOI of DOIs, multiple DOIs (see Figure 1), or other better solutions effectively.

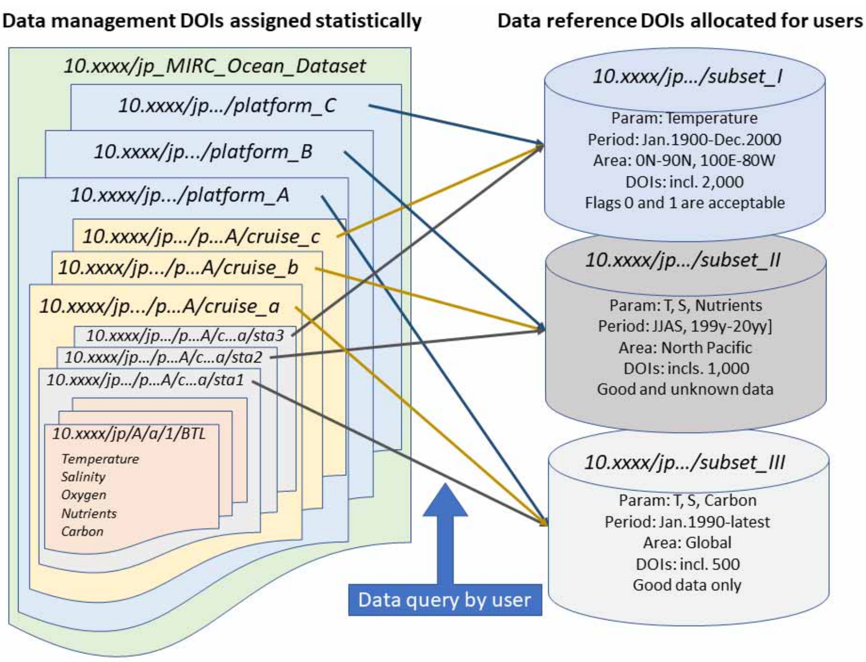


Figure 1 Schematic view of DOI assignment for oceanographic data by data centers.

Location vocabularies of the French Fisheries Information System

Armelle Rouyer (armelle.rouyer@ifremer.fr), William Lafarge (william.lafarge@ifremer.fr), Céline Vignot (celine.vignot@ifremer.fr)

Ifremer (France)

The French Fisheries Informations System (SIH)

The French Fisheries Informations System (SIH) is a national scientific network for the observation of resources and all professional fishing fleets on board. It is managed by Ifremer and provides knowledge for research and expertise, making it possible to contribute to sustainable exploitation. Numerous informations from fishing declarations, observations at sea or during landing is processed and stored in « Harmonie » database, using common vocabularies recommended by FAO, ICES, European Union and French Government.

They are used by Sacrois, a data cross-validation tool. It is an algorithm for consolidating declarative data. Its objective is to estimate at best: Who fishes? Where? How long? With which fishing gear? With what fishing effort? What species are fished? In what quantity? And for what value?

Location vocabularies

Location vocabularies of the SIH are used for all the data processing from the data entry until the map visualization. In the database, locations are organised in 94 levels and most of them has a geographical layer associated in a shapefile format (polygon or point) digitalised in the 2010's.

They are separated in three categories within several hierarchies in order to provide data aggregated at different scale according to the need:

- Land areas: these vocabularies enable to map and collect the data by countries (ISO), administrative areas (IGN), ports (United Nations Location Code).
- Areas at sea: these vocabularies are used for the reports to the European Commission or the French Government but also to follow the fishery activity at field especially for French overseas. They are based on fisheries organisations such as: Food and Agriculture Organization (FAO), International Council for the Exploration of the Sea (ICES), General Fisheries Commission for the Mediterranean (GFCM).
- Regulated areas at sea: these vocabularies are determined by limits defined in laws such as the Exclusive Economic Zone, EU Marine Strategy Framework Directive, Marine Protected Areas, …

Location entities and geometries

A location entity is identified by a code, a name and an id level. When it exists, a geometry is banked in the database. A link with the parent entity is recorded with the ratio of relationship calculated between the surface of the child area and the parent area.

Inside a location level, polygons must not cross each other and between levels the common limits between parent and child must correspond and be the same in order to respect the hierarchy.

In 2023, the geographical layers (shapefiles) have been updated from the different sources (FAO, ICES, SHOM, Marine Regions…) and corrected to align the limits between areas. The open source geographic informations system tool QGIS has been used.

Open Data

Most of these shapefiles will be accessible in open data in a catalog of Sextant a marine and coastal Geographic Data Infrastructure (Ifremer) or via an API on the website of Ifremer's Fisheries Informations System. Thus, they follow the INSPIRE directive and standards of the Open Geospatial Consortium (OGC) and they will be referenced by a Digital Object Identifier (DOI).

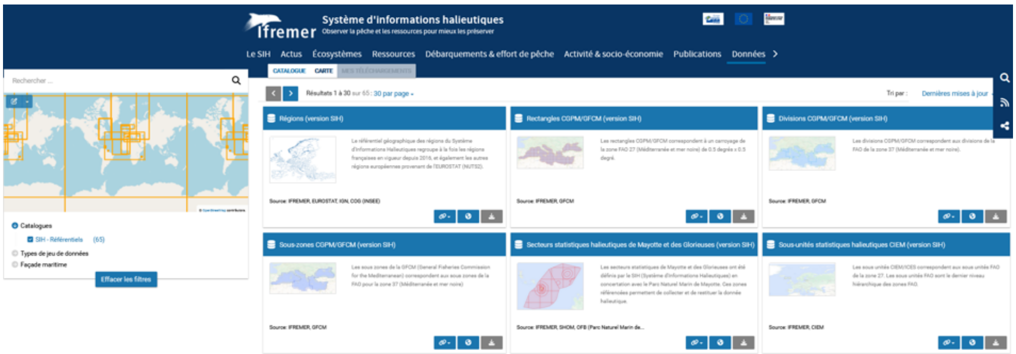


Figure 1 Catalog of geographical vocabularies of the French Fisheries Informations System.

Metadata are given in forms following FAIR principles (Findable, Accessible, Interoperable, Reusable). Sources and the genealogy have been completed with the informations found but a lack of traceability still exist for some of them. The hierarchy is also visible on each landing page and it's easy to navigate from a level to another related. Geographical data can be downloaded or viewable thanks to Web Map Service (WMS) and Web Feature Service (WFS).

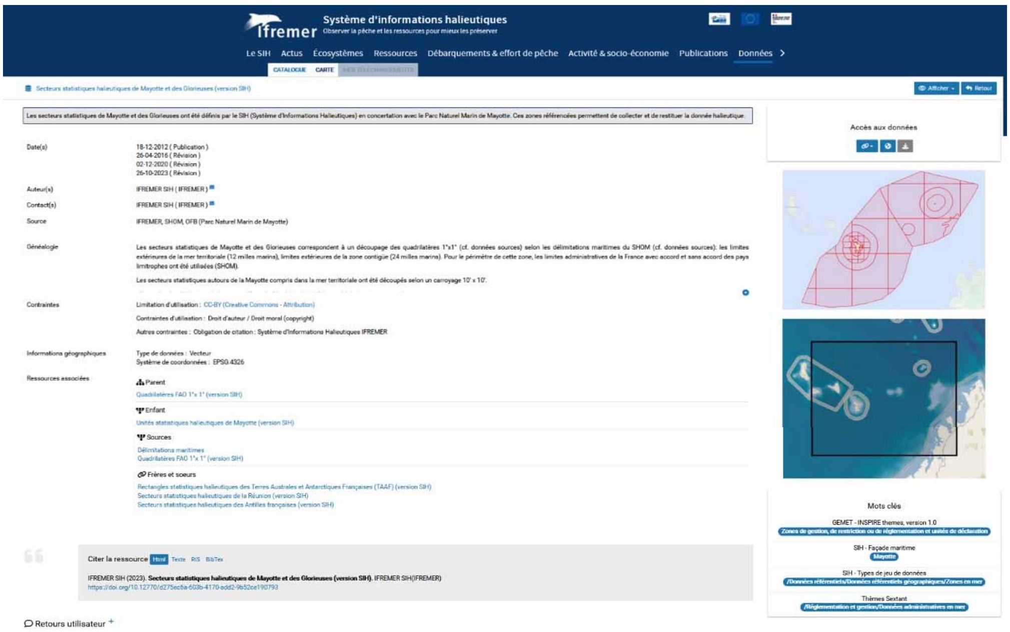


Figure 2 Example of a metadata form.

FAIR WISH project – automizing IGSN registration workflows from existing databases

Linda Baldewein[1] (linda.baldewein@hereon.de), Mareike Wieczorek[2] (Mareike.Wieczorek@awi.de), Alexander Brauser[3] (alexander.brauser@gfz-potsdam.de), Simone Frenzel[3] (sfrenzel@gfz-potsdam.de), Ulrike Kleeberg[1] (ulrike.kleeberg@hereon.de), Birgit Heim[2] (Birgit.Heim@awi.de), Kirsten Elger[3] (kelger@gfz-potsdam.de)

[1]Helmholtz-Zentrum Hereon (Germany)
[2]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (Germany)
[3]GFZ German Research Centre for Geosciences (Germany)

The International Generic Sample Number (IGSN) is a globally unique and persistent identifier for material samples. Within the FAIR WISH project, a generic sample registration template called the *FAIR SAMPLES Template* has been developed. The modular template was developed as an all-in-one solution for the varying use cases, with the ability for users to select only the variables needed for their specific sample type. It has been applied to three different use cases at different levels of digitalization of their metadata.

Figure 1 Overview of FAIR WISH use cases of Russian-German expeditions, Ketzin pilot site and Biogeochemical Sample Database and how they increase in level of digitalisation. All uses cases from all three partners register their IGSNs at GFZ German Research Centre for Geosciences.

The *FAIR SAMPLES Template* was developed as Microsoft Excel spreadsheet and is the source for semi-automated generation of standardized XML files required for the IGSN registration and IGSN landing pages. The development strategy of the template was explicitly followed to meet the practice of researchers who mainly organize their sample descriptions in tables. However, in the use case of the Biogeochemical Sample Database, the metadata has already been digitized. The Hereon expedition database is a fully normalized relational database that describes the diverse metadata associated with biogeochemical campaign samples, so called long-tail data.

It has been developed since 2016 and contains more than 12000 individual samples collected in land and sea-based campaigns since June 2017.
The *FAIR SAMPLES Template* is used as the basis for selected all mandatory and optional metadata fields for IGSN registration corresponding to metadata entries in the biogeochemical campaign database. In total, a list of 33 different fields is selected. The IGSN itself is hierarchical, which is considered when describing not only the samples in the database, but also their parents of the sample type site, where the sample was collected. Thus, the database to IGSN mapping is split into two hierarchy levels. For each mapping table, each IGSN field name is mapped to a database field if the value is variable across all samples. Some fields, e.g., the *current_archive*, are constant and pre-defined.
In an automated scripting process, the metadata from all samples and sites are extracted from the database, mapped to the corresponding IGSN field name, expanded by the constant IGSN fields and written into *the FAIR SAMPLES Template*. The template filled with metadata of more than 12000 IGSNs is sent to the registering agent at GFZ. The IGSNs are quality controlled, harvested, and registered. The landing pages of all samples can be found at https://dataservices.gfz-potsdam.de/igsn/.
Reversely, the IGSNs are stored in the biogeochemical campaign database and published alongside the data measured at these samples.
The usage of IGSNs for samples bridges one of the last gaps in the full provenance of research results. The FAIR WISH use case of the biogeochemical campaign database at Hereon shows that registering IGSNs for fully digitized samples can be retroactively implemented through automated processes.

Management of spatial data integrity including stakeholder feedback in Maritime Spatial Planning - A Perspective from Ireland

Sarah Flynn[1] (sarah.flynn@marine.ie), Andrew Conway[2] (andrew.conway@marine.ie), Tara Keena[2] (tara.keena@marine.ie), Tom Woolley[3], Kellie Heney[1] (kellie.heney@marine.ie), Caitriona Nic Aonghusa[2] (caitriona.nicaonghusa@marine.ie), Eoin O'Grady[2] (eoin.ogrady@marine.ie)

[1]Compass Informatics (Ireland)
[2]Marine Institute (Ireland)
[3]Department of Housing, Heritage and Local Government (Ireland)

Maritime Spatial Planning (MSP) is a strategic and dynamic process that aims to balance the different demands for using the sea. In Ireland, the Department of Housing, Local Government and Heritage (DHLGH) retains overall responsibility for implementation of MSP, while the Marine Institute provides government with MSP-related data management support. Ireland's Maritime Spatial Plan is the National Marine Planning Framework (NMPF). MSP is data intensive, having to draw upon the best available data coming from many different sources. There are, therefore, multiple challenges in managing spatial data for inclusion in Marine Plans. These challenges include the need for data integrity to drive reproducibility; as well as providing contextual information to enable end users to increase understanding as well as the potential for the data to be reused independently.
The approach adopted for Maritime Spatial Planning in Ireland [Flynn, et al., 2023] demonstrates how repeatability can be achieved for data products and the underlying processes necessary to drive data integrity, to ensure the best data is available for decision-making. It is built upon a framework previously developed in the Marine Institute to deliver data management processes for MSP in Ireland [Flynn et al., 2020]. As newer and better-fitting data was identified as the process evolved, mechanisms were also developed to integrate valuable stakeholder feedback and to provide a uniform way to deliver, manage and update datasets. This approach illustrates an interwoven process of updating datasets in a marine plan and the benefits of simultaneously integrating stakeholder consultation feedback and developing repeatable data management processes.

Stakeholder Feedback

Robust marine planning requires extensive and thorough stakeholder engagement. The aim is to integrate input and responses to ensure the MSP process incorporates relevant expertise and views of key audiences invested in the maritime area. The Irish MSP Stakeholder engagement process offered a unique opportunity to provide valuable context towards how data should be used, and what for; including how it should it be combined. The process was iterative, with each cycle driven by stakeholder feedback. By integrating feedback from stakeholders upstream into MSP data products on an ongoing basis, data may be utilised and interpreted appropriately by downstream stakeholders and decision makers. Throughout the MSP process in Ireland, the assimilation of the stakeholder participation and availability of data and information was key to the successful delivery of the NMPF.

MSP data management

To support efforts concerning the governance, reuse and provenance of data, a quality management framework for data was developed for Ireland's Marine Planning data processes [Leadbetter et al., 2020]. It served to improve data and knowledge generation; manage and share data and develop new products and services where appropriate. Underpinned by this quality management framework, the collection, collation, validation, and analysis of new and existing spatial datasets enabled the creation of a variety of maps, made publicly available for use in policy, including for marine planning. As these datasets were used to support policy, it is vital that they are managed in a coordinated manner to maximise the integrity of the data available, complying with relevant legislations (INSPIRE), best practice guidelines and licensing conditions; putting in place the framework for Maritime Spatial Planning to enable decisions that are consistent, transparent, sustainable and evidence based. The cycle of MSP Stakeholder and related data processes are represented in Figure 1.

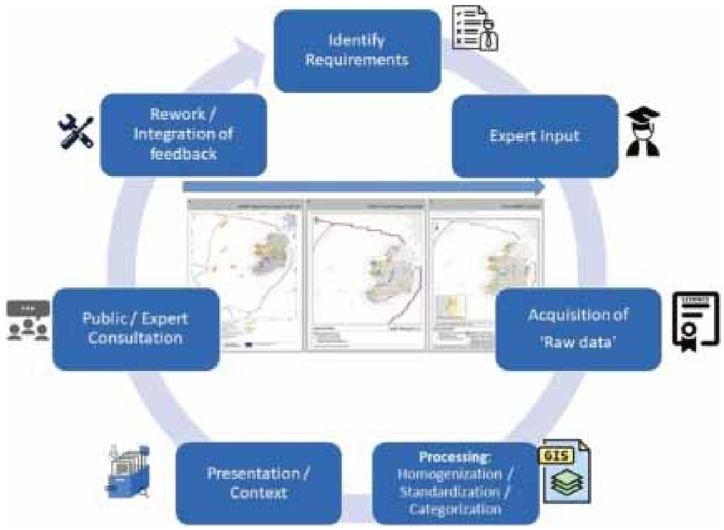


Figure 1 The Quality Process employed to integrate stakeholder feedback and generate and update the maps throughout Ireland's MSP process.

Overall, Stakeholder engagement and robust data management are two fundamental elements for successful maritime spatial planning. They must co-occur and evolve in tandem throughout the MSP development, finalisation and implementation stages. Without them the adoption and implementation of plans may be delayed or halted altogether. Examples within the literature of systematic integration of stakeholder engagement responses into maps in a marine plan while concurrently updating spatial datasets plan are scarce. This paper highlights an innovative approach to managing the integrity of marine spatial data over time, while also incorporating the concept of usability and data quality and stakeholder feedback under the guise of 'best available data' in the establishment of an evidence base.

References

Flynn S., Meaney Will., Leadbetter A., Fisher J., Nic Aonghusa C., (2021). *Lessons from a Marine Spatial Planning data management process for Ireland.* International Journal of Digital Earth, 14:2, 139-157, https://doi.org/10.1080/17538947.2020.1808720

Flynn S., Tray E., Woolley T., Leadbetter A. et al., (2023). *Management of spatial data integrity including stakeholder feedback in Maritime Spatial Planning*. Marine Policy 156, pages 105799

Leadbetter A., Carr R., Flynn S. et al., (2020). *Implementation of a Data Management Quality Management Framework at the Marine Institute*, Ireland. Earth Sci Inform 13, 509–521. https://doi.org/10.1007/s12145-019-00432-w

Towards FAIR data management of German marine seismic data

Janine Berndt[1] (jberndt@geomar.de), Gauvain Wiemer[2] (wiemer@allianz-meeresforschung.de), Daniel Damaske[3] (daniel.damaske@pangaea.de), Janine Felden[4] (janine.felden@awi.de), Robin Heß[4] (robin.hess@awi.de), Pina Springer[1] (pspringer@geomar.de)

[1]GEOMAR, Helmholtz Centre for Ocean Research Kiel (Germany)
[2]DAM, German Marine Research Alliance (Germany)
[3]MARUM, Center for Environmental Sciences, University of Bremen (Germany)
[4]Alfred-Wegner Institute, Helmholtz Center for Polar- and Marine Research, Bremerhaven (Germany)

Introduction

Making data openly accessible following the FAIR (findable-accessible-interoperable-reusable) principles is essential if their value is to be maximized and if they shall become available for direct access to computing systems for example for machine learning. As a result, most funding agencies request strategies for making all scientific data available under FAIR principles.
The German Marine Research Alliance (DAM) coordinates research data management activities in the field of marine research in Germany. One example is the Underway Research Data project which aims to improve and standardize the systematic data collection and data evaluation for expeditions with German research vessels. So far, this has only been applied to data from permanently installed sensors. Since the work- and dataflows implemented within the Underway Research Data project are generic, seismic data could easily be added to the portfolio of data transferred onshore via the Mass Data Management (MDM). All it needed was the development of a metadata standard and a standard operation procedure for seismic data.
Seismic data have been notoriously difficult to describe as they tend to be heavily processed and some of the processing routines are proprietary. Here, we present a first step towards standardization of metadata for marine reflection seismic data on board large German marine research vessels.
As part of the National Research Data Infrastructure for Earth System Sciences (NFDI$_4$Earth), GEOMAR has initiated the pilot study *German Marine Seismic Data Access* for the implementation of the FAIR principles for 2D multi-channel reflection seismic data. The pilot study was funded for one year in cooperation with the University of Hamburg and the University of Bremen and is now being continued as part of the DataHub initiative within the research field Earth an Environment of the Helmholtz Association and supported by the Underway Research Data project coordinated by the German Marine Research Alliance (DAM). We have implemented a standard operation procedure and a basic metadata standard for the long-term archival of raw 2D multichannel seismic data collected on board German research vessels. The metadata standard has been tested for different set-ups as each institute and university uses different gear and recording systems on board.
From a socio-cultural point of view, this process has already proven to be extremely beneficial as it has brought the scientific community together. In addition, it makes a direct data transfer of the raw data from the research vessel to the repository PANGAEA possible via a mass data storage device. This process is fully transparent to the users and supports them to fulfill their data management duties.

Data flow

On board, seismic raw data will be recorded on a single-shot basis, while an ASCII file for each shot will be generated containing all the necessary metadata. The data file and its related ASCII

file will then be transferred via the Mass Data Management (MDM) system onshore and will be subject to quality control before being semiautomatically archived at the repository PANGAEA. PANGAEA is a worldwide established repository fulfilling the FAIR principles. Before the publishing of raw seismic data in PANGAEA, the metadata are already available at institutional data access platforms, e.g., GEOMAR's ocean science information system OSIS. OSIS serves as a central information hub for various marine research activities like expeditions, simulations, and experiments with links to publications, reports, maps, samples, data, and code repositories but also as an internal data exchange platform for ongoing projects. Currently, a new version is in development to enhance interoperability.
Ship tracks along which seismic data has been acquired will be visualized within the German Marine Data Portal (marine-data.de) for the public to ensure user-friendly findability and access, especially for the seismic research and industry community.

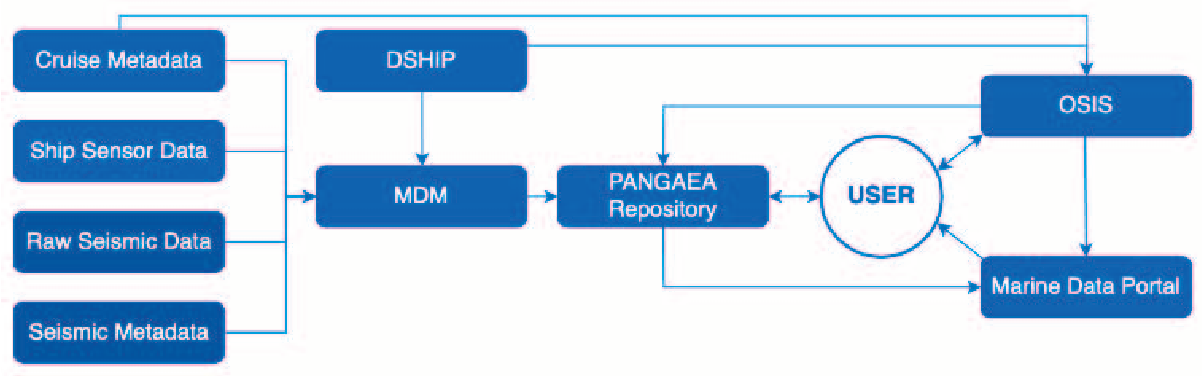


Figure 1 Schema for data transfer developed within the DAM underway research data project.

The German Marine Data Portal provides user-friendly, centralized access to marine research data, reports, and publications from a wide range of data repositories and libraries in the context of German marine research. It supports machine-readable, data-driven science and the development of cloud computing in science. The Marine Data Portal also provides different map-based viewers to explore and find data. The ship tracklines along which seismic data have been acquired will be visualized along with metadata in the viewer after being archived at PANGAEA.

Conclusion and Outlook

Metadata standardization for raw reflection seismic data collected on German research vessels facilitates the direct transfer of these data from research vessels to the PANGAEA data repository for their publication according to the FAIR principles. Tight coordination within the German seismic community with data centers abroad and the Open Subsurface Data Universe (OSDU) industry initiative will ensure that the development is sustainable and may serve as a template for future seismic data archival initiatives. The next steps include the development of metadata standards for processed seismic data, 3D seismic data, sub-bottom profiler data, and ocean bottom seismic data.

References

Berndt J., Hübscher C., Keil H., Lehmann A., and Berndt C., (2023). *German Marine Seismic Data Access*. NFDI$_4$Earth road map, Zenodo, https://doi.org/10.5281/zenodo.7875451
Mehrtens H., Berndt J., Getzlaff K., Lehmann A., and Lorenz S., (2023). *From local to global:*

Community services in interdisciplinary research data management. EGU General Assembly 2023, Vienna, Austria, 24-28 Apr 2023, EGU23-6873, https://doi.org/10.5194/egusphere-egu23-6873

Felden J., Möller L., Schindler U., Huber R., Schumacher S., Koppe R., Diepenbroek M., and Glöckner F.O., (2023). *PANGAEA – Data Publisher for Earth & Environmental Science*. Scientific Data, https://doi.org/10.1038/s41597-023-02269-x

Real-Time Data Transfer and Management for the RV Belgica Using FROST OGC SensorThings API

Nicolas de Ville de Goye (ndeville@naturalsciences.be), Thomas Vandenberghe (tvandenberghe@naturalsciences.be), Nils Van den Steen (nvandensteen@naturalsciences.be)

Royal Belgian Institute of Natural Sciences (Belgium)

Introduction

Oceanographic Research Vessels, like the new RV Belgica, fulfil their scientific role by performing campaigns with specific scientific investigation and by continuously measuring the properties of the atmosphere and the water. Continuous data serve a wide range of users: the involved researchers, the global marine research community and private companies, the in-house dissemination tools[3] and various Data infrastructures that RBINS constitutes and contributes to (e.g., GOSUD, INSPIRE, SeaDataNet, ...). To meet the requirements of the different users a fully operational sensor-to-client data flow has been developed for all bound sensors that serves rich, standardized and quality-controlled data in near-real time.

Data flow

Figure 1 shows the different components used for the automated data acquisition. Before the actual data dissemination to the different clients, the data needs to 1) get transferred from vessel to shore and imported into a replica of the vessel database, 2) be transferred to the database with standardization and metadata enrichment, 3) be quality controlled, 4) be sub-sampled and transferred to the production database.

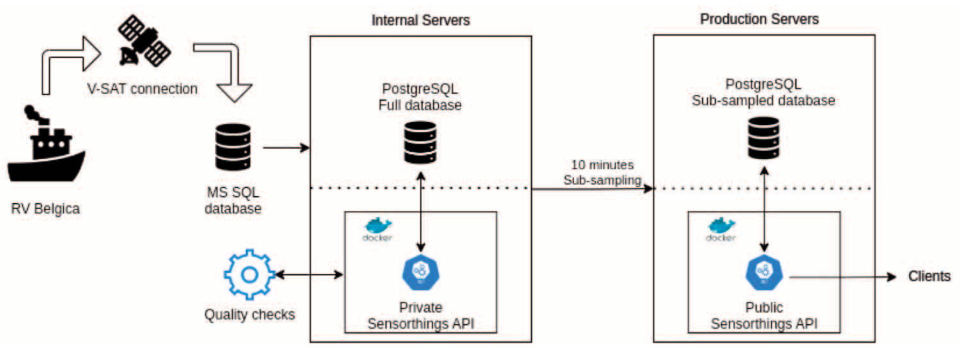


Figure 1 Automated Sensor-to-client data flow for en-route data processing and dissemination.

1. Around 500 observations are measured every 10 seconds on-board. This includes meteorological, navigation, operational and physical/chemical parameters. A compressed selection of those parameters is sent via V-SAT connection every 10 minutes to an on-shore FTP server. Those data are imported into a replica of the Microsoft SQL vessel database, which is pruned regularly to save on storage. This database is vendor-specific and only serves the on-board data acquisition software but doesn't store the data in any standardized way. A second database is therefore needed for long-term storage.
2. A Java/Hibernate application selects the data in the database, standardizes and geo-references them. The data are then imported into the full PostgreSQL database using the

[3] https://odnature.naturalsciences.be/odanext/en

OGC SensorThings API endpoint. The database schema follows the SensorThings data model and is enriched with the necessary metadata (NERC-controlled vocabularies P01, P02, P06, L05, L22 and L20). The Fraunhofer FROST implementation provides a properties Jsonb column that can be easily used to incorporate NERC vocabularies.

3. A Python application connects to the SensorThings API endpoint and performs automated quality checks on the location, range, gradient and spikes. The L20 quality flags are updated according to the test results. If the data passes the different tests, it gets a "probably good" flag. The "good" flag is only granted when a human check is performed.
4. For performance reasons only a downsampled version of the data is exposed publicly. The time-series frequency is reduced to 10 minutes data, using only data that passed the quality checks, which is enough for most data clients. Full-frequency data access is exposed only internally at RBINS.

Implementation principles

The variety of user-profiles interested in our data is the best incentive to only store standardized data. We must be able to easily generate datasets for the different users with a minimum level of data transformations. For this reason, the FAIR principles have been followed during every step of the data architecture development. In previous projects, we developed an open-source solution for the provision of INSPIRE-compliant metadata and data using Geonetwork and Geoserver (https://metadata.naturalsciences.be). These applications are closely following the Open Geospatial Consortium Standards (amongst others) and benefit from an active community. We decided to continue using the OGC Standards family with the Sensor Web Enablement (SWE) Standards (SensorML, SensorThings and Observations & Measurements).
Two Standards are defined in SWE, Sensor Observations Service (SOS) and SensorThings API. Based on our previous experience with SOS[4], we decided to select the dockerized FROST SensorThings API open-source implementation (https://sensors.naturalsciences.be/sta). The data model is simple, exposed using JSON and comes with powerful REST capabilities (time and spatial filtering, join operations, etc.). For data dissemination, two approaches are used in parallel, the datastreams and the multi-datastreams. Datastreams are simple time-series, useful if you are interested only in a few parameters, but not handy if you need to retrieve a lot of parameters simultaneously. For that we created ad-hoc multidatastream end-points that gather the interesting data in one place (e.g., all data for the GOSUD repository or the RV Belgica website).

Conclusions

A fully automated near-real-time vessel-to-client data transfer has been implemented for the en-route data of the new RV Belgica. It includes the transfer itself, metadata enrichment, data standardization, quality checks and data dissemination. The infrastructure has been implemented using mainly existing open-source solutions, although some small components have been developed internally (data normalization and quality control). The quality control package is available on GitHub as open source[5]. The FROST OGC SensorThings API proves to be a simple and reliable standard for the management and dissemination of sensor data, including various metadata (quality flags, geo-referencing, sensor information, etc.). One of the main issues is to ensure that the API performances for data retrieval (including filtering actions) are acceptable over time regardless of the database size. We decided to publicly expose only a sub-sampled version of the data for that purpose.

Poster session

[4] de Ville de Goyet, N., Fettweis, M. and Vandenberghe, T., (2021). On using a Sensor Observation Service as an INSPIRE-compliant download service. Bollettino di Geofisica, 12, p.243.

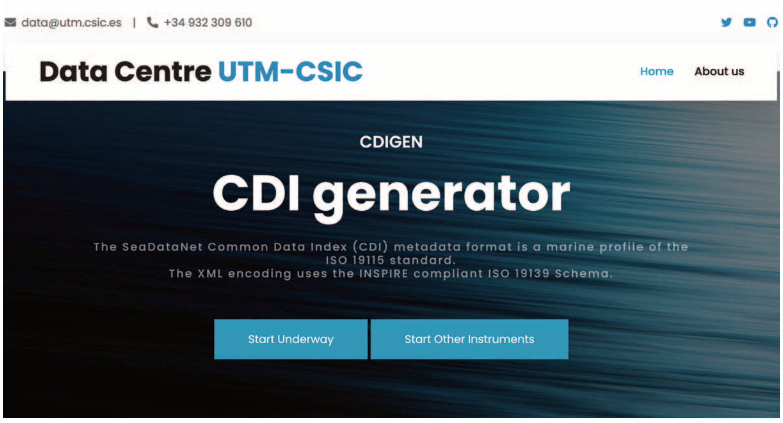
[5] https://github.com/naturalsciences/qualityAssuranceTool

UTM-CSIC Data Centre CDI generator

Marc Casas (mcasas@utm.csic.es), Elsa Martínez (emartinez@utm.csic.es)

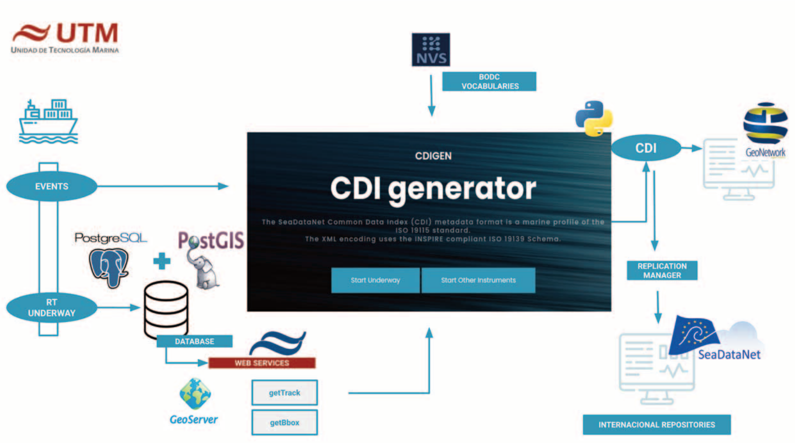
UTM-CSIC (Spain)

The Data Center of the Marine Technology Unit of the Spanish Research Council (UTM-CSIC) is responsible for managing a large volume of spatial data generated during oceanographic campaigns carried out on CSIC research vessels that are part of the Singular Scientific-Technical Infrastructure "FLOTA" of the Ministry of Science and Innovation (MICINN).
To carry out this management, the data center has a Spatial Data Infrastructure that consists of a Catalogue of Oceanographic Cruises and datasets (with data from 1991 to the present), a Geoportal with basic OGC geoservices that allows for consulting maps by cruises and creating your own maps, as well as other web services and applications that enable viewing, performing quality control, and downloading underway data such as temperature, salinity, or meteorological data obtained during the cruises. Likewise, there is also a download service for all types of raw data, in this case only for restricted users.



Different types of entries are shown in our data catalogue and uploaded to SeaDataNet: Cruise Summary Reports (CSRs) for reporting on cruises or field experiments at sea, and Common Data Indexes (CDIs) for reporting the different types of data that have been obtained during the cruises. As creators of these metadata entries, we have developed Python scripts to assist in their generation. Our tool directly references previously uploaded and approved cruise CSRs on the SeaDataNet Catalogue to populate new Common Data Index (CDI) files, using a standardized CDI modeled with Mikado. Additionally, it processes a CSV log file containing all onboard events to create a CDI file for each event of the oceanographic campaign. Furthermore, the script makes use of different web services developed by our Data Centre to extract the cruise GML track and bounding box information and insert it into the CDIs files for continuous data. These CDIs adhere to common SeaDataNet (SDN) vocabularies, maintained by BODC, ensuring interoperability and adherence to FAIR data principles in XML metadata descriptions.
In the field of oceanographic metadata generation, researchers and data managers have several options at their disposal. Tools such as Mikado, ODV, NEMO, BODC Tools, OceanBrowser, and EMODnet MEDI offer varying degrees of functionality and ease of use for creating metadata compliant with SeaDataNet standards. While each tool presents its own strengths and capabilities, technicians may seek a solution that streamlines the process and ensures user-friendly, automated metadata generation. In response to this need, we developed CDIgen, a tool specifically designed

to generate oceanographic metadata in a user-friendly and automated way. CDIgen not only simplifies the metadata creation process but also enhances efficiency, allowing technicians to produce large quantities of metadata seamlessly, thus facilitating the management and sharing of oceanographic data.
The CDIgenerator has proven highly beneficial, significantly automating manual tasks and minimizing human errors in metadata descriptor creation. Given the substantial volume of spatial data generated during oceanographic campaigns, this tool greatly enhances the productivity and efficiency of our Data Center technicians. Now, we can generate vast quantities of metadata in mere milliseconds, ready for submission to the SDN catalogue. Moreover, all vocabularies, organizations, and CSR references are updated daily via system services, with manual updates available through a designated button. The Python scripts are meticulously organized and extensively commented for usability by any Python user. Additionally, all necessary scripts and libraries will be maintained over time and remain accessible on our GitHub repository (https://github.com/utmdata), where community suggestions and contributions are welcomed and encouraged.
Initially developed for Google Colab, the script enables any user, regardless of Python proficiency or tools, to run the code easily. Furthermore, from this Python script, we developed a web application using Flask, simplifying the process further. Now, users can easily create and download metadata files from any location with an internet-connected browser, without additional specific software requirements.
This web application will continue to evolve and be improved over time. Although currently tailored to UTM Data Centre, our ultimate objective is to create a fully customizable CDI generator. This would enable Data Centers worldwide using SDN standardized vocabularies to customize, create, and download their CDIs, thereby simplifying the submission process to the SeaDataNet CDI catalogue. Our aim is to foster community engagement and facilitate participation in the SDN community by offering an intuitive, user-friendly approach.
In essence, our endeavors at the Data Center of the Marine Technology Unit reflect a commitment to not only managing vast amounts of oceanographic data but also to streamlining processes and fostering collaboration within the community. Through the development of robust tools and continuous refinement of methodologies, we strive to uphold the FAIR principles in data management. As we look towards the future, we remain dedicated to enhancing our capabilities, embracing new technologies, and working alongside people worldwide to advance our understanding of the world's oceans. Together, we aim to build a more interconnected and empowered community, united in our pursuit of efficient data management practices and promoting the sharing of oceanographic data.



UTM Data Centre web services: fleet app case of use

Xavier Rayo (xrayo@utm.csic.es), Juan Luis Ruiz (jlruiz@utm.csic.es)

UTM-CSIC (Spain)

Oceanographic research faces a crucial challenge in efficiently managing data collected from diverse research vessels. These vessels, each equipped with unique sensors and operating in many marine environments, generate vast amounts of valuable data. However, traditional data management approaches struggle to integrate and query data from multiple vessels simultaneously. Researchers often need to access data from various vessels together to gain comprehensive insights and make informed decisions.

In our case, the underway data from oceanographic vessels managed by the Marine Technology Unit of the CSIC is received in real time at the Data Centre. Upon arrival, it undergoes real-time processing to be disseminated and used in different applications developed by the Data Centre, as well as by those users who wish to do so through our web services. These web services provide users with the ability to access real-time data and query our stored historical data, facilitating their integration into visualization applications. We have different options to serve the data, from databases stored in Postgresql and served through OGC services through Geoserver, to more personalized web services for our customized applications, such as:

- getLast: Last datagram received from the vessels.
- getPoint: Real time information in both KML format and JSON.
- getSerie: It concatenates the ongoing files (NAV, MET, TSS) in order to obtain a CSV between two dates.
- getLine: Track between 2 dates in JSON format.
- getTrack: Cruise track in WKT, geoJSON and GML. An option for selecting the number of nodes you need in order to satisfy the limitation of Mikado software when generating the GML.
- getBBox: Bounding box calculated from the tracks of the cruises stored in the database.

Our OGC services are WMS (*Web Map Service*) and WFS (*Web Feature Service*) services that allow us to query the oceanographic cruise tracks, the position of different types of stations and operations carried out on board the vessels, which can also be consulted through our data portal http://data.utm.csic.es/.

It's worth noting that the process of capturing UDPs in Bash, saving series in daily files, and the web services themselves are developed in PHP. This setup ensures the system remains lightweight. Additionally, the data can be utilized in JSON format.

Fleet app

Most of the current oceanographic data management and visualization applications often focus on individual vessels, leading to fragmented data access and limited integration capabilities. While some systems offer vessel-specific data querying, there is a lack of solutions that enable simultaneous querying of data from multiple vessels and visualizing them.

The Fleet application allows simultaneous viewing of the positions of various vessels in our fleet, providing real-time access to underway data stored on our servers. For real-time data and historical queries, the application relies on two of our customized services previously mentioned, *getPoint* and *getSerie*, enabling users to query time series data. This information can be visualized through graphs for desired variables and can also be downloaded to the user's local disk for further processing. It's important to mention that the data available for viewing is always open data from ongoing cruises, and data acquired during cruises may not always be available for download.

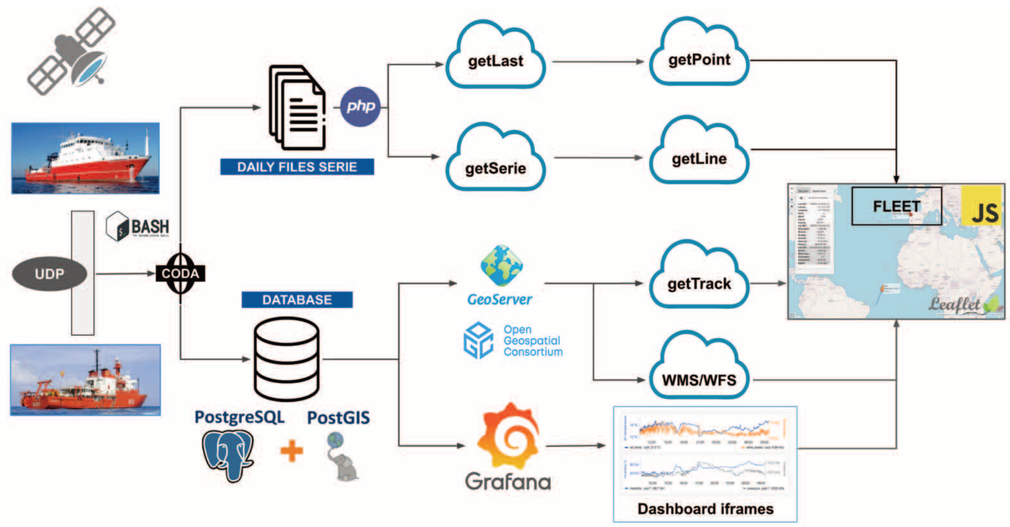
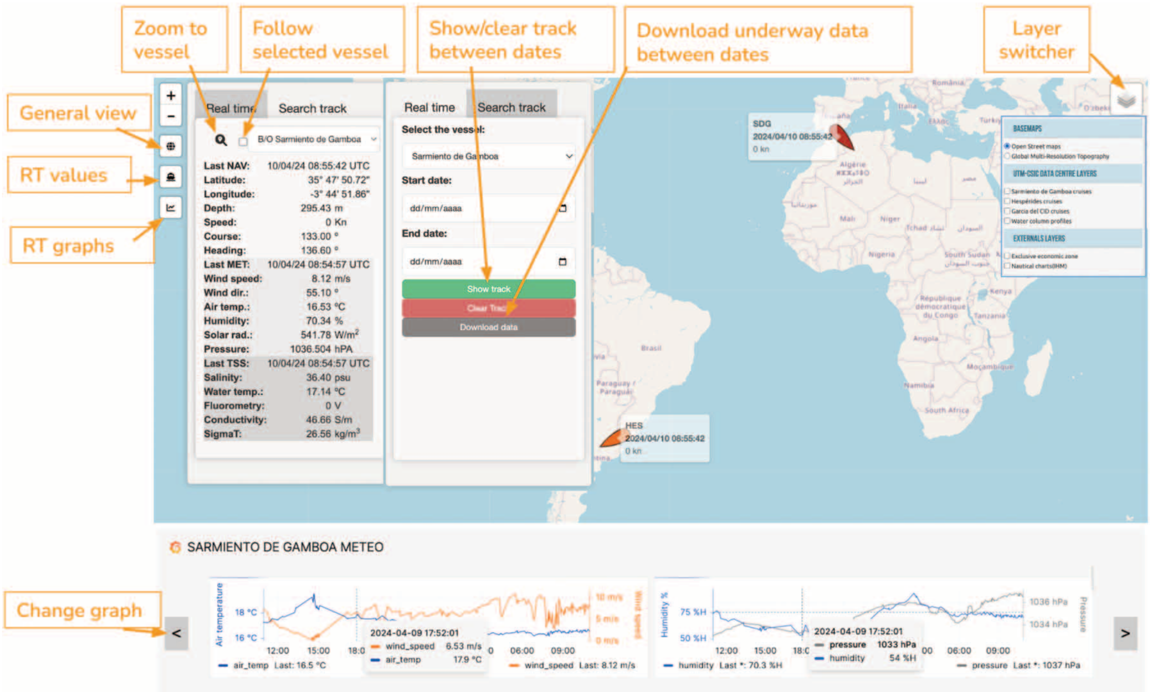


Figure 1 Schema showing the system from the UDP's to the web app's.

The application is developed using the very lightweight map library Leaflet.js along with some of its plugins and custom-developed tools, integrated with the powerful data querying and visualization opensource web mapping tool Grafana.



Adaptive Archives for Reproducibility in Complex Software Environments

Edvin Fuglebakk (edvin.fuglebakk@hi.no), Anthony Hennessey (anthony.hennessey@hi.no)

Institute of Marine Research (Norway)

Reproduction and corroborative analysis

Quantitative scientific results rely on data collection, scientific models and methods for estimation of model parameters. They also rely crucially on the correct implementation of models and methods, which usually means the implementation of a computer program. Correctness of scientific software cannot be assured, and general experience with software has taught the modern scientist to expect the presence of occasional bugs. This is a situation familiar to the scientist, as scientific knowledge in general relies on corroboration, rather than proof of correctness. As for other scientific knowledge, it is important that data products can be subject to continuous criticism and corroboration, even after it has entered into the scientific record. For time-series data one can easily envision a computation to be questioned decades after it was first computed. Concerns about correctness of implementation are naturally addressed by adding tests to the software producing a data product. Similarly, concerns about the applicability of data, methods or models are often answerable by implementing variants of the computation that produced the original result. For instance, the robustness of results to model assumptions may be investigated by making changes to the scripts and running them again. In the end the analysis must also be re-run to correct the record, in the event that errors in results have been revealed. It is therefore invaluable to archive not only the computed results that form a data product, but also the input data, execution parameters, and scripts used. Archives need to support reproducibility in this classical sense of corroborating results, not just the ability to recompute exactly the original result. See Goodman et al. for a clarifying discussion on concepts of reproducibility.

Support for such reproducibility is in principle provided by popular tools for managing computation of data products, such as StoX (Johnsen et al.)[6] or the Transparent Assessment Framework (TAF)[7].

Reproduction with complex environments

A major complication arise however, once we appreciate that modern software practices tends to delegate most of the implementation to underlying libraries or dependencies. These libraries are all part of the software that provides scientific knowledge. Library updates may fix or introduce bugs that may change previously calculated quantities, and checking the consequence of updates is a natural extension of delegating the implementation to an external library in the first place. One has to maintain a software environment that is up to date, at the cost of continuously adapting archived code as bugs are revealed or library maintainers stop providing updates compatible with the updated environment.

Adaptive archives

In practice, simply having stored code and input data along with data products does not ensure reproducibility a year or two later. They need to be adapted to an updated environment. We seek to facilitate this by making use of continuous integration support provided by the DevOps

[6] https://stoxproject.github.io/StoX/

[7] https://www.ices.dk/data/assessment-tools/Pages/transparent-assessment-framework.aspx

plaform GitLab[8]. Conceptually we maintain three environments: stable, testing and unstable. These environments encompass R, a curated set of R packages, and various system dependencies. The stable environment is made available on all platforms across the institute and is the environment users experience day-to-day. This is kept up to date with library updates, but at a reasonable frequency that allows some time for checking the consequences these updates have for existing data products. The *testing* and *unstable* environments principally exist for use in the in the GitLab pipelines (as docker containers). In the context of this discussion *unstable* is most interesting; each day all available environment updates are automatically incorporated into *unstable* and every night all projects are tested in the *unstable* containers, it is these tests that expose how updates across the software ecosystem impact re-computability. Failures during nightly tests against *unstable* are alerted to the IT support team who screen them for technical issues and alert remaining issues to the project owner. Updates are introduced into the stable environment only after project owners and IT support has had some time to make necessary adaptation. Figure 1 illustrates this process.

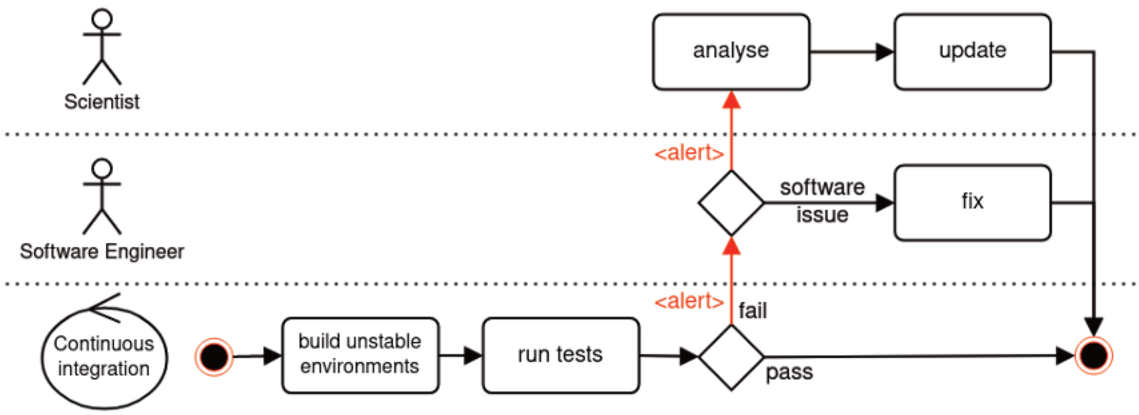


Figure 1 Conceptual representation of the nightly tests for re-computability.

The function of the continuous integration platform is both to automatically generate the *unstable* test environments and to automatically run tests in each of these environments. The test themselves are not prescribed by this setup and can be anything that is scripted and run within the environments. For the purposes of monitoring the data products these tests implements a comparison between computed and archived data products, so that any updates that changes the result of calculation is detected, and project owners are alerted.

Integration with other reproducibility systems

The framework is general enough to support anything implemented in R, which includes two reproducibility systems that we rely in at the Institute of Marine Research: StoX and TAF. We take full advantage of the support for estimation, documentation and transparency provided by these systems. In addition, by archiving copies of StoX projects and TAF projects in our gitlab repositories, we also gain the ability to monitor their compatibility with new software environments, pinpointing exactly when updates are causing problems and keeping our data products at all times ready to be modified, and re-run for corroborative analysis.

[8] https://about.gitlab.com/

References

Goodman S.N. et al., (2016). *What does research reproducibility mean?* Science Translational Medicine, 8, 341. https://doi.org/10.1126/scitranslmed.aaf5027

Johnsen, E. et al., (2019). *StoX: An open source software for marine survey analyses.* Methods in Ecology and Evolution, 10(9), 1523-1528. https://doi.org/10.1111/2041-210X.13250

Shipboard ADCP data: from acquisition to standardization and distribution in European IRs

Julie Gatti (julie.gatti@ifremer.fr), Françoise Le Hingrat (francoise.le.hingrat@ifremer.fr),
Michèle Fichaut (m.fichaut@free.fr), Léa Franc (lea.franc@ifremer.fr),
Pascale Lherminier (pascale.lherminier@ifremer.fr), Tina Odaka (tina.odaka@ifremer.fr)

Ifremer (France)

Most of the Research Vessels in the French Oceanographic Fleet (FOF) are equipped with Shipboard ADCP (SADCP) mounted on the hull and acquire underway ocean current measurements, from the coast to the open sea.
We propose to describe the SADCP workflow (Figure 1) from the data acquisition to the dissemination through European portals.

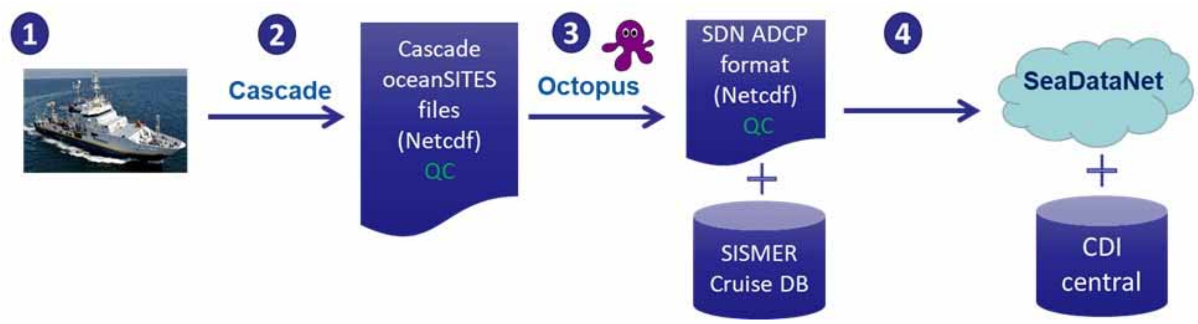


Figure 1 Workflow of SADCP data from FOF Research Vessels.

SADCP data (post-)processing and qualification
Guided by the necessity of using high-quality current data for its research projects, the Laboratory for Ocean Physics and Satellite remote sensing (IFREMER/LOPS) developed a software called CASCADE to compile, correct and qualify the data provided by the SADCP of research vessels. A SADCP acquires currents measured along each of its beams. A software is needed to transform beam velocities into earth coordinates and remove the motion of the ship to obtain: ocean velocity, using ancillary measurements such as heading and position. CASCADE software has been used for several years in an operational context by the SISMER French National Oceanographic Data Center. It can be used easily on *STA (for "Short Term Average") or *LTA STA (for "Long Term Average") files generated by the acquisition RDI software VMDAS (in terrestrial coordinates). CASCADE converts the *TA files in a single NetCDF survey file and cleans the data according to adjustable parameters. It is possible to diagnose and correct for a misalignment or a bad amplitude of the SADCP. It also adds useful ancillary variables: barotropic tide and bathymetry. Data can then be filtered or averaged along specific sections or stations. Graphic outputs of many kinds are displayed and saved to check the processing and illustrate a data report (Figure 2).

SADCP data standardization: Octopus converter
In the framework of EuroGO-SHIP, a converter has been developed to convert SADCP data at OceanSITES netCDF format (as the output of the CASCADE software) to the standardized SeaDataNet NetCDF TrajectoryProfile format. It has been integrated into the SeaDataNet

Octopus software which is already specialised in conversions to SeaDataNet format from different input formats like EGO for glider data, MGD77 for Magnetism, Gravimetry, and Depth data sets. The SeaDataNet NetCDF (CF) format is encoded in CF-compliant NetCDF together with the usage metadata relying on SeaDataNet standards (common Vocabularies, Quality Flag scale). All French SADCP data issued from CASCADE software will be converted to SeaDataNet format, beginning by a demonstration on the OVIDE-A25 section which has been carried out biennially since 2002.

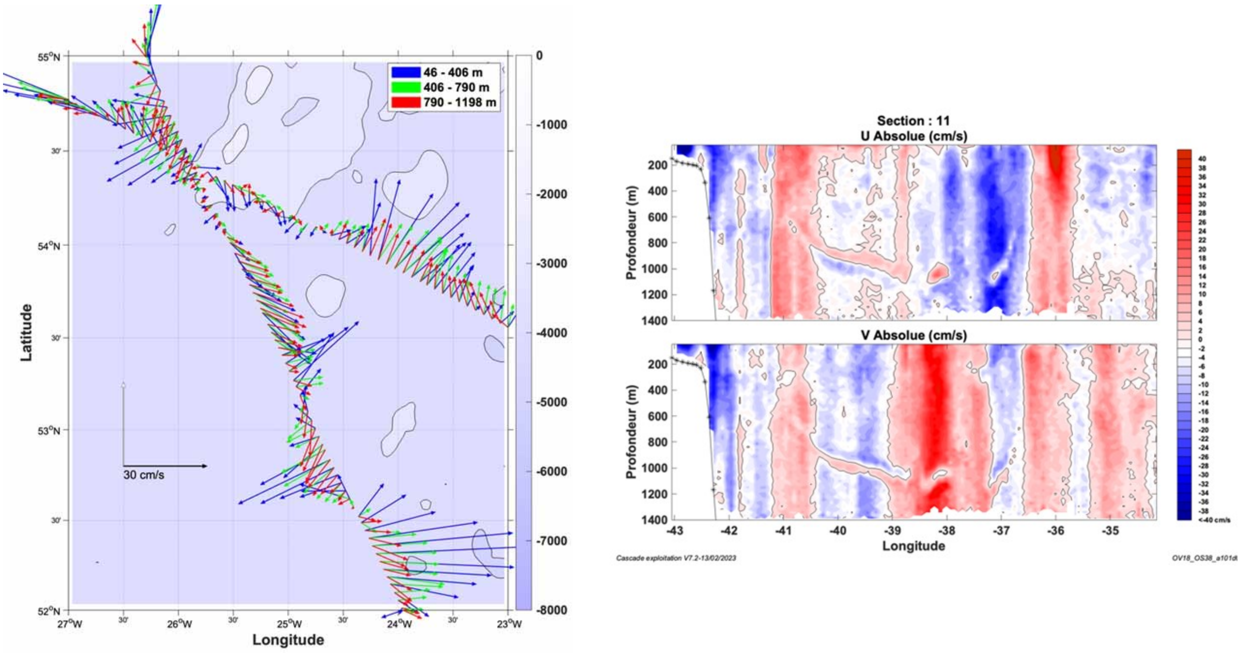


Figure 2 Example of SADCP 38kHz data visualisation from OVIDE 2018 survey using CASCADE software. Left panel: vectors of current in 3 different depth layers. Right panel: section of zonal (U) and meridional (V) current near Greenland.

SADCP data distribution into SeaDataNet infrastructure and an EuroGO-SHIP portal

Standardized metadata (Common Data Index or CDI) will be created to disseminate these standardized SADCP's files to the European SeaDataNet research infrastructure. The CDI service gives users a highly detailed insight in the availability and geographical spreading of marine data sets, that are managed by the SeaDataNet data centres. Moreover, it provides a unique interface for requesting access, and if granted, for downloading data sets from the distributed data centres across Europe.

In a second step, a webportal will be set up to access directly to GO-SHIP data through the SeaDataNet infrastructure and will allow the users to visualize them.

Perspectives

The next step will be to write a converter from CODAS format to SeaDataNet NetCDF TrajectoryProfile format. This will allow to enlarge the ADCP sources of ocean currents to data processed by CODAS by French and other international research institutes.

Interoperable services as foundation for data access and processing

Alexandra Kokkinaki[1] (alexk@noc.ac.uk), Peter Thijsse[2], peter@maris.nl), Tjerk Krijger[2] (tjerk@maris.nl), Marc Portier[3] (marc.portier@VLIZ.be), Gwen Moncoiffé[1] (gmon@noc.ac.uk), Paul Weerheim[2] (paul@maris.nl)

[1]British Oceanographic Data Centre - BODC (UK)
[2]MARIS (The Netherlands)
[3]VLIZ, Flanders Marine Institute (Belgium)

The problem in context

One of the aims of the Blue-Cloud 2026 and FAIR-EASE projects is to offer users Virtual Labs and a Virtual Research Environment (VRE) in which they can process and visualise marine- and other domain datasets from a wide range of data infrastructures. A key component of both projects is the Data Discovery and Access service (DDAS), which provides easy access to diverse distributed datasets offered by Blue Cloud Data Infrastructures (BDIs) and other data providers relevant to the FAIR-EASE project. These datasets have been harmonised to adhere to the ISO19115 standard, thanks to the geoDAB broker, which is an essential component of DDAS. The geoDAB broker connects to a variety of heterogeneous data services, that comply with standards like the Catalogue Service for the Web (CSW), Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and ISO 19115/19139) and collects metadata records heterogeneously encoded and harmonises them to meet the ISO19115 metadata model. Although the web services comply with the above-mentioned standards, machine to machine access is not yet enabled, as their descriptions are not harmonised nor are they machine actionable. This hinders web service discoverability, accessibility interoperability and reusability.

Subsetting services

A similar situation exists with subsetting services. Both FAIR-EASE and Blue-Cloud 2026 follow user requirements for more direct data access and data processing, offering services to facilitate access to subsets of data, where possible. Often original datafiles are too large or too many, and contain too many parameters the user may not be interested in. The advancement in subsetting software has led to an increase in BDIs providing subsetting services. In order to facilitate harmonised, machine-actionable, and ultimately scalable access to subsetting services, it's imperative that these services also provide harmonised and machine-actionable descriptions, aligned with the above-mentioned data services.

Research question

In the FAIR-EASE and Blue-Cloud projects we encountered a variety of machine-to-machine data access services that require harmonisation. This led us to focus on the following main research question:

"Define a schema to describe web services offered by a diverse range of service providers using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple providers."

This will lead to more FAIR services as it will increase their discoverability and automatic utilisation. Consequently, software applications capable of universal discovery and consumption of compliant services will emerge, simplifying development efforts. This question is equally relevant to the interoperability of EOSC services where no generally accepted solution exists.

An initial analysis of existing ontologies and gaps could lead to a set of recommendations. This will be presented during the session.

Service semantic artefacts

In order to achieve FAIR data access services, these must be semantically described using FAIR ontologies/vocabularies in a standardised manner, so that the following information is made available: what the service does; how it works; how to access it. This must not be mistaken with FAIRsFAIR's FAIR Assessment Framework for Data Services (1) which is a set of guidelines used to assess on a high level how well data services support FAIR data, and not the FAIRness of the actual services.

A preliminary list of available vocabularies, ontologies and standards for describing services has been drafted and includes OWL-S[9], OpenAPI Specification[10], smartAPI[11] project, Dublin Core Metadata Initiative[12] (DCMI), Data Catalog 3 (DCAT 3)[13], Hydra[14] and schema.org[15]. Schema.org guidance on service descriptions is provided by both the Earth Science Information Partners (ESIP) science-on-schema.org (SOSO)[16] cluster and Ocean Data Information System (ODIS)[17] although they cannot yet cover our needs for machine to machine (M2M) interoperability and actionability.

Application in the frame of FAIR-EASE

As part of the wider FAIR-EASE asset catalogue, the aim is to compile a list of data access services with standardised descriptions that will not only help developers navigate the API documentation and software libraries, but also enable machines (generic clients) to effectively select and execute the targeted data-requests. In a subsequent phase, this process might be further automated so that a generic request from a Virtual Research Environment (VRE) can select an appropriate data access service along with instructions on how to access it, resulting in a list of specific subset URLs for seamless retrieval and further processing.

References

Koers H., Herterich P., Hooft R., Gruenpeter M., & Aalto T., (2020). *M2.10 Report on basic framework on FAIRness of services (1.0)*. Zenodo. https://doi.org/10.5281/zenodo.4292599

Poster session

[9] https://www.w3.org/submissions/OWL-S/

[10] https://spec.openapis.org/oas/v3.1.0

[11] https://smart-api.info/

[12] https://www.dublincore.org/

[13] https://www.w3.org/TR/vocab-dcat-3/

[14] https://www.hydra-cg.com/spec/latest/core/

[15] https://schema.org/

[16] https://github.com/ESIPFed/science-on-schema.org

[17] https://oceaninfohub.org/odis/

Linked Open Data in ISPRA: Semantic web for Marine Observations

Marco Picone (marco.picone@isprambiente.it), Elio Giulianelli (elio.giulianelli@isprambiente.it), Giulio Settanta (giulio.settanta@isprambiente.it), Gabriele Nardone (gabriele.nardone@isprambiente.it), Carlo Lo Re (carlo.lore@isprambiente.it)

Istituto Superiore per la Ricerca e la Protezione Ambientale - ISPRA (Italy)

Linked open data in ISPRA

Environmental and oceanographic data plays a pivotal role in creating sustainable knowledge chains by providing valuable insights for reasearchers, decision-makers, communities, and citizens. However, disparate datasets that require manual manipulation on spreadsheets can add significant complexity, generate more questions than answers, and hinder the creation of useful knowledge. The large volume of data, driven by the new technologies applied to the observation (such as new sensors, satellite observations, and enhanced transmission and storage capacities), the need to monitor numerous events along coastlines, and the growing trend of "open data," has led to a substantial influx of data on the web. However, increasingly, we are faced with a large number of "values" lacking a clear informative description of what they represent and how to manage them. It is necessary to organize this information to avoid the "infopandemia" that only generates chaos.

The application of semantic technologies and the FAIR principles is recommended by a significant number of agreements, guidelines, and laws in the European context. In this framework, the philosophy of Linked Open Data (LOD) is based on the idea of using specific web resources and connecting them through similarity or equivalence mechanisms, facilitating the improvement of the initial information set according to rules expressed by Tim Berners-Lee in 2006 and then by the World Wide Web Consortium (W3C) standards. Linked ISPRA is a project of ISPRA (Italian Institute for Environmental Protection and Research), initiated in 2016 based on the development of an application, according to World Wide Web Consortium (W3C) specifications, to produce and publish LOD. The Linked ISPRA platform (https://dati.isprambiente.it), also developed through participation in two CEF Digital programs (Open-IACS project and WHOW project), currently provides access to several national environmental datasets under the terms of the CC-BY 4.0 license through a technology stack that is entirely developed with open-source components. To publish the data, an OWL conceptual model has been defined to describe involved entities and the relationships between some datasets. In this work, we focus on the published oceanographic information archives from the marine monitoring networks managed directly by ISPRA. The information is related to the National Wave Network, which defines the physical state of the Italian seas through 15 oceanographic buoys, and the National Tide Gauge Network, composed of 36 measurement stations along the Italian coast, to observe sea levels and meteorological parameters. Both networks have operated almost continuously since the eighties, and data are published in real-time.

The Semantic web for marine monitoring networks

The semantic web provides a common framework that allows data to be shared and reused by humans and machines across applications. This data management is strongly enhanced by the application of knowledge representation languages that use concepts, classes, and relations typically modeled throughout ontologies, vocabularies, and thesauri. In this framework, ISPRA develops different ontologies collected into the "ispra-network" repository. For example, the "ispra-top" module defines general classes (Concept, Location, Collection) useful for the entire data environment, while “ispra-plc” is created to describe all possible locations (administrative units, seas, rivers). The “ispra-emf” models the classes and properties for environmental monitoring facilities (including marine monitoring networks) according to the related INSPIRE data model. In particular, the ontology module represents the MonitoringObject INSPIRE class in various possible alternatives, such as MonitoringFacility and MonitoringNetwork. The module also presents classes and properties describing observations, collections of observations, sensors, platforms, networks, observation parameters, procedures, values, etc. The module is inspired by the SSN and SOSA ontologies, extending them when necessary for ISPRA applications, such as the use of observation series through the ObservationSeries class. The same approach is used to represent indicators and collections of indicators.

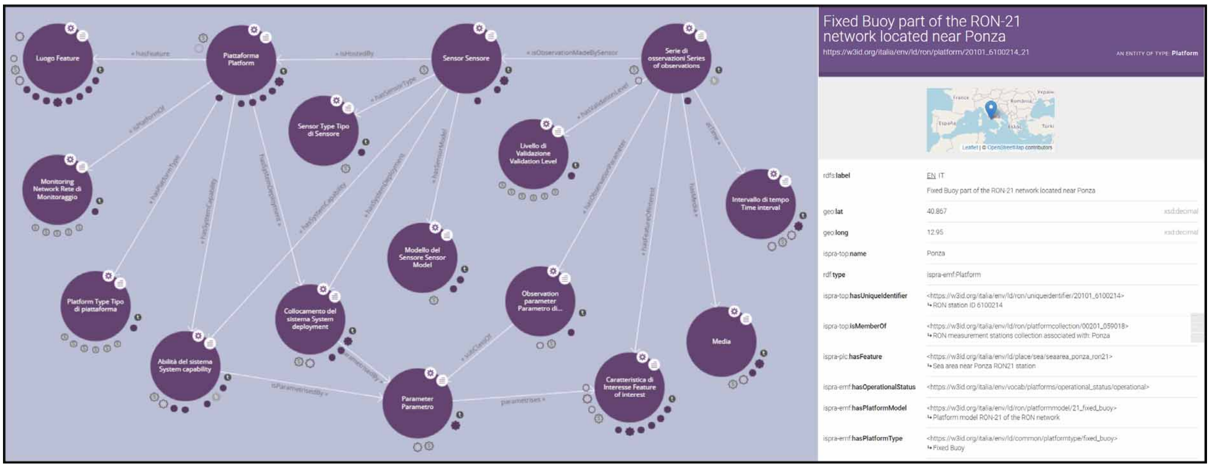


Figure 1 A brief representation of the main classes described by the ispra-emf ontology and an example of data representation (for Ponza buoy).

Several vocabularies have been published to describe the data validation level, the type of measurement station, the operational status, the water bodies, and the water quality. According to the developped ontologies and vocabularies, all data and metadata related to the marine monitoring networks are published.
An automatic process, described in Figure 2, allows the publication of all historical data and real-time information collected by the monitoring stations. In particular, the triplification phase transforms the static information (values and metadata) into meanings and relations by creating a list of simple declarative statements in the form of “triples”.

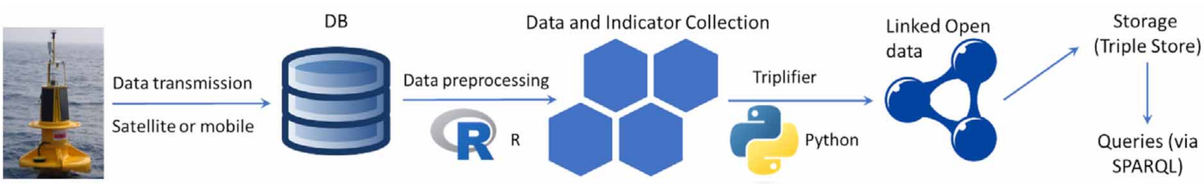


Figure 2 Data process and software used for LOD production.

Each piece of information is represented through a subject, a predicate (property), and an object, all of which are URIs, making resources on the web easily findable and linkable. While the representation as triples may seem simple in human language, the sequence of multiple URIs is

hard to read for standard users and is more suitable for machine applications. A significant effort in ISPRA has been applied to make the triples readable for standard users.
Users have the ability to directly navigate the data, investigating the relationships among several resources published by ISPRA, and also moving to resources belonging to other domains and institutions that are linked. Additionally, it is possible to query the necessary information through the SPARQL endpoint (https://dati.isprambiente.it/sparql).

CoreTrustSeal: Accreditation from the perspective of a National Oceanographic Data Centre (NODC)

Mark Hebden (mahe@noc.ac.uk), Lesley Rickards (ljr@noc.ac.uk)

National Oceanography Centre - NOC (UK)

Introduction

Independent certification is a well-recognized means for a data repository to demonstrate trustworthiness to potential users of its services. Certification can also provide a useful internal framework for data repositories to review their current strengths and weaknesses, benchmark against similar organizations and be used as a tool to plan continuous improvement. For a certification framework to be widely adopted it needs to be inclusive - for example, be affordable to smaller organizations and to bridge data domains. It also needs to be respected by the target audience, ideally being driven by that community.

The British Oceanographic Data Centre (BODC), managed and operated by the National Oceanography Centre (NOC), is the UK's national repository for preserving and distributing oceanographic and marine data. BODC is a component of the UK's Natural Environment Research Council's (NERC) Environmental Data Service (EDS) and a Data Archive Centre (DAC) within the UK's Marine Environmental Data and Information Network (MEDIN). Globally, BODC operates as a National Oceanographic Data Centre (NODC) within the International Oceanographic Data and Information Exchange (IODE) programme of the Intergovernmental Oceanographic Commission (IOC) of UNESCO.

During 2020, BODC commenced a project to prepare an application for CoreTrustSeal – an internationally-recognized framework for data centre accreditation and one which, crucially, has endorsement from key networks that BODC operates in. This paper describes the current CoreTrustSeal framework, outlines the associated accreditation process and highlights the wider benefits of acquiring accreditation.

CoreTrustSeal framework and background

The origin of CoreTrustSeal was a Research Data Alliance (RDA) Repository Audit and Certification Working Group (2013-2016) with members drawn from the World Data System (WDS) Scientific Committee and the Data Seal of Approval (DSA), both of which had developed core/foundation level accreditation schemes, the former mainly from an Earth and Space sciences background and the latter from Humanities and Social Sciences. The goal was to harmonize basic certification requirements and procedures, ultimately setting the stage for a global shared framework including other standards. The DSA and WDS were both basic, lightweight mechanisms for repository assessment with many commonalities. The outcomes of the Working Group were a set of common requirements, including guidance, backed up by a set of common procedures. CoreTrustSeal's 16 requirements cover the categories of Organisational Infrastructure, Digital Object Management and Information Technology and Security as shown in Table 1 below. Background and context information is also requested. A Testbed was also developed to allow 'real-world' evaluation of the requirements and procedures.

In 2017 the DSA and WDS launched a new certification organization, CoreTrustSeal, based on the DSA–WDS Core Trustworthy Data Repositories Requirements. This subsequently became a legal entity under Dutch law, replacing the DSA certification/WDS Regular Members certification. CoreTrustSeal is an international, community based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures governed by a

Standards and Certification Board composed of 12 elected members representing the Assembly of Reviewers. Currently (November 2023) there are about 100 certified repositories.

CoreTrustSeal Requirements 2023 – 2025		
R0. Background Information & Context		
Organisational Infrastructure		
Mission & Scope (R01)	Rights Management (R02)	Continuity of Service (R03)
Legal & Ethical (R04)	Governance & Resources (R05)	Expertise & Guidance (R06)
Digital Object Management		
Provenance and authenticity (R07)	Deposit & Appraisal (R08)	Preservation plan (R09)
Quality Assurance (R10)	Workflows (R11)	Discovery and Identification (R12)
Reuse (R13)		
Information Technology & Security		
Storage & Integrity (R14)	Technical Infrastructure (R15)	Security (R16)

Table 1 Current CoreTrustSeal accreditation framework.

BODC's CoreTrustSeal experience

In total, gaining CoreTrustSeal accreditation took BODC around 2.5 years – from an initial planning and assessment phase through to certification by CoreTrustSeal. A coordinator was assigned to the project and a small working group of experts formed, each championing a small number of CoreTrustSeal Requirements. At this stage any perceived weaknesses were either addressed or captured for resolution as part of BODC's longer term programme of work. Self-assessment against the various Requirements informed the level of maturity that was stated during the application process. The CoreTrustSeal accreditation process is very much an evidence gathering exercise to demonstrate fulfilment of these Requirements. Known shortfalls are acceptable if the Data Centre demonstrates a commitment to resolution ahead of re-accreditation. Written applications to CoreTrustSeal are made *via* the online CoreTrustSeal Application Management Tool (AMT). Upon submission (and payment of a review fee) two anonymous reviewers are assigned to the review process and an initial review-feedback loop initiated. Ultimately, BODC underwent just the two review cycles, with supplementary evidence, provided following preliminary review, sufficient to attain CoreTrustSeal.

Conclusions and post-accreditation activities

BODC has now entered a post-accreditation phase. The process of acquiring CoreTrustSeal has identified areas for ongoing continuous improvement (including areas needing greater transparency), which will be considered as part of planning for re-accreditation in 3 years' time. Additional next steps include:

- Applying for WDS membership
- Renewal of IODE accreditation (based on CoreTrustSeal with provision of additional evidence needed to meet IODE requirements)
- Contribution to the wider CoreTrustSeal community by joining the Assembly of Reviewers
- Sharing BODC's CoreTrustSeal experience with others

Marine data management is a dynamic discipline and one that is becoming increasingly complex as technologies evolve and the demand for information grows in response to societal and environmental needs. Data are often costly to acquire and unique, making their preservation essential. The expectation of National Oceanographic Data Centres (NODCs) is therefore great, as is the need for maintaining credibility from data providers and end users. NODC accreditation, from a trustworthy source, is a useful tool for conveying that trustworthiness.

Reference

Rickardsn L., Vardigan M., Dillo I., Genova F., L'Hours H., Minster J.B., Edmunds R., & Mokrane M., (2016). *DSA–WDS Partnership: Streamlining the landscape of data repository certification*. SciDataCon 2016, Denver, CO., 2016. https://doi.org/10.5281/zenodo.252417

Information systems and tools to support Antarctica communities

Rachele Bordoni[1] (rachele.bordoni@grupposcai.it), Giulia Dapueto[1] (giulia.dapueto@grupposcai.it), Francesco Misurale[1] (francesco.misurale@grupposcai.it), Beatrice Maddalena Scotto[1] (beatrice.scotto@grupposcai.it), Petra ten Hoopen[2] (peopen@bas.ac.uk), Jean-Baptiste Sallée[3] (jean-baptiste.sallee@locean-ipsl.upmc.fr), Andrew Meijers[2] (andmei@bas.ac.uk), Ruth Mottram[4] (rum@dmi.dk), Antonio Novellino[1] (antonio.novellino@grupposcai.it)

[1]ETT S.p.A., Department of Earth, Environment and Life Science, University of Genoa (Italy)
[2]British Antarctic Survey, NERC (UK)
[3]Laboratoire d'Océanographie et du Climat (France)
[4]Danish Meteorological Institute (Denmark)

Introduction

Located at the nexus of Earth's climate and oceanic processes, the Southern Ocean plays a central role in regulating global climate systems and carbon cycling. Its importance demands international cooperation to comprehensively observe and understand this vast and complex region. In response to these needs, the Southern Ocean Observing System (SOOS) was established in 2011 and acts as a unifying force, fostering collaboration and streamlining data delivery across stakeholders.

SOOS endorses programs and projects that support SOOS goals. Since 2017, SOOS and EMODnet Physics are collaborating for facilitating data sharing and data discovery [Bricher et al., 2018], where EMODnet Physics provides the data management framework to integrate existing data products from multiple data resources. This collaboration was strengthened in 2020 when a Memorandum of Understanding (MoU) for long term partnership was signed.

While SOOS relies on EMODnet for in situ data management, two more projects, SO-CHIC and OCEAN:ICE, are participating in the development of data access and data discovery by supporting and contributing to the development of the SOOSmap. SOOSmap, launched in October 2017, is an interactive online platform for visualising and distributing oceanographic data. It offers curated and standardised data from various oceanographic and Antarctic research programs globally. Users can access circumpolar data from a large number of nations and different scientific disciplines. Central elements of the SOOS data vision encompass tools for discovering data and metadata, coordinating fieldwork, and establishing connections between data collections and analysis tools [Gorringe et al., 2021].

The Southern Ocean Carbon and Heat Impact on Climate (SO-CHIC), a project funded by the European Union's Horizon 2020 research and innovation programme (grant agreement N° 821001), is dedicated to monitoring heat and carbon budgets. It actively promotes collaboration with various European, national, and international initiatives to tackle the challenges posed by climate change.

The Ocean-Cryosphere Exchanges in Antarctica (OCEAN:ICE) project, supported by the Horizon Europe research and innovation programme and UK Research and Innovation (grant agreement N° 101060452), seeks to enhance our comprehension of the Antarctic Ice Sheet's impact on global climate, deep water formation, and ocean circulation. Moreover, OCEAN:ICE seeks to reduce uncertainties in projections of freshwater fluxes and sea-level rise associated with melting, while also improving ocean observations and providing free and open access to the data generated.

Together, these collaborative efforts represent a significant step toward a more comprehensive understanding of the Southern Ocean dynamics and their profound implications for global climate and oceanic processes.

Methodology

Central to the mission of both SO-CHIC and OCEAN:ICE is the establishment of a data management infrastructure, including both backend and frontend components, in line with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, ensuring seamless accessibility and utility for stakeholders and the scientific community. This robust data management framework is set up through the use of contemporary tools and adherence to relevant standards, facilitating swift adoption and widespread availability of the generated data across prominent marine data integration entities such as SOOS, EMODnet and the Copernicus Marine Environment Monitoring Service (CMEMS), among others.

In this workflow, the SOOSmap (Figure 1a) serves as a user-friendly interface for external users to seamlessly access project data. Leveraging and powered by the SO-CHIC map viewer (Figure 1b), SOOSmap integrates data from SO-CHIC, supplemented by relevant project-specific information. The ongoing maintenance and development efforts are also directed towards the OCEAN:ICE project, with the understanding that OCEAN:ICE relies on SOOSmap for data visualisation. New project data will be organised on the backend and seamlessly linked into SOOSmap. OCEAN:ICE is utilised for monitoring and furthering the developments initiated by SO-CHIC, contributing to the continuous enhancement of SOOSmap features.

Figure 1 a) SOOSmap viewer, b) SO-CHIC map viewer.

The SO-CHIC map viewer allows users to explore both raw observational data and processed or aggregated versions over different time periods and spatial regions. While SO-CHIC partners are the primary users and providers, the tool integrates data from multiple sources in a catalogue format. Users can conveniently access data packages/products and explore available parameters for each platform. Selection of parameters and timescales results in pop-ups showing data product cards with key features. Further exploration includes pinpointing measurements on the map and detailed platform information. The user can access and download data in different formats either directly from the product cards or via the SO-CHIC ERDDAP data server. The data can also be accessed from EMODnet geoviewer.

The OCEAN:ICE portal will be the result of collaborative efforts between developer and user teams, who will prepare test datasets for both backend and web user interfaces to improve accessibility and usability.

Conclusion

The collaboration between SOOS, EMODnet, SO-CHIC and OCEAN:ICE significantly increased the level of open access, interoperability, and data transparency, as well as empowered stakeholders and fostered collaboration and innovation in the community.

References

Bricher P., Novellino A., Gorringe P., Alba M., Zhang J., & Proctor R., (2018). *SOOSmap brings circumpolar Southern Ocean data to a computer near you*. EGU General Assembly Conference Abstracts, p. 15262.

Gorringe, P., Bricher, P., Alba, M., Bonofiglio, L., & Novellino, A., (2021). *Southern Ocean data: A community effort to build a data ecosystem.* Bollettino di Geofisica, 12, 159.

https://www.soosmap.aq/, https://map.sochic-h2020.eu/

WEB GIS to access ARGO floats Black Sea database

Elena Zhuk (alenix@gmail.com), Natalya Markova (n.v.markova@gmail.com)

Marine Hydrophysical Institute, Rassian Academy of Science (Russia)

The free access data archives of the ARGO project (https://fleetmonitoring.euro-argo.eu/) is a source of data for the Black Sea ARGO database. It includes measurements obtained from profiling floats that are both ongoing and completed. The data includes satellite positioning information, observational oceanographic parameters, technical details and settings for each launched float as well as metadata.
The current work makes use of the ARGO float data array, which contains data from 2005 to 2022. The ARGO data are archived in separate files with metadata that offer information on the drift depth to the profile point, as well as files in the netCDF and CSV formats that provide data about the location, float ID, time of the profiling, and measured parameters. Nearly every float has the drift depth information for each cycle.
The database's structure was developed after the floats' data structure was analyzed, and python modules were built for information parsing and database table filling. Postgresql is used as a DBMS. The database consists of metadata table (trajectories) that has key fields connection to the measurement data table (profiles).
A user friendly web interface has been developed with all the features required for managing the database, figuring out Lagrangian velocities, samplings, and displaying profile data.
It available at http://bod-mhi.ru/ff/

Using the web user interface, you can:

- Select data by ID float. Here, the map shows a trajectory made up of every profiling cycle for the whole time that was accessible. Additionally, you can get profiling data at a chosen time; a pop-up window displaying the profile will appear.
- Select data based on the designated time interval and ARGO float identifier: The only profiles that appear are those that fall into the specified time frame.
- Select data by rectangle region, time intervals, water depth;
- Calculate and display the Lagrangian velocities for floats at a certain water layer;

An example of an interface for data sampling and viewing is shown in Figure1 and Figure 2. The estimation of Lagrange velocity in water layers was described in Markova and Bagaev [2016]. The use of the GIS software module for ARGO data analysis significantly simplifies the work with oceanographic data for the Black Sea region and their scientific analysis. Simultaneously, the software module presupposes that velocities can be promptly recalculated upon receipt new observational data.

Acknowledgements
The Black Sea Argo database and web interface for it was developed within the framework of the MHI RAS state task on themes No. FNNN-2023-0001 "Ensuring climate and bio-geochemical monitoring of carbon fluxes in the Black Sea base on long-term observations data and numerical simulation results" ("Carbon") and No. FNNN-2021-0003 "Development of operational oceanology methods based on interdisciplinary research of processes of the marine environment formation and evolution and on mathematical modeling using data of remote and contact measurements" ("Operational oceanology")

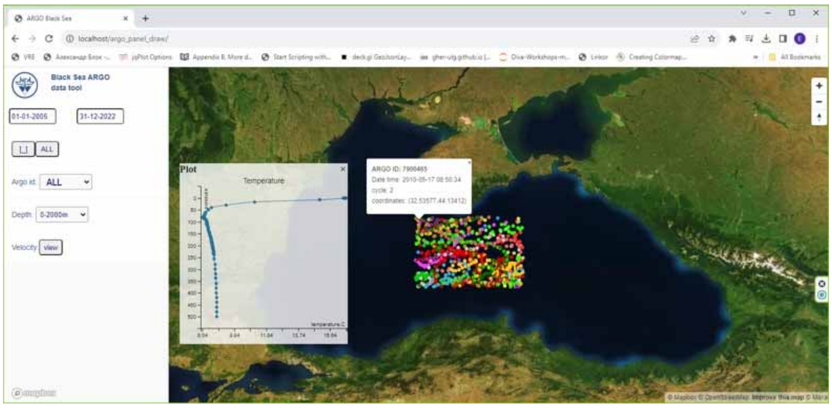


Figure 1 User Interface, data selection by rectangle region.

Figure 2 User Interface, data selection for layer 600-800 m.

References

ARGO portal, https://fleetmonitoring.euro-argo.eu/

Markova N.V., Bagaev A.V., (2016). *The Black Sea Deep Current Velocities Estimated from the Data of Argo Profiling Floats*. Physical Oceanography, n.3, p. 23-35; https://doi.org/10.22449/1573-160X-2016-3-23-35

Validated oceanographic Near-Real-Time data to support civil protection

Alexia Cociancich, (acociancich@ogs.it), Marina Lipizer (mlipizer@ogs.it), Elena Partescano (epartescano@ogs.it), Alessandro Altenburger (aaltenburger@ogs.it), Marco Jokic (mjokic@ogs.it), Erik Geletti (egeletti@ogs.it), Maria Eugenia Molina Jack (mmolinajack@ogs.it), Matteo Vinci (mvinci@ogs.it), Alessandra Giorgetti (agiorgetti@ogs.it)

Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)

Near-Real-Time (NRT) data from operational oceanography (e.g., from in situ meteo-oceanographic buoys, lagrangian and floating sensors, ...) provide fundamental information to address environmental emergencies and for civil protection needs. Timely availability of validated data is crucial and NRT data Quality Control (QC) is required to provide usable information to address societal needs.

For over 10 years the Italian National Oceanographic Data Center (NODC) has been managing NRT data from fixed stations positioned in the Adriatic Sea [https://nodc.ogs.it/geoportal; Partescano et al., 2017].

The network of fixed monitoring stations (Figure1), which contributes to the Italian Fixed-point Observatory Network for marine environmental monitoring – IFON [Ravaioli et al., 2017] has been mainly developed in the framework of a multi-year collaboration with the Regional Civil Protection Service (Protezione Civile Friuli Venezia Giulia) which has facilitated the installation of buoys, wave meters and river current meters in the north-east of Italy; in addition there are other buoys in the north and south Adriatic which are an integral part of international infrastructures such as EMSO (European Multidisciplinary Seafloor and water-column Observatory), ICOS (Integrated Carbon Observation System), JERICO (Joint European Research Infrastructure of Coastal Observatories) and ITINERIS (Italian Integrated Enviromental Research Infrastructures System).

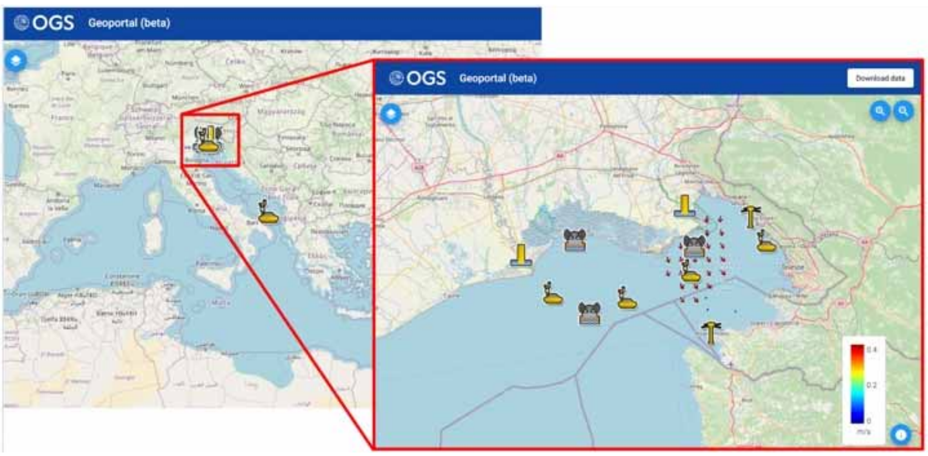


Figure 1 NODC network of fixed monitoring stations.

Relying on continuous monitoring, having data with a latency of a few minutes, and evaluating the quality of the data collected are the key elements and the challenges for the efficient risk prevention activity in the area. In case of errors in data transmission or data quality, it is important to have efficient alerts, analyze the causes, and act promptly to address problems.

Due to the characteristics of the data transmission systems, i.e., considering that the platforms are often located in areas where telephony does not have optimal coverage, it is important that the raw data is as compact as possible, in order to avoid problems relating to packet loss; for this reason the tendency is to use strings according to NMEA (National Marine Electronics Association) standard protocol and format or similar. With this format, the meaning of the data can be deduced exclusively from the position that the field occupies within the string, and this strategy exposes the flow to a practical problem regarding the correct mapping of the fields of this format with the corresponding parameters and the related units of measurement: sometimes, during maintenance operations, instrument configurations are changed compared to previous deployments, and this leads the system to the risk of inserting incorrect data into the database.

The NODC, taking advantage of the resources provided by Google Drive and Google App Scripts, has developed a code which, thanks to a simple configuration file, decodes the messages received from the instruments directly in the cloud in XML format, making them instantly available to the staff involved in the activity. The files are converted and enriched with the metadata necessary for a first visual verification and for automatic insertion into a relational database.

Once the data has been entered into the Near-Real-Time database, SQL procedures, performed on an hourly basis, are responsible for checking the ingested data values based on specific thresholds, defined according to climatological ranges in the specific area, and assigning an evaluation of the quality of the data itself. The data received and the related associated quality flags are made immediately publicly available through the ERDDAP service https://nodc.ogs.it/erddap).

The periodic check of the flags assigned to the data allows to have rapid feedback on the quality of the data and possibly formulate hypotheses on the anomalies found.

Acknowledgment

This work is supported by IR0000032 ITINERIS - Italian Integrated Environmental Research Infrastructures System

References

Partescano et al., (2017). *From heterogeneous marine sensors to sensor web: (near) real-time open data access adopting OGC sensor web enablement standards.* Open geospatial data, softw. stand. 2, 22. https://doi.org/10.1186/s40965-017-0035-2

Ravaioli et al., (2017). *The Italian Fixed-point Observatory Network for marine environmental monitoring – IFON*. In IMEKO TC19 Workshop on Metrology for the Sea, MetroSea 2017: Learning to Measure Sea Health Parameters (pp. 22-27). (IMEKO TC19 Workshop on Metrology for the Sea, MetroSea 2017: Learning to Measure Sea Health Parameters; Vol. 2017-October). IMEKO-International Measurement Federation Secretariat.

The need and use of marine data systems monitoring

Keila Lima (keila.lima@hvl.no), Rogardt Heldal (rogardt.heldal@hvl.no), Tosin D. Oyetoyan (tdo@hvl.no)

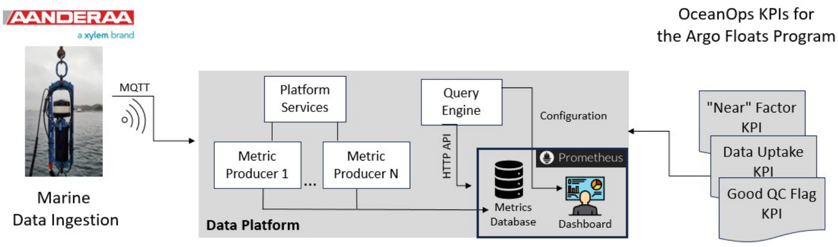
Western Norway University of Applied Sciences (Norway)

Introduction

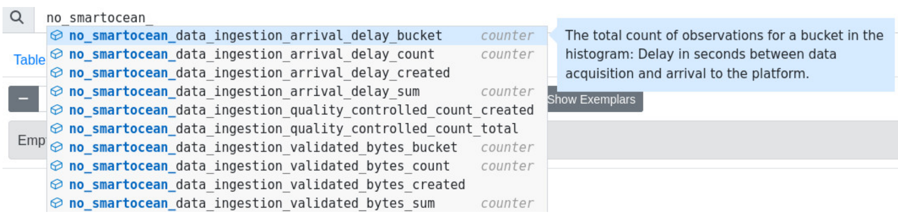
Marine data is increasingly becoming a critical asset for decision-making that impacts human lives in energy or food production sectors. As the majority of systems now operate around data for decision-making. Whether this process is done automatically by some artificial intelligence component or with human intervention, it is crucial to understand its quality to ensure decisions are made based on correct information. The EU has been working on regulations to avoid low-quality data in the application that can affect citizens' rights and to ensure that high-risk AI systems comply with such requirements as a factor to determine the trustworthiness of their systems [1]. While these requirements do not apply in sectors operating in the oceans yet, there is increasing concern regarding this topic and the need to comprehend oceans' role in the ecosystem for better decision-making and regulation of its use [Tan et al., 2022]. In the marine domain, many challenges can contribute to faulty data being injected into these decision-making systems, from sensor malfunctioning due to limited battery and the rough sea conditions to communication errors in the many unreliable channels used for data delivery. In addition to these technical limitations, deploying and maintaining sensors at sea is costly. Thus, sensors' operation must be maximized regarding "good" data production or timely detection and mitigation of "bad" data in the data processing pipelines. Furthermore, the volume of marine data shared is increasing due to the adoption of the FAIR principles not only at regional and international levels in programs such as the CMEMS or private initiatives like the Ocean Decade Corporate Data Group from IOC to facilitate the sharing of data from marine industries. Consequently, automated mechanisms must be in place during systems' operation to identify the adequacy of ingested data to its intended use [Moses, 2022], especially if data comes from external systems. We have ongoing research and development (R&D) studies addressing data quality and automatic monitoring of marine data and platforms used to share marine data. In the next section, we describe some of the R&D results.

Marine Data Quality Dimensions and Metrics Monitoring

The required quality of data can vary among different marine application use cases. Models such as the ISO/IEC 25012 for quality of data products can be used to define concrete dimensions to evaluate data. While most marine quality control procedures focus on accuracy, consistency, or completeness of measurements [Tan et al., 2022], other ones, such as currentness, availability, precision, or accessibility (to name a few), should be considered according to their relevance to the use case [Nguyen et al., 2023]. Within the SFI Smart Ocean project (sfismartocean.no), our research investigates which key performance indicators (KPIs), and metrics must be collected and monitored to understand if data has an adequate quality level for different use cases in the marine domain. We have developed an observability subsystem for our marine data platform (smartoceanplatform.github.io) to start monitoring its data and ingestion service. This monitoring establishes KPIs as thresholds for the data quality metrics being measured during the platform's operation. The subsystem used a model-driven approach to interact with an open-source monitoring toolkit to store the data platform's runtime metrics generated upon data ingestion. Furthermore, three data quality-related KPIs from the OceanOPS international initiative (ocean-ops.org/board?t=argo) were selected to conduct the case study to assess data collected and delivered to the platform (see Figure 1). As a result, we could identify violations of these KPI targets in runtime. Lastly, the observability subsystem implemented an existing multi-level model framework for quality evaluation systems. With our model-driven software engineering approach, the platform operators only need to model the domain, abstracting from learning query language rules and constraints for specifying the KPIs. Additionally, metrics and target violation detection are handled in the background by the solution using the models as configuration files (low code solution) for the automatic parametrization of dashboards, alerts, and notifications.



a) KPI engine prototype experiment setup from [Lima et al., 2023].



b) Data-related metrics collected during the experiment runtime from [Lima et al., 2023].

Figure 1 Marine data observability in the Smart Ocean Platform using KPIs and the Prometheus Monitoring toolkit.

Lastly, integrating monitoring components that can observe the behavior of the data pipeline can be a means to implement awareness of this problem. If we consider the MAPE-K architecture for self-adaptive systems [Kephart et al., 2003], we can note that monitoring is the first component in the implementation of self-adaptive behavior (the aspired “smartness”). Triggering alarms, retrieving contextual information, or correcting data upon detecting “bad” data could be an example of such behavior. Having confirmed this need and implemented an engine to detect KPI target violations, future work will focus on automatically detecting and mitigating the causes, improving the decision-making based on the system. The metadata will be crucial for providing contextual information to understand the root causes of errors.

References

Kephart J.O., et al., (2003). *The vision of autonomic computing*. Computer 36.1: 41-50.

Lima K. et al., (2023). *Marine Data Observability Using Kpis.* Zenodo, https://doi.org/10.5281/zenodo.8223694

Moses, B., (2022). *The rise of data observability: architecting the future of data trust*. Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining.

Nguyen N.T., et al., (2023). *Synthesized data quality requirements and roadmap for improving reusability of in-situ marine data*. 31st International Requirements Engineering Conference. IEEE.

Tan Z., et al., (2022). *Quality control for ocean observations: From present to future*. Science China Earth Sciences: 1-18.

[1] EC. *Regulatory Framework Proposal on Artificial Intelligence*. Digital Strategy EC (2022). Online.

Aligned development of digital twins across communities - bridging the gaps between different communities

Justin Buck (justin.buck@noc.ac.uk), Alexandra Kokkinaki (alexk@noc.ac.uk)

National Oceanography Centre - NOC (UK)

Introduction

The National Oceanography Centre, (NOC) UK, collaborate at national, European and international levels on the development of digital twins. A key driver in digital twins development is the goal of achieving interoperable and aligned digital twin developments across communities such as research, industry, government, etc. NOC has established collaborations across the spectrum of digital twins developers and contributed to interoperability efforts at varying scales. This paper will showcase those efforts and highlight key developments in achieving interoperability of digital twin components that are contributing to the goals of interoperable systems across the national to international scales.

National alignment

At the National scale the NOC has led the NERC funded Information management Framework for environmental digital twins (IMFe) and Piloting the IMFe projects. These projects have developed a set of recommendations [Siddorn et al., 2022] on how to develop digital twins across the environmental community based on the concept of digital commons. The Piloting an IMFe project has tested the recommendations using a Haig Fras marine protected area demonstrator including the development of an exemplar asset register as a prototype for an interoperable description of digital twins and their assets. At the community level NOC is collaborating with DT Hub that brings together 4,000 digital twin stakeholders and the Turing Research Innovation Cluster for environmental digital twins.

European alignment

At a European level the NOC are a partner in the FAIR-EASE project which aims to enhance Earth system, environment, and biodiversity observation and modeling through customized and integrated services. This involves improving components in collaboration with user communities, the European Open Science Cloud (EOSC), and research infrastructures. The conceptual architecture the IMFe proposes is based on the concept of a digital commons, embracing a narrow middle philosophy to its design and applying systems thinking principles intrinsically in the design. We are making sure that the components and assets developed in the FAIR-EASE project, will be interoperable both with EOSC and the IMFe. The current focus for FAIR-EASE is the interoperability of data and services.

International alignment

At an international level NOC are a member of the UN decade Digital Twin of the ocean (DITTO) and the associated DITTO TURTLE action that focuses on the interoperability of digital twins. TURTLE bring together key internal bodies like the open geospatial consortium with global actors such as the EU ILIAD project to provide a forum to facilitate convergence on digital twin development. The NOC-led IMFe project is a formal project under the DITTO programme.

Resulting challenges and steps towards alignment

The broad reach of NOC activity on digtial twins development has given a perspective across the activity that allows us to draw conclusions on key activities and standards to be aware of when developing digital twins. These include the use of linked open data following the OGC models, the adoption of the narrow middle concept proposed within the asset commons and the need to build crosswalks and mappings where multiple standards exist in different communities to achieve interoperability. We will present these recommendations along with the key networks to engage with.

EMSO ERIC Metadata Harmonization Efforts

Enoc Martínez[1] (enoc.martinez@upc.edu), Maurice Libes[2] (maurice.libes@osupytheas.fr), Claudia Fratianni[3] (claudia.fratianni@ingv.it), Elena Partescano[4] (epartescano@ogs.it), Alexia Cociancich[4] (acociancich@ogs.it), Sara Pensieri[5] (sara.pensieri@cnr.it), Francesco Paladini de Mendoza[6] (francesco.paladinidemendoza@cnr.it), Ivan Rodero[7] (ivan.rodero@emso-eu.org)

[1]Universitat Politècnica de Catalunya (Spain)
[2]OSU Pytheas CNRS (France)
[3]Istituto Nazionale di Geofisica e Vulcanologia – INGV (Italy)
[4]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[5]Consiglio Nazionale delle Ricerche - Istituto per lo Studio degli Impatti Antropici e la Sostenibilità nell'ambiente marino - CNR – IAS (Italy)
[6]Consiglio Nazionale delle Ricerche - Istituto di Scienze Polari - CNR-ISP (Italy)
[7]EMSO ERIC Central Management Office (Italy)

Introduction

Data collection efforts in the marine domain have expanded significantly over the years. However, a fundamental challenge persists - harmonizing data and metadata across the community. Data harmonization is the process of ensuring that data collected from various sources are compliant with FAIR data management principles, which articulate the attributes data need for enabling and enhancing reuse by machines and humans. In the marine domain, the heterogeneity of data formats, quality control procedures, and metadata protocols has resulted in a fragmented and often inefficient landscape. This fragmentation not only hinders scientific progress but also complicates the implementation of effective policies and management strategies for the world's oceans.

EMSO ERIC Data Services

EMSO ERIC (European Multidisciplinary Seafloor and water-column Observatory) is a pan-European distributed research infrastructure whose main role is to provide high-quality long-term time series. It comprises 14 fixed-point multidisciplinary Regional Facilities (RF) in key seafloor and water-column environmental sites.

Since EMSO ERIC is a distributed infrastructure, its data services are built following a decentralized approach. Every regional facility is in charge of hosting and curating its data, usually following different standards and methods due to historical reasons, national requirements, and the type of data stream (near-real-time or delayed-mode). Data is shared across EMSO ERIC using a federation of ERDDAP servers [Queric et al., 2021]. The rest of the EMSO ERIC data infrastructure relies on this ERDDAP federation to feed data and metadata to the rest of the services.

However, federating data is not enough for a data infrastructure. To build data services and products on top of a federated data source, it is of the utmost importance that the underlying data, the ERDDAP federation in this case, provides interoperable data and metadata in a coherent and standardized manner.

Building on previous initiatives

Over the years, there have been many different initiatives on data harmonization in the ocean observing communities, such as OceanSITES, Climate and Forecast (CF), SeaDataNet, and British Oceanographic Data Centre (BODC) vocabularies, among others. The EMSO ERIC

Metadata Specifications build on these experiences with the addition of further attributes to be compliant with the need for traceability of the data. Therefore, the EMSO ERIC metadata specifications can be regarded as a superset of OceanSITES specifications, compliant with CF standards, with some additional metadata coming from SeaDataNet specifications, adding meaningful information to the distributed datasets.
The specifications extensively use controlled vocabularies, solving the problem of ambiguity associated with data, clarifying the meaning of common concepts, and enabling long-term data interoperability and reusability through more precise definitions and machine-readable online references. Domain-specific metadata attributes rely on the NERC Vocabulary Server (NVS), the European Directory of Marine Organisations (EDMO) and the Research Organization Registry (ROR) are used for organizations, and SPDX is used for licensing.

Dataset verification

Instead of building a static text-based document, the specifications are managed in a GitHub repository written in a markup language accessible to both humans and machines[18]. Taking advantage of the machine-actionable nature of the specifications, a verification tool has been implemented[19]. Based on a set of compliance tests defined in the specifications, this tool can ensure that the fields are correct both syntactically and semantically.
On the one hand, several tests are executed to ensure the syntactic correctness of the datasets - data type, all required fields are filled, etc. On the other hand, using proper vocabulary terms is ensured to achieve semantic interoperability. For instance, the unit definition must be a term within the NVS P06 vocabulary, and the data license must be a term within the SPDX repository.
As output, this verification tool provides an overall score of the harmonization as a percentage. Additionally, it provides an evaluation of all individual terms required within the metadata, providing tips to correct errors and improve the overall harmonization.
Based on the ERDDAP federation of fully harmonized data across EMSO ERIC regional facilities, data products to foster multidisciplinary scientific research will be produced.

Conclusion

This article provides insight into EMSO ERIC data harmonization processes to ensure FAIRness of data by overcoming the complexity and heterogeneity of different data and metadata formats. The use of semantic artifacts such as vocabularies and standards together with the use of ERDDAP as web-based service is a key element to achieving a high level of FAIRness, speeding up information and knowledge generation process.
These challenges were addressed by focusing on the use of common data formats and metadata standards, managing distributed data through an ERDDAP federation and using NetCDF files with standardized metadata from BODC, CF, OceanSites. Thus, EMSO ERIC provides homogeneous FAIR data.

Acknowledgements

The authors would like to acknowledge the effort and commitment of the EMSO ERIC Data Management Service Group, without the collaboration of its members this work would not have been possible. This work has been partially funded by the European Commission under the Geo-INQUIRE project (grant agreement 101058518).

[18] https://github.com/emso-eric/emso-metadata-specifications
[19] https://github.com/emso-eric/metadata-harmonization

References

Antoine Queric et al., (2021). *EMSO ERIC Data Services: managing distributed data through an ERDDAP federation*. IMDIS 2021 - International Conference on Marine Data and Information Systems, 12-14 April, Online, 2021.

Convergence and Innovation; how to build and break the marine information silos

Piotr Zaborowski[1] (pzaborowski@ogc.org), Rob Atkinson[2] (ratkinson@ogc.org), Bente Lilja Bye[3] (bentelil@hotmail.com), Arne Berre[4] (Arne.J.Berre@sintef.no), Raul Palma[5] (rpalma@man.poznan.pl), Marco Amaro Oliveira[6] (marco.a.oliveira@inesctec.pt)

[1]Open Geospatial Consortium (Poland)
[2]Open Geospatial Consortium (Australia)
[3]BLB (Norway)
[4]SINTEF (Norway)
[5]PSNC (Poland)
[6]INESC (Portugal)

Marine and Met-Ocean communities, working with one ocean and its local flavors, have been considerably successful in the implementation of FAIR principles. Through the years, under the leadership of vital national and international agencies and academia, common practices were developed, integrating marine and geospatial concepts and significantly contributing to environmental data interoperability. Several technologies emerged that are more or less aligned with each other, including most essential (Application Programming Interfaces) APIs and formats widely adopted. In addition to the growing legacy potential, it creates a high threshold for the entrants, limiting the potential data refinement, which could generate added value. The reality of the service provision brings additional levels related to the profile of the implementation and data constraints. The work on the Iliad APIs, DITTO, and FMSDI helped to understand the challenges of the long-term development of widely endorsed practices and standards. It allowed for the tailoring of data-level building blocks that could help further expand the marine data portfolio.

Multi-layer building blocks

There is already a revolution happening with cloud-native formats like Zarr, COG, GeoParquet, APIs and federation based on openEO/OGC Processes that leverage capabilities of the distributed infrastructure minimizing data exchange needs significantly.

Conversely, subdomains and digital twins' implementations go far beyond the agreed semantic conventions agreed interdisciplinarily for the outcomes and the services[20] descriptions simultaneously. Simple, successful standards has dozens of variations (take STAC[21] extensions).Similarly, in the Iliad scenarios, stakeholders and governance analysis, approximately half of the properties in the results are covered within CF vocabulary, and the other half is pilot and physical model specific. With growing specializations and implementations, a number of such cases require a more dynamic but disciplined way of building consensus and combine these elements together.

As the way forward the authors propose an approach that uses multi-level building blocks that can be shared among various specializations, allowing more accessible data fusion from various domains. Multi-level composition is related to the various aspects of interoperability. In particular, geospatial domain levels are traditionally related to spatial conceptual models and vocabularies, logical models and serialization patterns on one end. Behavioral description of the

[20] European Commission, Joint Research Centre, Soille, P., Lumnitz, S., Albani, S., Proceedings of the 2023 conference on Big Data from Space (BiDS'23) – From foresight to impact – 6-9 November 2023,

[21] Austrian Center, Vienna, Soille, P. (editor), Lumnitz, S. (editor), Albani, S. (editor), Publications Office of the European Union, 2023, https://data.europa.eu/doi/10.2760/46796

data exchange and API functions, application components implementing these levels, and subsystems responsible for the ecosystem functions.

In the example of various representations of observations and models, proposed building blocks are represented within the Ocean Information Model (OIM). OIM is the ontology building links between various representations of catalogs, data structures but also between, spatiotemporal domain and sensors and observations, AI models, and simulations. The linkages to serializations like JSON-FG, CoverageJSON, NetCDF/Zarr, GeoParquet metadata are defined as the spatial building blocks building bridges between these various representations. One of the example cases is the Observation Features Collection composed of ISO/OGC Feature and Observation and SOSA/SSN concepts.

Formal specializations and compositions do not break lower-level interoperability with tools, so the data can still be produced and, more importantly, consumed, white exchange toolkit helps to define the specializations and validate the services in the Continuous Integration/Continuous Deployment (CI/CD) pattern. Data exchange can be further optimized by re-using the pattern of dynamic-based separation. On the other hand, it supports full syntactic (based on schemas like JSON schema/SHACL) and semantic validation of the implementations.

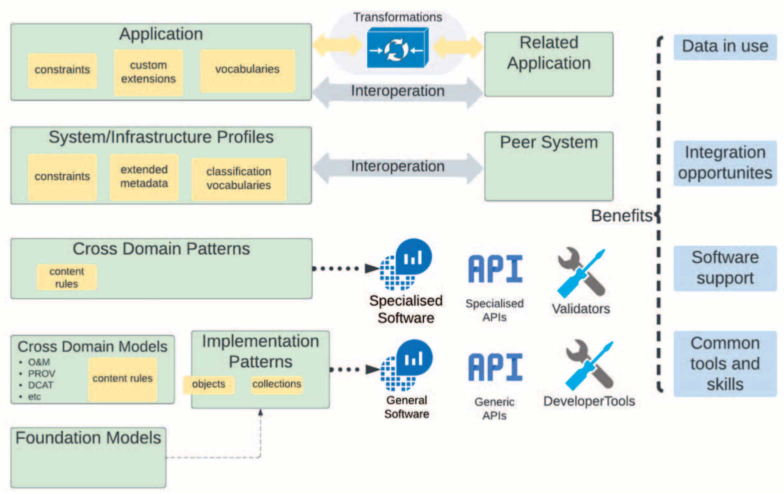


Figure 1 Layered interoperability tools and related value proposition.

Conclusions and looking forward

Good practices of data fusion are still a tacit knowledge of the practitioners, limiting the potential leverage of the marine data ecosystem. Collection of these bits of practices, their implementations, and data samples is required to leverage community efforts' outcomes. Formalisms in the variations of representations and building logical linkages between them are already widely endorsed by conventions on the data structures and meanings. However, a dynamic environment naturally goes beyond the consensus so the consensus process shall be agile and stable at the same time not to limit developments built on recent innovations and ensure quality of the data management. The proposed approach of stacked extensions, proven successful in general IT, can be combined with a lightweight 'applied semantics' without scaring most developers immediately. It is model and implementation agnostic so can be applied to current and future standards. Minimized ambiguity is built and used now, but it can be also future-proof for machine-based discovery. Still, the mapping between entities and supporting tools is the area of research and automation in the definition-validator-service-analytics provision chains between initiatives to foster reusability of the research and analytics.

Credit for large complex datasets

James Ayliffe[1] (jamayl@noc.ac.uk), Martina Stockhause[2] (stockhause@dkrz.de), Shelley Stall[3] (SStall@agu.org), Deb Agarwal[4] (DAAgarwal@lbl.gov), Justin Buck[1] (Justin.Buck@noc.ac.uk), Caroline Coward[5] (caroline.m.coward@jpl.nasa.gov), Chris Erdmann[3] (CErdmann@agu.org)

[1]British Oceanographic Data Centre - BODC, National Oceanography Centre - NOC (UK)
[2]Deutsches Klimarechenzentrum (Germany)
[3]American Geophysical Union (United States of America)
[4]Lawrence Berkeley National Laboratory (United States of America)
[5]NASA Jet Propulsion Laboratory (United States of America)

In the Environmental sciences, data are often complex, multi-disciplinary and expensive to collect. The use of data repositories is a key feature in managing, maintaining and publishing the data that scientists have spent time to collect, process and analyse. In a modern world and an increasing amount of complex, diverse data; journals and funders are requiring, as a condition of publication, that data are to be published in an appropriate repository without restriction. The expansion in the amount of data and the granularity of datasets means the ease of being able to cite these data in a suitable manner is becoming more complicated between communities, scientists, funders, journals, repositories and global indexers.
The Research Data Alliance (RDA) [Agarwal D. et al., 2023] complex citations working group was set up to address citing large numbers of existing objects to identify the challenges in complex data citation and support a means to measure the impact of these objects. A container termed a "reliquary" has been identified as a possible solution to facilitate citation for different known examples. A reliquary would be a container that would hold large and varied numbers of citations.
Currently there are four use cases: the Ameriflux network, the British Oceanographic Data Centre (BODC) and two from the Intergovernmental Panel on Climate Change (IPCC). These use cases have provided scenarios that cover complex citations, developing the scenario with some examples of how a reliquary would look. The RDA working group are now looking for comments to ensure that a working solution is provided that offers a framework by which complex citations may be achieved. This poster will look to advertise our the group's use cases, and provide reliquary examples. This promotion of our work will seek to inform the IMDIS community of the work whilst seeking additional support and ensure the solution that is proposed will allow diverse users of citations to understand the value of formalised publications of data, whilst meeting the needs of the scientific method.

References

Agarwal D., Stall S., Stockhause M., Wyborn L., (2023). *Research Data Alliance Complex Citations Working Group*. https://www.rd-alliance.org/groups/complex-citations-working-group#:~:text=This%20working%20group%20proposes%20to,objects%20to%20be%20properly%20assigned

Poster session

Next steps in creating FAIR sea level data

Andrew Matthews (antt@noc.ac.uk), Elizabeth Bradshaw (elizb@noc.ac.uk), Chanmi Kim (chanmi.kim@noc.ac.uk), Mark Hebden (mahe@noc.ac.uk), Lesley Rickards (ljr@noc.ac.uk)

National Oceanography Centre - NOC (UK)

Introduction

Our knowledge of changes in the height of the ocean over the past two centuries is dependent on the global network of tide gauges. Since 1933, the Permanent Service for Mean Sea Level (PSMSL), based in the Liverpool site of the UK's National Oceanography Centre (NOC), has been tasked with the responsibility of gathering together as much of this tide gauge data as possible and distributing it in a common format as an easy-to-use single dataset. This monthly and annual mean sea level dataset has proved to be a vital data source in oceanography and geodesy, and is a key dataset in the Intergovernmental Panel on Climate Change's reports on sea level.

The length of these records sometimes makes providing fully FAIR metadata challenging: tide gauge locations change over time as harbours are redeveloped, and sensors change as new technologies are discovered. In some earlier records, there are uncertainties about how and where recordings were made.

The PSMSL operates within a framework of many international organisations with complex relationships. In particular. the PSMSL is one of the data centres of the Intergovernmental Oceanographic Commission's Global Sea Level Observing System (GLOSS). These centres, together with all the global tide gauge operators, form a network of sea level providers. We have been working to ensure that data moves smoothly from sensors to users, whatever the purpose: from observations available in real time to monitor tsunamis, to long term records such as PSMSL's, used to produce projections of future sea levels.

Creating FAIR metadata for tide gauges

The tide gauge community has been working on standards for metadata for well over a decade and we have in the past tried to be exhaustive in the definitions. However, in order to provide some concrete guidance, we want to establish the minimum required to share Near Real Time tide gauge data, not just to share the data machine-to-machine, but that also makes sense to a user accessing the data. We have been working on standardising metadata for tide gauges, through GLOSS, the EuroGOOS tide gauge task team, the Horizon 2020 EuroSea project and other European and national initiatives. We are developing a metadata crosswalk between various different standards (DataCite Metadata Schema, SeaDataNet CDI, OceanSITES, WMO WIGOS etc.) to help define this minimum metadata standard.

As part of this process, we have been asking tide gauge operators to populate the tide gauge metadata catalogue https://eutgn.marine.ie/, developed by the Irish Marine Institute as a deliverable of the EuroSea project. Under EMODnet Physics, we have been discussing next steps for the catalogue, such as importing information into GitHub to improve updates and reduce duplication of effort.

The PSMSL has been involved in creating new CF standard names for tidal parameters, both observed and calculated, we have specified new attributes for datum and benchmark metadata as part of a pilot project to deliver enriched NRT tide gauge data via an ERDDAP instance.

Credit where credit is due

One of the top challenges we face is being able to track who is using PSMSL data, and being able to tell our data suppliers how their data are being used. We want to be able to give credit to everyone involved in the data lifecycle. As a start to this process, we want to publish the PSMSL dataset with a Digital Object Identifier (DOI).

We need to develop unique persistent identifiers (PIDs) for tide gauge data, including for sensors, stations and sites. Having unique identifiers would help data aggregators confirm the level of processing that a dataset has undergone, reduce data duplication, and offer products from the same location for multiple purposes. We need to build on the use case we contributed to the Research Data Alliance's Persistent Identification of Instruments Working Group.

Data aggregation and consolidation

GLOSS currently has several data centres, each specialising in a different product (real time, delayed mode etc.). This system, while functional, can cause confusion, so GLOSS plans to launch a unified data portal aggregating data from all centres. To facilitate this, each centre is planning to make their data available through an ERDDAP instance, and we are defining a shared metadata model. Details are still under negotiation, but it will need to implement the CF and ACDD metadata conventions and be mappable onto models used by the Copernicus Marine Service and EMODnet Physics.

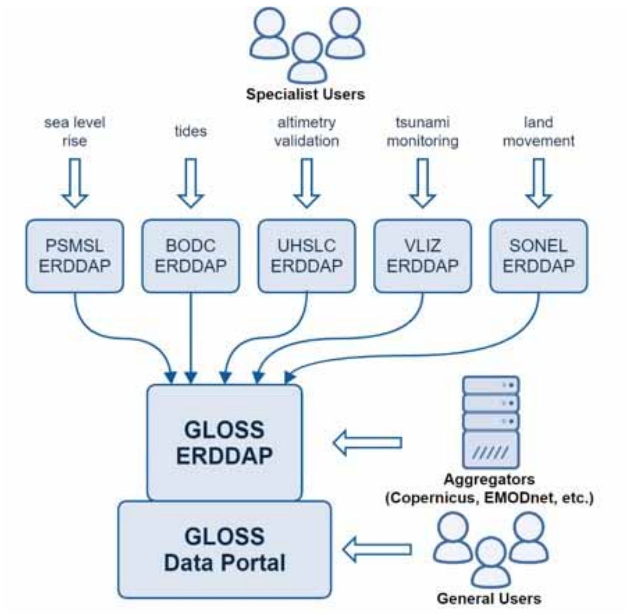


Figure 1 The proposed improved structure for GLOSS data, along with suggested access routes for example use cases.

The structure is illustrated in Figure 1. A central GLOSS ERDDAP will be created, aggregating the data from each individual data centre, and giving API access to large aggregators. It will also act as a data server for a new GLOSS data portal providing simpler access to general users. More specialist users will still be able to access the individual data centres as before.

This aggregated data source would bring together the work on PIDs, controlled vocabularies, and metadata models, and would result in an access point for tide gauge data that is FAIR for all users.

Development of Information system to support studies on carbon flux in the Black Sea

Elena Zhuk (alenixx@gmail.com)

Marine Hydrophysical Institute, Russian Academy of Science (Russia)

An information system is built to forecast and make calculations of CO2 fluxes. It consists of data access and analyzes tools and a database. The project database was created using contact measurements obtained during MHI expeditions and freely available oceanographic data. So it accumulates hydrological and hydrochemical MHI's data for the Black Sea, freely access oceanographic data from EMODNET (European Marine Observation Data Network) [EMODNET portal https://emodnet.ec.europa.eu/en/] and WODB (World Ocean Database) and data on the concentration of organic and inorganic carbon, carbon fluxes in suspended matter. The information system provides access to and analyzes observational data and mathematical modeling results, which are presented in heterogeneous formats for different time periods. The structure of the system is shown in Figure 1.

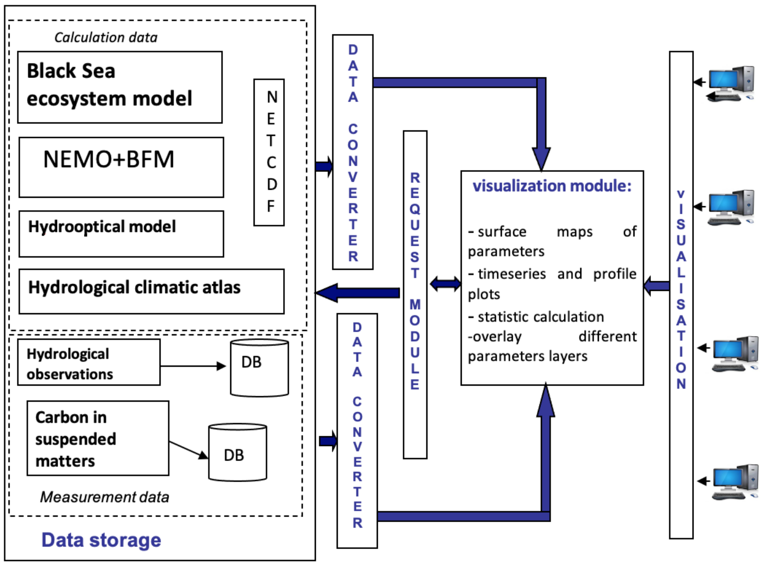


Figure 1 The structure schema of Carbon information system.

The following data is integrated into the system: an array of carbon dioxide and dissolved inorganic carbon fluxes according to the Black Sea ecosystem model (132x238 nodes); a range of 73 physical and biochemical parameters, including organic and inorganic carbon, according to the NEMO+BFM model of the Black Sea (66x129 nodes); an array of primary production according to a model based on the hydro-optical characteristics of the Black Sea (4x4 km, 6 regions); primary data on measurements of hydrological and hydrochemical parameters from the MHI oceanographic data base (1923–2023); temperature and salinity of the climatic array (1951–2018, 10′x15′) [Belokopytov, 2018].
The key component of the information system's development is the Django framework. The contact measurements are stored in a PostgreSQL database, the calculated data are kept in the netCDF files and stored as structured archive. JavaScript libraries were used in the development of the user interface, and Open Layers was applied to provide map service tools. Example of User Interface is shown at Figure 2.

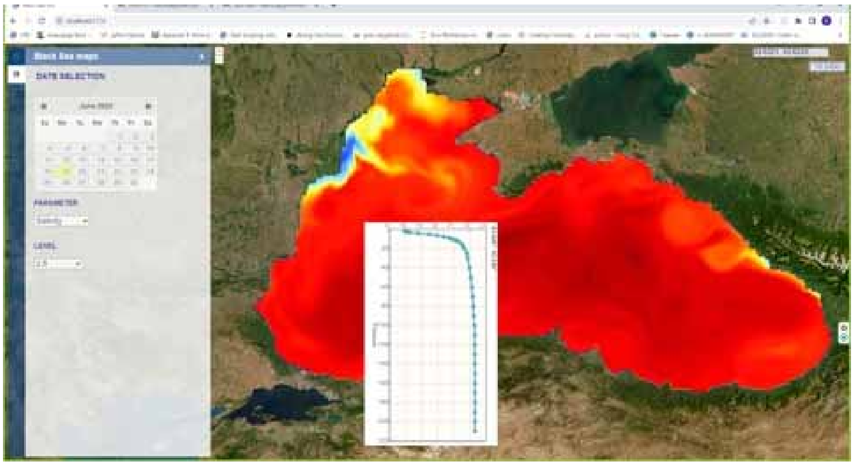


Figure 2 The example of calculated data visualization in user interface of Carbon information system.example use cases.

Acknowledgements

The Carbon information system is developed in frame of MHI RAS' state task No. FNNN-2023-0001 Ensuring climate and bio-geochemical monitoring of carbon fluxes in the Black Sea based on long-term observations data and numerical simulation results (Carbon).

References

Belokopytov V.N., (2018). *Retrospective Analysis of the Black Sea Thermohaline Fields on the Basis of Empirical Orthogonal Functions*. Physical Oceanography – 2019 – No 25(5). – P. 380-389. https://doi.org/10.22449/1573-160X-2018-5-380-389

SESSION PRODUCTS ORAL PRESENTATIONS

Fisheries acoustics and deep learning

Nils Olav Handegard[1] (nilsolav@hi.no), Leif Edvard Bildøy[2] (leif.edvard.bildoy@km.kongsberg.com), Olav Brautaset[3] (obr@nr.no), Tomasz Furmanek[1] (tomasz.furmanek@hi.no), Arne Johannes Holmin[1] (arnejh@hi.no), Ingrid Utseth[3] (utseth@nr.no), Ketil Malde[1] (ketil.malde@hi.no)

[1]Institute of Marine Research (Norway)
[2]Kongsberg Discovery (Norway)
[3]Norwegian Computing Centre (Norway)

Introduction

Acoustic surveys provide data for fisheries management. The current data processing pipeline involves a number of interactive steps that rely on human intervention, and automating the pipeline are believed to increased efficiency and consistency.

In recent years research in AI, big data, and deep learning have developed powerful technologies that have the potential to learn from historical data and automate data processing that was previously the domain of human experts [Malde et al., 2019; Rubbens et al., 2023]. Implementing data processes in terms of automated (computerized) steps permits the scalability needed to process large data streams generated by autonomous observation platforms and for reanalyzing large data archives.

The objective of this contribution is to describe our approach for implementing AI and automated data processing for acoustic water column data, with a particular focus on acoustic target classification, which is the process of allocating acoustic backscatter to a species or group of species. This is used as input for fisheries assessment models.

Data for training models

Modern data processing methods can work with high resolution data, and convolution-based methods work very efficiently on n-dimensional data. We were looking for a structure that could keep the original resolution in the data to the extent possible and at the same time align the different frequency channels in a multidimensional array with time, range, and frequency as dimensions. This allows for simple interface to machine learning frameworks, e.g., Keras, pyTorch etc. Interoperability requires a well-defined data convention, where the variables are defined following a standard. We have strived to follow the process for developing a data convention for acoustic data in ICES. Some mandatory fields from the convention are not present in the historical data and must be manually added, which requires substantial effort when preparing the historical data. Often the cost of mandatory fields are not assessed when developing data conventions, and this precluded strict adherence to the data convention in our case. We believe it will take further iterations before we can confidently state that we have true interoperability. Although it is desirable to have a clean data set to work on, it requires several iterations to stabilize the data quality. The challenge with the "perfect" data approach is that it takes too long before we can start training models. The effort required to convert historical data to a standardized and open format should not be underestimated.

Deep learning for acoustic target classification

We started using the U-Net architecture [Ronneberger et al., 2015], a popular convolutional neural network (CNN) for segmentation tasks. We successfully trained the U-net algorithm for ATC on data from the Norwegian Sand eel survey [Brautaset et al., 2020]. For the Sand eel case only 0.2% of the data samples were annotated to the foreground classes; the key for success was to address class imbalance. We have further developed methods to address class imbalance [Pala et al., 2023], and we have used semi-supervised models where less training data are required [Choi et al., 2021; 2023].

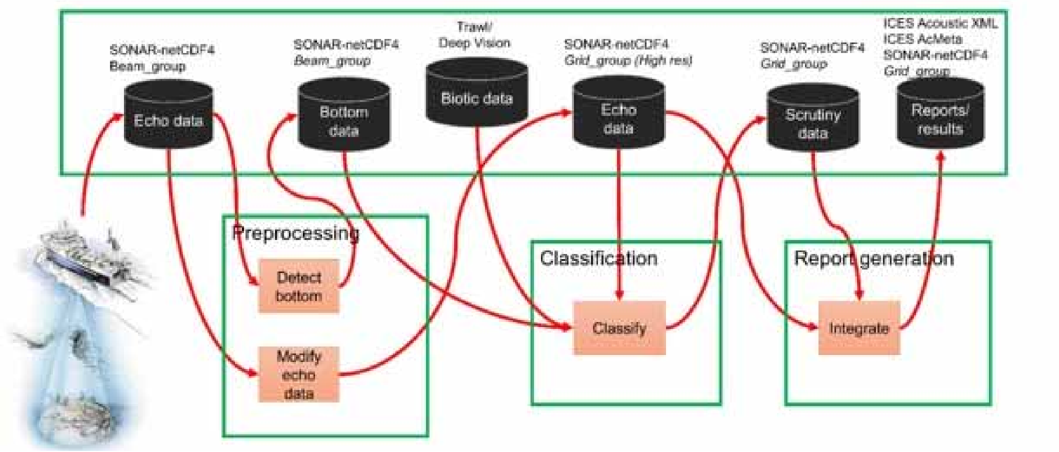


Figure 1 Proposed data processing pipeline for acoustic water column data.

ML ops

ML operations, i.e., practices for reliably and efficiently training, deploying, and maintaining machine learning models in production, are essential for deploying methods. To ensure portability we rely on containerisation, and each processing step are packaged in a Docker container that serves as a step in the full pipeline. Challenges with distributed systems are defining standardized APIs and orchestrating their operation. This is where Kubernetes provides a good companion to automate deployment, scaling and management of the containers. With this we can also leverage the remote update capability of the Kongsberg Blue Insight platform, where remote updates works with a cloud-edge paradigm and a cloud-based edge management service can control the edge side of the Kubernetes cluster.

Concluding remark

To successfully tap the potential of modern data processing methods, several steps need to be in place. This includes data preparation, securing infrastructure for training models, choosing and training the models, testing the performance and deployment.

References

Brautaset O., Waldeland A.U., Johnsen E., Malde K., Eikvil L., Salberg A.-B., and Handegard N.O., (2020). *Acoustic classification in multifrequency echosounder data using deep convolutional neural networks*. ICES Journal of Marine Science, 77: 1391–1400.

Choi C., Kampffmeyer M., Handegard N.O., Salberg A.-B., Brautaset O., Eikvil L., and Jenssen R., (2021). *Semi-supervised target classification in multi-frequency echosounder data*. ICES Journal of Marine Science. https://doi.org/10.1093/icesjms/fsab140 (Accessed 12 August 2021).

Choi C., Kampffmeyer M., Handegard N.O., Salberg, A.-B., and Jenssen, R., (2023). *Deep Semisupervised Semantic Segmentation in Multifrequency Echosounder Data*. IEEE Journal of Oceanic Engineering: 1–17.

Malde K., Handegard N.O., Eikvil L., and Salberg A.-B., (2019). *Machine intelligence and the data-driven future of marine science*. ICES Journal of Marine Science.

Pala A., Oleynik A., Utseth I., and Handegard N.O., (2023). *Addressing class imbalance in deep* *learning for acoustic target classification*. ICES Journal of Marine Science: fsad165.

Ronneberger O., Fischer P., and Brox T., (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*, arXiv:1505.04597 [cs.CV]. https://arxiv.org/abs/1505.04597 (Accessed 15 November 2018).

Rubbens P., Brodie S., Cordier T., Destro Barcellos D., Devos P., Fernandes-Salvador J.A., Fincham J.I., et al., (2023). *Machine learning in marine ecology: an overview of techniques and applications*. ICES Journal of Marine Science: fsad100.

Interpolation of Plankton Continuous Recorder data using a neural network technique

Charles Troupin[1] (ctroupin@uliege.be), Alexander Barth[1] (abarth@uliege.be), Pierre Hélaouët[2] (pihe@mba.ac.uk)

[1]University of Liège (Belgium)
[2]Marine Biological Association (UK)

Summary
We present the results obtained by the application of a neural network method, DINCAE, on *Calanus finmarchicus* and *Calanus helgolandicus* observations collected by the *Continuous Plankton Recorder*. The main features (spatial distribution, seasonal cycle) are well reproduced in the interpolated fields, even if the amplitude of the variation is dampened by the gridding method.

Data and methods
Continuous Plankton Recorders (CPR) are high-speed plankton samplers towed by ships of opportunity. The same sampling method has been used for decades, making the dataset particularly relevant to study long-term changes.
We focus on *Calanus finmarchicus* and *Calanus helgolandicus*, which constitute key contributors to the zooplanktonic ecosystem. They are both known to be influenced by environmental parameters, such as the bathymetry, the surface sea water temperature [Hélaouët and Beaugrand, 2007].
The uneven sampling effort is highlighted in Figure 1, where the main routes are clearly visible. The dataset contains a total of 250021 records (time, position and counts of different taxa and taxonomic groups).
The *Data-Interpolating Convolutional Auto-Encoder* (DINCAE) method [Barth et al., 2022] is applied to create gridded fields from the in-situ observations. Environmental data are also considered in the neural network. The code is written in Julia and available from https://github.com/gher-uliege/DINCAE.jl.
Several *hyperparameters* have to be set in order to run DINCAE. Some of them were set by hand, i.e., by using standard values, while the most relevant parameters (learning rate, number of epochs and Laplacian penalty coefficient) are selected using a cross-validation approach.

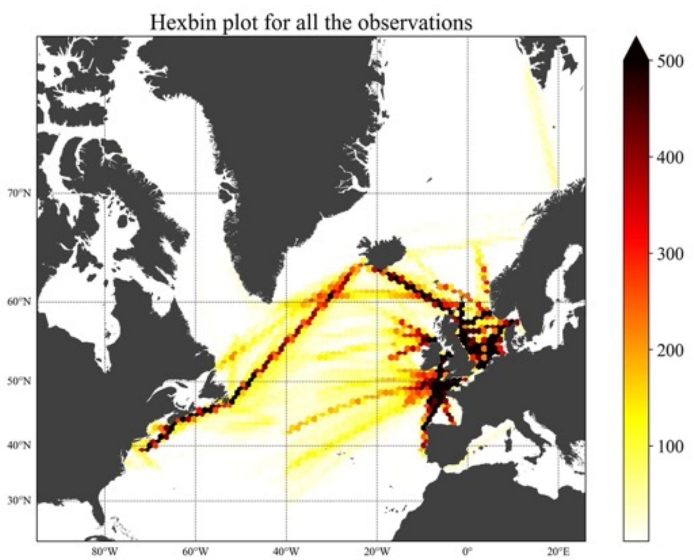


Figure 1 Spatial distribution of Calanus Finmarchicus and *Calanus Helgolandicus*.

Results and validation

Several experiments have been conducted with different configurations:

- input data preparations: applying a transformation (logarithm) or a quality control (removing values above the 95th percentile), considering different time periods (climatology, monthly fields, monthly fields with 2 years before/after the year of interest, …);
- environmental data: bathymetry, sea surface temperature, distance to closest coast.

The validation is performed by discarding 10% of the observations, performing the DINCAE analysis and then comparing the values of the gridded field at the locations of the discarded data with the discarded values.

The results display the main features (spatial distribution and seasonal cycle) once the hyperparameters have been tuned and when the bathymetry and the sea surface temperature are considered in the neural network.

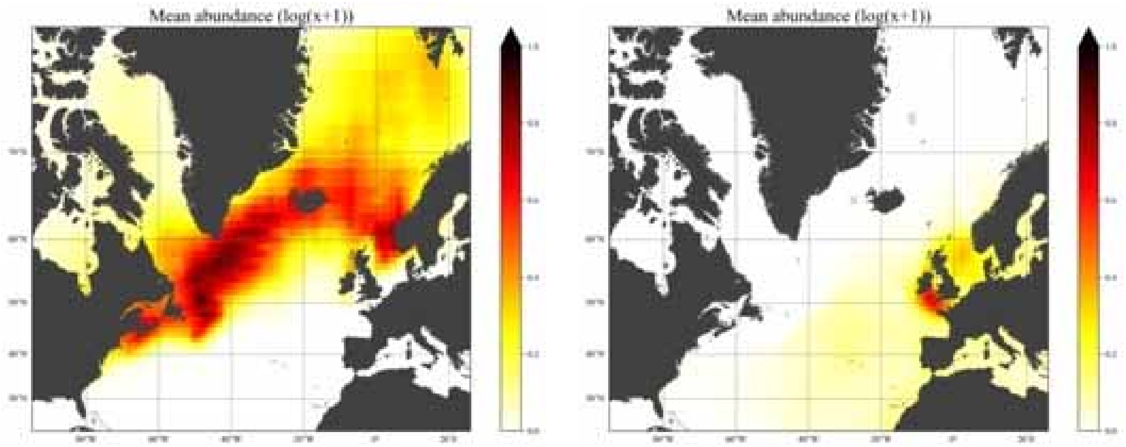


Figure 1 Mean fields (all years and months) of *Calanus finmarchicus* and *Calanus helgolandicus* produced with DINCAE.

Conclusions

This work presents a first attempt to interpolate in situ, zooplankton data using a neural network technique. Future work will focus on the following aspects:

1. Further optimisation of the hyperparameters, since here the attention was specifically on 3 of them: (learning rate, number of epochs and Laplacian penaly coefficient.
2. Interpolation of other taxa or other taxonomic groups, with a particular interest on the *Total Dinoflagellates* or the *Total Copepods*.
3. Use of other environmental variables such as the chlorophyll concentration and its horizontal gradient.

References

Barth A., Alvera-Azcárate A., Troupin C., & Beckers J.-M. (2022). *DINCAE 2.0: multivariate convolutional neural network with error estimates to reconstruct sea surface temperature satellite and altimetry observations*. Geoscientific Model Development, 15(5), 2183–2196. https://doi.org/10.5194/gmd-15-2183-2022

Helaouët P. and Beaugrand G. (2007). *Macroecology of Calanus finmarchicus and C. helgolandicus in the North Atlantic Ocean and adjacent seas*, Marine Ecology Progress Series, 345, 147–165, https://doi.org/10.3354/meps06775

Compiling FAIR sea ice core data: a reuser perspective

Anna Simson[1] (simson@mbd.rwth-aachen.de), Anil Yildiz[1] (yildiz@mbd.rwth-aachen.de), Sebastian Mieruch[2] (sebastian.mieruch@awi.de), Julia Kowalski[1] (kowalski@mbd.rwth-aachen.de)

[1]MBD, RWTH Aachen University (Germany)
[2]Alfred-Wegener-Institute (Germany)

Compiling FAIR data for reuse scenarios

Vast amounts of environmental data are available from a multitude of data repositories. Most of these repositories adhere to and rely on the FAIR principles. Therefore, this data should be available for all kinds of reuse scenarios such as comparative studies, integration into complex simulation workflows, and for virtual testbed infrastructure needed to test and develop environmental robotics. These scenarios require large data foundations that often call for the compilation of many individual, potentially heterogenous data sets. Reuse scenarios differ from one another, so that the respective data compilation must be purpose-driven and scenario-tailored. While a single data set may be readily reusable, the integration of the same data set into a data compilation poses challenges to effortless reusability. We investigate these challenges and propose solutions from the perspective of an agnostic reuser, i.e., a person who was not involved in data acquisition. We start by discussing an intuitive approach to compiling data and report on its drawbacks and limitations with an example of a comprehensive sea ice core data compilation. We finally present solutions to facilitate reuse.

Intuitive approach to compile FAIR data

FAIR data principles provide guidelines to enable future reuse. They include i) assignment of searchable metadata that clearly relates to the data described, ii) description of data and metadata with community-controlled vocabularies, iii) adherence to community standards for data sharing, and iv) provision of contextual information by neglecting assumptions about potential reuse scenarios. If these guidelines are met, the intuitive approach to compile data by a data reuser might be:

1. Definition of reuse scope, i.e., the data and metadata required to conduct the reuse scenario.
2. Screening of data repositories on the metadata level for reuse scope matching data sets.
3. Aggregation of selected data sets in a format compatible with the reuse scenario.

The feasibility of this approach depends on the *FAIRness* level of the available data sets, and this is often related to the availability and use of community-established standards and best practices for controlled vocabularies, ontologies, structures and formats of data and metadata as well as for their corresponding files. Several challenges arose while compiling sea ice core data sets hindering effortless reuse and leading to solutions for deviations from the intuitive approach.

Sea ice core data compilation: challenges and solutions

We compiled data from 287 sea ice cores orginating from several measurement campaigns. All of these cores are findable and accessible via three data repositories: Zenodo, Australian Antarctic Data Centre and Pangaea.

1. Definition of reuse scope

Reuser's perspective is defined by the reuse scenario. A potential scenario could be the prediciton of sea ice core temperature profiles based on coordinates, date, and sea ice type. Parts of this scope may be unavailable from data repositories. In addition, there can be a different understanding between data collector and reuser regarding the distinction of what classifies as data and metadata. Consequently, parts of the required scope are often not included in the data sets, but may be mentioned in abstracts, comments, or readme files that the reuser has to find beyond the searchable metadata. The reuser has to adjust the scope sequentially during the search process.

2. Screening of data repositories

Data repositories have different data and file structures, as well as different search and filtering capabilities. Depending on the file and data structure of the repository, the data and metadata associated with one sea ice core is often distributed across multiple data sets and files, for example, temperature and salinity data of the same core may be stored in two different data sets. The reuser has to find all relevant data for each sea ice core and adapt the search strategy for each repository. The search often extends to resources that provide contextual information, such as articles and reports. The data files may be provided in a non-machine readable way, so that the reuser has to decide on its suitability by manually checking and potentially adjusting the data file. When extracting data and metadata from the assembled data sets, there can be inconsistencies between and within the included resources, for instance, varying label names or values for a measurable. Here, the reuser needs to transparently interpret by gathering and citing all relevant data and metadata and potentially consider further contextual resources such as application notes.

3. Aggregation of selected data sets

Instead of directly compiling suitable data sets, the reuse scope for each sea ice core has first to be compiled from all relevant data sets and possibly enriched from other resources. For this purpose, a flexible ".yaml" file is instantiated for each sea ice core, which allows individual referencing of the different resources. After combination at the individual sea ice core level, compilation can be performed.

The compiled data will be made availabe for further reuse through the interactive online visualization and analysis tool webODV (Ocean Data View; https://mvre.webodv.cloud.awi.de).

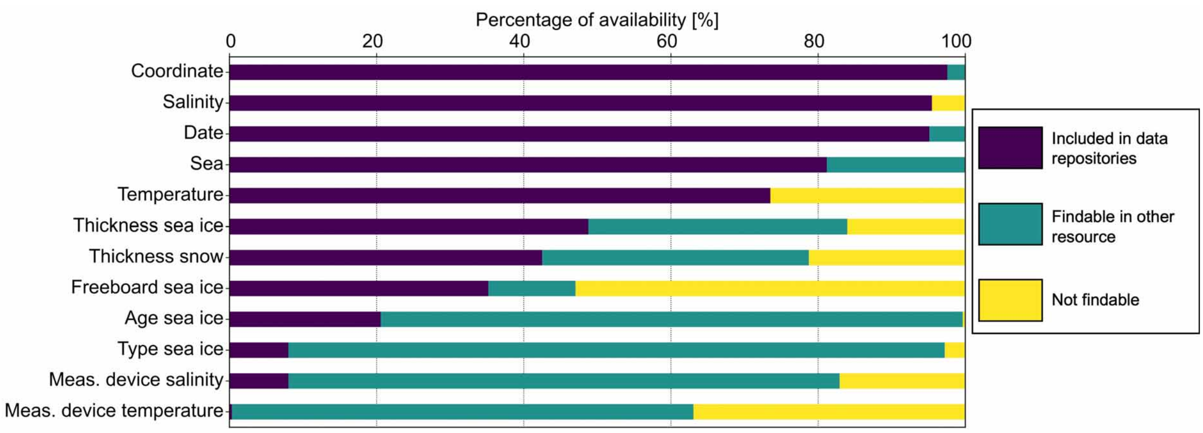


Figure 1 Availability and origin of the scope (y-axis) for all sea ice cores of the compilation.

Conclusion

We conclude that the degree of reusability of FAIR data sets depends on the type of reuse scenario: while the reuse of a single dataset for replotting may be straightforward, reusability becomes a challenge for scenarios requiring the compilation of large amounts of heterogeneous data for complex reuse scenarios with extensive and detailed scopes. This is also reflected in the fact that current trends in data-driven and model-based simulation show the need for a more data-centric approach, which recognizes that the sustainable preparation and management of the data involved is a challenge in itself. This challenge cannot be solved by raising the *FAIRness* standards of individual data sets, mainly because the reuse scenarios are typically not defined at the time of data acquisition. Rather, we need a conceptual streamlining of the (potentially agnostic) compilation process, as the first steps in this direction are presented here. We believe that similar workflows in the future can be set up more efficiently by following the suggested solutions. Future work will be devoted to technological solutions that support the presented conceptual approaches.

2023 release of World Ocean Database and World Ocean Atlas

Alexey Mishonov[1,2] (alexey.mishonov@noaa.gov), Tim Boyer[1] (tim.boyer@noaa.gov), Ricardo Locarnini[1] (ricardo.locarnini@noaa.gov), Hernan Garcia[1] (hernan.garcia@noaa.gov), Dan Seidov (ret.)[1] (dan.seidov@gmail.com), James R. Reagan[1] (james.reagan@noaa.gov), Christopher Paver[1] (christopher.paver@noaa.gov), Olga Baranova[1] (olga.baranova@noaa.gov), Scott Cross[1] (scott.cross@noaa.gov), Courtney Bouchard[1] (courtney.bouchard@noaa.gov), Ebenez Nyadro[1,3] (ebenezer.nyadjro@noaa.gov), Alexandra Grodsky[1,2] (alexandra.grodsky@noaa.gov), Dmitry Dukhovskoy[2] (dmitry.dukhovskoy@noaa.gov)

[1]Univ. of Maryland/ESSIC (USA)
[2]NCEI/NOAA (USA)
[3]Mississippi St. Univ. (USA)

The World Ocean Database 2023 (WOD23) and the World Ocean Atlas (WOA23) have been recently published by the National Centers for Environmental Information (NCEI/NOAA). The updated WOD23 contains all in situ ocean profiles assembled and processed by the Ocean Climate Laboratory team at the NOAA/NCEI through December 2022. This includes all Argo measurements, with quality control as of April 1, 2023. Additionally, WOD23 contains updated quality control flags based on the quality improvements made during the construction of the WOA23. WOD23 includes more than 3 million new profiles when compared to the WOD18 and continues to be the world's most extensive collection of ocean profile data, which is updated four times per year and available without restriction.

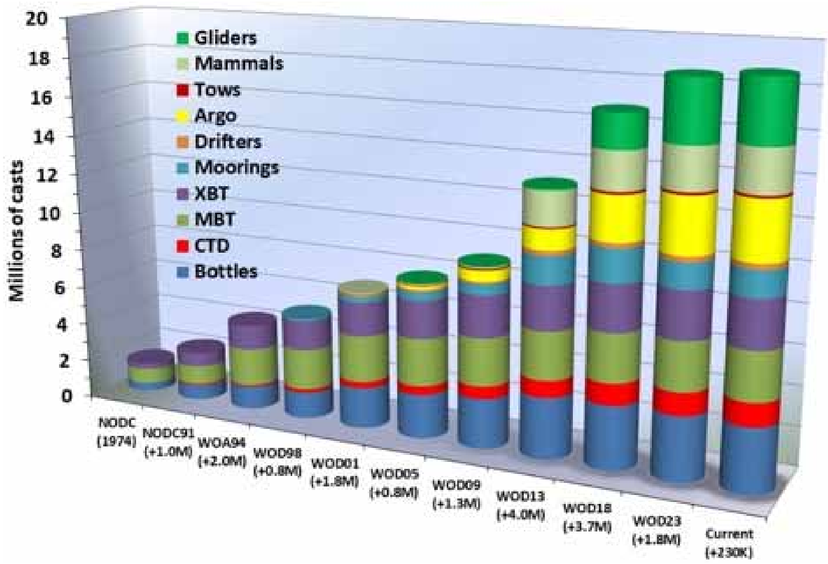


Figure 1 World Ocean Database: growth of the seawater Temperature data holding (Current: 19-Jan-2024).

The new release of the WOA23 is based on the analyses of all in situ ocean profiles data contained in WOD23. Temperature, salinity, and other variables in the WOA23 have been analyzed in a consistent, objective manner on one-degree and, in the case of temperature and salinity, quarter-degree latitude-longitude grids at standard depth levels from the surface to a maximum depth of 5500m. The procedures for "all-data" climatologies are identical to those used in the World Ocean Atlas, 2018.

Temperature and salinity climatologies are comprised of seven “decadal” climatologies for the following time periods: 1955 - 1964, 1965 - 1974, 1975 - 1984, 1985 - 1994, 1995 - 2004, 2005 - 2014, and 2015 - 2022 as well as three successive 30-year “climate normal” periods: 1971 - 2000, 1981 - 2010, and 1991 - 2020. Additionally, the temperature and salinity climatologies contain a ‘decav’ climatology which is the average of all seven individual decades mentioned before. O_2, AOU, and O_2 saturation climatologies are prepared for the 1965 - 2022 time period as well as 30-year 1971 - 2000 “climate normal”. The O2 climatology and related fields also contain, in addition to bottle data, sensor data from CTDs and Argo profiling floats, the first time this has been included in the WOAO2 series. All of the nutrient climatologies cover the 1965 - 2022 time period.
The WOA23 is composed of smoothed fields from analyses of historical means, based on relatively few observations (in certain areas). We believe, however, that useful information about the oceans can be gained through our procedures and that the large-scale and even mesoscale features (in quarter-degree analyses) are representative of the real ocean.

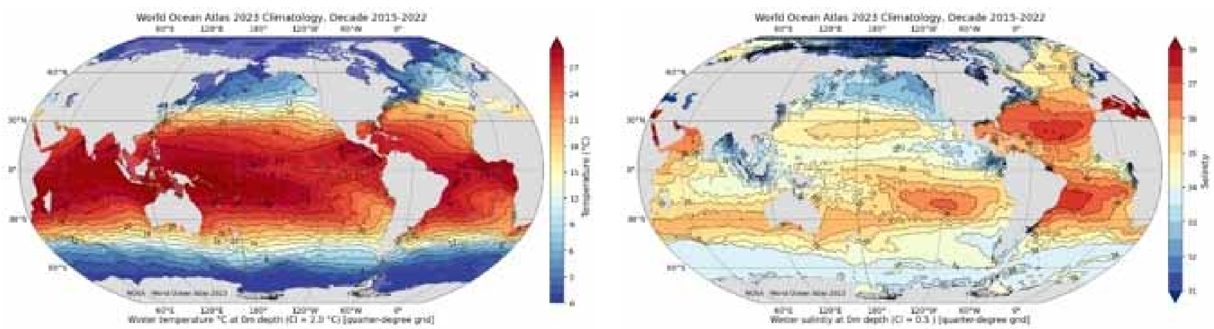


Figure 2 WOA23: winter surface Temperature (left) and Salinity (right) for 2015 - 2022 time period on quarter-degree grid.

Maps and data files of WOA23 include the following types of horizontal fields representing annual, seasonal, and monthly spatial distributions of analyzed data and data statistics as a function of selected standard depth levels on one-degree or quarter-degree grids:

1. Objectively analyzed climatology fields (grid squares with less than 3 data values available for objective analysis marked on maps by a white shading).
2. Statistical mean fields used in the objective analysis.
3. Data distribution fields - number of observations used in the objective analysis per grid square.
4. Standard deviation fields binned per grid square.
5. Objectively analyzed standard deviation fields (for some variables).
6. Standard error of the mean fields binned per grid square.
7. Standard error of the analysis.
8. Difference between observed and analyzed fields binned per grid square.
9. Difference between seasonal/monthly fields and the annual mean field.
10. The number of mean values within the radius of influence for each grid box (data files only). This is not represented as stand-alone online maps, but the results are used in 1) and 2) maps (as above) to shade the grid boxes with fewer than three mean values within the radius of influence.

WOA23 introduces two new ancillary fields: (i) the objectively analyzed standard deviation (5 as above; available for temperature, salinity, O2, AOU, and %O2 saturation). Its computed using the same objective analysis, but the input values are the one- or quarter-degree statistical standard deviations. We further refine the input standard deviations by only incorporating standard deviation grids in which there were at least six observations within that grid box. (ii) the standard error of the analysis (7 as above; available for all variables), which is an estimation of the uncertainty in the objective analysis. It is computed by calculating the standard error of the differences between the objectively analyzed and statistical mean fields within the 2nd radius of influence (669km for one-degree, 267km for quarter-degree).

The complete set of all climatological maps, objectively analyzed fields and associated statistical fields at all standard depth levels as well as complete set of data files and documentation are available online.

Stakeholders' engagement as a key instrument to devise the future strategy for the Observatorio Costeiro da Xunta de Galicia

Clara Almécija[1] (calmecija@cetmar.org), Pablo Álvarez-Chaver[1], Ignacio Gonzalez-Liaño[1], Eva Méndez[1], Silvia Piedracoba[1], Cristian Simoes[1], Silvia Torres[1] (utmar@cetmar.org), Marta Vazquez[1], Silvia Allen-Perkins[2], Garbiñe Ayensa[2] (gayensa@intecmar.gal), Silvia Calvo[2], Pedro Montero[2] (pmontero@intecmar.gal), Jesús Calviño[3], Isaac Gómez-Piñeiro[3] (isaac.gomez.pineiro@xunta.gal), Santiago Salsón[3], Juan Taboada[3] (coordinador-predicion.meteogalicia@xunta.gal)

[1]Centro Tecnológico del Mar - Fundación CETMAR (Spain)

[2]Instituto Tecnolóxico para o Control do Medio Mariño de Galicia - INTECMAR (Spain)

[3]Dirección Xeral de Calidad Ambiental, Sostenibilidad y Cambio Climático - Meteogalicia (Spain)

Observatorio Costeiro da Xunta de Galicia

The Observatorio Costeiro da Xunta de Galicia is located in the NW Iberian Peninsula, where the Galicia's extensive coast is vital for human activities as fisheries, aquaculture, renewable energies, and maritime tourism. The observatory applies operational oceanography to this area, aiming the continuous data collection, interpretation and dissemination for managing marine resources, addressing environmental and climate change challenges, providing valuable data and aiding in managing coastal resources and maritime safety.

It has been operating for 15 years, by the collaboration of three public institutions: DXCASCC-MeteoGalicia, Intecmar, and CETMAR focusing on meteorological observations, marine environment quality control, and marine resource management. The observatory engages these three areas by the use of various systems and networks to gather real-time data on ocean and atmospheric conditions. It also maintains numerical forecast systems and develops value-added products and services based on the collected information, aiding different stakeholders such as marine industry sectors and environmental monitoring. All data is available at www.observatoriocosteiro.gal and on the main European Data Aggregators as CMEMS, EMODnet and SeaDataNet. Furthermore, the Observatorio Costeiro Xunta de Galicia belong to Observatorio RAIA (www.marnaraia.org), a leading case of European Cross-Border cooperation in the Galicia-North Portugal Euroregion.

Observing systems encompass 6 automated platforms in Rías Baixas, 43 oceanographic stations, more than 50 coastal weather stations, and an external HF radar network, that provide real-time data on various parameters like temperature, salinity, currents, and meteorological conditions. Some of these series started more than 30 years ago, being some of the longest and most extensive ones, globally. All this information is necessary for several different uses like aiding in monitoring ship traffic schemes or suppling crucial data for scientific research.

The forecasting system involves numerical models that predict sea level, currents, temperature, wind, and waves, always validated through observations and with high reliability. Likewise, on order drift models can be run to help maritime authorities to manage accidental pollution or maritime accidents. Additionally, the observatory develops value-added products and services like daily ocean-meteorological forecast bulletins, specialized applications for specific end-users, contributing to marine industry sectors, pollution control, and environmental monitoring.

Interaction with Stakeholders

The study conducted 30 meetings with stakeholders to evaluate the awareness, utilization, and usefulness of observatory information and identify potential additional information valuable to end-users. These stakeholders encompassed various sectors like fishing, aquaculture, tourism, GIS professionals, competent authorities, and marine safety, reflecting diverse connections with the sea and the Blue Economy.

These meetings showcased the extensive oceanographic capabilities of the Observatorio RAIA in Galicia, detailing spatial and temporal distribution, parameters, units, and instructions for accessing data available on websites. They also revealed a comprehensive list of new requirements essential for stakeholders to enhance their activities, gathered from stakeholder feedback.

Discussion and conclusion

The results outline the familiarity and usage of Observatory RAIA information, detailing data known and utilized by stakeholders. Additionally, it highlights data considered vital for stakeholders' future activities and presents new requirements identified for oceanographic information.

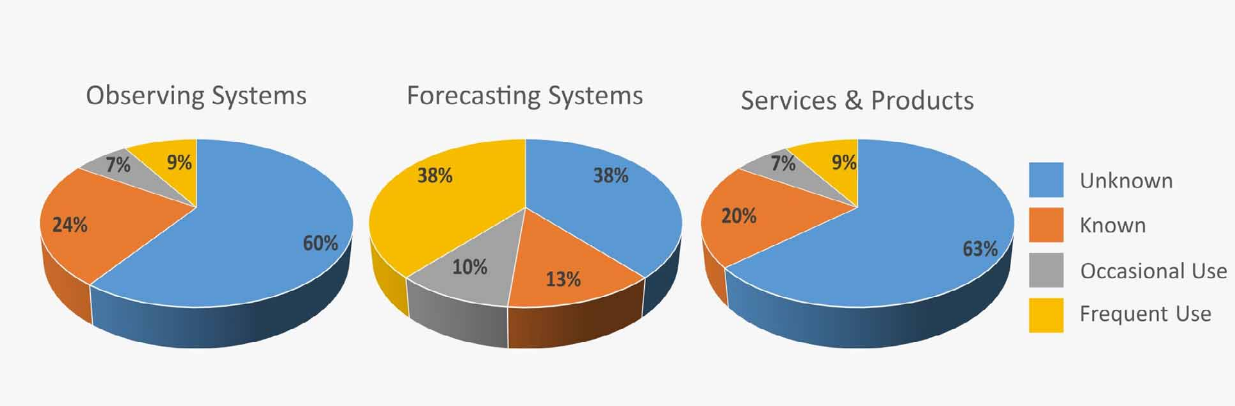


Figure 1 Percentage of Observing Systems (left), Forecasting Systems (centre) and Services and Products (right) that are Unknown (blue), Known (orange), Occasionally Used (grey) and Frequently Used (yellow) by the external stakeholders.

Regarding the observing systems, models, services, and products in Galicia, external interviewees demonstrate awareness of around 50% of available information, with roughly one fifth of it being occasionally used (10% frequently). Notably, nearly 60% of facilities associated with Observatorio Costeiro da Xunta de Galicia are well-known, and about one-third are commonly used, especially predictive models such as atmospheric and wave models. Observing systems like oceanographic buoys, weekly CTD stations, biogeochemical sampling, and weather stations are widely utilized, except for the recently launched HF RADAR information, which has intuitive and user-friendly viewers.

Some previously unused data has significantly interested to the stakeholders, being deemed as a highly valuable data for future use. Oceanographic buoys, oceanographic models, external HF Radar viewers for current, wave, and upwelling index, and Plan Camgal—an administration tool for marine accident contingency plans—stand out among these data sets, all belonging to Observatorio Costeiro da Xunta.

Furthermore, the most demanded oceanographic information identified by stakeholders includes specific temperature and salinity in certain areas to detect local particularities or biological events, as well as time series related to mussel rafts affected by toxic Harmful Algal Blooms (HABs). Improved wave and wind models, and the integration of different viewers into a unified one linking metadata and accessible, usable, and reliable data, are also highlighted.

It's noteworthy that despite the existence of some requested data, there's a clear need to enhance efforts in disseminating and making this information more accessible to reach a wider audience of potential end-users. Training initiatives to encourage the utilization of available information and services to improve professional outcomes and quality of life are imperative. Additionally, efforts to enhance data accessibility, standardization, and harmonization are essential to ensure their usability.
This study concludes that the stakeholders' engagement significantly influences the strategic planning of the Observatorio Costeiro da Xunta de Galicia. Their input guides the observatory to address societal needs and requirements, representing the initial steps towards their active involvement and future engagement.

Acknowledgements
Thanks to Interreg VI-A España–Portugal Program 2021-2027 to fund CAPTA project and to European Maritime, Fisheries and Aquaculture Fund (2021-2027) and Xunta de Galicia support the Observatorio Costeiro Xunta de Galicia.

EMODnet near real time river data and land boundary condition services

Francisco Campuzano[1] (francisco.campuzano@colabatlantic.com), Antonio Novellino[2] (antonio.novellino@ettsolutions.com), Patrick Gorringe[3] (Patrick.Gorringe@smhi.se), Caio Fonteles[1] (caio.fonteles@colabatlantic.com), Luís Figueiredo[1] (Luis.figueiredo@colabatlantic.com), Enrico Quaglia[2] (enrico.quaglia@grupposcai.it), Francesco Misurale[2] (francesco.misurale@ettsolutions.com), Marco Alba[2] marco.alba@ettsolutions.com

[1]+ATLANTIC CoLAB (Portugal)
[2]ETT S.p.A. (Italy)
[3]Swedish Meteorological and Hydrological Institute (Sweden)

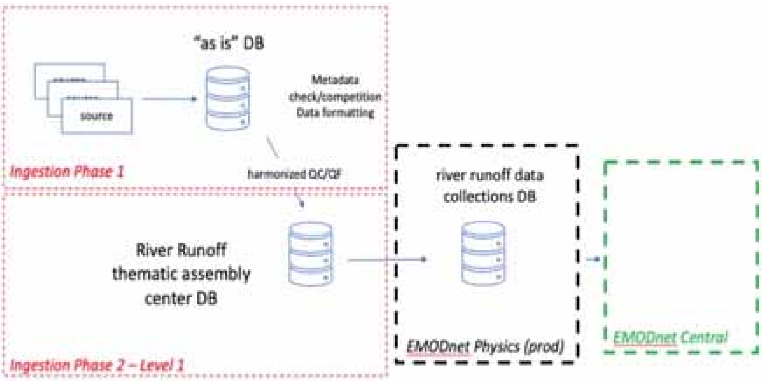
Introduction

EMODnet Physics is one of the seven domain-specific projects of the European Marine Observation and Data network (EMODnet) that has successfully designed, organised and run operational services providing ocean and coastal physical data and data products. EMODnet Physics, together with EMODnet Data Ingestion are sharing effort to facilitate the access to operational data streams. In this framework, to accomplish to stakeholders needs and include crucial information on river runoff data, EMODnet Physics and Ingestion had to design a new and dedicated river-data management workflow. While river climatology and research quality data are linked in from the Global Runoff Data Center, the near-real-time operational River runoff product is operated in collaboration with +ATLANTIC CoLAB.

River data outflow data management

The following scheme (Figure 1) presents a simplified schema of the operational river runoff data publishing pipeline. Sources (river basins administrative authorities) are contacted to establish an operational data flow from their services towards EMODnet river data node. River outflow may be a direct outflow measurement or a derived value from river water level (according a conversion law based on the river section). Some stations also provided temperature and river water level. Observations are combined with harmonized metadata based on existing vocabularies (Table 1).

Figure 1 Simplified workflow for river runoff data ingestion.



Currently EMODnet Physics is assigning a unique id to river stations and connecting with the id from GRDC station number. A new convention based on the country, river name and river station is being developed and will be soon put in place.

Metadata field	Vocabulary exists	Link to vocabulary	Vocabulary governance
Platform id	*		EMODnet Physics
Station		https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/grdc_reference_stations_zip.html?nn=201698	GRDC
Owner/provider Institution	Yes	https://edmo.seadatanet.org/	SeaDataNet
variable names	Yes	http://vocab.nerc.ac.uk/collection/PXX/current/ where XX=02;01;07	BODC:NVS
unit	yes	https://vocab.nerc.ac.uk/collection/P06/current/	BODC:NVS
Quality Flag Scheme	yes	http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf	OceanSites
Time	yes	ISO8601	ISO
Datum	Yes	WGS84	ISO
Country	yes	ISO3166	ISO
License	Yes	https://creativecommons.org/	CC
INSPIRE	Yes	ISO 19115	ISO/INSPIRE

Table 1 Applied standards and vocabularies.

River data service

The EMODnet near-real-time river data service is a product designed to satisfy the needs of the coastal community. Currently, it integrates more than 600 active stations in Europe, North America and South America from 35 providers of 17 countries (Figure 2). The continuous development of EMODnet Physics River Node, reflects a crucial step towards comprehensive coastal management and effective support to EMODnet stakeholders.

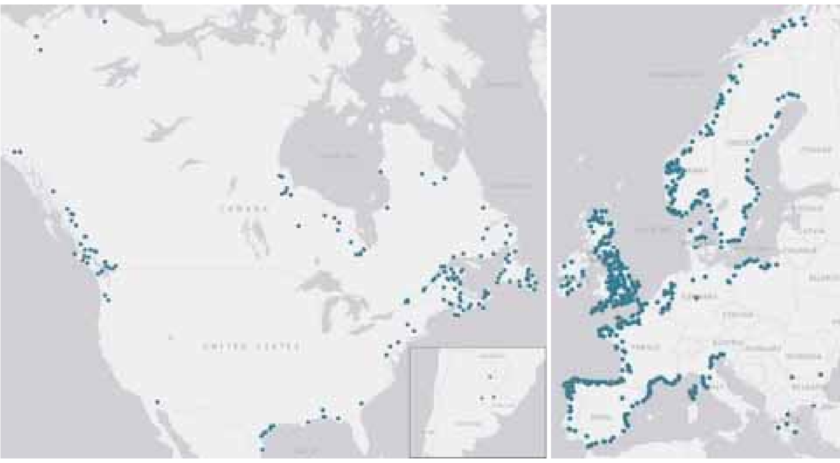


Figure 2 Current coverage of near-real-time river stations.

To improve land boundary conditions, an estuarine proxy that incorporates local tides and ocean and atmospheric operational conditions has been designed (Figure 3). This numerical model proxy, based on the MOHID model (http://mohid.com), produces more realistic hourly salinity concentrations and discharged volumes reaching the coastal area with low computational cost.

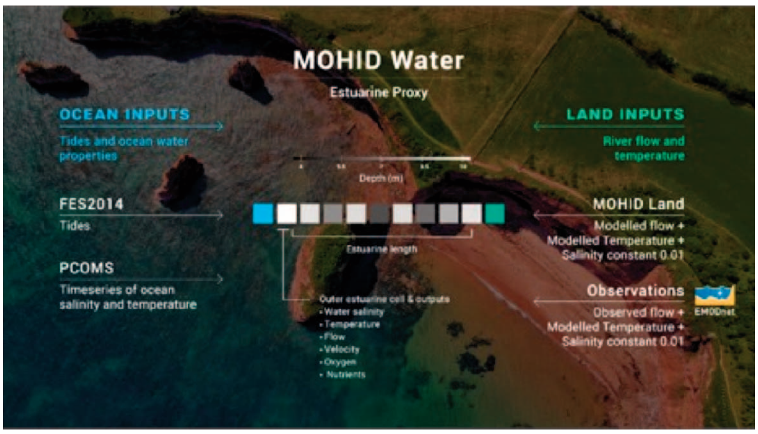


Figure 3 Diagram of the estuarine proxy and forcings.

Conclusions

The significance of accounting for river inputs in oceanographic analysis becomes evident, as rivers bridge the gap between land and ocean, influencing circulation patterns, biotic diversity, and essential processes such as eutrophication. The ongoing development of EMODnet Physics River Node, with its focus on near real-time outflow data from river mouth stations, reflects a crucial step towards comprehensive coastal management and effective support to EMODnet stakeholders.

Broad ranges for quality control of coastal data related to marine eutrophication: the Mediterranean Sea case study

Marina Lipizer[1] (mlipizer@ogs.it), Maria Eugenia Molina Jack[1] (mmolina@ogs.it), Megan Anne French[1] (mfrench@ogs.it), Alessandra Giorgetti[1] (agiorgetti@ogs.it), Anastasia Iona[2] (sissy@hnodc.hcmr.gr), Marilena Tsompanou[2] (mtsompanou@hcmr.gr)

[1]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)

[2]HCMR/HNODC Hellenic Centre for Marine Research, Hellenic National Oceanographic Data Centre (Greece)

Introduction

The assessment of marine environmental status at regional and sub-regional scales, as required by EU and regional policies, needs data derived from multiple, sometimes heterogeneous, data sources. To support data comparability, coherent data management and accurate and transparent quality control (QC) procedures are crucial [Molina Jack et al., 2020]. However, such procedures for heterogeneous datasets are not uniform or consolidated. Common data QC procedures involve automatic checks including "global, regional, and climatology range tests" to verify if data lie within expected extremes encountered in the oceans, in a particular region or period [SDN, 2010; U.S. Integrated Ocean Observing System V2.1, 2018]. Although concentration ranges have been proposed for data QC checks for open waters, due to complex land–sea interactions and intense spatio-temporal variability, data validation for coastal waters is still challenging.

This study aims to evaluate broad ranges of biogeochemical parameters in coastal waters to help validate data collected from multiple sources and managed by large scale data infrastructures, using data provided by EMODnet Chemistry.

Methods

This study uses data collected in the framework of scientific research projects and national monitoring activities driven by several EU and regional policies, which are collated, standardized, and harmonized within the framework of EMODnet Chemistry. Data and metadata are managed according to standard SeaDataNet protocols using common vocabularies, metadata, data formats, and data QC procedures [Giorgetti et al., 2020]. The Mediterranean Sea dataset[22] has been validated according to a shared protocol proposed within EMODnet Chemistry and the SeaDataNet scientific community [SeaDataNet, 2010; Lipizer et al., 2023]. As a result, all data have been assigned a quality flag (QF) based on the SeaDataNet QF scheme (SeaDataNet, L20 vocabulary). For the purpose of this study, profiles and time series were collated into a regional dataset spanning from 1911 to 2022.

To evaluate the broad ranges of parameters in marine waters influenced by terrestrial inputs by adopting a common approach for the entire Mediterranean Sea, "coastal influenced waters" were selected for four sub-basins of the Mediterranean: Western Mediterranean, Ionian and Central Mediterranean, Adriatic, and Aegean–Levantine seas according to the following approach:

- spatial diffusion of continental waters was evaluated according to sea surface salinity climatology provided by in situ data (SeaDataNet[23]) and by reanalysis[24] (Copernicus);

[22] Mediterranean Sea - Eutrophication and Acidity aggregated datasets 1911/2022 v2022

[23] https://www.seadatanet.org/Products/Climatologies

[24] https://marine.copernicus.eu/

- official layers of internal and territorial waters (12 nautical miles, nm) were used to filter data of "coastal influenced waters" as continental inputs often extend beyond explicit coastal waters.

Stations located in large semi-enclosed areas such as aquaculture areas were excluded. For each of the four sub-basins, the coastal dataset (i.e., coastline to 12 nm) was further subdivided into an upper water layer (0 – 74.99 m depth), and a middle deeper layer (75 – 199.99 m depth), and a deep layer (200–bottom depth). To propose concentration ranges for QC purposes, the data with QFs 1 ("good"), 2 ("probably good"), 6 (<limit of detection), and Q (<limit of quantification) for each depth layer were used to calculate statistics including the 50th percentile (i.e., median), 75th percentile, 99th percentile, and maximum values for dissolved oxygen, nitrate, nitrite, ammonium, phosphate, silicate, chlorophyll-a and dissolved inorganic nitrogen.

Results and conclusions

The general trends were common to most parameters (the results for nitrate, as an example, are displayed in Figure 1), with the highest concentrations and variability observed in the upper layer and in the Adriatic Sea, and the lowest in the Aegean and Levantine seas, which are recognised as the most oligotrophic parts of the Mediterranean.
This study allowed concentration ranges to be proposed for the four sub-basins of the Mediterranean for use in data validation. The proposed methodological approach could also be adopted for other marine regions to identify and update existing sub-region specific concentration limits for data validation.
The continuous improvement of QC is essential for policy related assessments and the development of research products that contribute to EU policies (Marine Strategy Framework Directive or Marine Spatial Planning) to prevent marine biodiversity loss and manage marine resources sustainably.

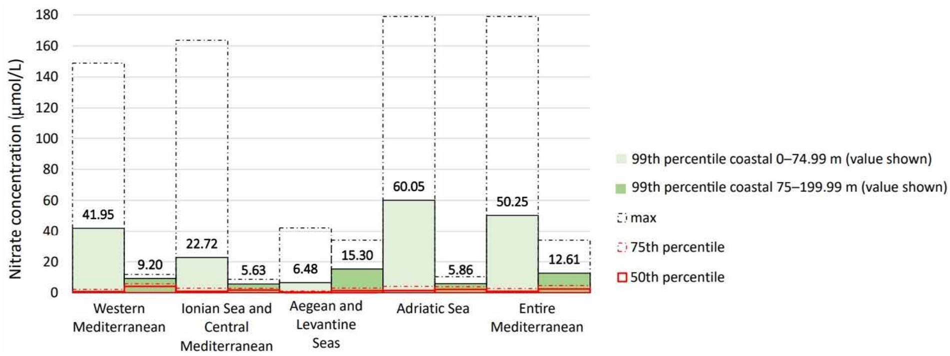


Figure 1 Example of results for nitrate concentration.

References

Molina Jack M.E. et al., (2020). https://doi.org/10.3389/fmars.2020.571365
Giorgetti A. et al., (2020). https://doi.org/10.3389/fmars.2020.583657
Lipizer M. et al., (2023). https://doi.org/10.13120/8XM0-5M67
U.S. Integrated Ocean Observing System V2.1.2018. *Manual for Real-Time Quality Control of Dissolved Oxygen Observations*. A Guide to Quality Control and Quality Assurance for Dissolved Oxygen Observations in Coastal Oceans. https://cdn.ioos.noaa.gov/attachments/2018/08/QARTOD_DOSecondUpdate_final.pdf
SDN SeaDataNet (2010). *Data Quality Control procedures, v.2.0*. https://www.seadatanet.org/Standards/Data-Quality-Control

EMODnet Bathymetry - High Resolution Seabed Mapping – increasing the resolution of the digital bathymetry for European seas

Thierry Schmitt[1] (thierry.schmitt@shom.fr), Dick Schaap[2] (dick@maris.nl),
George Spoelstra[3] (george@ggsgc.eu), Benoit Loubrieu[4] (benoit.loubrieu@ifremer.fr),
Cecile Pertuisot[4] (Cecile.Pertuisot@ifremer.fr), Martin Verlaan[5] (Martin.Verlaan@deltares.nl),
Ricard Campos[6] (ricard.campos@coronis.es), Knut Hartmann[7] (hartmann@eomap.de)

[1]Shom (France)
[2]MARIS (The Netherlands)
[3]GGSGc, (The Netherlands)
[4]Ifremer (France)
[5]Deltares (The Netherlands)
[6]Coronis Computing S.L. (Spain)
[7]EOMAP (Germany)

Introduction

EMODnet Bathymetry is part of the European Marine Observation and Data Network (EMODnet) initiative. EMODnet aims at assembling European marine data, data products and metadata from diverse sources in a uniform way. The EMODnet Bathymetry project is active since 2008 and has developed Digital Terrain Models (DTM) for the European seas, which are published at a regular interval, each time extending coverage, improving quality and precision, and expanding functionalities for viewing, using, and downloading.

Gathering and managing Bathymetric Data sets

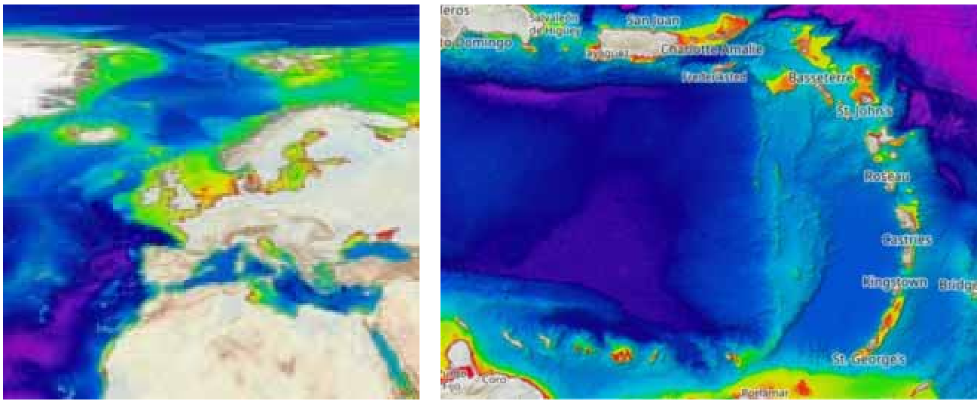
The DTMs are produced from surveys and aggregated data sets (Composite DTMs) that are referenced with metadata adopting the SeaDataNet Catalogue services. SeaDataNet is the pan-European infrastructure for marine and ocean data management, developed and operated by NODCs in Europe, and connecting already more than 115 major oceanographic data centres around the European seas. Bathymetric survey data sets are gathered and populated by national hydrographic services, marine research institutes, and companies in the SeaDataNet CDI Data Discovery & Access service. Currently, this amounts to more than 42.000 datasets, while more than 270 composite DTMs and Satellite Derived Bathymetry (SDB) products are included in the SeaDataNet Sextant catalogue service. The SDB data are based upon Landsat-8 and Sentinel-2 satellite images and fill gaps in coverage of the coastal zones, in particular in several Mediterranean countries. In total, data sets were received from 63 data providers. Further gathering is on-going and a major selection of these datasets will be used for preparing a new release of the EMODnet Digital Terrain Model (DTM) for all European waters, which is planned for end 2024.

Current Digital Terrain Model (DTM) Version 2022

The current EMODnet DTM was released in Early 2023 and has a grid resolution of 1/16 * 1/16 arc minutes (circa 115 * 115 m), covering all European seas including part of the Arctic Ocean and Barents Sea. Moreover, there is a DTM for part of the Caribbean region. This European DTM is based upon circa 21.790 in situ datasets and composite DTMs. Overall, this DTM contains approximately 12.3 billion grid nodes, organized in 113892 columns and 108132 rows (seabed and terrestrial coverage included) (for comparison: GEBCO has 933 million grid nodes for worldwide coverage, at a resolution of circa 500m*500m).

Migration to EMODnet Central Portal

Begin 2023, the information and products of EMODnet Bathymetry have been migrated from the dedicated EMODnet Bathymetry portal to the new EMODnet Central Portal, where products from all EMODnet Thematics are now available in one shopping window. The EMODnet Bathymetry DTM and other relevant map layers can now be viewed in the Central Portal Map Viewer service, whereby the maps are pushed from EMODnet Bathymetry by OGC WMS, WMTS, and WFS services. The Map Viewer also includes 3D viewing capability, based upon Cesium. Metadata information about each product is given by the Central Portal Products Catalogue which is synchronized with the OGC CSW catalogue of EMODnet Bathymetry by regular harvesting.



Figures New 2022 DTMs for European Seas and Caribbean region.

Published products for viewing and downloading

Next to the DTM, EMODnet Bathymetry provides access to: 1) a layer with digital Satellite Derived Coastlines for the Lowest Astronomical Tide (LAT), the Mean Sea Level (MSL) and the Mean High Water (MHW) tidal reference levels; 2) a layer with High Resolution hotspots, consisting of a collection of circa 260 high resolution composite DTMs for selected areas. Their resolution varies between 1/32 and 1/512 arc minutes, depending on the local data policy of data providers; 3) a layer with Digital Terrain Model Quality Indicators for vertical and horizontal precision, survey age, purpose of the survey, and combined quality; and 4) an inventory of national baselines and coastlines collected from 34 national authorities. EMODnet Bathymetry is also managing the European contribution to the international Seabed 2030 project. The EMODnet DTM can be freely viewed and downloaded (by 64 tiles), and shared by OGC web services. The EMODnet Bathymetry products are very popular and in 2021 – 2023 more than 100.000 EMODnet DTM tiles were downloaded by circa 12.000 users, and more than 68 million OGC service requests were registered over the 2 years.

Work in progress

Currently, EMODnet Bathymetry is working on gathering and processing new survey data sets for the production of the 2024 Version of the EMODnet DTMs and updating the other EMODnet Bathymetry products.

The i-waveNET Decision Support System – user-driven aggregations and analysis of forecasts and observations

Joel Azzopardi (joel.azzopardi@um.edu.mt), Audrey Zammit (audrey.zammit@um.edu.mt), Adam Gauci (adam.gauci@um.edu.mt)

University of Malta (Malta)

The Malta-Sicily channel is a key segment of the Mediterranean's 'marine highway' due to its heavy ship traffic. It attracts a wide range of stakeholders, including port authorities who ensure maritime safety, environmental agencies protecting the health of the sea, and vessel operators who must navigate safely and efficiently. The local community also has a vested interest, particularly when their livelihoods are linked to the coastal seas' health and condition. All these stakeholders share a common need for accurate and timely marine data, like nowcasts and forecasts.

Online services like the Copernicus Marine Service (CMEMS) (https://marine.copernicus.eu), EMODnet (https://emodnet.ec.europa.eu/) and SeaDataNet (https://www.seadatanet.org) provide valuable data. Yet, local stakeholders would benefit from a one-stop shop whereby they can access relevant high resolution data (including coastal-scale forecasts and observations) with ease. More importantly, rather than raw data, stakeholders would prefer visualisations or information on specific points or areas. This need was partly addressed by previous local initiatives such as the CALYPSO HF-Radar interface in 2012 [Azzopardi et al., 2013], the KAPTAN mobile application [Drago et al., 2016], and the PORTO online service [Drago et al., 2022]. These services provide real-time access to visualisations of observations, nowcasts and forecasts but offer limited capabilities for users to focus on specific points/areas of interest.

The purpose of the i-waveNET Decision Support System (DSS) was to provide a one-stop shop of such data. Its sources include: satellite and forecast data products from CMEMS; high-resolution forecasts from a number of forecasts maintained by the i-waveNET partners (including hydrodynamics, waves, meteo, and biogeochemistry); high-resolution currents from the CALYPSO HF-Radar network; and real-time observations from a number of meteo stations, sea-level stations and wave buoys in the Malta-Sicily channel.

The i-waveNet DSS interface (available at https://data.ocean.mt) was developed to be intuitive for users to be able to extract specific data for their area of interest. Users are provided with an expandable data catalogue from which they can choose simultaneously a scalar dataset (e.g., temperature) and a vector dataset (e.g., currents) and these are shown overlaid on each other (see Figure 1a below). Map visualisations are provided on a GIS interface utilising OpenStreetMap and Leaflet.js. Thus, users can pan and zoom to view their area of interest for the selected date/time and depth level (where applicable). The level of detail of the plots varies according to the current zoom level. The visualisations are fully customisable – users are able to modify the colour scale, and the palette as needed.

The i-waveNET DSS enhances user interactivity by enabling the examination of depth profiles and time series at specific points on the map. By right-clicking on any point on the visualised map, users can access a context menu to select and view detailed depth profiles or time series data for that exact point. This data is displayed through interactive charts created with charts.js, which reveal precise values upon mouse hover. Moreover, the system provides access to real-time observational data from wave buoys, meteorological, and sea-level stations – stations are presented as icons on the map which the user can click to open visualisation dialogs as shown in Figure 1b below. To further enrich the user experience, there's an option to log in via Google SSO. This login feature unlocks additional capabilities, such as downloading the depth profile and time series data, allowing users to seamlessly integrate this valuable information into their workflows.

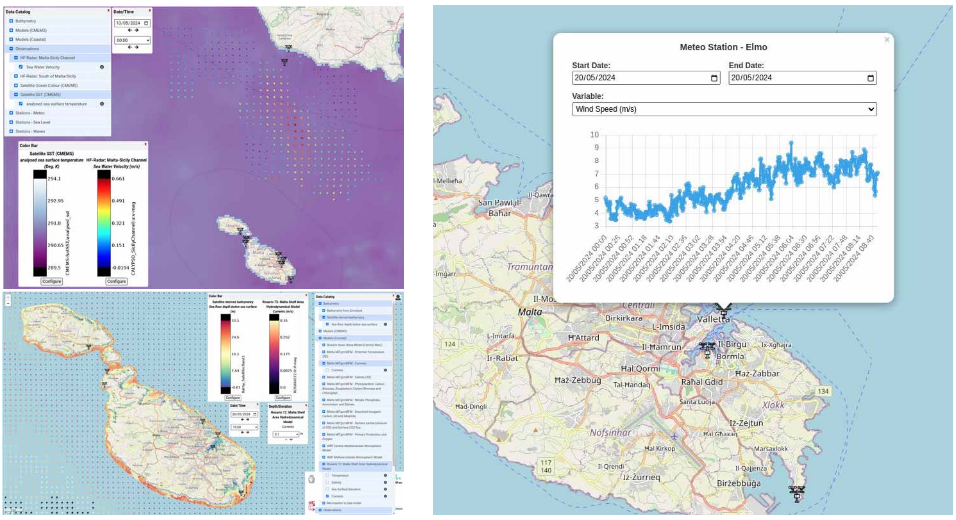


Figure 1 The i-WaveNET DSS interface: (a) map visualisations; (b) station time series visualisations.

The i-waveNET DSS was developed completely using open source technologies. The front-end is implemented using HTML5, CSS and Javascript. It is hosted on a Linux Apache2 server that proxies a python flask/gunicorn server. All server scripts are written in Python. The gridded datasets are hosted on an ncWMS2 server instance, and map visualisations are provided through WMS. The DSS also provides HTTP RESTful APIs to serve data to third party systems using JSON encoding. The i-waveNET DSS is currently operational and serves data from numerous models and observation stations. The addition of new data sources only involves an update from the system's admin interface, and the inclusion of this data on ncWMS (in case of gridded data). Currently, this DSS is undergoing enhancements to incorporate a climatology section providing climate and other value added services such as zonation information for the different categories of recreational water crafts as per Direction 2013/53/EU of the European Parliament.

Acknowledgements

The i-waveNET DSS was initially developed within the i-waveNET project which was funded by the Interreg European Regional Development Fund, Italia-Malta programme. The project was led by the Università degli Studi di Palermo (UNIPA), and the partnership included Università degli Studi di Catania (UNICT), Istituto Nazionale di Geofisica e Vulcanologia (INGV), Istituto Superiore per la Ricerca e la Protezione Ambientale (ISPRA), Consiglio Nazionale delle Ricerche (CNR) and Transport Malta (TM) together with the University of Malta.

References

Azzopardi J., Tarasova R., Gauci A.P., & Drago A., (2013). *The CALYPSO HF Radar Data Interface*. Rapport du Congrès de la Commission Internationale pour l'Exploration Scientifique de la Mer Méditerranée, 40, 844.

Drago A., Zammit A., Tarasova R., Gauci A., Galea A., Azzopardi J., Ciraolo G., Capodici F., (2016). *KAPTAN – A smartphone application for mariners*. Submitted to the International conference on Marine Data and Information Systems, Gdansk, Poland, 11-13 October 2016.

Drago A., Zammit A., Gauci A., Galea A., Azzopardi J., Galea de Giovanni R., Ciraolo G., Capodici F., Aronica S., Langiu A., Giacalone G., Fontana I., Sinatra R., Paradiso E., Campanella S., Ruvolo V., (2022). *PORTO Online – Featured data services in practice*. International Ocean Data Conference (IODC-1), Sopot, Poland, February 2022.

The International Quality-controlled Ocean Database (IQuOD)

Simona Simoncelli[1] (simona.simoncelli@ingv.it), Rebecca Cowley[2] (rebecca.cowley@csiro.au), Zhetao Tan[3] (tanzhetao19@mails.ucas.ac.cn), Rachel Killick[4] (rachel.killick@metoffice.gov.uk), Guilherme Castelão[5] (castelao@ucsd.edu), Lijing Cheng[3] (chenglij@mail.iap.ac.cn), Simon Good[4] (simon.good@metoffic.gov.uk), Tim Boyer[6] (tim.boyer@noaa.gov), Bill Mills[7] (william.mills-1@colorado.edu), Uday Bhaskar[8] (uday@incois.gov.in), Ricardo Locarnini[6] (ricardo.locarnini@noaa.gov)

[1]Istituto Nazionale di Geofisica e Vulcanologia - INGV (Italy)
[2]CSIRO Environment (Australia)
[3]Chinese Academy of Sciences (China)
[4]UK Met Office (UK)
[5]Scripps Institute of Oceanography (USA)
[6]NCEI (USA),
[7]University of Colorado (USA)
[8]INCOIS (India)

Advancing our understanding of climate variability and change and associated socio-economic and environmental impacts demands unbiased climate data of the highest quality, consistency, and completeness. Long-term historical records are critical to place modern ocean properties in context with the past, and to separate the influence of anthropogenic drivers of change from the modes of natural climate variability.

Subsurface ocean temperature and salinity are two of the Essential Climate Variables (ECVs) which provide a direct window into changes in the Earth's planetary heat balance, water cycle and sea level. Subsurface ocean observations are widely used to evaluate and initialize model simulations and are assimilated into ocean and climate models. The big challenge is thus maximising the potential of an irreplaceable collection (since 1772) of tens of millions of historical temperature profiles and worth tens of billions of dollars.

The IQuOD (International Quality-controlled Ocean Database, https://www.iquod.org/) effort is organized by the oceanographic community, and includes experts in data quality and management, climate modellers and the broader climate-related community. IQuOD aims to produce and freely distribute the highest quality and complete single ocean profile repository for use in ocean climate research applications. IQuOD is developing and implementing an internationally-agreed framework to achieve these goals and several Task Teams (TT) work in parallel to advance knowledge on different topics and incorporate new solutions into future IQuOD releases.

IQuOD follows FAIR (Findable, Accessible, Interoperable, Reusable) principles and a newly-formed FAIR data TT aims to ensure all tools and products generated conform to these principles. Key goals include maintaining the codes developed to be available and reusable to all, seeking feedback from users to improve them. To this purpose, a synergy has been set up with the Blue Cloud (BC) 2026 European project (https://blue-cloud.org/) that established a cyber platform, providing access to multi-disciplinary datasets from observations, analytical services and computing facilities for open web-based science. Some of the IQUOD tools will be tested and adapted within BC analytical workbenches for data intensive processes, generating qualified data collections merging data from various databases, among which the World Ocean Database (WOD, (https://www.ncei.noaa.gov/archive/accession/NCEI-WOD) and IQuOD. This cooperation will provide mutual feedback on data, metadata, tools and services.

The intelligent metadata TT has been focused on using machine learning to determine metadata for XBTs, instruments which commonly do not have make or model recorded. In IQuOD v0.1

[2018, (https://www.ncei.noaa.gov/archive/accession/0170893] release a simple decision tree style method is used, determining XBT type based on year, maximum depth and deploying country [Palmer et al. 2018, https://doi.org/10.1175/JTECH-D-17-0129.1]. Since then two further papers improving the correct XBT classification have been published, Leahy et al. [2018, https://doi.org/10.1175/JTECH-D-18-0012.1], using neural networks and Haddad et al. [2022, https://doi.org/10.1175/JTECH-D-21-0117.1], exploring multiple machine learning methods and recommending a sophisticated decision tree approach with an ensemble output allowing for better quantification of uncertainty (https://github.com/MetOffice/XBTs_classification). This newer work will be incorporated into a future IQuOD release.
The IQuOD v0.1 (2018) contains a 'first cut' of uncertainties for temperature data based on instrument type and manufacturer information [Cowley et al. 2021, https://doi.org/10.3389/fmars.2021.689695]. The uncertainties TT aims to improve and refine these measurement uncertainty estimates in several ways, including retrieval of historical information from voyage reports, published data archaeology efforts and creation of uncertainty budgets. It is also developing methods of calculating representativity errors for use in models and climate science applications. These methods will be applied to IQuOD future releases for not just temperature, but other variables such as salinity and oxygen.
Many international oceanic databases suffer from the existence of duplicate data and consider this a priority task. The goal of the duplicate & metadata integrity TT is to identify errors in metadata and remove duplicate profiles from WOD by a semi-automated process in combination with an expert manual check. At present, some criteria to define duplicates and errors have been established. The algorithm has been developed and applied to the 1975, 1995, 2011 WOD data and a benchmark dataset with duplicate flags is published [Song et al., 2023, https://doi.org/10.12157/IOCAS.20230821.001]. The mechanism of reporting the duplicates and wrong metadata has been established. A scientific paper is in preparation and the Python Package is at https://github.com/IQuOD/duplicated_checking_IQuOD.
Recently, a suite of automated QC tests has been published for application to data from the WOD [Good et al., 2023, https://doi.org/10.3389/fmars.2022.1075510]. These tools, written in Python, are a community effort from experts in the application of automated QC to ocean observations and are available at https://github.com/IQuOD/AutoQC.
In addition, expert QC [Castelao, 2020, https://doi.org/10.21105/joss.02063] is the point of difference between IQuOD and other ocean observation databases. IQuOD will supply flags applied by expert visual inspection of the data. Many datasets already exist (Hydrobase, XBT Ship of Opportunity datasets, GoShip/WOCE) where such QC flags have been applied. IQuOD will ingest these flags while an interface is being built for experts to access, view and QC the historical data that did not pass through it. This scientific QC may become a crowdsourced tool.
The Global Data Assembly Center TT applies the results of the various task teams to the WOD and serves the IQuOD data (https://www.ncei.noaa.gov/archive/accession/0170893), updated quarterly with uncertainties and intelligent metadata included both as static files and through a subset and delivery tool (https://www.ncei.noaa.gov/access/world-ocean-database-select/dbsearch.html). Duplicate data are removed as directed by the TT, improvements to the existing data and metadata from the Scientific QC TT are verified and applied. The auto QC and expert QC will be incorporated into the next version of IQuOD in the form of quality control flags.
Since nowadays education and outreach is a crucial activity for products uptake, training via the IODE Ocean Teacher Global Academy (OTGA, https://classroom.oceanteacher.org/) and other pathways will be developed by a dedicated TT in coming years to enable users to access the database and tools for their particular application.

Integrating deep learning and multiscale ecosystem connectivity of the marine habitat

Haitham Ezzy[1] (human.h89@gmail.com), Anna Brook[1] (a.brook@geo.haifa.ac.il), Dror Angel[2] (dangel@univ.haifa.ac.il)

[1]School of Environmental Sciences, University of Haifa (Israel)
[2]Recanati Institute of Maritime Studies, University of Haifa (Israel)

Introduction

Marine ecosystems are intricate tapestries of biodiversity and seascape structures, essential for the ecological balance and resilience of our oceans. As Costanza et al. [1997] articulated, these ecosystems are "among the most diverse, productive, and economically valuable, and they are among the most threatened by the impacts of humans." The characterization of these underwater realms is crucial for effective conservation, resource management, and understanding of historical anthropogenic impacts on marine habitats.

In the arena of marine science, artificial intelligence (AI) and three-dimensional (3D) modeling stand at the forefront of innovation. AI "has the potential to revolutionize our understanding of marine biodiversity patterns and processes" [Williams et al., 2019]. It sifts through complex datasets, unveiling patterns in biodiversity and ecological dynamics. In tandem, 3D modeling offers spatial precision in mapping underwater topographies, affording detailed examination of marine structures and associated biological communities.

The confluence of AI and 3D modeling represents a significant advancement in the way we understand and manage marine environments. These technologies not only enhance our visualization of the seascape but also provide insights into the structural and functional connectivity within these complex ecosystems. As Aguzzi et al. [2020] emphasize, "The application of multiparametric generation, storage, and automated hierarchic treatment of biological and environmental information is required to capture the spatiotemporal complexity of a marine ecosystem." This study harnesses state-of-the-art tools to delineate a detailed picture of underwater landscapes, thereby facilitating better management and protection of marine habitats, biodiversity and archaeological resources.

This study aims to address the following objectives/aims: 1) To Evaluate the Current State of Marine Biodiversity: we plan to utilize AI and 3D modeling to analyze the distribution of marine species across various seascapes to understand the influence of environmental factors on biodiversity; 2) To Advance Seascape Characterization Methods: we seek to enhance traditional seascape characterization methods by integrating AI-driven data analysis and 3D spatial modeling, providing a more comprehensive view of marine habitats; 3) To Develop Predictive Models for Conservation Efforts: Machine learning algorithms will be used to develop models that can predict changes in marine biodiversity and seascapes, supporting proactive conservation and resource management strategies; 4) To Optimize Monitoring and Management of Marine Ecosystems: we will demonstrate how AI and 3D modeling can improve the efficiency and precision of monitoring programs and inform the management of marine protected areas and other critical habitats.

Methodology- Data Acquisition

For our study, we have collected airborne LiDAR scans for bathymetry up to 30 meters deep, in both marine-protected and unprotected areas. We have also collected underwater videos recorded by fixed 360-degree cameras during various seasons, and we've utilized archived multibeam echosounder data from the 1990s. High-resolution 3D models have been created using photogrammetry from scuba-diving expeditions, vital for our marine structural and functional connectivity analysis. The multi-feature extraction process focused on the complexity and structural connectivity of the seafloor. This work leverages the synergy between LiDAR data and high-resolution 3D photogrammetry models, enabling us to reconstruct and enhance the resolution of MBES data using SRGAN. This innovative process not only improves the fidelity of seafloor mapping but also aids in understanding structural connectivity by revealing intricate habitat features that influence species distribution and ecological linkages.

Deep Learning for Biodiversity Assessment

For fish detection and classification, deep learning models, particularly CNNs, will be trained to recognize and identify fish species from the extensive video data collected. This process enables us to understand functional connectivity, as species identification is crucial for tracking movement patterns and habitat utilization, thereby linking these patterns to the structural aspects of the marine environment.

Multi-Scale Habitat Characterization

Multi-scale analysis of the habitat introduces a nuanced perspective to both structural and functional connectivity. By assessing spatial patterns and terrain heterogeneity, we can infer not just the physical configuration of the seafloor (structural connectivity) but also predict how these features influence species distribution and behavior (functional connectivity). This multi-tiered approach allows us to explore the dynamics of marine ecosystems, discerning how structural attributes of habitats shape the ecological functions within.

Integrating Connectivity in Seascape Ecology

In conclusion, we plan to bridge the gap between structural and functional connectivity through our analytical processes. The structural connectivity depicted by the seafloor's physical composition is mapped and modeled to reveal potential pathways and barriers for marine life. Subsequently, the observed fish species distributions and behaviors, derived from deep learning classification, illustrate the functional connectivity, showcasing how marine life interacts with the structural components of their habitat. By examining the interplay between these two forms of connectivity, our study aims to provide a comprehensive view of marine ecological dynamics, essential for conservation and management efforts.

Results

Preliminary outcomes from our SRGAN-enhanced MBES data show a promising decrease in RMSE to 14.22, improving upon the baseline RMSE of 20.4. These improvements signal the beginning of our journey to match the high resolution of LiDAR data. Our next steps involve further model refinement using a broader set of depth data, aiming to achieve LiDAR's level of detail in seafloor reconstructions. Our deep learning work for fish detection and classification is advancing, with efforts underway to amass a more varied dataset from different marine habitats and times. Presently, our model recognizes 70% of fish from five species; with dataset expansion and further training, we anticipate pushing this accuracy to 90% or more. This step is crucial for precise biodiversity assessments and effective conservation actions. Such progress underscores our dedication to enhancing marine data analysis and our research's adaptability for broader conservation uses.

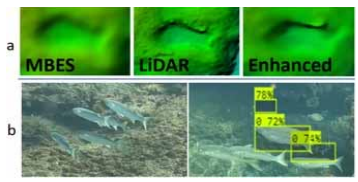


Figure 1 Comparative Analysis and Fish Detection in Marine Study. (a) Sequential images depict seafloor mapping progression: original MBES data, detailed LiDAR scans, and SRGAN-enhanced MBES resolution. (b) Fish detection via algorithm, with the left image as the unprocessed original and the right showing identified fish post-algorithm application with bounding boxes.

SESSION PRODUCTS POSTERS

MedSeaPod: a Citizen Science pilot project

Alberto Ribotti[1] (alberto.ribotti@cnr.it), Sergio Stagno[2] (stagno@netsoul.net), Mauro Montis[2] (montis@netsoul.net), Stefano Schirru[2] (schirru@netsoul.net), Alessandro Madeddu[3] (alessandro.madeddu@maraltro.com), Andrea Satta[1] (andrea.satta@cnr.it), Stefano Marras[1] (stefano.marras@cnr.it)

[1]Consiglio Nazionale delle Ricerche - Istituto per lo Studio degli Impatti Antropici e la Sostenibilità nell'ambiente marino - CNR – IAS (Italy)
[2]Netsoul srl (Italy)
[3]Maraltro slrs (Italy)

Extreme warming events known as heat waves or Marine Heatwaves have increased in frequency and magnitude worldwide, with detrimental impacts on coastal ecosystems and human activities [Frölicher et al., 2018; Oliver et al., 2018; Smale et al., 2019]. Sea Surface Temperature (SST) satellite data from the Copernicus Marine Service show an average summer surface temperature in the Mediterranean over 2.5 °C in 2018 and of 5 - 6 °C in 2022. The seas around Sardinia reached temperatures over 30 °C in summer 2022. However, these satellites do not have such a high resolution to cover coastal areas with a seabed below 50 m, nor to measure the main oceanographic parameters (temperature, salinity and pressure) between the surface and the bottom. In a recent study on Mediterranean heat waves in the period 1989-2018, Ciappa [2022] only uses NOAA's SST-AVHRR satellite data, thus not covering coastal areas. Similarly, Garrabou et al. [2022] uses SST data from CMEMS and NOAA OISST for the studied period 2015 - 2019 but complementing them with hourly high-frequency time series of temperature only acquired by underwater operators using HOBO data loggers and uploaded to the T-MEDNet database (http://www.t-mednet.org/). This database, created within the homonymous project, consists of data generally acquired every 5 m between the surface and 40 m depth in Mediterranean coastal areas. However, not all Mediterranean coasts are covered by the same number and frequency of samplings. In Sardinia there are only five samplings dating back to 2017, three located in the north-east and two in the south-east. Considering the high number of tourists on the island in summer, the amount of data that could be acquired in coastal areas would be considerable, with an important direct impact on the knowledge of the state of Sardinian coastal marine ecosystems and, indirectly, on the island's economy (mainly fishing and tourism). Every day local fishermen, yachtsmen, spearfishermen, swimmers, freedivers and divers dive in the waters of the Mediterranean Sea and are potential acquirers of scientifically useful data. Why not provide these people with a tool to actively contribute to monitoring the health of our sea? Citizen science and data crowdsourcing are the tool to facilitate the monitoring and protection of the marine environment and raise general awareness of the impact of climate change on the sea and marine life. Based on these considerations, on mid-September 2023 two small Italian-Sardinian ICT enterprises (Netsoul srl and Maraltro slrs), the Council of Carloforte (the village of the small island of San Pietro in Sardinia), and the institute in Oristano (Sardinia, Italy) of the National Research Council (CNR-IAS) started the 2-years Citizen science pilot project Mediterranean Sea People's Observatory Data (MedSeaPod), financed by the Sardinian Region through the Regional Development Programme 2020 - 2024.

What is

MedSeaPod is a research and development pilot project based on scientifically rigorous processes for collecting big data of localized, through an integrated GPS, sea temperature and pressure in Sardinian coastal marine areas. This will be done through the intense use of citizen

science activities localized around the small island of San Pietro, SW Sardinia. It aims to process, display and share the collected data, with the scientific community and policy makers and civil society, through an innovative and interactive data visualization system. So acquired data will be potentially used to develop new conservation programs for coastal ecosystems and raise awareness among the population about the ongoing effects of climate change through creative and effective communication approaches.
Three are main goals to achieve in this pilot project:

- create a dataset of temperature along the coastal areas of Sardinia that will constitute a benchmark for future measurements and analyzes by the global scientific community;
- generate an “active awareness” of the impact of climate change on marine ecosystems, to be done specifically in public opinion, in citizens and in political decision-makers;
- popularize scientific research. It means making the research easily understandable and shareable to make it perceived as essential to the environmental, social and economic well-being of all operators of the sea and citizens.

How MedSeaPod works

Sea temperature, water pressure and GPS data will be acquired by the use of small commercial sensors. Data will be periodically compared with those from CTD probes. These sensors will be distributed to different subjects as free divers, diving and SUP centers, underwater fishermen, snorkelers, local fishermen, and the Coast Guard. These people will be the active protagonists of our Citizen Science activities in the small island of San Pietro. Data will be automatically sent to a server after processing. Good data will be part of a dataset online with free access from users.

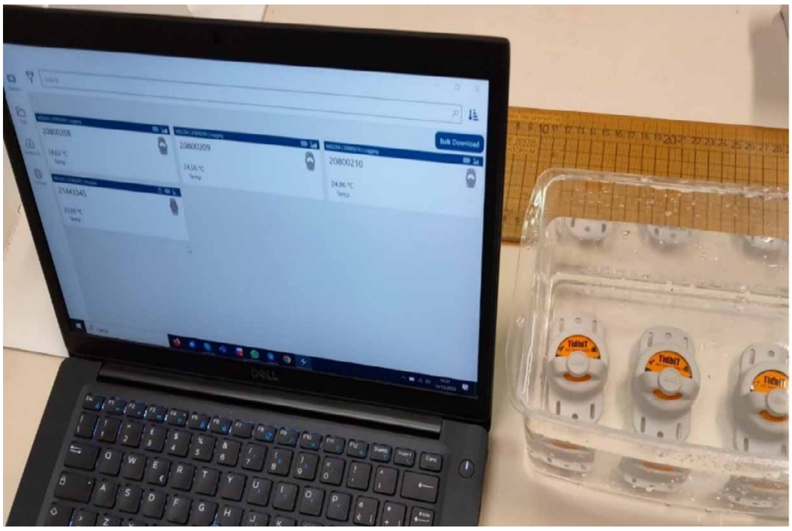


Figure 1 The three Onset HOBO 1-800-loggers tested in the CNR lab.

At mid-October 2023, after having evaluated different sensors, the choice was on the robust data loggers from Onset Corp. and TechnoSmart. Waiting for TechnoSmart’s data loggers with further GPS and pressure sensors planned in early 2024, the first tests occurred in the CNR laboratory in Oristano on three Onset HOBO 1-800-loggers with temperature sensor (Figure 1). This to evaluate criticalities in data acquisiton and start writing protocols in data collection, processing and web applications.

Expected results and future steps

MedSeaPOD data will be used as a benchmark for analyzing ongoing climate change and predict future climatic scenarios, create the first coastal thermal map in Sardinia and validate coastal numerical models. In the future, the project could be exported to other coastal areas and improve the acquisition system with other sensors (i.e., dissolved oxygen and conductivity).

References

Ciappa A.C., (2022). https://doi.org/10.1016/j.pocean.2022.102828
Frölicher T.L., Fischer E.M., Gruber N., (2018). https://doi.org/10.1038/s41586-018-0383-9
Garrabou J., Gómez-Gras D., Medrano A. et al., (2022). https://doi.org 10.1111/gcb.16301
Oliver E.C.J., Donat M.G., Burrows M.T. et al., (2018). https://doi.org/10.1038/s41467-018-03732-9
Smale D.A., Wernberg T., Oliver E.C.J. et al., (2019). https://doi.org/10.1038/s41558-019-0412-1

Data-driven habitat suitability modelling to support decision making in sustainable pelagic fisheries

Ward Standaert (ward.standaert@vliz.be), Rutendo Musimwa (rutendo.musimwa@vliz.be), Martha Stevens (martha.stevens@vliz.be), Steven Pint (steven.pint@vliz.be), Gert Everaert (gert.everaert@vliz.be)

VLIZ, Flanders Marine Institute (Belgium)

Introduction

Due to the withdrawal of the United Kingdom from the European Union, the Belgian fishing fleet lost access to part of their fishing waters and experienced diminished catches. The Belgian fishing fleet currently targets demersal fish [Landbouw en Visserij, 2021]. One of the initiatives to overcome the loss of fishing grounds is to provide information about alternative fishing grounds and niche fisheries. This study aims to contribute to the question of where and when pelagic fishing by the Belgian fishing fleet could be reintroduced by looking at the seasonal distribution of Atlantic herring (*Clupea harengus*), an economically important pelagic fish species in the Northeast Atlantic.

Habitat suitability models have been widely used to derive species-environment relationships and predict the geographical distribution of species [Thorson et al., 2016]. We developed a habitat suitability model for adult Atlantic herring (*Clupea harengus*) in the Northeast Atlantic. To take herring's spawning distribution into account, we also built a model for herring larvae.

Materials & methods: data

Occurrence data of Atlantic herring were retrieved from the International Council for the Exploration of the Sea (ICES) Database of Trawl Surveys (DATRAS, 13 294 adult records) and the International Herring Larvae Surveys (IHLS, 7 902 larvae records) in the Northeast Atlantic, restricted to 48°N – 62°N and 12°W – 10°E and from 2000 to 2020.

Sampling bias was considered by filtering occurrences in geographical space [Vollering J. et al., 2019]. A minimal distance of 10 nautical miles (NM) was taken as it is the recommended distance between valid hauls in the trawl surveys [Andersens, 2010]. Initial results showed spatial autocorrelation in the model residuals which was minimised by an additional filtering technique in environmental space, using the Malahanobis distance. [5]. A total of 400 adult and 400 larval occurrences were maintained to fit the two habitat suitability models.

Environmental variables were retrieved from the European Marine Observation and Data Network (EMODnet) and Copernicus Marine Service, including depth (https://emodnet.ec.europa.eu/en/bathymetry), Sea Surface Temperature (SST, https://doi.org/10.48670/moi-00021), Sea Surface Salinity (SSS, https://doi.org/10.48670/moi-00021), zooplankton (https://doi.org/10.48670/moi-00020) and phytoplankton concentration (https://doi.org/10.48670/moi-00058), nearby windfarm presence (https://emodnet.ec.europa.eu/en/human-activities) and seabed characteristics (https://emodnet.ec.europa.eu/en/seabed-habitats).

Preprocessing of these environmental variables involved aggregation to a 10NM-by-10NM grid per month to match the spatiotemporal resolution of the occurrence data [de Oliveira et al., 2014]. To account for multi-collinearity in the predictors, the least ecologically relevant variable of variables pairs with a Variance Inflation Factor (VIF) larger than 10 was discarded from analysis [Sillero et al., 2021].

Materials & methods: model

A machine-learning method, called maximum entropy (MaxEnt) model, was used to develop species-environmental relationships (using R-package *dismo* [Hijmans et al., 2023; Phillips et al., 2006]. Background points were sampled randomly and restricted to the ICES areas of the occurrences. The number of background points was set at 10 times the number of presences [Hysen et al., 2022]. Applying the appropriate model settings can enhance model interpretability and avoid overfitting [Merow et al., 2013]. Therefore, different MaxEnt model settings (feature class and regularisation multiplier) were evaluated using the corrected Akaike's Information Criterion (AICc) as a selection criterion (R-package *ENMeval*, Kass et al. [2021]). Model performance was evaluated using the Area Under the Curve (AUC) of the Receiver Operating Characteristic plot and the True Skill Statistic (TSS) using a 5-fold cross-validation [Báez et al., 2020; Liu et al., 2013].

Results

The larval model performed best (AUC of 0.9 and TSS of 0.7 compared to AUC of 0.7 and TSS of 0.3 for adults). Depth was the most important variable in both models (63 and 37% of total variable importance for adults and larvae respectively), followed by SST (13%) and SSS (11%) for adults and SST (15%), seabed substrate (14%), zoo- (13%) and phytoplankton (13%) concentration for larvae. For both life stages, habitat was most suitable in shallower waters compared to deep waters and at low sea surface water temperatures. If pelagic fishing by the Belgian Fishing fleet were reintroduced, fishing in the Belgian Part of the North Sea would involve the least shipping costs. In this area, habitat was found most suitable during winter months for both adults and larvae. Due to this overlap and to protect spawning stock, fisheries should be effectively managed. During other seasons, the Belgian fishing fleet would have to travel further, to the greater North Sea, to find Atlantic herring.

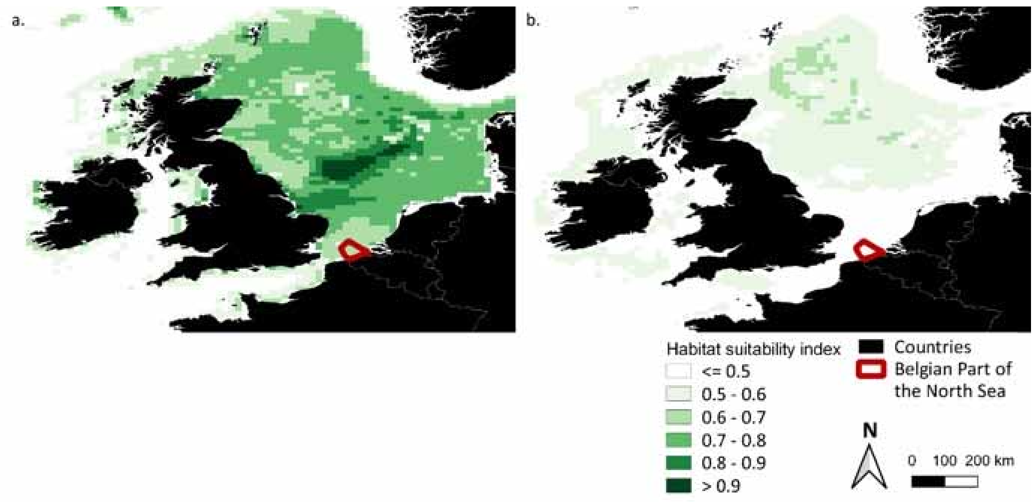


Figure 1 Predicted habitat suitability for Atlantic herring in the Northeast Atlantic in (a) January and (b) July averaged over 2000 - 2020. Darker shades of green depict areas that are more suitable during that month.

Acknowledgements

This work was supported by the Brexit Adjustment Reserve project (BAR) [0051] and the EU Public Infrastructure for the Digital Twin of the Ocean (EDITO-Infra) [project No 101101473].

References

Landbouw en Visserij, (2021). *Departement.*

Thorson J.T. et al., (2016). *In Methods in Ecology and Evolution* 7, 990–1002.
Vollering J. et al., (2019). *Ecography* 42, 1717–1727.
Andersens H.C., (2010).
de Oliveira G. et al., (2014). *Ecography* 37, 637–647.
Sillero N. et al., (2021). *International Journal of Geographical Information Science* 35, 213–226.
Zuur A.F. et al., (2010). *Methods Ecol Evol* 1, 3–14.
Hijmans R. et al., (2023).
Phillips S.B. et al., (2006). In *International Journal of Global Environmental Issues* 6, 231–252 (Inderscience Publishers).
Hysen L. et al., (2022). *Ecol Inform* 72.
Merow C. et al., (2013). *Ecography* 36, 1058–1069.
Kass J.M. et al., (2021). *Methods Ecol Evol* 12, 1602–1608.
Báez J.C. et al., (2020). *Ecol Evol* 10, 175–184.
Liu C. et al., (2013). *J Biogeogr* 40, 778–789.

Modern responsive web application with maps, challenges and their solution (example: bathing waters)

Damir Ivanković (ivankovic@izor.hr)

Institute for Oceanography and Fisheries (Croatia)

Sea bathing water quality on beaches in the Republic of Croatia – web application

In order to achieve the prescribed bathing water quality standards (Croatian National Ordinance on the Quality of Marine Bathing Water and EU Directive on the Management of Bathing Water Quality No. 2006/7/EC), measures for the management of marine bathing waters are hereby established. The monitoring of bathing water quality is carried out by a legal entity authorised to carry out monitoring activities in the field of environmental protection in accordance with the Environmental Protection Act. A web application was developed at the Institute of Oceanography and Fisheries in 2009 for the purpose of collecting, validating, managing and presenting this data to the public (presented as a poster at IMDIS 2010). After many upgrades and special mobile version of this application, it is time for a new unified and optimised responsive web application.

Introduction

Modern web applications are interactive and consist of two basic parts: the data itself and the program code. In our case, these two parts are separate from each other. The application itself is customized and provided as a static web page with JavaScript. All data is sent separately and interactively in JSON format when requested. Even the strings for the multilingual interface are sent as JSON. We try to use existing services and open-source solutions. So, we use OpenStreetMap tiles and layers as the default base layer and switch from Google Maps api to Leaflet. Additionally, we have a local Geoserver instance to provide tiles for one of the basic map layers.

Challenge number one - page layout

Different devices have different screen resolutions: from desktop computers with multi-K (HD) resolutions (HD, 2K, up to 4K), to mobile devices with vertical orientation and even lower resolutions starting at 720p. Our approach was to use the full-screen view (without a fixed maximum width) for horizontal orientation (desktop and tablets) and the switchable “Info” and “Map” views for vertical orientation (smartphones). In the desktop/tablet layout (horizontal), the total width is split in a ratio of 40/60 for the “Info” and “Map” areas.

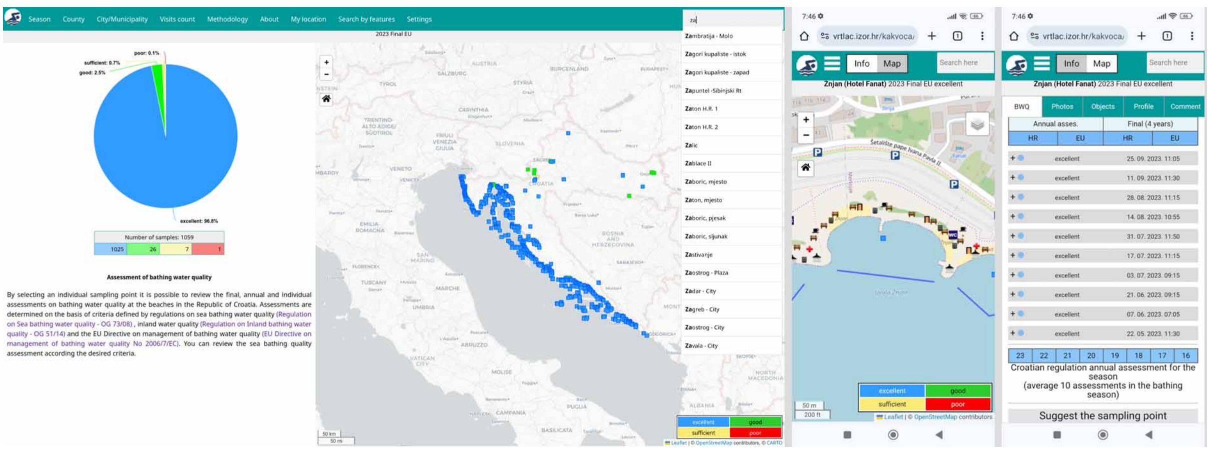


Figure 1 Desktop and alternate mobile layouts.

Figure 1 shows the page layout for the desktop and mobile case. In the case of a medium tablet resolution, the layout is the same as in the desktop case, only the top menu is "compressed" and is displayed when clicked. The info part of the view is divided into five tabs, each of which shows a group of information about a specific selected beach. The main div elements have a fixed position on the screen, which is essential for accurate map display and menu and search accessibility. Fixed div positions also solved problems with changing the viewport on mobile browsers (hiding the address bar), which can lead to inaccurate page rendering.

Challenge number two - user interface / finding information

For any application (web applications are no exception) it is important that the user can find the desired information easily and intuitively. We apply the most important principles: a simple, one-step main menu and the ability to find a specific location in at least three ways. The user can click on the place marker on the map, search by the name of the place or city or find the places nearby by positioning. It is also important that the map follows the user's actions in the info area and vice versa.

Challenge number three - performance and optimizations

It is important for the web application that the loading time is reasonably short. It is also important that the amount of information transmitted to the user is as small as possible. To achieve this goal, we have decided that we will use client-side processing as much as possible and that we will transfer all data asynchronously in JSON format, including the interface strings. The application itself contains only basic elements and JavaScript (and can be loaded from a fixed-pages web server - Apache). During initialization, the application recognizes the user's language preferences and then loads interface strings and fixed text (the application is multilingual). The second JSON is loaded from the database and contains a list of locations with the latest bathing water ratings and object codes. If the user wants to filter the locations and e.g., find locations with showers, this search is performed locally on the client side, such as searching by name or by city name. When the user "opens" a specific location, a special additional JSON is loaded that contains detailed information about this location (ratings, pictures, beach profile).

Conclusion

With the continuous growth of internet access speed (5G, fibre), as a side effect, we can often find web pages sub-optimised with many non-essential JavaScript code and multimedia content (e.g., multi-MB image displayed in small dimensions). Also, the global trend is to load and use many cloud-based services as part of a web application. Our approach was different. We focused on efficiently presenting bathing water quality information to our users, often using mobile traffic while roaming to minimise the data load and also the load on our servers. We have not forgotten to include the possibility of feedback (submitting comments with optional photo upload). We hope that our application fulfils its purpose and raises awareness for a responsible use of marine resources. The application can be found at https://vrtlac.izor.hr/ords/kakvoca/bwdn

Habitat suitability modelling of pelagic fish using a mechanistic approach: Insights for Sustainable Decision-Making in the Belgian part of the North Sea

Rutendo Musimwa (rutendo.musimwa@vliz.be), Ward Standaert (ward.standaert@vliz.be), Martha Stevens (martha.stevens@vliz.be), Steven Pint (steven.pint@vliz.be), Gert Everaert (gert.everaert@vliz.be)

VLIZ, Flanders Marine Institute (Belgium)

Introduction

The economic landscape of the North Sea is significantly impacted by pelagic species, making it crucial to thoroughly understand their migration patterns and habitat preferences to support sustainable fisheries management. After Brexit, Europe shifted its focus from demersal fisheries, prevalent before Brexit, to pelagic fisheries, primarily due to the closing of the United Kingdom's fishing grounds for foreign fishermen. Belgian fishermen, for instance, have been affected as they can no longer fish for flatfish e.g., *Solea solea* in UK waters. Consequently, the exploration of pelagic fisheries in the Belgian Part of the North Sea (BPNS) has gained prominence.

This transition necessitates ecological models like the mechanistic niche models which provide invaluable insights into environmental drivers and population dynamics. These models predict suitable habitats using the mathematical description of the species' niche by understanding their interaction with the environment. Furthermore, this tool can incorporate climate-induced habitat suitability, delving deep into the intricate relationship between changing environmental conditions and the distribution of pelagic fish. In the face of the challenges posed by a changing climate, this multifaceted research provides insights that pave the way for developing informed and sustainable management strategies.

Mechanistic Model: Data and model

The mechanistic modelling approach employed in this study was adapted from Westmeijer et al. [2019] for modelling pelagic fish distribution in the Belgian Part of the North Sea (BPNS). The study selected target fish species based on their economic importance and regional occurrence (i.e., Clupea *harengus* (Atlantic herring), *Scomber scombrus* (Atlantic mackerel), and *Dicentrarchus labrax* (European Seabass). Informed by fish niche ecology, we collected key influencing variables (as detailed in Table 1). To ensure data consistency and accuracy, a rigorous preprocessing phase was undertaken, involving the handling of missing data and standardizing units. Species-specific response curves were derived from a thorough review of existing literature and expert knowledge (see Figure 1), and these insights were integrated into our models using advanced population modelling techniques. The study employed fuzzy logic principles, which permit habitat suitability values to span any real number between 0 and 1, to explore the full spectrum of habitat conditions and their influence on fish distribution. These insights were integrated into the mechanistic models, allowing for predicting habitat suitability indices (HSI) under various conditions. To assess the impact of climate change on these species, the study utilised climate prediction data for temperature and salinity from Bio-Oracle [Assis Jet al., 2018; Tyberghein et al., 2011].

Variable	Units	Clupeaharengus	Scomber scombrus	Dicentrarchus labrax
Sea surface temperature	°C	x	x	x
Sea surface salinity	PSU	x	x	x
Bathymetry	m	x	x	x
Human artefacts	m	x	x	x
Seabed energy	/	x		
Seabed substrate	/	x		
Sea surface velocity	m s-1		x	
Euphotic depth	m		x	
Dissolved oxygen	mmol m-3		x	
Diet	g C m-2	x	x	

Table 1 Variables included per species.

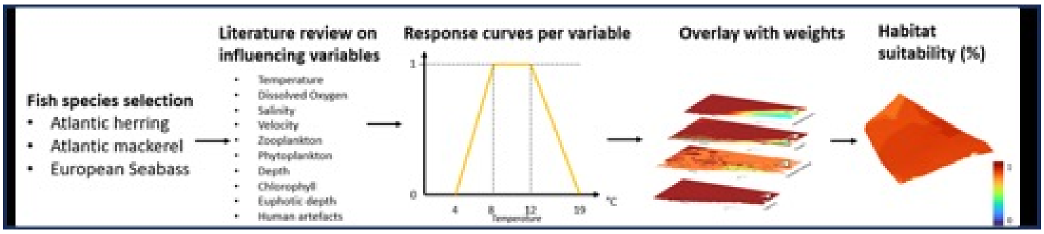


Figure 1 Work plan for the mechanistic approach with an example of one of the response curves produced.

All habitat suitability maps were uniformly resampled at a resolution of 100 m x 100 m and subsequently analysed using R [R Core Team, 2018] Different models were developed for spawning and non-spawning adults, focusing on temporal variations achieved by modelling each month separately. What-if scenarios based on climate change scenarios were used to simulate the habitat suitability of the selected species. Our analysis covers the period from 2020 to 2090 and was first applied to the Belgian part of the North Sea and then extended to the European Seas. We validated our models with fish occurrence data from various literature and OBIS data (www.obis.org).

Results

Warming oceans and changing salinity levels due to climate change were projected by the study to push economically important fish species in opposite directions. Warmer waters will likely make northern areas more suitable, while changes in salinity may push these fish south as seen for Atlantic herring (Figure 2). Under the worst-case scenario, the suitable habitat for all three fish species could be moved over 500 kilometres north due to rising temperatures. This northward movement due to climate change has been documented [Hu L et al., 2022;

Overholtzet al., 2011; Perry et al., 2005]. Our study unveiled that, climate-driven migrations of pelagic fish in European seas create a dynamic but an unresolved environment with the post-Brexit landscape. To navigate this intricate challenge, innovative solutions and renewed international commitment are essential for securing a sustainable future for both the marine environment and dependent communities.

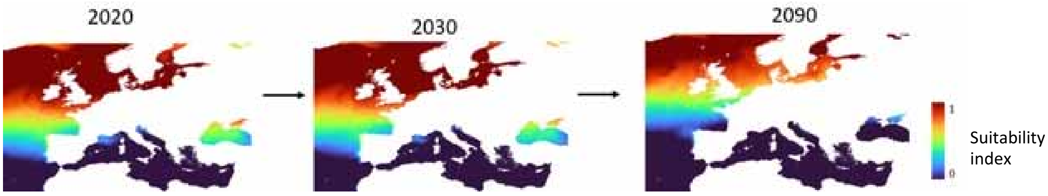


Figure 2 Habitat suitability predictions for Atlantic herring for current (2020) and future (2030 & 2090) due to sea surface temperature SSP585 pathway predictions. Habitat suitability areas are shown to migrate more to northern latitudes. HSI Red is high habitat suitability and blue is low habitat suitability.

Acknowledgements
This work was supported by the Brexit Adjustment Reserve project (BAR) [0051] together with the Flanders Marine Institute (VLIZ) and the EU Public Infrastructure for the European Digital Twin project (EDITO-Infra).

References

Westmeijer G., Everaert G., Pirlet H., De Clerck O., Vandegehuchte M.B., (2019). *Mechanistic niche modelling to identify favorable growth sites of temperate macroalgae*. Algal Res 2019;41:101529.

Assis J., Tyberghein L., Bosch S., Verbruggen H., Serrão E.A., De Clerck O., (2018). *Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling*. A Journal of Macroecology 2017;27:277–84.

Tyberghein L., Verbruggen H., Pauly K., Troupin C., Mineur F., De Clerck O., (2011). *Bio-ORACLE: A global environmental dataset for marine species distribution modelling*. Global Ecology and Biogeography 2012;21:272–81.

R Core Team, (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing [Internet] 2023 [cited 2023 Aug 11]; Available from: https://www.R-project.org/

Hu L sha, Dong Y wei, (2022). *Northward shift of a biogeographical barrier on China's coast*. Divers Distrib 2022;28:318–30.

Overholtz W.J., Hare J.A., Keith C.M., (2011). *Impacts of interannual environmental forcing and climate change on the distribution of atlantic mackerel on the U. S. northeast continental shelf.* Marine and Coastal Fisheries. 2011;3:219–32.

Perry A.L., Low P.J., Ellis J.R., Reynolds J.D., (2005). *Climate change and distribution shifts in marine fishes*. Science (1979) 2005;308:1912-1915.

Whale of a Time: The JCDP Vision for Standardization and Universal Access of Cetacean Data

Carlos Pinto[1] (carlos@ices.dk), Neil Holdsworth[1] (Neilh@ices.dk), Nikki Taylor[2] (Nikki.Taylor@jncc.gov.uk), Niki Clear[2] (Niki.Clear@jncc.gov.uk)

[1]ICES (Denmark)
[2]JNCC (UK)

Introduction

Cetaceans (whales, dolphins and porpoises) are highly mobile species that inhabit an environment that is challenging for us to observe and understand. Their distribution and behaviour further hamper efforts to monitor and research these cryptic species, making monitoring methods labour intensive and expensive. Cetacean data are important for drawing accurate conclusions on population size, dynamics and changes over time.

These data are collected by a wide range of organisations for a variety of purposes, but all with the same goal of a clearer understanding of cetacean abundance and distribution in their area of interest. Independently, much of this data is limited in capacity to perform analyses at relevant spatial and temporal scales, but together, have the potential to significantly increase our understanding of these cryptic creatures.

Historically, there has been no mechanism to facilitate access to all existing effort-related cetacean monitoring datasets in the north-east Atlantic. The process collating available data is labour intensive and inefficient, and often involve project specific data access agreements. However, there are multiple examples of analyses using collected data in this way have proven to be highly effective and under-pin key national and international population and indicator assessments.

The Joint Cetacean Data Programme (JCDP) aims to better facilitate the process of accessing and utilising cetacean survey data by collating the growing evidence-base across the north-east Atlantic into a single accessible resource. These data have a wide range of applications and have the potential to support analyses at a range of spatial and temporal scales.

Vision

The JCDP vision is to promote and facilitate cetacean data standardisation and maximise value through collation and enabling of universal access, the JCDP aims to achieve this vision through:

- Development of an international platform to host cetacean survey datasets from the north-east Atlantic.
- Development of a data standard to guide data collection and storage to enable a high-quality collation of data.
- Provision of regularly updated open access data products for use in strengthening cetacean science and subsequent decision-making.
- Facilitating access to the collated dataset for use in bespoke analyses.

History

The International Council for the Exploration of the Sea (ICES) was contracted by Joint Nature Conservation Committee (JNCC) on behalf of Defra, to build the Portal including the database and associated infrastructure to explore and access the data, which is hosted in the ICES Data Centre. The database comprises data that meets the agreed data standard from multiple data providers collected through ship-based and aerial methodologies. To develop this platform ICES and JNCC facilitated a series of workshop with the stockholders to define the viable mininum product and agree a data standard for the JCDP. Using an AGILE methodology, both teams (JNCC and ICES) developed the platform, data standard and supporting guidance documents using springs. The platform was launched in December of 2022.

The JCDP Platform

The JCDP platform consists of thee éléments; the JNCC Information Hub, hosted on the JNCC website (https://jncc.gov.uk/jcdp/), the JCDP Data Portal, hosted on the ICES Data Centre (https://cetaceans.ices.dk), and a linked metadata catalogue hosted on GeoNetwork.

The JCDP Information Hub provides background information about the programme and hosts key resources including the data standard, adopted by Marine Environmental Data and Information Network (MEDIN) in the UK and Ocean Best Practices, along with guidance and resources for uses on how to submit, use, and collect at-sea cetacean data which are in line with the JCDP.

The JCDP Data Portal is a website that allows data providers to upload cetacean survey data into the JCDP database and enables data users to explore data products and download datasets. Currently, the cetacean database holds 323 surveys, covering a total of 255,062 kilometers of survey effort. It encompasses data for 31 unique species of cetaceans and spans the years from 2005 to 2022, with a total of 16 469 observations.

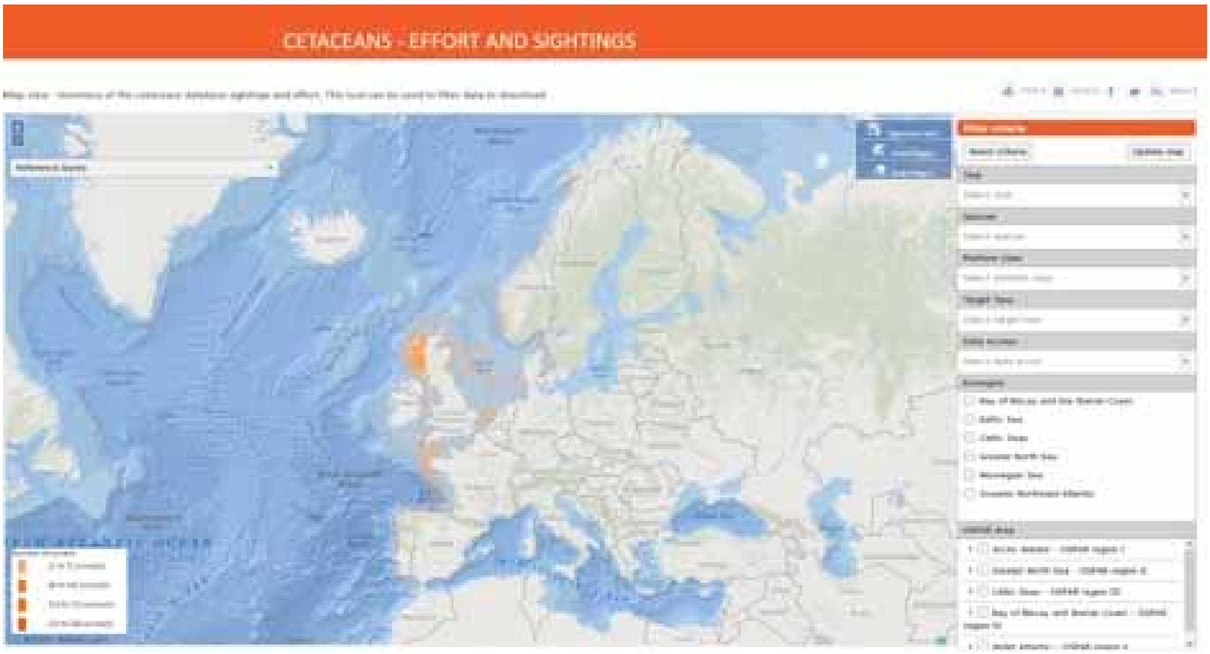


Figure 1 JCDP Data portal.

Governance

A dedicated working Group on the Joint Cetacean Data Programme (WGJCDP) is responsible for the JCDP database– and works in line with the JCDP vision. WGJCDP drives continued input of standardized data into the database and develops analyses and data products in line with identified priorities across the cetacean research and policy community. WGJCDP will work in collaboration with the Working Group on Bycatch of Protected Species (WGBYC) and the Working Group on Marine Mammal Ecology (WGMME) to provide robust and relevant outputs in support of ongoing priorities and needs such as e.g., abundance and distribution trends and areas of persistent densities.

Conclusion

The Joint Cetacean Data Programme (JCDP) is a vital initiative that aims to centralize and standardise cetacean data from the north-east Atlantic, with the potential to be expanded into other regions where there is a need. By providing a unified platform for data collection and access, JCDP enhances our understanding of cetaceans, which are challenging to observe, offering valuable insights into their populations and behaviour. Collaboration with other working groups strengthens the utility of the data for research and policy development. This consolidated effort can bring valuable insights into cetaceans' abundance, distribution, and trends, benefiting both science and conservation efforts.

Updated hi-resolution Northern North Pacific Regional Climatology

Alexey Mishonov[1,2] (alexey.mishonov@noaa.gov), Dan Seidov (ret.)[2] (dan.seidov@gmail.com), Olga Baranova[2] (olga.baranova@noaa.gov), Tim Boyer[2] (tim.boyer@noaa.gov), Scott Cross[2] (scott.cross@noaa.gov), Courtney Bouchard[2] (courtney.bouchard@noaa.gov), Ebenezer Nyadjro[2,3] (ebenezer.nyadjro@noaa.gov), Jennifer Webster[2] (jennifer.webster@noaa.gov)

[1]Univ. of Maryland/ESSIC (USA)
[2]NCEI/NOAA (USA)
[3]Mississippi St. Univ. (USA)

NCEI has recently updated hi-resolution Northern North Pacific (NNP) regional climatology to version 2, refreshing its growing portfolio. This update replaced previous NNP regional climatology version 1, which was released in 2016 and was based on World Ocean Database/World Ocean Atlas 2013 edition. Updated version of the NNP regional climatology is based on recently released World Ocean Atlas 2023 edition. NNP regional climatology, version 2 have been enhanced with new data collected during recent years as well as historic data recovered from earlier unaccounted observations added to the 2023 release of the World Ocean Database (WOD23).
The Northern North Pacific is a resource rich region, which includes Gulf of Alaska and Bering Sea, with abundant fisheries and other material assets that can be significantly impacted by ocean climate variability. Ocean processes in the Northern North Pacific play an important role in long-term ocean climate change affecting all aspects of marine ecosystems.

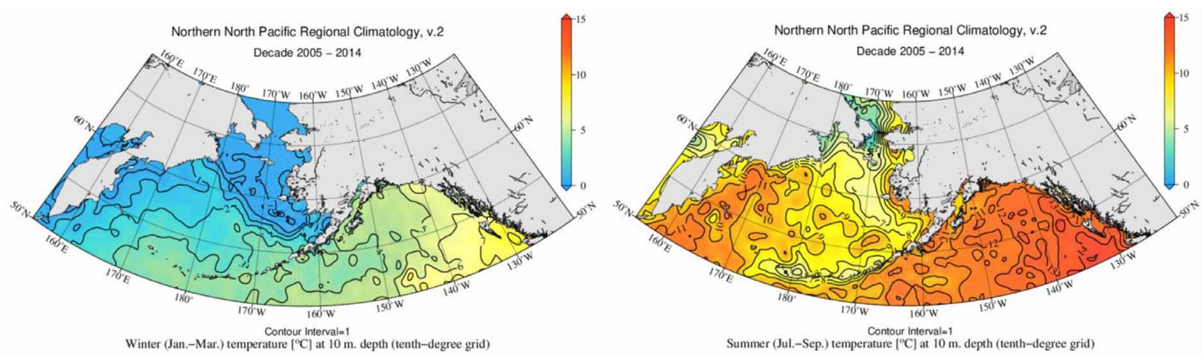


Figure 1 NNP RC: Winter (left) and Summer (right) Temperature for 2005 - 2014 decade on 1/10-degree grid at 10m depth.

The NNP regional climatology, version 2 is a collection of high-resolution quality-controlled temperature and salinity fields retrieved from WOD23 on standard depth levels from the sea surface to 5,200 m depth covering the period from 1955 to 2022. Annual and seasonal fields for the entire period as well as for each of the seven decades (1955 - 1964, 1965 - 1974, 1975 - 1984, 1985 - 1994, 1995 - 2004, 2005 - 2014, and 2015 - 2022) are compiled. In addition to the usual WOA23 fields computed on one- and quarter-degree horizontal grids, hi-resolution NNP regional climatology contains all fields computed on the tenth-degree horisontal grid as well. As examples, 2005 - 2014 decadal fields at the near surface layer for winter and summer seasons complied on one-tenth-degree longitude-latitude grid is shown on Figure 1 for seawater temperature and on Figure 2 for salinity. Full suite of the statistical fields for all grids is present in the NNP regional climatology too, following the WOA standards.

The NNP regional climatology, version 2 is instrumental for assessing ocean climate changes in this region of the North Pacific Ocean over the extended time period of seven decades and can be used in various climate studies, numerical ocean and climate modeling, environmental research projects, and other related applications. Updated hi-resolution NNP regional climatology, version 2 is an enhanced tool to study the long-term oceanic trends in this dynamic area.

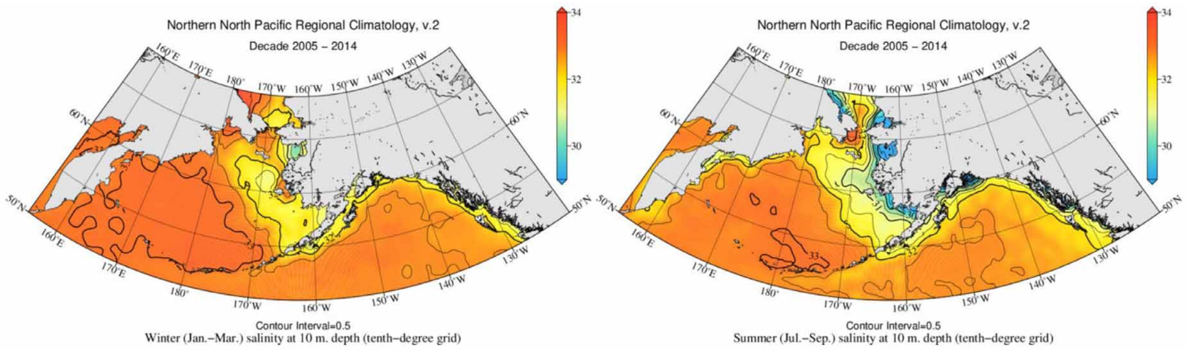


Figure 2 NNP RC: Winter (left) and Summer (right) Salinity for 2005-2014 decade on 1/10-degree grid at 10m depth.

The complete set of all climatological maps, objectively analyzed fields and associated statistical fields at all standard depth levels, presented in ASCII comma separated value (CSV), ArcGIS compatible, and netCDF formats as well as documentation are available online at the NNP regional climatology, version 2 web site https://www.ncei.noaa.gov/products/northern-north-pacific-regional-climatology.

Food web models to support decision making in sustainable pelagic fisheries

Martha Stevens (martha.stevens@vliz.be), Steven Pint (steven.pint@vliz.be),
Rutendo Musimwa (rutendo.musimwa@vliz.be), Ward Standaert (ward.standaert@vliz.be),
Gert Everaert (gert.everaert@vliz.be)

VLIZ, Flanders Marine Institute (Belgium)

Introduction

Ecosystem-based management requires comprehensive, quantitative assessment methods of the benefits and burdens of activities in the blue economy. Such an ecosystem-based decision framework is yet to be developed for the Southern Bight of the North Sea. Ecological models can bridge this gap, by integrating available data into decision-making tools for stakeholders and policymakers. For sustainable fisheries management in particular, food web models can be used to explore the effects of prospective policy changes.

Food web models describe the predator-prey interactions between a set of functional groups [Cohen, 1978; Drossel et al., 2003; Pimm, 1982]. This can then be used to answer questions ranging from fisheries production[4], to pollution [Mackay, 1989] and climate change effects [Zhang et al., 2017]. By simulating potential future scenarios, the impact of altered anthropogenic activities or potential policy changes can be investigated. However, to provide accurate predictions, a food web model must be tailored to reflect the local ecosystem. Currently, the most specific models available for our study area, the Southern Bight of the North Sea, have a spatial extent ranging from Skagerrak to the English Channel [Stäbler et al., 2016; Püts et al., 2020]. This area can be divided based on bathymetry and differing oceanographic conditions into the deeper central North Sea and the shallow Southern Bight of the North Sea [ICES, 2022]. Although the distinct environmental characteristics of the Southern Bight are well-described, to date, no food web model has been tailored to address questions concerning the sustainable management of fisheries for this particular ecosystem.

Methodology

Using Ecopath with Ecosim, a software to create mass-balanced food web models, two snapshots of the Southern Bight ecosystem have been developed for 1991 and 2018 respectively. These models describe the dietary relationships between 32 functional groups ranging from harbor porpoise to fish, invertebrates, plankton and detritus (Figure 1). Biomass estimates for these groups were obtained from various International Council for the Exploration of the Sea (ICES) reports (specified in technical report [Pint et al., (2023]). For species where this data was not available, biomass was either (1) estimated according to the method of Sparholt [1990], (2) extracted from scientific literature, or (3) estimated based on other models with overlapping study areas [Stäbler et al., 2016; Mackinson et al., 2008]. The relationships between functional groups are based on species diets, which were adjusted from Stäbler et al. [2016] based on fish stomach records from the UK Centre for Environment, Fisheries and Aquaculture (CEFAS). Other model parameters such as productivity and consumption rate, were estimated using empirical formulas [Möckeln et al., 1976; Pauly, 1980]. The impacts of fisheries fleets on ecosystem dynamics have also been incorporated in both models. Our approach also includes recreational fisheries, which have often been overlooked in the development of food web models even though these activities are highly impactful for certain species. For example, 27% of the total removals of European sea bass can be attributed to recreational fisheries [Hyder et al., 2018]. Hence, data on recreational fisheries activities in Belgium from Verleye et al. [2023]

were extrapolated to the Southern Bight of the North Sea. Whereas commercial fisheries landings and discard data were obtained from the Scientific, Technical and Economic Committee for Fisheries (STECF) and ICES reports. Finally, to obtain a model in equilibrium, i.e., an ecosystem where no more than the available biomass is being consumed, both models have been mass-balanced using pre-balance diagnostics as suggested by Link et al. [2010] and Heymans et al. [2016].

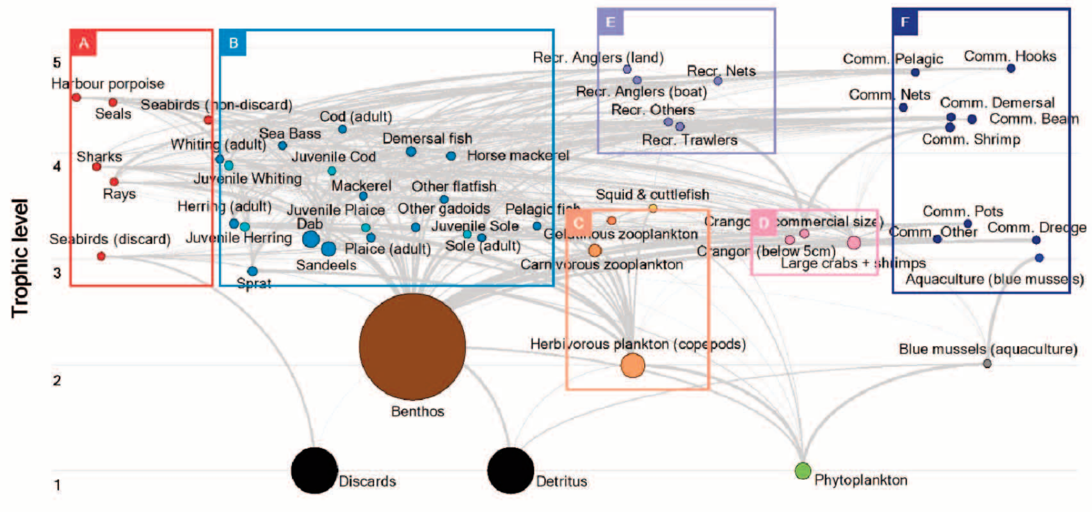


Figure 1 The 2018 Southern Bight of the North Sea food web model. Circles represent functional groups with their size indicating biomass. Rectangles highlight related functional groups or fisheries fleets: A. predators, B. commercial fish species, C. plankton, D. crabs and shrimp, E. Recreational fisheries and F. commercial fisheries.

Conclusion

A comparison of the 1991 and 2018 Ecopath models for the Southern Bight of the North Sea developed in this study shows that both the biomass and trophic level of catch have decreased, suggesting that the exploitation between these years has not been sustainable [Pauly et al., 1998]. Next, these models can serve as a base to run time-dynamic simulations in EcoSim, which allows the exploration of prospective policy impacts on food web dynamics. Insights obtained from these models can provide guidance for ecosystem-based fisheries management in this economically and ecologically important marine region.

References

Cohen (1978). Princeton Uni Press.
Drossel et al., (2003). In Handbook of Graphs and Networks 218–247, John Wiley & Sons Ltd.
Pimm, (1982). In Food Webs 1–11, Springer NL.
Gaichas et al., (2010). *Can J Fish Aquat, Sci* 67, 1490–1506.
Mackay, (1989). *Great Lakes Res*, 15, 283–297.
Zhang et al., (2017). *Proc R Soc B: Biological Sciences*, 284.
Stäbler et al., (2016). *Ecol Modell*, 331, 17–30.
Püts et al., (2020). *Ecol Modell* 431.
ICES, (2022).
Pint et al., (2023).
Sparholt, (1990). *Cons. int. Explor. Mer* 46.
Mackinson et al., (2008). *Sci Ser Tech Report*.
Möckeln et al., (1976). *Ornis Scandinavica* 7.
Pauly, (1980). *Cons. int. Explor. Mer* 39, 175–192.

Hyder et al., (2018). *Fish and Fisheries* 19, 225–243.
Verleye et al., (2022).
Link, (2010). *Ecol Modell* 221, 1580–1591.
Heymans et al., (2016). *Ecol Modell* 331, 173–184.
Pauly et al., (1998). *Science* (1979) 279.

The challenge to map the global seafloor – The Nippon Foundation-GEBCO Seabed 2030 Project

Pauline Weatherall (paw@noc.ac.uk), Helen Snaith (h.snaith@noc.ac.uk), Christopher Thompson (chromp@noc.ac.uk), Mark Hebden (mahe@noc.ac.uk)

British Oceanographic Data Centre -BODC (UK)
National Oceanography Centre - NOC (UK)

Bathymetry (the depth and shape of the seafloor) is a fundamental parameter in the study of the oceans and critical to help our understanding of ocean circulation patterns - affecting climate and weather patterns, tides, wave action, sediment transport, tsunami wave propagation and underwater geo-hazards and much more. However, it is estimated that around 75% of the global seafloor remains unmapped using modern echo-sounding techniques.

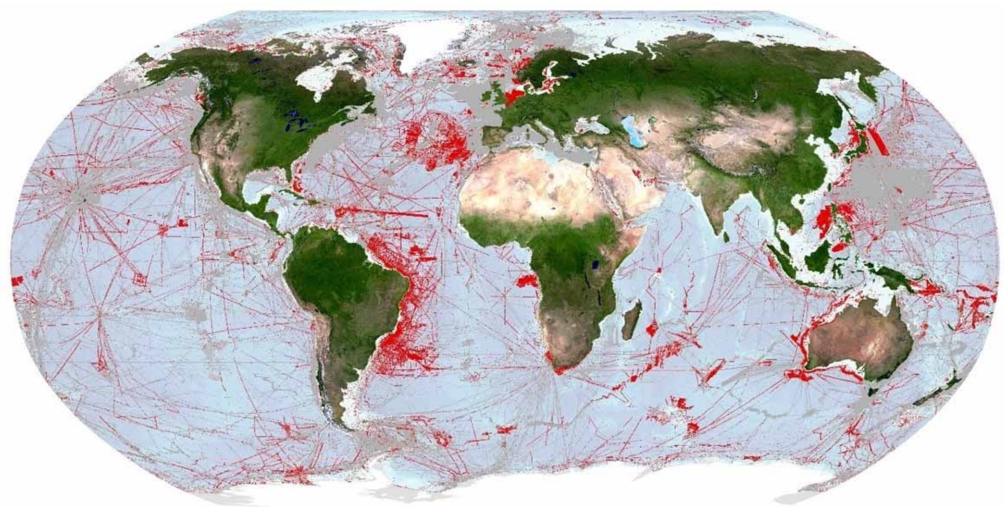


Figure 1 The grey and red regions show areas of the seafloor that are considered mapped within the GEBCO_2023 Grid. Grey regions are areas mapped in 2020 and the red regions are the additional coverage included in 2023.

Operating under the joint auspices of the IHO and IOC of UNESCO, The General Bathymetric Chart of the Oceans (GEBCO) has been developing and making available bathymetric data sets and charts since 1903. In 2017, GEBCO partnered with The Nippon Foundation of Japan to setup The Nippon Foundation-GEBCO Seabed 2030 Project with the focussed goal of producing the definitive, high resolution bathymetric map of the entire World Ocean by 2030.
Seabed 2030 has been officially recognised as a programme of the United Nations Decade of Ocean Science for Sustainable Development – in support of UN Sustainable Development Goal 14 (SDG14) – a commitment to conserve and sustainably use the ocean, seas and marine resources for sustainable development.

Defining how much of the seafloor has been mapped
At the start of the Seabed 2030 project a set of depth-dependent 'resolution goals' were defined to help determine how much of the ocean floor has been mapped Mayer et al. [2018]. This is effectively an estimation of data density. With each release of the GEBCO grid, the Project's progress is computed. The percentage mapped has increased from 6.7% in 2018 to 24.9% in 2023.

There is still much to do to achieve Seabed 2030's goal of 100% of the seafloor mapped. An international effort is needed to identify and bring together existing bathymetric data held within national and international data collections; the hydrographic and scientific communities; industry and individual data collectors through crowdsourced bathymetry initiatives.
To help with its work, Seabed 2030 has setup four Regional Centers (Arctic and North Pacific, Atlantic and Indian Ocean; South and West Pacific and Southern Ocean) to champion mapping activities; assemble and compile bathymetric data and collaborate with existing mapping initiatives in their regions.
The British Oceanographic Data Centre (BODC) of the National Oceanography Centre (NOC) hosts Seabed 2030's Global Center (GDACC). It works to produce, maintain and deliver the GEBCO global bathymetric grid. The IHO Data Center for Digital Bathymetry (IHO DCDB) acts as a data repository for the project.

Outputs from our work

A new global GEBCO bathymetric grid is released each year. The main data set is currently a global terrain model for ocean and land, providing elevation data, in meters, on a 15 arc-second interval grid.
The four Seabed 2030 Regional Centers work to compile contributed bathymetric data sets into grids for their specific geographic areas of responsibility. These regional grids are then supplied to the Global Center, for use in compiling the global data set.
For areas outside the polar regions (primarily south of 60°N and north of 50°S), the regional grids are supplied in the form of 'sparse grids', i.e., only grid cells that contain data values are populated. For the polar regions, complete grids are provided due to the complexities of incorporating data held in polar coordinates.
To generate the global GEBCO grid, the sparse regional grids are included on a base grid, SRTM15_plus, which is provided by one of Seabed 2030's partners: Scripps Institution of Oceanography (SIO). This base grid is largely generated through a combination of predicted bathymetry (from satellite altimetry) and bathymetric data observations.
The global GEBCO grid is accompanied by a Type Identifier (TID) Grid. This indicates the type of source data that the corresponding cell in the bathymetric grid is based in, i.e., measured data or predicted/interpolated data.

Accessing the GEBCO grid

The GEBCO grid is made available to download via GEBCO's web site (https://www.gebco.net) as:

- global files, in geographic co-ordinates: as either a single netCDF format file or as a set of 8 tiles, in Esri ASCII raster or data GeoTiff formats.
- bathymetric grids for user-defined geographic areas.
- bathymetric grids for polar regions in polar projection co-ordinates.
- imagery based on GEBCO's grids for user-defined areas.
- Web Map Services (WMS).

Reference

Mayer L., Jakobsson M., Allen G., Dorschel B., Falconer R., Ferrini V., Lamarche G., Snaith H., Weatherall P., (2018). *The Nippon Foundation—GEBCO Seabed 2030 Project: The Quest to See the World's Oceans Completely Mapped by 2030*. Geosciences 2018, 8, 63. https://doi.org/10.3390/geosciences8020063

Digital Animal Sound Archive: a collaborative repository for bio-acoustics

Thomas Vandenberghe[1] (tvandenberghe@naturalsciences.be), Robin Brabant[1] (rbrabant@naturalsciences.be), Yves Laurent[1] (ylaurent@naturalsciences.be), Claire Brabant[2] (claire.brabant@natagora.be), Bob Vandendriessche[3] (bob.vandendriessche@rlhp.be), Wout Willems[3] (wout.willems@natuurpunt.be), Steven Degraer[1](sdegraer@naturalsciences.be), Ruth Lagring[1] (rlagring@naturalsciences.be)

[1]Royal Belgian Institute for Natural Sciences - RBINS (Belgium)
[2]Natagora (Belgium)
[3]Natuurpunt (Belgium)

A wide variety of animals produce acoustic signals or calls, that are in many cases species-specific. The use of these animal sounds in biological and ecological studies is widespread as they can be used to study species distribution, phenology, ecology and behaviour of organisms that are often visually elusive (e.g., marine mammals, bats). This results in extensive individual collections (tens of terabytes range) that are scattered in many different locations (e.g., scientific institutes, universities, environmental consultants, citizen scientists). A critical aspect of being able to learn from such large and varied acoustic datasets is providing consistent and transparent access that can enable the integration of various analysis efforts. Considering the data sizes, processes are hard to scale up. The overall objective of the Digital Animal Sound Archive (DASA) is to set up a robust data model, and a user-friendly web interface enabling Belgian bio-acoustic workers to collect, archive and explore biological acoustic data and accompanying metadata. The main partners in the project are RBINS and Natagora and Natuurpunt, two nature conservation and citizen science NGOs. Similar projects are ongoing abroad, and reaching out to these initiatives to share experience will be an integrated part of the DASA project. Therefore, specialists from the Muséum national d'Histoire naturelle (MNHN) in Paris and the British Trust for Ornithology (BTO) are part of the Follow-up committee.
The added value of this digital collection is manyfold: (1) to serve as a digital archive, (2) to add to the collections hosted by the Royal Belgian Institute of Natural Sciences (RBINS), (3) to serve as a reference collection of species and their behavioural acoustic calls, (4) to offer a validated dataset for the development of automated identification software tools and (5) to serve as a dataset for new ecological studies on dispersal of species and habitat preference.
As a proof of concept regarding the applicability of the data model and web interface the developments will first be focused on the recordings of bat sounds. The first release will be open to bat calls collected both on land and at sea. RBINS curates large offshore windmill nacelle bat monitoring datasets, which this project will finally expose.

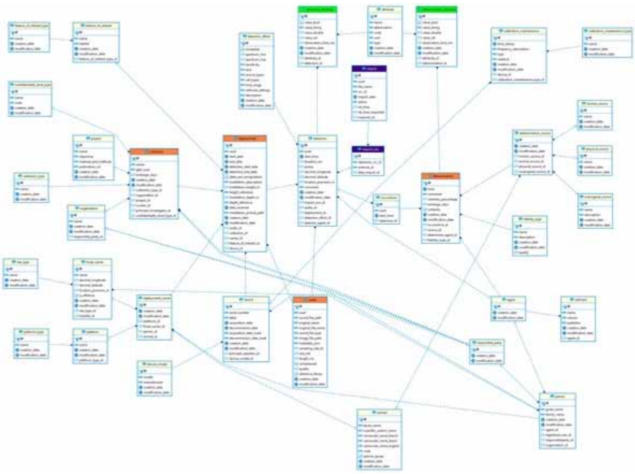
Scope
Bio-acoustic data systems can be categorized by the way the origin of the sound is determined. These aspects can be combined with each other:

1. Sound collection aspect: determination by a submitter, before submission into an archive.
2. Citizen science aspect: determination by a website user (validator).
3. Incremental: determination in addition to that of the submitter.
4. Primary: first determination because the submitter didn't do it.
5. Reference sound catalogue: validated sound files are tagged as high-quality and offered as a separate category. These files can then be used in a sound library or as classifier training material.

6. Real-time classifier aspect: web service that, upon receipt of the sound file, makes a determination using a classifier (storage is optional). The determination may or may not be returned to the submitter.
7. Delayed classifier aspect: a (primary or incremental) determination by a classifier that is executed at a later time (asynchronously). This is done in bulk on a complete archive.

The DASA project will implement aspects 1, 2 and 3 and build a foundation to include the use of automatic classifiers. Multi-species group classifiers augment the value of audio recordings since they valorise 'by-catch sounds' such as crickets and birds. They are of greater value than most bat classifier software which tend to label all non-bat sounds as "noise", whatever the origin. As a first step towards an analysis pipeline, the Tadarida-D and C software packages, together with a basic bat sonotype classifier [Roemer et al., 2021], have been dockerized, so that we can better distinguish between "noise" of biological, mechanical or acoustic origin. This leads to better storage space use, as "true noise" can be filtered beforehand. At this moment RBINS has reserved 50TB of local storage for the project, including backup. Long-term storage is foreseen on tape.

Figure 1 Data model.



Data model

The data model has been developed with fully normalised entities, to capture metadata to the fullest extent, for those who can provide it. After all, not all data providers have the same incentive for diligent metadata annotation. Many fields are optional to cater to volunteers and take the considerable workload needed for metadata annotation into account, especially for historical data and noise origin differentiation. The model is open to any species group and can describe long-term marine mammal surveys and sounds of physical or human origin as well (e.g., wind turbine piledriving at sea).

To develop the data model, we incorporated ideas from citizen science systems such as observations.org, Xeno-Canto.org and reference sound catalogs such as La Sonothèque and ChiroVox. Many aspects of the Tethys bio-acoustic data model have been taken over, such as the Deployment and DetectionEffort concepts. Tethys has been developed to manage the metadata from marine mammal detection and localization studies. The GBIF DwC Occurrence concept and the composition trough EventCore (deployment + detection) are easily retrievable from the data model.

Data acquisition

The project operates in three phases: 1) initial data acquisition from the partners in 2023, 2) web application development in 2024, and 3) a post-project operational phase where data is acquired via the web application (2025). In order to allow data transfer over http, the Tus.io protocol is strongly considered. To capture the metadata, a metadata Excel form has been created. The metadata form is more lenient than the data model and will be partially replaced by the web application. Finally, occurrence data will be disseminated to GBIF. Data owners can select a CC license; sensitive observations (commercial interests or rare species) can be embargoed at dataset or observation level.

Reference

Roemer C., Julien J.-F., & Bas Y., (2021). *An automatic classifier of bat sonotypes around the world*. Methods in Ecology and Evolution, 12, 2432 – 2444. https://doi.org/10.1111/2041-210X.13721

Towards an accuracy assessment service for surface ocean currents from global ocean re-analysis and forecast products

Björn Backeberg (bjorn.backeberg@deltares.nl), Sonia Heye (sonia.heye@deltares.nl), Martin Verlaan (martin.verlaan@deltares.nl)

Deltares (The Netherlands)

Global ocean reanalysis and forecast products are routinely used in operational oceanography and downstream applications. For example, they can provide initial conditions and boundary conditions for high-resolution coastal models that simulate the behavior of the ocean in nearshore regions, where the interaction between the open ocean and the coastline is complex. Additionally, they provide initial conditions and ocean current data needed for particle trajectory modeling to simulate the movement of particles, such as drifters, plankton, pollutants, or oil spills, in the ocean.

The accuracy of ocean reanalysis and forecasting systems depends on several factors, including the performance of the underlying Ocean General Circulation Model, the data assimilation system and the quality and coverage of the assimilated observations. And although global ocean reanalysis and forecast products have global coverage, there can be regional differences in their accuracy. Therefore in the context of using these data for initial conditions and boundary conditions for high-resolution coastal models or in particle trajectory modelling applications it is important to quantify, geograhically, the accuracy of the different global ocean reanalysis and forecast products.

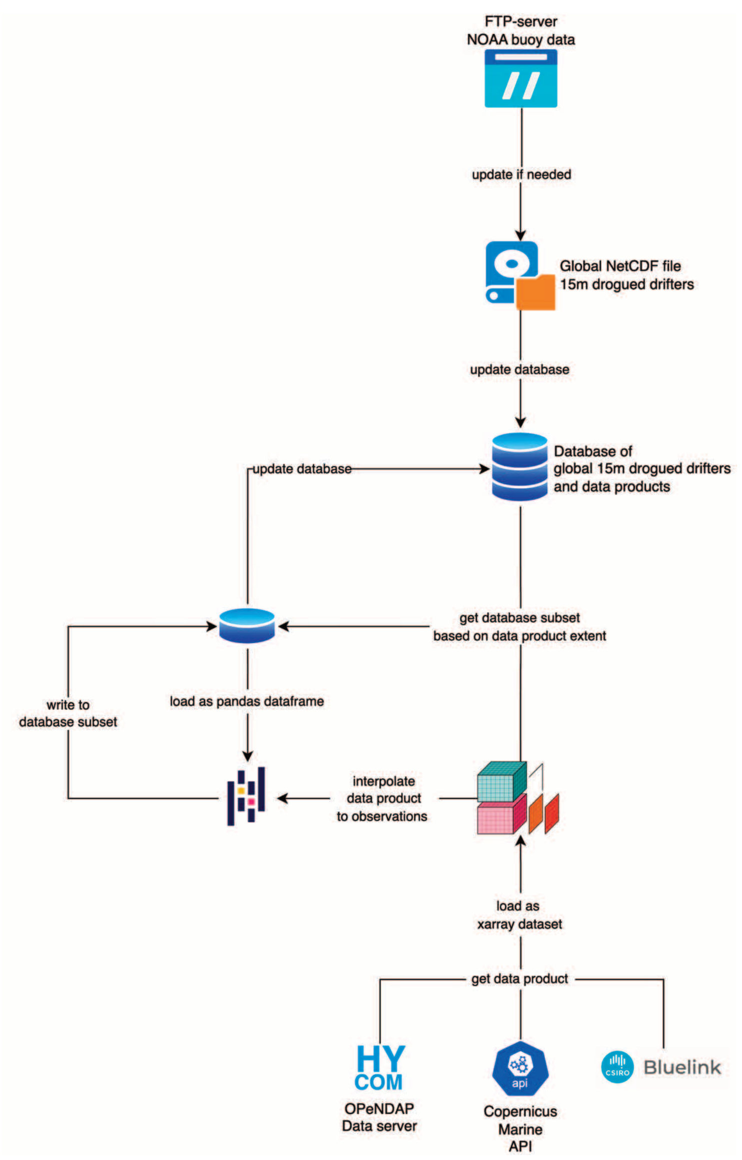
Additionally, since ocean reanalysis and forecasting systems typically assimilate along track altimeter sea level anomalies, satellite sea surface temperatures, sea ice concentrations, and *in situ* temperature and salinity vertical profiles, it is important to use independent (unassimilated) observations to validate the accuracy of these systems. Hence, as a starting point, here we focus on using ocean surface current and sea surface temperature measurements from surface drifters provided by the global drifter programme to evaluate the accuracy of global ocean re-analysis and forecasting systems.

The overarching objective of this work is to develop an interactive web application that offers an accuracy assessment service for, amongst others, coastal modellers and particle trajectory modellers to make informed decision about which global ocean reanalysis and forecast product to use when running simulations in different parts of the global ocean.

In developing an interactive web application that offers a fast, easy-to-use service for end-users, several data and processing challenges need to be addressed. These include investigating strategies for extracting ocean current velocities and sea surface temperatures from multiple ocean reanalysis and forecast products at the corresponding drifter observation location for more than 20 million observation points globally as well as testing technologies to calculate and visualise validation metrics on the fly for user-defined areas and periods of interest.

Here we present the preliminary results of our work where we explored Python processing libraries such as Xarray, Pandas, Dask and Zarr for handling and analyzing large datasets efficiently and effectively. Xarray supports the management of multi-dimensional arrays and labelled data structures facilitating seemless organisation, manipulation and analysis of vast volumes of ocean data, whilst Pandas enables data transformations and computations, making it easier to extract valuable insights from the data. Dask, a parallel computing library, plays an important role in scaling and distributing the processing pipeline across multiple CPU cores or even clusters, significantly speeding up the data processing workflows, while Zarr, a library designed for efficient storage and retrieval of chunked, compressed data, is integral for optimising the storage and retrieval of large amounts of data, ensuring the the web application is both resource-efficient and responsive. By combining the capabilities of these libraries, we are working towards a toolkit and processing pipeline for an interactive web application that efficiently extracts ocean current velocities and sea surface temperatures from multiple reanalysis and forecast products for more than 20 million drifter observation points across the globe.

The initial processing pipelines developed here have the potential to be expanded upon by extending to other data sources for validating global ocean reanalysis and forecast products. For example other observing platforms such as ocean gliders, moorings and (unassimilated) satellite remote sensing could be included to enable end-user making well-informed decisions when selecting global ocean reanalysis and forecast products for their simulations in diverse regions worldwide.



A case study: Transitioning the Bio-Oracle dataset version 3 to ERDDAP

Salvador Jesús Fernández Bejarano[1] (salvador.fernandez@vliz.be), Vinícius Salazar[2] (viniws@gmail.com), Jorge Assis[3] (jorgemfa@gmail.com), Bart Vanhoorne[1] (bart.vanhoorne@vliz.be), Frederic Leclercq[1] (frederic.leclercq@vliz.be), Lennert Tyberghein[1] (lennert.tyberghein@vliz.be), Olivier De Clerck[4] (olivier.declerck@ugent.be), Ester Serrão[3] (eserrao@ualg.pt), Heroen Verbruggen[2] (heroen.verbruggen@gmail.com), Lennert Schepers[1] (lennert.schepers@vliz.be)

[1]VLIZ, Flanders Marine Institute (Belgium)
[2]University of Melbourne (Australia)
[3]University of Algarve (Portugal)
[4]Ghent University (Belgium)

Understanding habitats changes is key in the context of climate change and global warming. Restoration strategies are optimized based on the identification of biodiversity hotspots and potential species niche shifts. However, the techniques that allow to gather this policy-relevant information require of harmonised, interoperable, high-quality environmental datasets, which are not always readily available in the marine realm [Tyberghein et al., 2012; Assis et al., 2017]. Such datasets are essential for species distribution modeling and detecting habitat changes. Initially released in 2012, with a major update in 2017 incorporating benthic data layers and future climate scenarios aligned with IPCC predictions, Bio-ORACLE now introduces Version 3. This version includes substantial updates not only in data but also in access methodologies, transitioning from a file-based system to an ERDDAP server. Our experience illustrates a case study of transitioning an existing database into ERDDAP and, from that, we offer a concise set of recommendations for teams considering to do the same.

Bio-Oracle version 3

The Bio-Oracle database (https://bio-oracle.org/) was traditionally disseminated as raster files, downloadable from the Bio-ORACLE website or retrieved using the R package sdmpredictors. Bio-Oracle encompasses 18 environmental variables such as ocean temperature, salinity or concentration of phytoplankton. Version 3 enhances the resolution of the dataset on multiple levels: spatial resolution increases to 0.05 degrees, and future time predictions expand to ten decades per variable, encompassing the years 2000 - 2010, 2010 - 2020, and so forth until 2090 - 2100 [Assis et al., under review]. Future climate change scenarios are updated to include the latest Shared Socioeconomic Pathways (SSP) from the IPCC Sixth Assessment Report.
To address the challenge of heavy data files resulting from increased resolution, Bio-Oracle Version 3 adopts NetCDF grids. This binary file format allows subsetting smaller portions of the dataset, essential for efficient data handling and analysis. Each NetCDF accommodates environmental variables with dimensions of longitude, latitude, and time, aligning with the Climate Forecast conventions. This also facilitates integration into an ERDDAP instance, freeing the client from dealing with large data files by outsourcing the slicing and other capabilities to the server side. ERDDAP offers a consistent way to download subsets of scientific datasets, make graphs, and maps, with a focus on simplifying scientific data retrieval.

Publishing via ERDDAP

A dedicated ERDDAP server was installed at the Flanders Marine Institute Data Centre to host Bio-Oracle Version 3 (https://erddap.bio-oracle.org). Our ERDDAP instance is encapsulated within a Docker container, ensuring a self-contained and reproducible deployment environment. Its configuration is available at <https://github.com/bio-oracle/bio-oracle-erddap>. We built upon the IOOS 'gold standard' ERDDAP configuration (https://github.com/ioos/erddap-gold-standard). Custom scripts for automated data processing are integrated in our ERDDAP server configuration, facilitating seamless updates and maintenance. Automation reduces manual intervention, minimizes errors, and ensures the dataset's timely availability for data consumers.

Accessing Bio-Oracle v3

ERDDAP provides both a web interface and a REST API, allowing server-side operations on datasets. Users can retrieve specific slices of data, optimizing bandwidth use. ERDDAP supports up to 20 file types for downloading data, including NetCDF, CSV, GeoTiff, and PNG plots. This flexibility is advantageous for users interested in specific regions, realms, or climate change scenarios and decades. ERDDAP API's consistency across datasets simplifies integration with other data sources. For instance, the Irish Marine Institute meticulously curates a registry of ERDDAP projects and deployments at <https://github.com/IrishMarineInstitute/awesome-erddap>. This registry, systematically accessed by versatile clients like rerddap in R, facilitates comprehensive data searches across diverse ERDDAP deployments [Chamberlain S., 2023]. The rerddap R package offers functionalities for listing and downloading data within the R environment, while Erddapy serves as its Python counterpart [Fernandes F., 2017]. In addition to these general-purpose clients, we have developed dedicated R and Python extensions, namely 'pyo_oracle' and 'biooracler,' tailored for seamless exploration and retrieval of Bio-Oracle data. These dedicated packages streamline the assimilation of Bio-ORACLE data into prevailing bioclimatic modeling workflows. They are available at <https://github.com/bio-oracle/pyo_oracle> and <https://github.com/bio-oracle/biooracler> respectively. The development of these API clients underscores our commitment to accommodating a broad spectrum of scientific frameworks and programming languages, nurturing collaborative and interdisciplinary research endeavors. Furthermore, the dataset is poised for dissemination across novel data platforms and initiatives, including the European Digital Twins of the Ocean, the Microsoft Planetary Computer, and Google Earth Engine. Leveraging a renowned standard server like ERDDAP will be instrumental in facilitating data sharing with these initiatives.

References

Assis J. et al., (2017). *Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling*. Global Ecology and Biogeography, 27, 277–284.

Assis J. et al., (under review). *Bio-ORACLE v3.0. Pushing marine data layers to the next-generation scenarios of climate change research*. Global Ecology and Biogeography.

Chamberlain S., (2023). *rerddap: General Purpose Client for 'ERDDAP' Servers*. https://docs.ropensci.org/rerddap/, https://github.com/ropensci/rerddap

Fernandes F., (2017). *Python interface for ERDDAP*. https://ioos.github.io/erddapy/, https://github.com/ioos/erddapy

Tyberghein L. et al., (2012). *Bio-ORACLE: A global environmental dataset for marine species distribution modelling*. Global Ecology and Biogeography, 21, 272–281.

Using Extrema estimates for Quality Control: rationale and recent developments for the MinMax approach

Jérôme Gourrion[1] (jerome.gourrion@pokapok.org), Delphine Leroy[1] (delphine.leroy@pokapok.org), Christine Coatanoan[2] (christine.coatanoan@ifremer.fr)

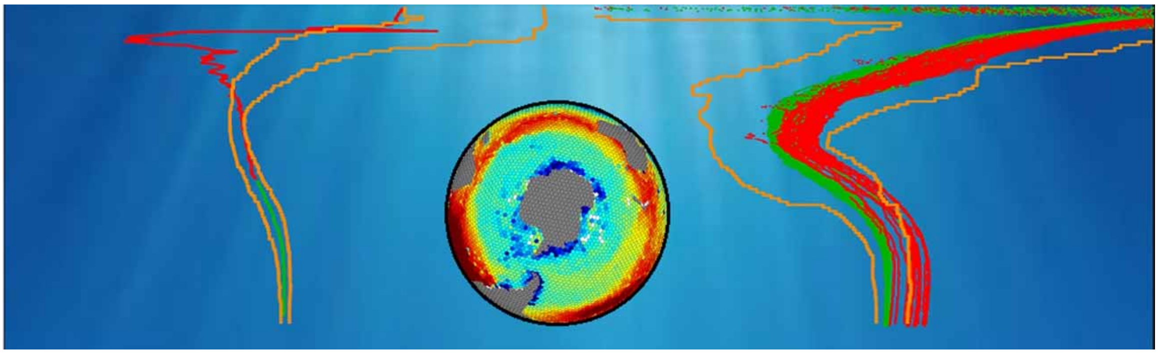
[1]POKaPOK (France)
[2]Ifremer (France)

Among various categories of Quality Control (QC) procedure types, the *local range* one does check whether the observed value lays inside a validity interval built upon the historical knowledge of the local variability: "have we ever seen such a high/low value in the geographical neighborhood of the present observation?".

If the idea is trivial, all the complexity (and possible latter success) is contained in the validity intervals estimation and its underlying statistical assumptions; the stronger the assumptions, the lower the amount of work to estimate the validity bounds; the weaker the assumptions, the better the robustness of the procedure. Of course, different trade-off leads different result, performance and robustness.

In previous approaches, the validity bounds are estimated either from first and second order moments of the local statistical distribution, or from quantiles. In the MinMax approach, the validity bounds are directly inferred from minimum and maximum estimates (Gourrion et al., 2020). Further, the bounds are modified 1) as an attempt to account for the non-sampled variability and improve the method robustness and 2) in order to meet potential operational constraints in terms of risk of erroneous classification (in a fully automated implementation) or daily amount of alerts to be visualized and confirmed by an operator (in a semi-automated mode).

In this communication, the recent developments to improve the applicability domain of the method (extension downto 5500 db), its robustness (bounds modification) and performance (application to the T/S space, parameter anomaly rather than full parameter values) will be presented together with method evaluation in regards of the previous approaches.



Reference

Gourrion J., Szekely T., Killick R., Owens B., Reverdin G., & Chapron B., (2020). *Improved Statistical Method for Quality Control of Hydrographic Observations*. Journal of Atmospheric and Oceanic Technology, 37(5), 789-806. https://doi.org/10.1175/JTECH-D-18-0244.1

Poster session

Geographical Information System Oceanographic for Marine Strategies Development

Luis Miguel Agudo Bravo (luismi.agudo@ieo.csic.es), Maria Olvido Tello-Antón (olvido.tello@ieo.csic.es), Paula Gil Cuenca (dulce.mata@ieo.csic.es), Gerardo Bruque (gerardo.bruque@ieo.csic.es), Dulce Mata Chacon (dulce.mata@ieo.csic.es)

Spanish Institute of Oceanography (Spain)

The Spanish Institute of Oceanography-Council of Spanish National Research Council (IEO-CSIC)

The IEO-CSIC, is an organisation dedicated to the study of the marine environment, especially in relation to scientific knowledge of the oceans and the sustainability of marine resources and the marine environment, it has been collecting information since its foundation in 1914. As a result of this systematic and massive collection of data, the IEO has become the organisation that holds the largest volume of historical data on the marine environment in Spain. The the Oceanographic Data Center (CEDO) and the SIG MARINO of IEO are the responsibles for managing the spatial data acquired on board oceanographic vessels and for making these data as interoperable as posible, enabling its use and reuse for the knowledge, dissemination and conservation of the marine biodiversity.

The Marine Strategy Directive

The Marine Strategy Framework Directive (MSFD) (Directive 2008/56/EC) is the European Union initiative that seeks to achieve Good Environmental Status (GES) of marine ecosystems through the regulation of human activities. The Directive has been transposed in Spain through Law 41/2010. The main objective of the MSFD is to ensure that human activities do not harm the marine environment and to develop marine strategies to protect and preserve the seas and their ecosystems. The IEO-CSIC plays an important role in research and access to data on fisheries resources, marine ecosystems and aquaculture.

The Monitoring Programmes MSFD

The monitoring programmes for the Marine Strategy Framework Directive consist of sequential oceanographic surveys carried out sequentally with a established periodicity. These programmes aim to gather data for the 11 descriptors defined within the MSFD. The GIS team of MSFD is responsible for managing the spatial data related to each descriptor and ensuring that the data can be used interchangeably for research and decisión-making advice. Marine data acquired, which includes information about 11 descriptors: biodiversity data, invasive non-native species, commercially exploited species, eutrophication, alteration hydrological conditions, pollutants and their environmental effects, marine litter, underwater noise, integrity of the deep sea, trophic networks or food chain. The ultimate goal is to disseminate this integrated data for further analysis and research.

The IEO GIS and Spatial Data Infrastructure (IDEO)

The IEO has developed the MARINE GIS System for spatial data management and analysis. The IEO Spatial Data Infrastructure (IDEO) facilitates access to IEO data, through Marine Data Viewer (http://www.ideo-base.ieo.es/Home) and provides access to metadata through its own data catalog (www.datos.ieo.es), thereby facilitating discovery of IEO's marine spatial data sets.

Workflow for providing access to information acquired in marine strategy monitoring programmes

The marine Gis team of MSFD is in charge of coordinating GIS work, as well as, advise all research groups involved in the treatment and flow of marine geographic information in the IEO. This team receives the data acquired in the monitoring programmes and, from that, creates georeferenced databases, guaranteeing the interoperability of this databases in compliance with the specifications of the INSPIRE Directive (2007/2/CE). The received data in different formats are transformed into spatial datasets and processed in different ways before being integrate into the GIS Geodatabases. In the proccess made, it is also very important the task of quality control of topological consistency of data. This involves detailes task management and adaptation to predefined data models following international standard and protocols stablished by GIS team and MITECO. It is also important to check field structures, its content and improving their representation and spatial processing. Once the datasets have been processed and validated, they are integrated into the spatial databases. All dataset have associated metadata development in the IEO Catalog (www.datos.ieo.es).

This team aims to disseminate information among the technical and scientific staff by facilitating the use of tools that allow the study of numerous and complex variables in the framework of Marine Strategies project. To achieve these objectives, WMS services have been developed for visualization and consultation the data of monitoring programs, and the WFS service for downloading it. The GIS team is in charge of updating and maintaining the web cartographic services both WMS and WFS. These services are used in the IEO data viewer.

This IEO data viewer also provides functions for visualizing the information generated within the framework of MSFD. With the integration of all information generated in the framework of monitoring programmes for Marine Strategies Development in the marine data viewer of IEO, the workflow is completed.

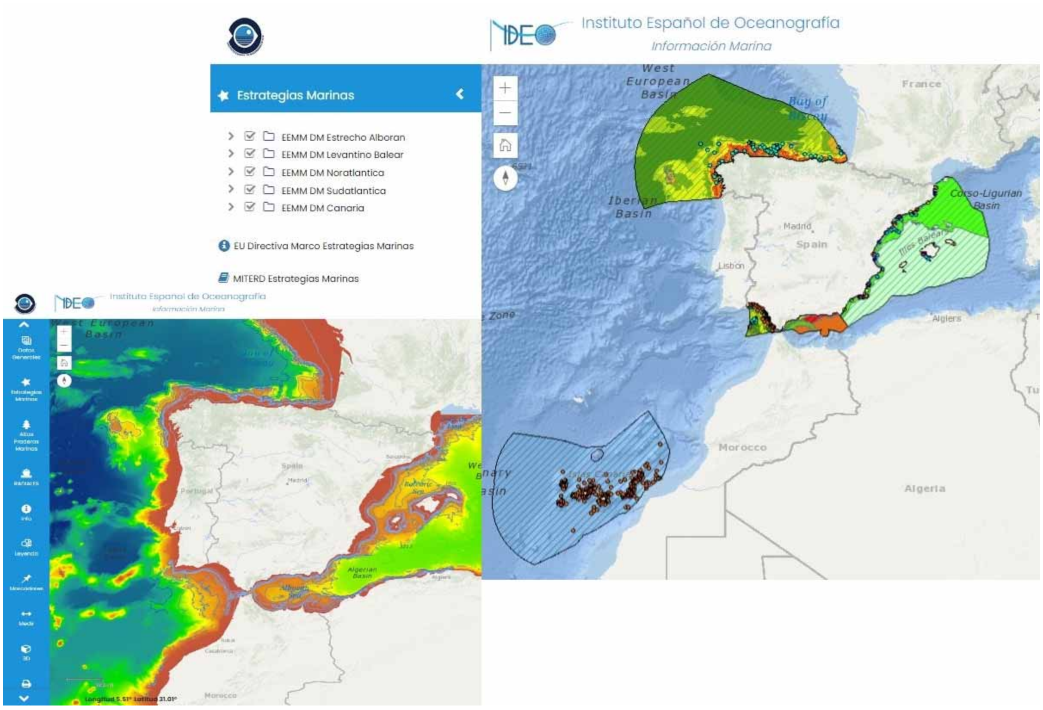


Figure 1 Marine Reference Information Viewer.

OSPARs regional data and information management to support delivery of its Quality Status Report 2023

Chris Moulton (chris.moulton@ospar.org), Eleanor Dening (eleanor.dening@ospar.org)

OSPAR Commission (UK)

The OSPAR Assessment Portal (OAP[25]) and the OSPAR Data and Information Management System (ODIMS[26]) are two key tools in the management and dissemination of OSPAR data and information, and the delivery of OSPARs regular work as well as its decadal assessment, the Quality Status Report (QSR). These tools played a critical role in the delivery of the recently published, OSPAR Quality Status Report 2023[27].

OSPAR's Quality Status Report 2023 is a comprehensive assessment of the environmental health and status of the North-East Atlantic Ocean and of human activities interacting with it. The QSR feeds directly into the United Nations Decade of Ocean Science for Sustainable Development as "Action No.513 - OSPAR Quality Status Report 2023"[28].

The QSR 2023 is made up of more than 120 assessments and covers various aspects, including, biodiversity, habitats, and human activities that impact the marine environment. It examines the presence of contaminants and pollutants, such as chemicals and microplastics, and assesses their potential effects on marine life and ecosystems. Additionally, the report evaluates the status of different marine species, from fish and seabirds to mammals and plants. It looks at population, distribution, and trends, identifying any changes or threats to these species. This information helps scientists and policymakers understand the overall health and resilience of the oceanic ecosystem. Another crucial aspect addressed in the QSR 2023 is the impact of climate change and ocean acidification on the marine environment. It examines the changes in ocean temperature, acidity, and sea-level rise, among other factors, and assesses their effects on marine life and ecosystems.

Overall, the QSR 2023 serves as an important tool for scientists, policymakers, and the public to understand the current state of the North-East Atlantic, the challenges it faces, and the measures needed to ensure its long-term health and sustainability.

Underpinning the QSR 2023 is a huge amount of data and associated data products which are the result of years of collective work from more than 400 scientists and experts, to define, collect, administer, prepare and analyse. The position of data and data products within the preparation of the QSR is presented in figure 1, below.

Data are hosted in ODIMS and made available in accordance with the OSPAR Rules of Procedure (OSPAR Agreement 2013-02[29]) which states:

"*OSPAR is committed to making as much information as possible publicly available, consistent with achieving other similarly important goals of public policy. The framework for this is set out in Article 9 of the OSPAR Convention.*"

Following this, OSPAR data made available via ODIMS, are licensed according to Creative Commons Zero[30] and OSPAR information presented via OAP are licensed according to Creative Commons BY-4.0[31]. Further information is available via the OSPAR Information[32] and Data[33] Policies.

[25] https://oap.ospar.org/

[26] https://odims.ospar.org/

[27] https://oap.ospar.org/en/ospar-assessments/quality-status-reports/qsr-2023/

[28] https://oceandecade.org/actions/ospar-quality-status-report-2023/

[29] https://www.ospar.org/documents?d=34012

[30] https://creativecommons.org/publicdomain/zero/1.0/

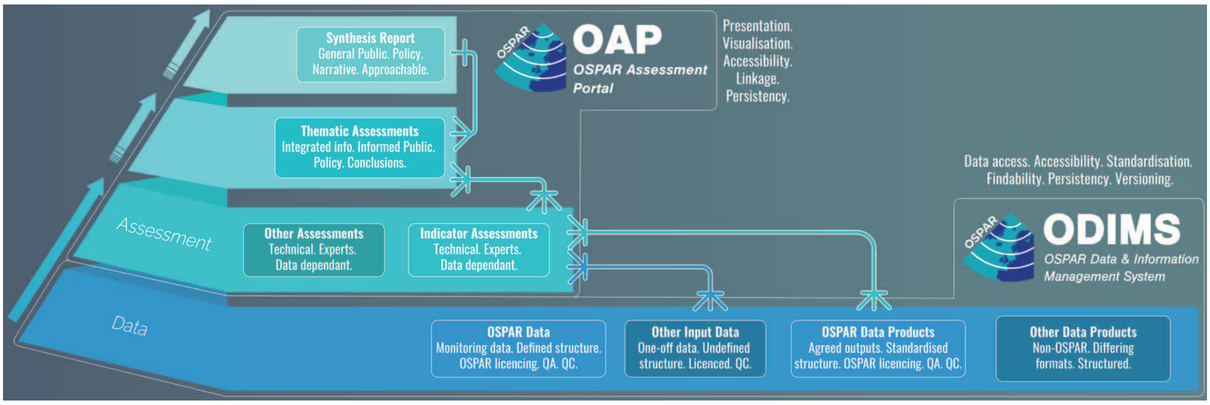


Figure 1 How data sets and products underpin the QSR2023 and its deliverables.

ODIMS was initially developed as an online tool providing a single point of access to all data gathered through OSPAR's Joint Assessment and Monitoring Programme (JAMP) as well as being critical in delivering the Monitoring and Assessment element of the North-East Atlantic Environment Strategy 2030[34] (NEAES). With an increasing level of detail on the actions and process, underneath the JAMP is the Coordinated Environmental Monitoring Programme (CEMP). The CEMP details the overall aims and concepts of delivering comparable data from across the OSPAR Maritime Area, which can be used in assessments to address the specific products raised in the JAMP.
As the underlying collection methodologies and structure are detailed, the content can be widely and reliably reused. An example of this is the integration of QSR2023 assessments in Member State reporting under European Directive 2008/56/EC - Marine Strategy Framework Directive[35]. To date, ODIMS contains over 900 individual data submissions, from 42 data streams[36]. Data are from the full complement of OSPAR thematic areas; including, data on environmental pressures, environmental status, area-based management, and measures. Not all data are managed by the Secretariat and ODIMS allows for connection to external data managers via compliant webservices. Via ODIMS, all data with a spatial component can be overlaid by the user to create individual custom maps.

[31] https://creativecommons.org/licenses/by/4.0/
[32] https://oap.ospar.org/en/data-policy/
[33] https://odims.ospar.org/en/data_policy/
[34] https://www.ospar.org/convention/strategy
[35] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32008L0056
[36] https://odims.ospar.org/en/datastreams/

Development of the ICES Fecundity and Atresia reporting format and inventory

Maria Makri[1] (maria.makri@ices.dk), Anna Osypchuk[1] (anna.osypchuk@ices.dk), Cindy van Damme[2] (cindy.vandamme@wur.nl), Evanthia Bournaka[3] (evbo@dhigroup.com), Carlos Pinto[1] (carlos@ices.dk)

[1]International Council for Exploration of the Sea – ICES (Denmark)
[2]Wageningen Marine Research (The Netherlands)
[3]DHI (Denmark)

Introduction

Estimates of spawning stock biomass (SSB), particularly through the estimation of reference points from the stock-recruitment curves, are one of the basic tools used in the assessment of the status of main commercially important fish stocks. The combination of egg production estimates from ichthyoplankton surveys with fecundity and batch fecundity data from fish sampled at appropriate time periods, can yield accurate SSB and potential SSB estimates to be used in the design and implementation of stock management plans. The estimation of fish fecundity is also crucial in providing predictions on the timing and scale of spawning, thus enabling assessment of stock condition and the provision of fishing advice. ICES has for around a decade, hosted a fish eggs and larvae database, with data collected from ichthyoplankton surveys carried out in the North Atlantic and the Mediterranean Sea. To provide full access to the data for SSB estimation from egg production methods, ICES has built a new database and a respective data inventory, to host estimations of fecundity, batch fecundity and atresia of major commercially important fish species.

Reporting Format

The reporting format and the inventory were developed by the ICES Data Centre in cooperation with the expert group on Mackerel and Horse Mackerel Egg surveys (WGMEGS). The format was designed in accordance with the manual followed by WGMEGS for the estimation of daily and annual egg production of Mackerel and Horse Mackerel [1]. According to the sampling protocol, after removing ovaries from the fish, both whole mount and sectioned ovary samples are screened for spawning markers and/or early atresia. Information on the development stages of oocytes, the presence of the hydration stage in oocytes, the presence of Post Ovulatory Follicles (POFs) and estimated Spawning and Atresia markers is collected. Based on the most advanced oocyte stage, it is decided whether the ovary sample will be used for fecundity, batch fecundity and/or atresia estimation, and the samples are flagged accordingly. The collected data are submitted to the Fecundity and Atresia database through the website interface (Figure 1) in xml or csv files. The submission files are organised in the following records: a) Cruise Information, b) Haul Information, c) Fish Measurement, c) Whole mount screening data, d) Histology screening data and, if applicable, e) Fecundity data, f) Batch Fecundity data and g) Atresia data [2]. Prior to data upload, the data undergo quality checks to ensure file integrity, format compliance, and overall quality control. These quality control checks are constantly improving, and the data manager can insert new quality control measures as needed. The database underwent testing in October 2023 and is scheduled to open for data submissions on the 1st of January 2024.

Data policy and data Extraction – Visualization

Users will be able to download data using the inventory and will be able to filter the data by Ecoregion, Country, Year, Species or Survey (Figure 2). Data will be delivered in csv format, with a separate csv file for each of the record types. Once the database is launched and data are uploaded, users will be able to spatially visualize different parameters of the data, such as oocyte stage or hydration stage. The data will be available under a CC BY 4.0 license, and in accordance with the ICES Data Policy [3].

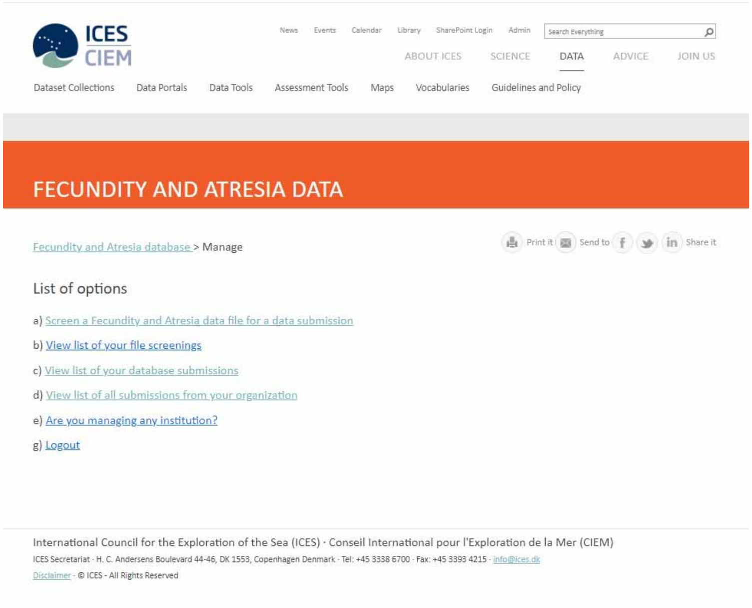


Figure 1 Inventory for screening and submitting data.

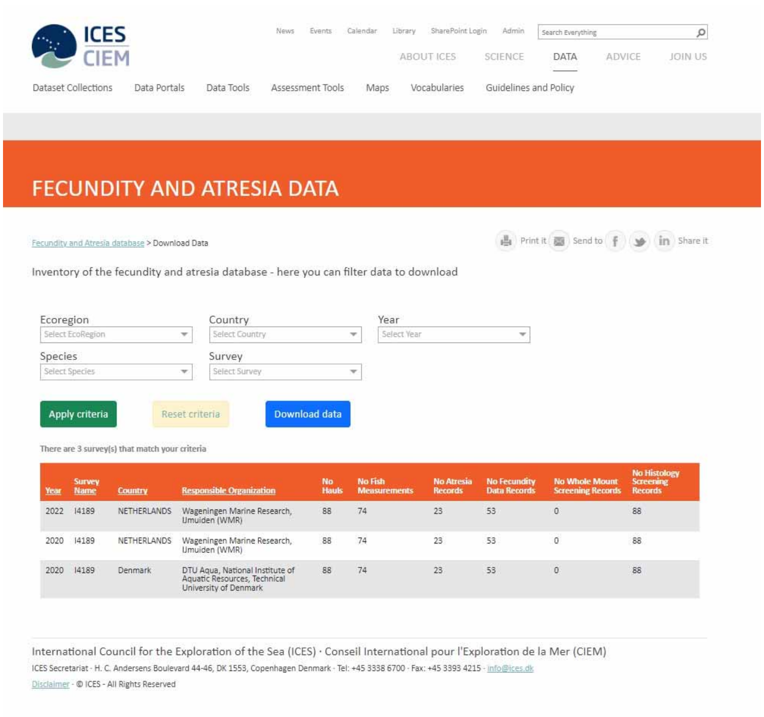


Figure 2 Inventory for downloading data from the fecundity and Atresia database (https://fecundityandatresia.ices.dk/Inventory).

Conclusions

The goal of the Fecundity and Atresia database is to provide a common platform for experts of the wider scientific community to share and access data according to the FAIR data principles [4].

It is expected to constitute a useful tool in assessing temporal and spatial shifts in fecundity and spawning of major fish stocks, thus contributing to their understanding and sustainable management. Data from fish surveys, such as the mackerel, horse mackerel, sardine, and anchovy egg surveys, are expected to populate the database, enabling the expert group to later calibrate, and analyze the data. In summary, this paper outlines a comprehensive approach to collecting, storing, and managing fecundity data.

Links

[1] https://ices-library.figshare.com/articles/report/SISP_5_-_Manual_for_the_AEPM_and_DEPM_estimation_of_fecundity_in_mackerel_and_horse_mackerel/19051103
[2] https://datsu.ices.dk/web/selRep.aspx?Dataset=157
[3] https://www.ices.dk/data/guidelines-and-policy/Pages/ICES-data-policy.aspx
[4] FAIR Principles - GO FAIR (https://www.go-fair.org/fair-principles/)

Production of machine learning datasets to predict ocean water temperature along the coast of the Korean Peninsula

Sung-Dae Kim (sdkim@kiost.ac.kr), Hyuk-Min Park (hyukmin@kiost.ac.kr)

Korea Institute of Ocean Science and Technology (Rep. of Korea)

Along with marine heat wave phenomena worldwide, high ocean temperature phenomena frequently occur in coastal areas around the Korean Peninsula. In order to predict the ocean high temperature phenomenon using deep learning technology, we collected oceanographic data and meteorological data around the Korean Peninsula and produced learning datasets for deep learning. Ocean temperature observation data was basically collected at 10 locations where abnormal marine temperatures frequently occur, and meteorological observation data, satellite data, and meteorological reanalysis data were also collected. The collected data was processed into machine learning datasets through quality control and interpolation of missing data. A prediction model for ocean high temperature was established using Python-based libraries (Tensorflow and Keras), and the model was tested under various training conditions. When the prediction period was fixed at 7 days, prediction performance was compared depending on the difference in learning period. Choosing a learning period of 7 to 8 days gave better prediction performance than choosing a learning period of 9 to 14 days. Model performance was tested using various combinations of learning datasets, and the best results were obtained when SST, reanalysis wind, and real-time water temperature data were used. Seven of the 10 test points showed predictive performance with a correlation coefficient of 0.9 or higher even after 7 days. However, considering RMSE, it is believed that the model and learning datasets can be used to predict ocean water temperature at five points.

Data collection and production of learning datasets

Ten real-time observation stations were selected to predict abnormal high temperature phenomena along the coast of the Korean Peninsula. Among the real-time observation points of the National Institute of Fisheries Science (NIFS), 7 points where relatively long-term data could be secured, and 3 points among the Korea Meteorological Administration (KMA) 's offshore buoys were selected (Figure 1). Daily averaged water temperature data was collected at NIFS points, and daily averaged meteorological observation data was collected at KMA points. Error data was removed from the collected data through the first QC using a program and the second QC using the naked eye. Missing data or data removed during the QC process were filled in during interpolation process (Figure 2).

In order to reflect the atmospheric influence on ocean water temperature, data on temperature, precipitation, humidity, barometric pressure, and sunshine hours were collected from 137 meteorological observation points of KMA across the Korean Peninsula. In addition, in order to use satellite SST data over the Korean Peninsula and the East China Sea as learning materials, NOAA's OISST data was collected. Climate reanalysis wind data provided by the Max Planck Institute for Meteorology's CDO (Climate Data Operators) was collected and used as meteorological data affecting sea surface temperature changes.

Test of deep learning time series prediction models

An ocean water temperature prediction test was performed using the LSTM model, which shows excellent performance in predicting time series data. In this study, a model based on Python and Tensorflow was established, and Anaconda and Jupyter notebook were used as development environments. Keras, a library that reduces the difficulty of deep learning development, was also used. In testing the ocean water temperature prediction model, the prediction period was fixed at 7 days, and the model performance was tested by changing the learning dataset combination, learning period, and detailed running conditions (Table 1).

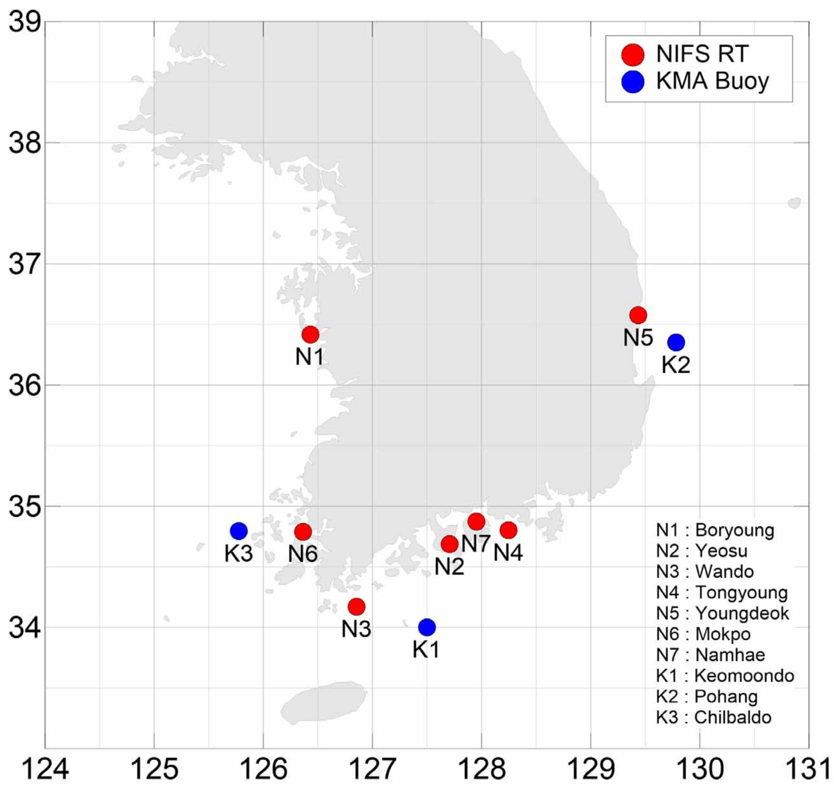


Figure 1 Observation points for prediction.

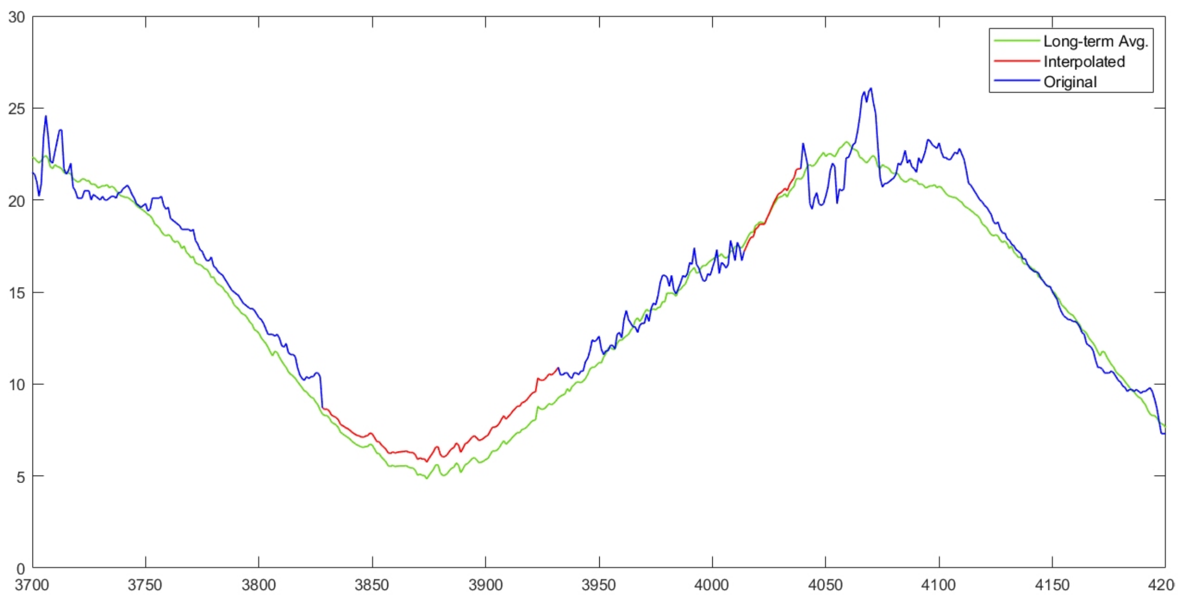


Figure 2 Comparison of original data and interpolated data.

Test	Training Datasets					
	East Chnia Sea OISST	East Chnia Sea Reanal. U wind	Korean Penisul. OISST	Kor. Penins. Re-anal. U wind	KMA Air temp., Sun hours	Observed SST
Test 1	o	o				o
Test 2			o	o		o
Test 3			o		o	o
Test 4			o	o	o	o

Table 1 Combination of learning datasets to predict ocean water temperature.

Looking at the results of each test conducted on Yeosu (N2) and Namhae (N7), where high temperature phenomenon occurred in August 2021, the correlation coefficient of Test 2 is overall high regardless of the study period (Figure 3). Test 2 also shows the highest performance in other regions. As a result of testing the model while changing the learning period in Yeosu (N2) and Namhae (N7) using the data items of Test 2, the predictability was slightly lower when the learning period was 9 days or more than when the learning period was 7-8 days (Table 2).

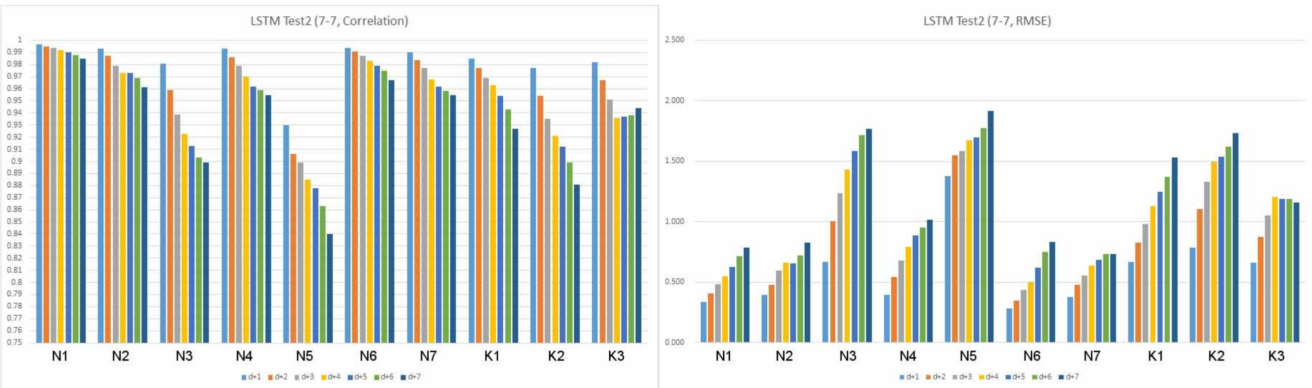


Figure 3 LSTM model performance using Test 2 combination.

Point	Learning Period	Prediction Period	d+1	d+2	d+3	d+4	d+5	d+6	d+7
Yeosu(N2)	7	7	0.993	0.987	0.979	0.973	0.973	0.969	0.961
	14	7	0.979	0.973	0.973	0.971	0.966	0.962	0.955
Namhae(N7)	7	7	0.990	0.984	0.977	0.968	0.962	0.958	0.955
	14	7	0.989	0.981	0.976	0.973	0.967	0.957	0.952

Table 2 Performance comparison according to learning period (Test 2).

Acknowledgement

This research was supported by Korea Institute of Marine Science & Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (20220033).

Leveraging Anomaly Detection for Automatic Oceanographic Data Quality Control

Ngoc Thanh Nguyen[1] (ntng@hvl.no), Rogardt Heldal[1] (rohe@hvl.no), Patrizio Pelliccione[2] (patrizio.pelliccione@gssi.it)

[1]Western Norway University of Applied Sciences (Norway)
[2]Gran Sasso Science Institute (Italy)

Oceanographic data is important to improve our understanding of the ocean, enabling sustainable and profitable explorations and exploitation of the enormous environment [Nguyen et al., 2023b]. Oceanographic data is prone to error due to several factors, like biofouling or high error rate of underwater acoustic communication [Nguyen et al., 2023a]. If low-quality data is used in decision making, the results will be misleading or suboptimal. Data Quality Control (DQC) must be done to ensure data reliability. There is an increasing need for automatic/real-time DQC [Kålvik et a., 2023]. Traditionally, DQC consists of multiple statistical tests like spike check, range check, constant value check, or gradient check [Tan et al., 2022]. This process is time-consuming and may take up to 6 months [Nguyen et al., 2023b] and is a major hindrance of sharing the data [Lima et al., 2022]. There are newly proposed AI-based solutions, which are promising to shorten the time needed for DQC significantly, like SalaciaML [Mieruch et al., 2021] or CoTeDe [Castelão, 2021]. However, they are not widely used in practice due to the complexity of the software.
In this paper, we introduce our novel anomaly detection algorithm called Adaptive Anomaly Detector (AdapAD). Similar to CoTeDe [Castelão, 2021], AdapAD detect anomalies and considers them as bad data. However, its mechanism is simpler and requires less input from users than the former. Specifically, users are expected to indicate one threshold to distinguish normality and anomaly with AdapAD, which requires knowledge of the data to be examined.
AdapAD is implemented by using Python with Pytorch (https://pytorch.org/) being utilized as processing mechanism for AI-based components, which are a normal data predictor and a threshold generator. Both components are based on Long Short-Term Memory (LSTM) and Fully-Connected Network (FCN) neural networks (see Figure 1 for the architecture) which are continuously updated to adapt to concept drift [Lu et al., 2018]. Concept drift is unforeseen changes of data distributions. For example, although temperatures change over time, we cannot know how significance of the changes.

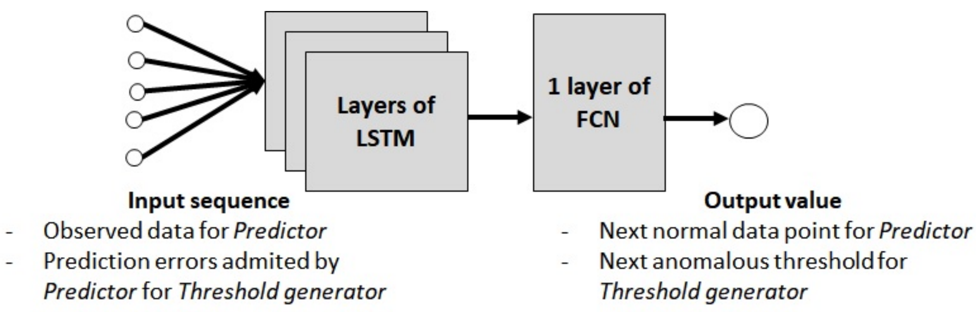


Figure 1 Architecture of AI-based components.

Thresholds to distinguish anomaly and normality are dynamically generated during the operation of AdapAD based on current states of the examined data. Therefore, human interventions during AdapAD's operation are minimized. The decision is taken by comparing differences between

the normal data predictor's output and the observed values. If the difference is larger than the in-use threshold, the examined data point is considered anomalous (i.e., bad data).
While developing AdapAD, we proactively sought input from nine domain experts from six marine organizations to ensure that the algorithm meets their needs regarding automatic DQC. AdapAD was tested on three real-world oceanographic datasets, with one being collected at the test site of the Austevoll municipality, west coast of Norway, within the scope of the SFI Smart Ocean project (https://sfismartocean.no/) and the others being obtained from the One Ocean Expedition (https://oneoceanexpedition.com/).
Technical evaluations showed that AdapAD achieved the precision of at least 85% on the three tested datasets and could make decisions in less than one minute per data point. We also performed a user acceptance survey with 11 marine data producers and consumers from the SFI Smart Ocean project to assess whether AdapAD is a viable solution for automatic marine DQC. We received a score of 4.4 (5 is highest) for the potential of using AdapAD for such the task. Aanderaa company recommended us converting AdapAD to the C/C++ programming language and optimizing its footprint to embed it into sensors for DQC at underwater data acquisition sources. AdapAD is used to perform DQC on the wave dataset of the OOE [Ølberg et al., 2024].

References

Castelão G.P., (2021). *A machine learning approach to quality control oceanographic data*. Computers & Geosciences, 155:104803.

Lima K., Nguyen N.-T., Heldal R., Knauss E., Oyetoyan T.D., Pelliccione P., and Kristensen L.M., (2022). *Marine Data Sharing: Challenges, Technology Drivers and Quality Attributes*. In International Conference on Product-Focused Software Process Improvement, pages 124–140, Springer.

Lu J., Liu A., Dong F., Gu F., Gama J., and Zhang G., (2018). *Learning under concept drift: A review*. IEEE Transactions on Knowledge and Data Engineering, 31(12):2346–2363.

Mieruch S., Demirel S., Simoncelli S., Schlitzer R., and Seitz S., (2021). *SalaciaML: A deep learning approach for supporting ocean data quality control*. Frontiers in Marine Science, 8:611742.

Nguyen N.-T., Heldal R., Lima K., Oyetoyan T.D., Pelliccione P., Kristensen L.M., Hoydal K.W., Reiersgaard P.A., and Kvinnsland Y., (2023a). *Engineering challenges of stationary wireless smart ocean observation systems*. IEEE Internet of Things Journal.

Nguyen N.-T., Lima K., Skålvik A.M., Heldal R., Knauss E., Oyetoyan T.D., Pelliccione P., and Sætre C., (2023b). *Synthesized data quality requirements and roadmap for improving reusability of in-situ marine data*. In 2023 IEEE 31st International Requirements Engineering Conference (RE), pages 65–76. IEEE.

Kålvik A.M., Saetre C., Frøysa K.-E., Bjørk R.N., and Tengberg A., (2023). *Challenges, limitations, and measurement strategies to ensure data quality in deep-sea sensors*. Frontiers in Marine Science, 10:1152236.

Ølberg J.T., Bohlinger P., Breivik Ø., Christensen K.H., Furevik B.R., Hole L.R., Hope G., Jensen A., Knoblauch F., Nguyen N.T. and Rabault J., (2024). *Wave measurements using open-source ship mounted ultrasonic altimeter and motion correction system during the one ocean circumnavigation*. Ocean Engineering, 292, p.116586.

Tan Z., Zhang B., Wu X., Dong M. and Cheng L., (2022). *Quality control for ocean observations: From present to future*. Science China Earth Sciences, pp.1-18.

Mareano User Focus and FAIR

Gjertrud Røyland[1] (gjertrud.royland@kartverket.no), Geir Arne H. Nordhus[1] (geir.arne.nordhus@kartverket.no), Ane Bang-Kittilsen[2] (ane.kittilsen@ngu.no), Kjell Gunnar Bakkeplass[3] (kjell.bakkeplass@hi.no), Aave Lepland[2] (aave.lepland@ngu.no), Hanne Bertelsen[1] (hanne.bertelsen@kartverket.no), Marte L. Strømme[3] (marte.stromme@hi.no)

[1]Norwegian Mapping Authority (Norway)
[2]Geological Survey of Norway (Norway)
[3]Institute of Marine Research (Norway

The Mareano programme has for the last 20 years been mapping Norwegian offshore areas (Figure 1). The main products are:

- Information about seabed characteristics and biotopes.
- Information about the distribution of benthic fauna, communities, biological diversity and production.
- Information on contaminants in the sediments.
- Detailed bathymetric maps.
- Online database and map services collecting information on Norwegian coastal and ocean regions.

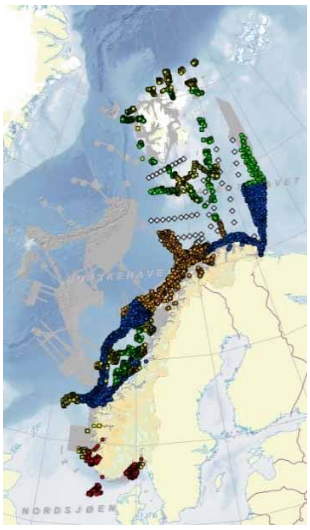


Figure 1 Mareano mapped areas.

The Institute of Marine Research, the Geological Survey of Norway and the Norwegian Mapping Authority comprise the Executive Group which is responsible for carrying out the Mareano field sampling and other scientific activities. The Programme Group, led by the Norwegian Environment Agency, has the executive responsibility for the MAREANO activities.
In order to meet the needs of the users, the Mareano program has appointed the User Focus and FAIR groups, with the the mandate to ensure that Mareano data, products and services are always focused on meeting the users' needs, be publicly known, easily accessible and to provide a good user experience.
In order to map the needs of the user, and continously work towards meeting their evolving needs, the User Focus group conducted two user surveys. During 2021 and 2022, 17 in-depth interviews were conducted on selected stakeholders representing main user groups, including Academia, fisheries, oil and gas exploration, environmental protection organisations, aquaculture

and environmental management. The work resulted in a report where the findings were presented in form of user journeys. In 2023, this work was followed up by a user survey in form of a online questionary, based on findings from the user journeys.
Both surveys gave rather similar results that give us good basis for designing work plans for User Focus group and giving advise for Mareano Program group. Some issues that came to light can be easily managed within today's technology. Others will require strategic decisions in relation to Mareno's objectives, and new technology.

Mareano data users
Mareano has to take into consideration that the users vary considerably in regards of data skills and what kind of information they are looking for. To better understand the needs, wishes and different ways Mareno data and services is used, we defined four user categories (Table 1).

Users	Basic technical skills	Advanced technical skills
Basic academic skills	Need ready-made maps, fewer categories, simple language.	Have competence for downloading and organizing data. Do not evaluate content as carefully.
Advanced academic skills	Prefer to use ready-made maps, familiar with technical language and benefits from detailed categories.	Is interested in and capable of using all available categories and data. Can process and analyze data themselves. Need rich content of data.

Table 1 Mareano user classification.

Result highlights:
- Everyone is interested in better coverage of data.
- Mareano should be one step ahead and prepared to provide data for marine areas where high activity is planned.
- All users agree on the benefits of collaboration and providing a common information basis for decision making.
- The users trust Mareano data.
- Users wish to combine Mareano data with data from different sources.

Mareano FAIR work
The Mareano FAIR group proposes solutions to the Program group to ensure that data, products and metadata are accessible and usable for users through online services in accordance with the FAIR principles, as well as the obligations regarding sharing of data and services through the Geonorge and NMDC (Norwegian Marine Data Center) infrastructures.
The FAIR status calculator (Figure 2) provides automatic assesment of FAIRness compliance of Mareano datasets in the national geographical metadata catalouge Geonorge. The FAIR-status score is calculated based on weighted criteria. Almost 100 Mareano data sets have been reviewed so far.
In 2024, the FAIR group plans to test out and adopt new OGC standards/API's that are organized for improved ease of use and better facilitated for analysis (filtering, overlay, modeling, dynamic data).
Future plans exists to tie the User Focus group and FAIR group more tightly together, to provide user evaluation and to guide users of the "FAIR-ified" data sets through workshops and webinars.

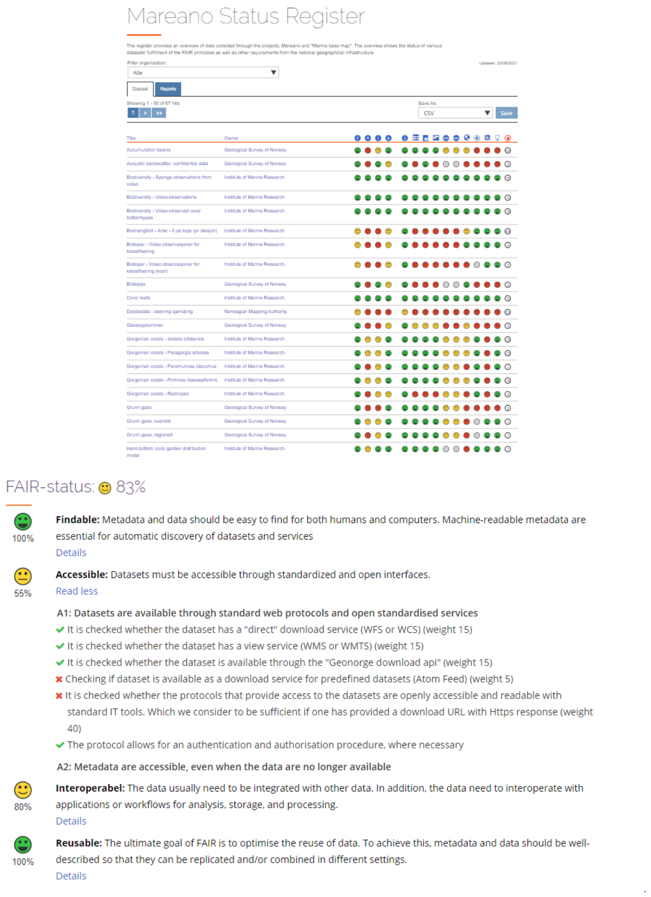


Figure 2 Mareano FAIR status register. https://register.geonorge.no/mareano-statusregister.

The potential of the Portuguese Coastal Monitoring Network (MONIZEE) as a validation source for ocean products

Luísa Lamas (luisa.lamas@hidrografico.pt), Vânia Lima (vania.lima@hidrografico.pt), Inês Martins (marina.martins@hidrografico.pt), Teotónio Barroqueiro (pires.barroqueiro@marinha.pt), Paulo Nunes (antunes.nunes@marinha.pt)

Instituto Hidrográfico (Portugal)

General Description

The Instituto Hidrográfico (IH) has national responsabilities for providing data and information essential for the safety in navigation, marine protection and sustainable exploration of the ocean. To that aim, IH has been observing the ocean for more than 60 years, namely on the Portuguese Economic Exclusive Zone (EEZ), which spans from the Iberian Peninsula into the central North Atlantic, covering the Azores Islands and the subtropical waters around Madeira Island. As a result, IH manages an extensive national database on oceanic data, a repository that grows daily. The Portuguese Coastal Monitoring Network (MONIZEE), operated by IH and other national partners, comprises a set of in-situ sensors (moored buoys and tide gauge stations) and remote sensing systems (coastal HF Radar) [Martins et al., 2010]. These sensors offer near-real-time oceanographic and meteorological data, including sea surface temperature, sea state, sea level, atmospheric pressure at sea level, and surface currents (Figure 1). These parameters are crucial to know, understand and predict the complex dynamics of the coastal ocean and provide essential support for naval missions, maritime activities, and model predictions. The MONIZEE network contributes to european infrastructures such as EMODnet Physics, EuroGOOS and JERICO and its data is integrated in the Copernicus Marine Service In Situ TAC.

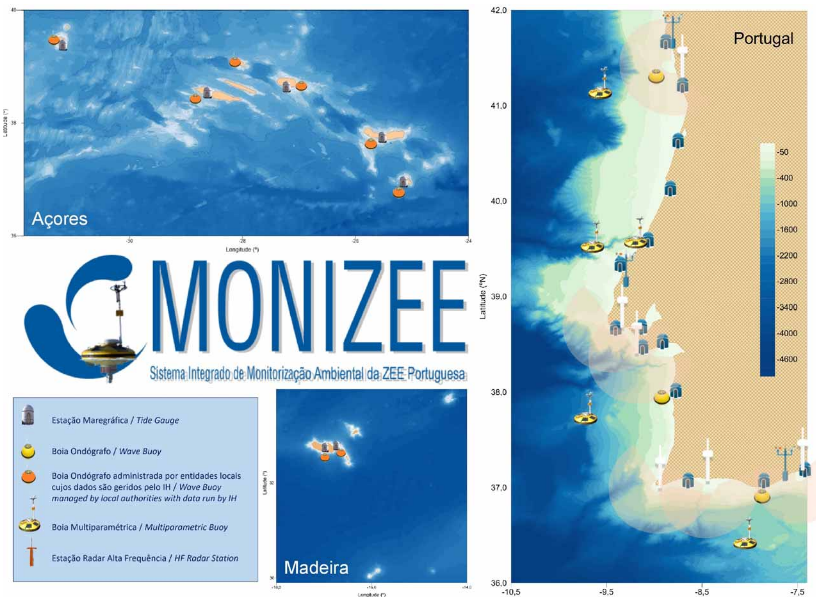


Figure 1 The Portuguese Ocean Monitoring Network (MONIZEE).

In recent years, IH has been working on the comparison between sea surface temperature (SST) measurements collected by moored buoys and from sensors onboard satellites, namely as beta-testers for the European Space Agency's Sea Surface Temperature Climate Change Initiative (ESA SST CCI) Analysis Product Version 2.0 and 3.0 [Lamas L., 2022], working on assessing the product's quality for oceanic and climatic studies over the Portuguese marine waters (Figure 2).

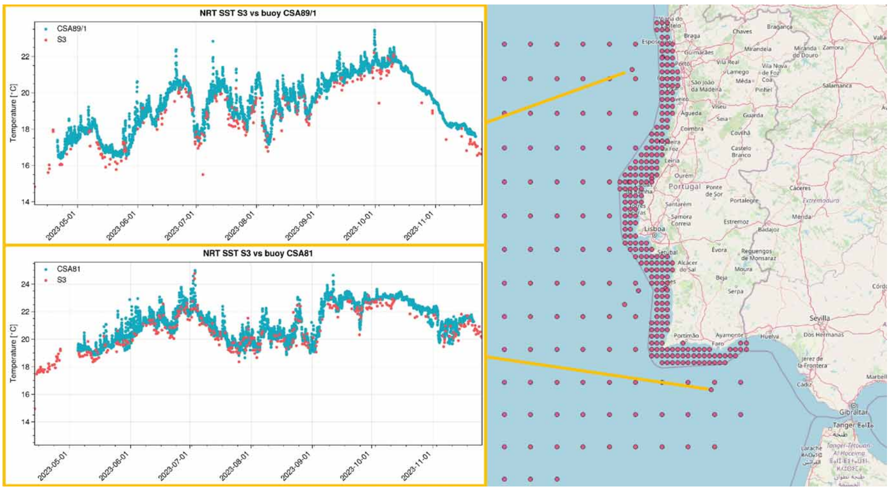


Figure 2 Statistical comparison between MONIZEE data and ESA SST CCI v3.0 product [Lamas, 2022].

As part of the Sentinel 3 Validation Team (S3VT), IH is undertaking the regional validation of SST products obtained from the Sentinel 3 (S3) satellites. This validation is focused specifically on the oceanic and coastal waters off Portugal, using in-situ data collected from moored buoys within the MONIZEE system. Since these in-situ data are not integrated into the calibration procedure of S3, they serve as a unique and independent source for validating SST products over an extensive area with limited available in-situ data.

IH's initiative within the S3VT involves not just statistical comparisons between S3-derived SST data and historical in-situ data from MONIZEE, but also the implementation of an operational near-real-time monitoring service accessible at Hidrografico+ (geomar.hidrografico.pt), the marine data infrastructure of IH (Figure 3). This service facilitates a systematic inter-comparison between near-real-time S3 data and in-situ measurements, providing valuable insights for monitoring purposes.

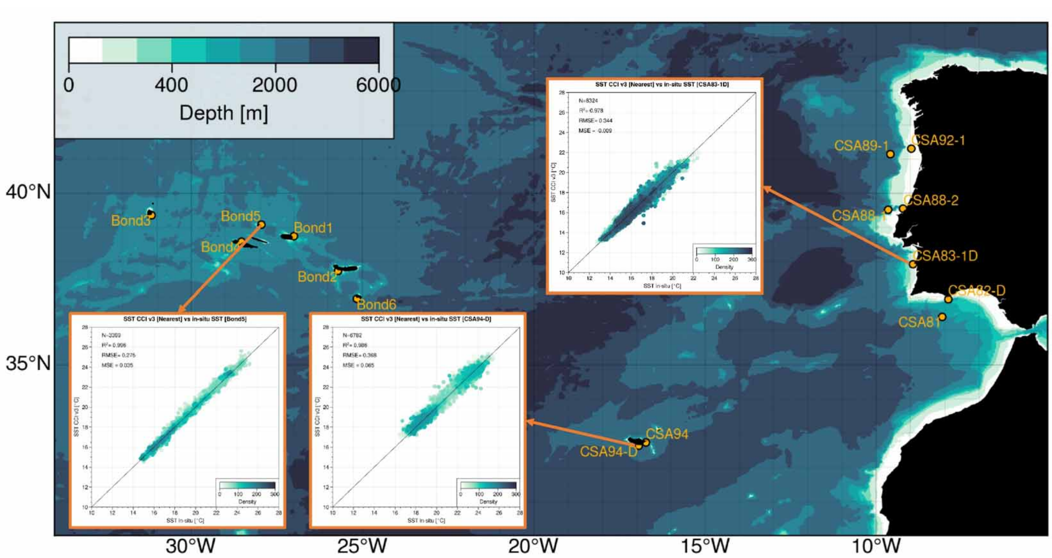


Figure 3 S3 Validation Service from Hidrográfico+.

IH's extensive historical ocean database together with near-real-time monitoring capabilities present a valuable opportunity to independently assess the quality of ocean products, made available by numerous international programs, such as Copernicus.

References.
Martins I., Vitorino J. and Almeida S., (2010). *The Nazare Canyon observatory (W Portugal) real-time monitoring of a large submarine canyon.* OCEANS'10 IEEE SYDNEY.
Lamas L., (2022). *Feedback from the ESA SST CCI v3.0 analysis product over the Eastern Atlantic.* Trail Blazer User Report.

A regional dataset of dissolved oxygen in the Western Mediterranean Sea (2004-2023): O2WMED

Malek Belgacem[1] (malek.belgacem@ve.ismar.cnr.it), Katrin Schroeder[1] (katrin.schroeder@ismar.cnr.it), Siv K. Lauvset[2] (sivk@norceresearch.no), Marta Álvarez[3] (marta.alvarez@ieo.csic.es), Jacopo Chiggiato[1] (jacopo.chiggiato@ismar.cnr.it), Mireno Borghini[1] (mireno.borghini@sp.ismar.cnr.it), Carolina Cantoni[1] (carolina.cantoni@ts.ismar.cnr.it), Tiziana Ciuffardi[4] (tiziana.ciuffardi@enea.it), Stefania Sparnocchia[1] (stefania.sparnocchia@cnr.it)

[1]Consiglio Nazionale delle Ricerche - Istituto di Scienze Marine - CNR – ISMAR (Italy)
[2]NORCE Climate and Environment (Norway)
[3]IOE (Spain)
[4]ENEA (Italy)

Ocean deoxygenation is a crucial measure of changes in the ocean due to climate conditions. Over the past two decades, oxygen concentrations in the ocean have been declining as a result of global anthropogenic pressures. The warming of the ocean and increased levels of CO_2 have caused significant changes in the distribution and structure of marine communities, ultimately altering the composition of ecosystems.

The decline in oxygen levels can be attributed to multiple factors. Firstly, the warming of the ocean reduces the solubility rate of oxygen, making it less available in the water. Additionally, increased stratification of water columns inhibits the diffusion of well-oxygenated surface waters to the intermediate and deep layers.

While the surface layer of the ocean remains well-oxygenated due to photosynthesis and oxygen dissolution from air-sea exchange, oxygen levels decrease with depth. Oxygen is consumed and regenerated through ventilation processes, leading to variations in the biological loop and distribution of biogeochemical tracers, which are still not fully understood.

The Mediterranean Sea has experienced significant impacts from these changes, particularly in recent decades with the occurrence of more intense and frequent marine heat waves. Deep-water formation in the Gulf of Lion, for example, has become less intense, resulting in a reduction of available oxygen in the intermediate water. Therefore, understanding oxygen variations in the Western Mediterranean Sea (WMED) is of utmost importance.

To investigate oxygen trends in the WMED, a reliable and consistent dataset is essential. The O2WMED (dissolved oxygen in the Western Mediterranean Sea) dataset has been developed, consisting of 1381 CTD (conductivity, temperature, and depth) stations collected during 24 cruises between 2004 and 2023 by the Italian National Research Council. We utilized data from cruises where regular calibration against Winkler titration was conducted. This dataset has not been published previously and will be made available in PANGAEA (in progress).

In this study, we provide a comprehensive overview of the oxygen status in the WMED and describe in detail the quality control scheme employed to generate a consistent data product from the Italian Cruises. The dataset underwent rigorous quality checks and thorough analysis to identify any potential biases or errors. This analysis included comparing deep profiles in the same area at different times, following the Global Ocean Data Analysis Project (GLODAP) protocol.

Any anomalies or inconsistencies in the O2WMED dataset were identified and rectified, ensuring the integrity of the dataset and contributing to existing datasets. The resulting product is well-suited for validating models and producing regional climatology for the WMED including data from European Infrastructures. This dataset will significantly enhance our knowledge of oxygen variations in the region, ultimately aiding in the understanding of ocean deoxygenation and its ecological consequences.

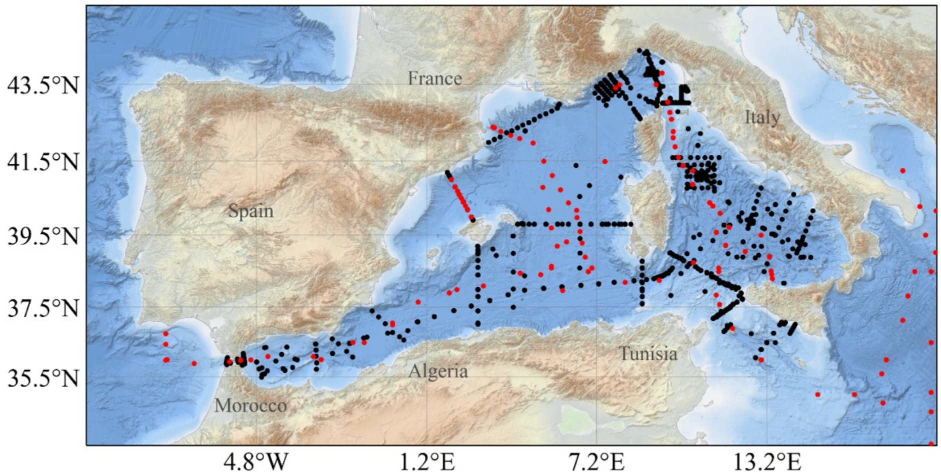


Figure 1 Map of the western Mediterranean Sea showing the dissolved oxygen CTD stations in black dots included in the O2WMED and the reference cruises in red dots.

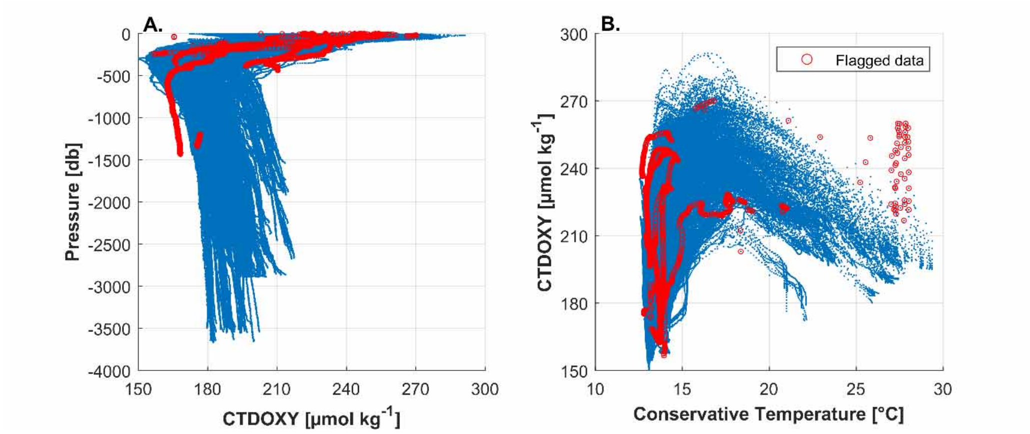


Figure 2 Scatter plots illustrating (A) Oxygen vs. pressure and (B) Oxygen vs. Conservative Temperature while red circles indicate observations flagged as questionable (Flag 3).

Poster session

Promoting equitable access and use of data in the EAF-Nansen Programme

Cristian Muñoz Mas[1] (cristian.munoz.mas@hi.no), Nikolaos Nikolioudakis[1] (nikolaos.nikolioudakis@hi.no), Erik Olsen[1] (erik.olsen@hi.no), Maik Tiedemann[1] (maik.tiedemann@hi.no), Kathrine Michalsen[1] (kathrine.michalsen@hi.no), Merete Tandstad[2] (Merete.Tandstad@fao.org), Kyriakos Kourkouliotis[2] (Kyriakos.Kourkouliotis@fao.org), Rishi Sharma[2] (Rishi.Sharma@fao.org), Pedro Barros[2] (Pedro.Barros@fao.org)

[1]IMR (Norway)
[2]FAO (International)

Introduction

Since 1975, the Nansen Programme has collaborated with 58 countries across Africa, the Indian Ocean, Central and South America and Southeast Asia (Figure 1). In 2006, the Programme committed to improving fisheries management in line with the ecosystem approach to fisheries (EAF), strengthening the capacities of fisheries institutions and generating scientific knowledge on marine resources and ecosystems. The current EAF-Nansen Programme is executed by the Food and Agriculture Organization (FAO) in close collaboration with the Norwegian Institute of Marine Research (IMR) and funded by the Norwegian Agency for Development Cooperation (Norad). It is one of the endorsed 'Decade Actions' under the United Nations Decade of Ocean Science for Sustainable Development (2021-2030). FAIR and CARE are complementary perspectives. FAIR principles primarily concentrate on data attributes that promote greater data sharing between entities, while overlooking power dynamics and historical contexts. The CARE Principles for Indigenous Data Governance prioritize people and purpose, acknowledging the pivotal role of data in driving Indigenous innovation and self-determination [Carroll et al., 2021]. FAIR and CARE principles can be used for discouraging colonial science while promoting equitable scientific practices across all the partner countries of the EAF-Nansen Programme.

Figure 1 Map of the geographic scope of the EAF-Nansen Programme during the period 1975-2023.

Challenges for implementing FAIR-CARE principles within the EAF-Nansen Programme

Ensuring FAIR principles has been a challenging task with the continuously expanding volume and diversity of data generated by the Programme, mainly through surveys with the research vessel Dr. Fridtjof Nansen, since 1975. The Programme, committed to FAO's principles on data ownership, aims at promoting partner countries' credit, collective benefit and capacity development and discouraging colonial science. Safeguarding data ownership and intellectual property rights while encouraging international collaboration and creation of global public goods without fostering data silos is a delicate equilibrium. The challenge intensifies when there is limited interoperability or a lack of data exchange between repositories, resulting in non-synchronised data repositories that pose a significant barrier to upholding FAIR principles. Findability emerges as arguably the most crucial. The increasing trend towards data transparency in scientific publications further complicates matters for implementing CARE principles. This trend amplifies the complexity by instigating a tug-of-war between the advantages of openly sharing data and the imperative to uphold countries' rights in achieving equitable development and sovereignty.

EAF-Nansen Programme Data Strategy

The EAF-Nansen Programme, in its new phase (2024-2028), is increasing the visibility of data and promoting their utilization while safeguarding the rights of the partner countries through an updated data strategy. FAIR-CARE operationalization will be a core element, and several actions are planned to address identified challenges:

- A comprehensive metadata catalogue, implementation of standards and best practices for managing (meta)data, and deployment of interoperable platforms for (meta)data distribution.
- Promoting training and education in data management to equip partner countries with the necessary tools and knowledge, allowing them to make the best-informed decisions possible for governing their data. This will facilitate the creation of a technical network of data stewards within the Programme to work and safeguard national data and to establish and maintain relationships with other data stakeholders that can support the generation of value from and for the countries owning the data and support the implementation of the FAIR-CARE principles within the Programme.
- Support data and information hubs of regional partners in alignment with the UN Ocean Decade key outcome for *"a transparent ocean with open and equitable access to data, information and technologies"* on a regional-global scale.
- Encourage synergies to promote capacity development opportunities for the EAF-Nansen Programme partners, and in accordance with the EAF-Nansen Data Policy, for sharing specific data and information with other relevant Ocean Decade Programmes (e.g., Seabed2030 or OceanData-2030 through Ocean InfoHub), also considering the proposed Actions in the Ocean Decade Africa Roadmap.
- Facilitate the creation and dissemination of adaptive methodologies to improve the best available knowledge in accordance with the existing capacities with the support of the Ocean Practices for the Decade Programme (OceanPractices).

Conclusion

As the open science movement is being embraced worldwide, operationalizing FAIR-CARE in the context of the EAF-Nansen Programme is critical for ensuring equitability in the process. A new data strategy, highlighting FAIR-CARE principles, will be implemented during the next period of the Programme (2024-2028). FAIR-CARE operationalization will require the engagement of all partner countries of the Programme, particularly in the co-design, co-development and implementation of the data strategy, and co-management of information generated. A dynamic Data Policy, revised frequently and as necessary, is the governing tool for the implementation of this strategy.

ReMAP innovative technical framework for the support of the European Maritime Spatial Planning process

Andrej Abramic[1] (andrej.abramic@ulpgc.es), Victor Cordero Penin[1] (victor.cordero@fpct.ulpgc.es), Yaiza Fernández-Palacios Vallejo[1] (yaiza.fernandezpalacios@ulpgc.es), Marta Ballesteros[2] (mballesteros@cetmar.org), Jose Luis Santiago[2] (jsantiago@cetmar.org), Yolanda Irawan Rincón[2] (yirawan@cetmar.org), Stefano Menegon[3] (stefano.menegon@cnr.it) Alessandro Sarretta[3] (alessandro.sarretta@cnr.it), Amedeo Fadini[3] (amedeo.fadini@ve.ismar.cnr.it), Alessandro Mulazzani[3] (alessandro.mulazzani@gmail.com), Deborah Shinoda[4] (deborah.shinoda@helcom.fi), Joni Kaitaranta[4] (joni.kaitaranta@helcom.fi), Andzej Milos[4] (Andzej.Milos@helcom.fi), Emilie Delaroche[5] (emilie.delaroche@shom.fr), Bastien Fouque[5] (bastien.fouque@shom.fr), Jean-Baptiste Suzanne[5] (jean-baptiste.suzanne@shom.fr), Adeline Souf[5] (adeline.souf@shom.fr), Daniele Brigolin[6] (dbrigolin@iuav.it), Fabio Carella[6] (fcarella@iuav.it), Isabel Maria Moreno Aranda[7] (Isabel.M.Moreno@cedex.es), Ricardo Haroun Tabraue[1] (ricardo.haroun@ulpgc.es)

[1]ECOAQUA Institute, University Las Palmas de Gran Canaria (Spain)
[2]CETMAR (Spain)
[3]Consiglio Nazionale delle Ricerche - CNR (Italy)
[4]HELCOM (Finland)
[5]Shom (Francia)
[6]IUAV (Italy)
[7]CEDEX (Spain)

The “blue acceleration” describes the race among diverse claims for using marine resources and space. Maritime Spatial Planning (MSP) has become crucial to balancing objectives, allocating uses, managing conflicts and guiding future development. In the European Union, the MSP Directive prompted the approval of MSP plans. Once the planning phase is completed, Member States and practitioners face implementation challenges, often lacking tools to support effective monitoring. Using a multidisciplinary approach, the project Reviewing and Evaluating the Monitoring and Assessment of Maritime Spatial Planning (ReMAP) focuses on setting-up strategies and approaches to review MSP plans. ReMAP develops data tools, models and reuses operational data infrastructures, allowing interoperability and enabling European Member States (MS) to share MSP data and assessment information. The main output is a set of online open access tools useful for the implementation and review of the adopted plans.
The framework builds on the findings of the Technical Expert Group on data for MSP, which in 2021 set a data standard for sharing maritime plans, based on INSPIRE principles, and currently applied with EMODnet to deliver European MSP data layer. Currently EMOdnet has harmonized 10 European MSP’s, including the plans within region of Baltic, North Sea, Atlantic and Mediterranean. Harmonized plans provide the opportunity to build on the interoperability. ReMAP proposes an innovative approach of modular analytics. The proposed modular analytics allow assessment of the environmental, economic and social MSP performance. Applying 10+1 simple (and reusable) analytics of a wide range of key aspects, jointly provide complex assessment on current MSP status.
A modularity approach is applied, which means that each analytical module is self-standing and can be run independently of other modules. An analytical module consists of a conceptual model, structured input information (detailed by a data model that follows INSPIRE principles) and pre-defined output information. To ensure feasibility and replicability, the analytical modules are

designed using a “Keep It Simple” approach - developing simple analysis replicable for any other plan that follows the MSP data standard.

This approach decreases the complexity of the assessments and, at the same time, increases the probability of being reused by a wide MSP community. Making the tools usable in real policy contexts requires a triangulation of methods: first, the use of case studies to test the tools in 3 use cases in Galicia (local), and two cross-border use cases: the Western Mediterranean sub-basin, where it will be applied with Spanish, French and Italian MSP; and for the Baltic basin. Second, the implementation of co-development and technology uptake processes engaging stakeholders in alpha and beta test workshops to strength the tool’s usability, credibility, legitimacy, and relevance. Third is the multidimensional approach, addressing environmental and socioeconomic dimensions, services, uses and interactions (Figure 1).

Figure 1 ReMAP technical framework that build on Technical Expert Group on MSP data.

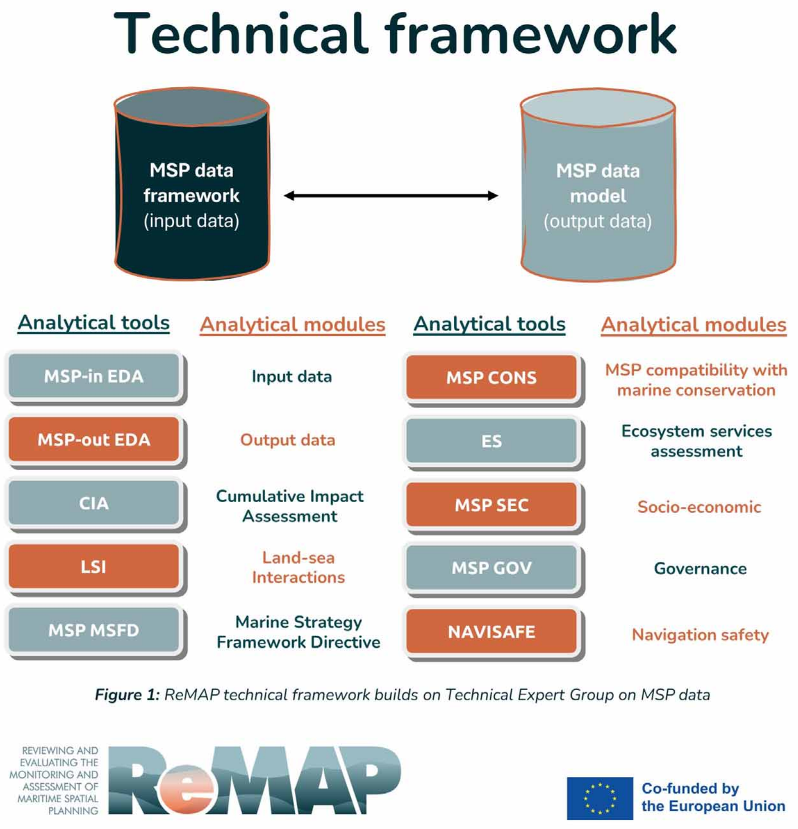
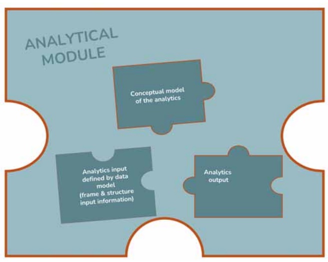


Figure 2 Analytical module structure.



After first year, the development of 10+1 analytical modules has been completed. Sharing the findings in terms of the tools, the interactions with multiple stakeholders and end users and the challenges associated with the modular and integrated approach of the framework will benefit the reflection of the IMDIS session *Data products for multiple stakeholders/end users including the Blue Economy/Blue Growth/Maritime Spatial Planning*.

SESSION INFRA ORAL PRESENTATIONS

The EMODnet Central Portal: how federated webservices led to a centralisation success story

Joana Beja[1] (joana.beja@vliz.be), Conor Delaney[2] (Conor.Delaney@emodnet.eu), Tim Collart[2] (tim.collart@emodnet.eu), Frederic Leclercq[1] (frederic.leclercq@vliz.be), Bart Vanhoorne[1] (bart.vanhoorne@vliz.be)

[1]VLIZ, Flanders Marine Institute (Belgium)
[2]EMODnet Secretariat (Belgium)

The European Marine Observation and Data Network, EMODnet, is service of the DG Mare (Directorate Generals of Maritime Affairs and Fisheries, of the European Commission), established in 2009, to provide open access to in situ marine data and derived data products, tools, and services. Its guiding principle to "collect once use may times" has been the backbone for all developments and evolution which led, in 2023 to a unified service through a single access point web portal. Since its inception and until 2022, EMODnet existed as a collection of nine (Bathymetry, Biology, Chemistry, Geology, Human Activities, Physics, Seabed Habitats, Central Portal, and Data Ingestion) web portals which provided access to their outputs in a similar way. Seven of those portals (thematic lots- Bathymetry, Biology, Chemistry, Geology, Human Activities, Physics, Seabed Habitats) have now been merged into the EMODnet Central Portal and work is ongoing to incorporate Data Ingestion, a work that is due completion in the spring of 2024.
The new EMODnet Central Portal provides unified data, metadata discovery, visualisation and download services, which are presented to users through a map viewer and metadata catalogue powered by a collection of backend webservices maintained by the thematic lots. The centralisation process was only possible due to the interoperability provided by these webservices, which comply with a set of geospatial (OGC, Open Geospatial Consortium) and OpenDAP (via ERDDAP API) data service standards.

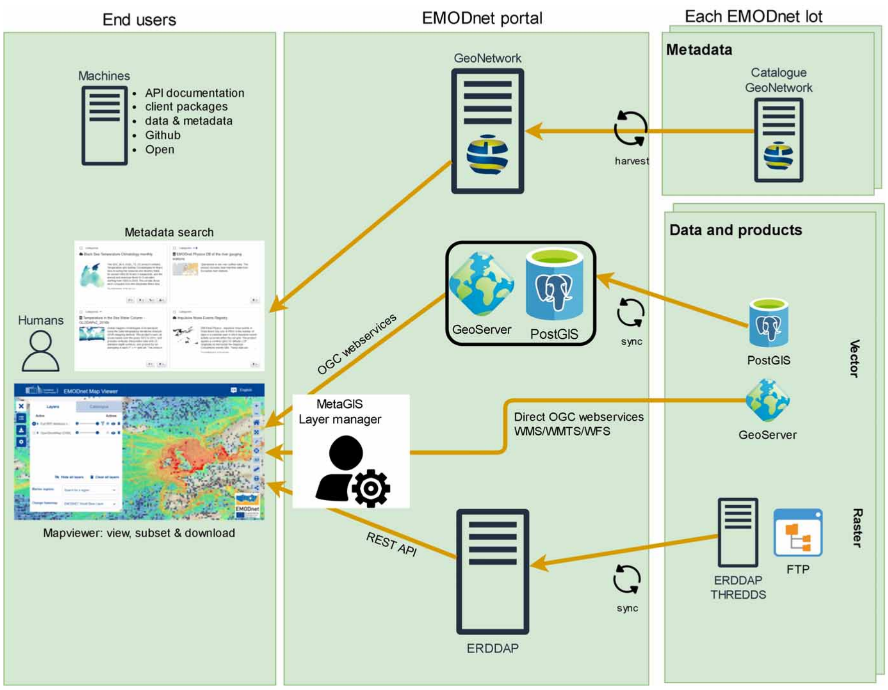


Figure 1 High level technical architectural diagram of EMODnet publishing services.

This process impacted not just the technical but also communication and administration aspects, leading to increased standardisation and consistency of both the publication and communication procedures of the various assets and work developed by the EMODnet community.

In this abstract, we present the EMODnet Central Portal technology stack, we discuss distributed architectures, webservice standards and how they were used to provide the functionality of the map viewer and the metadata catalogue. Finally, we discuss the positive and negative aspects of managing a distributed spatial data infrastructure and how we plan for further improvement in EMODnet services by means of the EDITO-INFRA Project, which will implement the core infrastructure of the EU Digital Twin Ocean.

The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) 2021/1139 of the European Parliament and of the Council of 7 July 2021 establishing the European Maritime, Fisheries and Aquaculture Fund and its predecessor, Regulation (EU) No. 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund.

EMODnet Biology: an EU service for the marine biology community and beyond

Joana Beja[1] (joana.beja@vliz.be), Leen Vandepitte[1] (leen.vandepitte@vliz.be),
Ruben Perez Perez[1] (ruben.perez@vliz.be), Benjamin Weigel[2] (benjamin.weigel@inrae.fr),
Marina Lipizer[3] (mlipizer@ogs.trieste.it), Bart Vanhoorne[1] (bart.vanhoorne@vliz.be),
Lennert Tyberghein[1] (lennert.tyberghein@vliz.be), EMODnet Biology[4]

[1]VLIZ, Flanders Marine Institute (Belgium)
[2]INRAE - EABX (France)
[3]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[3]EMODnet Biology Consortium

EMODnet Biology is the EU's operational in situ service for marine biodiversity data, with actions focusing mainly on data publishing, creation of data products and publication of informative material covering the European Seas and the following groups: (macro)algae, angiosperms, benthos, birds, fish, mammals, phytoplankton, reptiles and zooplankton. The data work is supported by the EurOBIS data infrastructure (database and European OBIS node) hosted by the Flanders Marine Institute which has frequent updates to allow for the management and publication of new types of data, from imaging to species tracking and omics. On the IMDIS 2021 edition, EMODnet Biology was finishing Phase III; we are now in Phase V and will report not only on the progress achieved since 2021 but also on what we plan to do until the end of the current Phase, ending in May 2025.

Data Publication

Since 2021 the number of datasets with published data has increased by more than 20%, from 1077 to 1301 metadata records. This brought the total number of biological occurrence records from 25.5 million to 35.6 million and additional biotic and abiotic parameters, collected alongside the biological data, from 30.7 million to 92.4 million. A detailed breakdown of acquired occurrence records for each group and region, per EMODnet phase, is illustrated in Figure 1, comprising almost 15 years of data acquisition. With 98.9% open data and only 1.1% being restricted, this demonstrates the success the initiative has had over the past 1.5 decades in expanding its network of data providers and promoting Open Science and FAIR principles.
Our data infrastructure is now ready to receive not only species observations, but data from automated imaging systems, e.g., Zooscan and tracking data. Developments are underway to allow for the publication of omics data, adhering not only to the international community standards and guidance but also complying with the EU's requirements.

Data Products

In April 2023 (end of Phase IV), the number of available products increased to 38, including three from external sources. The latest products have their code and methodology available on the EMODnet GitHub (https://github.com/EMODnet) allowing users to implement the approaches to their own data.
The EMODnet centralisation in 2023, forced changes and updates to internal procedures as well as to the way our products were built, in order to comply with the software and systems set in place for the EMODnet Central Portal. This required, e.g., the drafting of guidance on how to create biological products in a format that was compatible with the systems in place (e.g., ERDDAP, GeoServer). Despite the work and efforts made, integration wasn't always straightforward and althought following the established conventions (e.g., CF conventions),

biological products are often not as easy to manipulate or integrate with other data types (e.g., Physics or Chemistry).
Within Phase V specific efforts will be dedicated to promote EMODnet biological data and data products towards major data users in charge of marine environmental status asessments, such as Regional Sea Conventions and EU institutions responsible for the implementation of marine environmental directives.

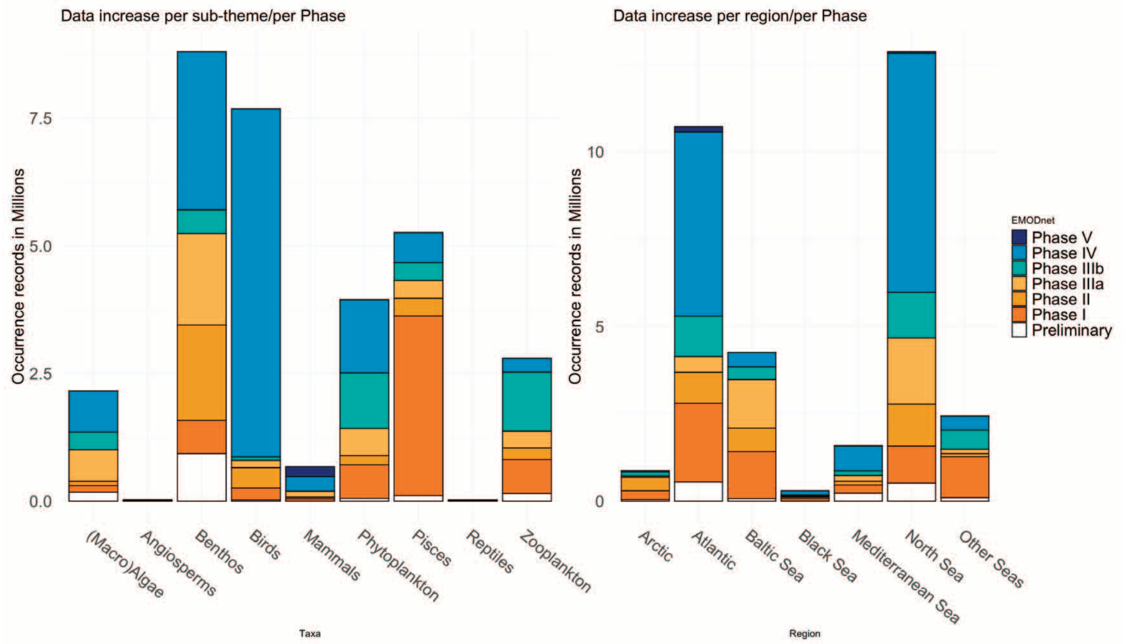


Figure 1 Data increase during each Phase per functional group (left) and European Sea region (right).

Outreach

An important line of work that has been maintained for a number of years consists in the sharing of our knowledge and experience by providing resources to the wider Marine Biology community and beyond. This work has been done by not only organising training workshops, developing an online and open 'Contributing datasets to EMODnet Biology' course, but also by a study on text mining software to capture ecological traits and sampling descriptors and developing R Packages (EMODnetWFS and EMODnetWCS) that allow for the search and download for all thematic lots data/products, using EMODnet's webservices. During Phase V we will continue with this line of work, by updating the existing data course, organising various workshops, creating training resources focusing on the R Packages and publishing various informative material.

Acknowledgements

This work has been financially supported by the EC DG-MARE (EMODnet European Observation and Data Network Lot n° I – Biology CINEA/EMFAF/2022/3.5.2/SI2.895681). The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) 2021/1139 of the European Parliament and of the Council of 7 July 2021 establishing the European Maritime, Fisheries and Aquaculture Fund and its predecessor, Regulation (EU) No. 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund.

Marine Litter data management in EMODnet Chemistry: how lessons learned in EU contribute to develop global-scale litter data management

Matteo Vinci[1] (mvinci@ogs.it), Maria Eugenia Molina Jack[1] (mmolinajack@ogs.it), Elena Partescano[1] (epartescano@ogs.it), Alexia Cociancich[1] (acociancich@ogs.it), Alessandro Altenburger[1] (aaltenburger@ogs.it), Alessandra Giorgetti[1] (agiorgetti@ogs.it), Francois Galgani[2] (francois.galgani@ifremer.fr), Amandine Thomas[2], (Amandine.Thomas@ifremer.fr), Georg Hanke[3] (georg.hanke@ec.europa.eu), Miquel Canals[4] (miquelcanals@ub.edu), Noriko Tamiya-Hase[5] (noriko_hase@env.go.jp), Luis F. Ruiz-Orejón[3] (luis.RUIZ@ec.europa.eu), Marco Palma[3] (Marco.PALMA@ext.ec.europa.eu), Artur Palacz[6] (a.palacz@ioccp.org), Jennifer Webster[7] (jennifer.webster@noaa.gov)

[1]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[2]Ifremer (France)
[3]JRC (Italy)
[4]Universitat de Barcelona (Spain)
[5]MOEJ (Japan)
[6]Institute of Oceanology, Polish Academy of Sciences (Poland)
[7]National Oceanic and Atmospheric Administration (USA)

Since 2016, the EMODnet Chemistry thematic lot has been actively working on marine litter data management at the EU level, according to the FAIR principles, with the main goal of supporting the implementation of Descriptor 10 of the Marine Strategy Framework Directive (MSFD). The National Oceanographic Data Centres in the EMODnet consortium are collaborating in coordination and agreement with the most relevant EU stakeholders and communities involved in this topic like the Joint Research Centre (JRC), the EU MSFD -Technical Group on Marine Litter (TG ML), the Regional Sea Conventions (RSC) and the Member States (MS). The topics addressed initially were coastline litter (beach), seafloor litter (from trawling), and floating microlitter. Lately, sediment microlitter was added to the data management workflow.
The consolidated pillars of this data management infrastructure relate to the following core components:

- the definition, sharing, and continuous update of agreed and harmonized guidelines for data formatting, based on the already consolidated community best practices and protocols when available;
- the setup and continuous improvement of a strong, interactive and efficient pre-ingestion data validation phase to ensure data quality, reliability over time and the streamlining of the data management workflow;
- further ad-hoc post-ingestion quality controls such as specific queries and dedicated GIS geoprocessing;
- the data-products generation workflow, which provides an additional check on data quality by detecting potential anomalies and a means of highlighting specific data features;
- close collaboration within an expert community in order to improve data quality and comparability through iterative feedback mechanisms.

In the last years the project focused several efforts on the development of tools and workflows able to generate marine litter aggregated collections, similarly to other topics managed by the consortium (e.g., eutrophication and chemical contaminants). This process, besides the publication of relevant datasets of interest for the stakeholders, represents a

further quality control step, providing further reliability and comparability to the published data. The participation of the consortium, with its data management expertise, in the Euroqcharm project represents another added value to the last year's activities. Euroqcharm is an EU Coordination and Support Action involving a set of high-level laboratory experts, with an ambitious goal of revising and harmonizing plastic pollution measurement methods. This subject is of the greatest relevance for the EU-TG ML and for the management of plastic pollution data globally. One of the main outcomes of this collaboration, as already highlighted in the interactions with EU-TG ML for baselines calculations, is the confirmation that "data management can provide a harmonized common framework for data reporting but it cannot alone address the imbalance or heterogeneity due to non-comparable monitoring methods".
Seafloor observation and imagery represent a novel approach for macro litter data collection, which brings a way forward for assessments and data management at EU and global levels. EU seafloor trawling data alone did not fully match the requirements for monitoring with large-scale coverage and the required comparability of data. In addition, trawling monitoring surveys can be used in rather limited areas and not by all countries; this could be another limitation. A combined monitoring approach integrating seafloor images and trawlings, where these are available, could be an efficient way forward.
The above points, resulting from the EMODnet Chemistry experience, constitute a legacy for active interactions with global stakeholders.
Over the past 5 years, a very relevant joint collaboration at the global level has been addressing the objective of contributing to the international legally binding instrument on plastic pollution promoted under UNEP, including in the marine environment. The G20 Action Plan on Marine Litter, coordinated by the Ministry of Environment of Japan (MOEJ), UNEP's Global Partnership on Plastic Pollution and Marine Litter (GPML) and the Integrated Marine Debris Observing System (IMDOS) global community are the main initiatives supporting this important goal.
The experience of EMODnet Chemistry at the European level provides a good baseline for the above mentioned international effort on plastic pollution. Participating in GPML Communities Of Practice, sharing available data and products with MOEJ to develop a global prototype for microlitter data management, and with IMDOS supporting their work and other initiatives advising as required (e.g., on the design of the Essential Ocean Variable for Marine Debris), the ultimate target of all these efforts is getting closer.

10 years of being a national research infrastructure - NMDC (Norwegian Marine Data Centre) providing seamless access to marine data

Helge Sagen (Helge.Sagen@hi.no)

Institute of Marine Research, Norway

Multidisciplinary data

Marine research is by nature multi-disciplinary, combining physical, geological, chemical, and biological knowledge and data. High quality and efficient marine research require easy and rapid access to marine data across institutions and disciplines. Gathering information about the marine environment is very expensive due to the dependence on vessels with high investment and operating costs. New technological developments allow a broader approach to sampling of the ocean using buoys, drifters, and various bottom mounted equipment, but these are also generally expensive to develop and operate. Partly due to cost, information about the sea is extremely limited compared to our knowledge about ecosystems on land. The relative scarcity of marine data makes it vitally important to make the data that does exist easily accessible for efficient usage in science.

Established national research infrastructure

An integrated infrastructure NMDC addressing interoperability, data documentation, archiving, providing data access has been running for more than 10 years, 2012 - 2023. The infrastructure simplifies the technical obstacles scientists encounter when finding and using data from various sources. In addition, the infrastructure provides a cost-effective interface for data providers. NMDC is ensuring proper stewardship for data, both in the short and long term, uncovering the potential hidden in historical data sets by documenting and storing them in a long-term archive. The project has undertaken efforts in data archaeology to mobilize data that was previously unavailable before the establishment of the research infrastructure.

Standardization of data

Establishing NMDC has shown the importance of standardization of data collection processes, data storage/exchange formats e.g. The marine data at each partner institution is managed differently. The NMDC infrastructure helps making data available and ensures that data undergo a standardization process including a metadata enriching process. The 16 partners have successfully worked together in making this possible and all partners have made their data go through a standardization process before being released in the infrastructure. The work has identified a long list of possible data that could be made available using the infrastructure. Currently more than 3000 datasets are available in the metadata catalogue.

Datset search

NMDC offers a search facility on the webpage www.nmdc.no being able to filter on geographical area, time, science keywords, data providers or free text. The result list is presented with the title and access to detailed information about each dataset. Requesting details about one specific dataset will activate the landing page of the dataset. The landing page is a formatted web page with the metadata of the dataset and necessary download information to get the data itself. It is possible to add datasets in the search list to a basket to make it easier to download more than one dataset at the time. The basket offers all datasets in one compressed file for easy download.

Trusted repository and accredited NODC

NMDC is a joint effort of Norwegian institutions and universities, coordinated by the Institute of Marine Research (IMR), to build a national research infrastructure for seamless access to documented research data. The data center at IMR holds the certificate Core Trust Seal to ensure good data management practices and became an accredited International Oceanographic Data and Information Exchange (IODE) National Oceanographic Data Centre (NODC) in 2023. The establishing of NMDC is funded by the Norwegian Research Council. The infrastructure consists of distributed data nodes that provide a local node with metadata, which is used for discovery of relevant data sets. The information about all the data sets available within NMDC is accessible through a web portal, where researchers or other end users can search for relevant data. The NMDC data portal supports the usage of Digital Object Identifiers, (DOI).

Further developments

An upgrade and further development of the infrastructure is planned to be in the forefront of the technical developments and advancements. The proposal for Norwegian Marine Data Space (NMDS) is under evaluation at the national research council funding for 2024 for a new 10-year period 2025 - 2034.

H2020 Eurofleets+ Data Management: an integrated data stewardship approach across multidisciplinary Transnational Access cruises

Hong Minh Le[1] (hmle@naturalsciences.be), Anastasia Iona[2] (sissy@hnodc.hcmr.gr), Elena Partescano[3] (epartescano@ogs.it), Dick Schaap[4] (dick@maris.nl)

[1]Royal Belgian Institute of Natural Sciences (Belgium)
[2]Hellenic Center for Marine Research (Greece)
[3]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[4]MARIS (The Nederlands)

Introduction

Eurofleets+ is an INFRAIA H2020 project started in 2019 and running over 57 months. Eurofleets+ (EF+) facilitated open free of charge access to a unique fleet of state-of-the-art research vessels from European and international partners which has been arranged by competitive transnational access (TA) calls. A Data Policy has been defined and adopted and EF+ project applies the FAIR and Open research data principles of the H2020 framework. Therefore, marine Data Management (DM) is an integral part of the EF+ approach and is implemented in synergy with SeaDataNet (SDN), a leading pan-European infrastructure for marine data management, involving NODCs (National Oceanographic Data Centers) as core partners. The EF+ project funded 28 cruises from which were collected data across various oceanographic disciplines and around several world oceans and seas. A total of 164 datasets have been collected with the aim of making them FAIR. Three NODCs (HCMR, OGS and RBINS) gave guidance and support to the Principal Investigators (PIs) for arranging that generated metadata and data are preserved, validated, harmonised and become populated in SDN infrastructure for re-use. The project highlighted the complementarity between the scientific expertise providing meaningful data and the experience of the NODCs of leading data stewardship in order to produce trustworthy open research data.

Data management approach

In order to assist the PIs in managing the collected data following FAIRness, creation of a Data Management Plan (DMP) was required prior to the cruises with detailed information on: the data planned to be collected, how to ensure its curation, preservation and sustainability and the data access license and embargo period. After the cruises, the Cruise Summary reports (CSR) were published to SDN CSR catalogue. Within 2 months after the cruise or as soon as available for samples requiring lab work, the data were gathered and deposited to the assigned Responsible Data Center (RDC). The RDC proceeds to the (meta)data curation and publication on European oceanographic data repositories (i.e., EMODnet Data Ingestion Portal, SDN CDI Service and SEANOE), in accordance with data usage license and moratorium defined in the DMP. Deposit is asked to be done by submitting the data and a minimum set of metadata to the EMODnet Data Ingestion Portal (DIP) which plays a role in EF+ as a central mailbox. Assigned RDC completes the submission with additional metadata for “as-is” publication (Phase I). The data is then further elaborated, where possible, with validation and harmonisation by the assigned RDC to generate CDIs, Common Data Index entries, published on the SDN service as Phase II. SEANOE is recommended for publication of ‘exotic’ datasets (e.g., eDNA, pictures) and/or to get a DOI.

Synthesis of the TA cruises data management

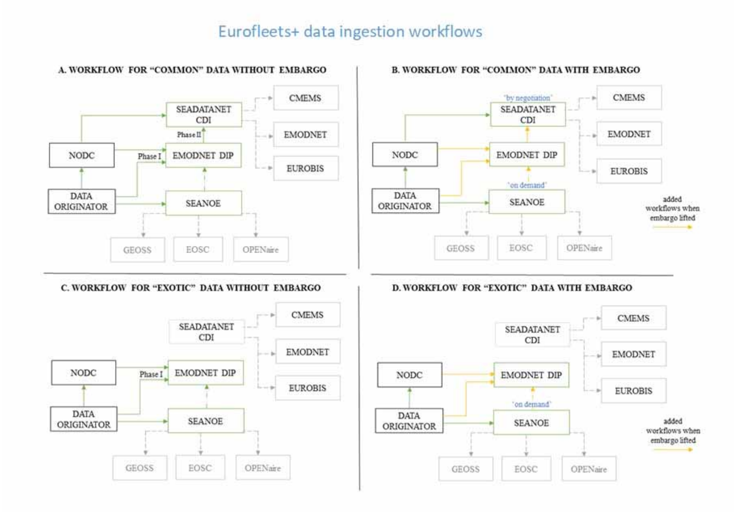
Over the 28 cruises, a total of 164 datasets have been generated. By the end of the project, 66% of the datasets have been deposited to the RDCs for preservation which represent more than 5.7 Tb of data files, mainly of geophysical data type. 10% of the datasets could not be transferred because they were still under analysis (lab work, post-processing, quality checks) but PIs committed to deliver them when they would be processed even beyond the project end. COVID situation was another reason that the data could not be delivered in time. RDCs will follow up with the PIs for remaining datasets. All PIs opted for open access licence which reflect their will to share the data. However, 80% of the PIs asked for an embargo period (up to 2 years). Eventually, around 25% of the datasets were published openly to EMODnet DIP and almost 15% were published on SEANOE. 30% of the total datasets have been curated and populated into the SDN CDI Service which corresponds to almost 4000 CDI entries generated. Monitoring descriptors for FAIRness have been described and EF+ DM led to 43% of the generated datasets to be published with enriched metadata on repositories assigning unique and persistent identifier which makes them Findable and eventually Openly Accessible after embargo end. 30% of the collected datasets were published following standard (meta)data transport format and using common vocabularies making then Interoperable and Re-usable at the end of their moratorium. Given the scale of the project, the amount of data generated and the time scale involved achieving present scores is considered as a very reasonable rate.

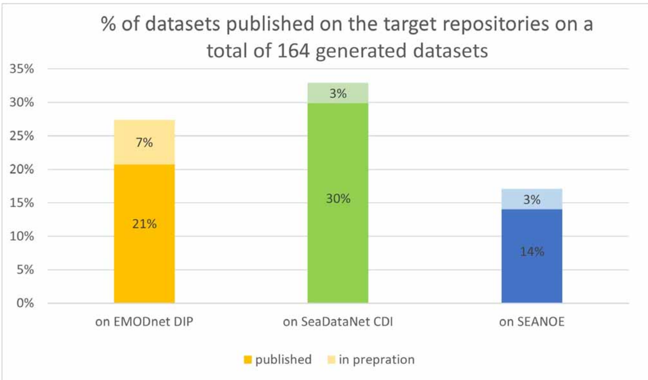
Lessons learned

The cruise DM process is driven by the involvement of the PIs (i.e., data-originator) and the RDCs (i.e., data managers). Where it has been observed that DM is not yet part of the scientific mainstream, clear and effective guidance should be provided by the RDCs. Communication is a key component of the collaborative work to provide trustworthy data for sharing and re-use. The RDC cannot do without the expert opinion of the data originators to assess the consistency and quality of the data nor to get the necessary meta-information to elaborate harmonised data publication. Assistance from RDCs is needed to support scientists in achieving data FAIRness by offering guidance during the creation of a DMP, by giving trainings, by explicating the workflows and the required (meta)data. It is essential that scientists provide well documented data (including lineage), processed rather than raw data and quality checked data. Effective DM needs a pro-active follow-up from RDCs as scientists are often mainly absorbed with their research.

Conclusion

The singularity of EF+ project is that its DM is centralized and coordinated by RDCs for all and each of the TA funded cruises in collaboration with the PIs and scientific crews. DMPs have been prepared prior to the cruises, CSRs created and collected data have been gathered, validated and harmonised for publication on the pan-European SDN platform, as much as possible. The DM progress and the EF+ monitoring indicators of FAIRness reached was what could be expected to be achievable given the scale of the project, the amount of cruises to follow-up, the quantity and diversity of data generated and the time scale involved. EF+ is under this perspective an ambitious pilot project as it deployed an integrated and uniform DM approach on a multiplicity of research projects with their own objectives, challenges and timescales. Main lesson learned is that successful DM is the result of a close cooperation between scientists and data managers as they are complementary. Achieving data FAIRness means bringing awareness to scientists on DM good practices and fostering collaboration and communication between data managers and data-originators.





Reference

Le H.M., Iona S., Partescano E., Schaap D., (2023). *Synthesis of the data management follow-up reports, lessons learned and recommendations*. Deliverable D4.15. Project Eurofleets+: An alliance of European marine research infrastructures to meet the evolving requirements of the research and industrial communities.
https://www.eurofleets.eu/project-information/deliverables/

EuroGOOS Data Policy - enhancing FAIR data for ocean science and operational oceanography

Thierry Carval[1] (Thierry.Carval@ifremer.fr), Dina Eparkhina[2] (dina.eparkhina@eurogoos.eu), Juan G. Fernández[3] (jfernandez@socib.es), Vicente Fernandez[4], Antonio Novellino[5] (antonio.novellino@grupposcai.it), Dominique Obaton[1] (Dominique.Obaton@ifremer.fr), Simona Simoncelli[6] (simona.simoncelli@ingv.it), and the DataMEQ Working Group

[1]Ifremer (France)
[2]EuroGOOS (EU)
[3]SOCIB (Spain)
[4]EuroGOOS (EU)
[5]ETT Solutions S.p.A. (Italy)
[6]Istituto Nazionale di Geofisica e Vulcanologia – INGV (Italy)

Operational oceanography brings various economic and societal benefits (EuroGOOS, 2016). It underpins significant decision and policy making in the marine domain, from ocean forecasts to maritime activities, search and rescue operations, ocean and climate modelling, and delivers vital services and information for a sustainable blue economy. It has been widely recognised that oceanographic data and metadata must be shared in a free, open, and timely way. This is key both for the co-production of operational oceanographic products and services and for the advancement of science and ocean knowledge. However, there are still gaps and barriers preventing open and timely marine data circulation.
The Intergovernmental Oceanographic Commission (IOC) of UNESCO, recognising the vital importance of timely, free and unrestricted international exchange of oceanographic data for the efficient acquisition, integration and use of ocean observations gathered by the countries of the world, in 2019 endorsed the IOC Oceanographic Data Exchange Policy. At its 32nd Session in June 2023, the IOC Assembly adopted the new IOC Data Policy and Terms of Use (2023).

The EuroGOOS Working Group on Data Management, Exchange, and Quality (DataMEQ) has developed a new EuroGOOS Data Policy to support the implementation of the IOC-UNESCO Data Policy at European level. The EuroGOOS Data Policy will help to establish or improve the European marine data circulation according to the principles of Findable, Accessible, Interoperable, and Reusable (FAIR) marine data and metadata. The EuroGOOS Data Policy recommends the incorporation of Data Management Plans that ensure the adoption of FAIR principles from the early stages of data production.
EuroGOOS recommendations include that core European marine data must be shared with clear, unambiguous data licences such as the Creative Commons CC-BY licence. The core data must include, at the minimum, the physical and biogeochemical Essential Ocean Variables (EOVs) needed for the Copernicus Marine Service, the European Marine Observations Data Network (EMODnet), the EuroGOOS Regional Operational Oceanography Systems (ROOS), and SeaDataNet. The data must be shared with comprehensive and standardized metadata information acknowledging and citing the data providers, funding programmes, and investigators.
Implementing the EuroGOOS Data Policy implies continuous FAIRness improvements among:

- Data formats, vocabularies and community-networks standards.
- Best practices for observation networks (quality control, versioned data and metadata).
- Online data services for human and machine to machine access (such as OpenAPI or ERDDAP), newer open and cloud based technologies.

- Ensure long term archive in certified repositories.
- Promote data lineage taking advantage of Persistent and globally Unique Identifiers (PIDs): from the data provider to the data aggregator and vice versa.

This presentation will introduce the EuroGOOS Data Policy and offer specific recommendations aimed at enhancing ocean data circulation and utilization in support of European and global policies, addressing societal challenges linked to the ocean.

Unlocking the Depths: ODIS Initiative and Polish Contributions to Global Oceanographic Data

Marcin Wichorowski[1] (ostra@iopan.pl), Lena Szymanek[2] (lszymanek@mir.gdynia.pl), Urszula Pączek[3] (upac@pgi.gov.pl), Michał Wójcik[4] (Michal.Wojcik@im.umg.edu.pl), Krzysztof Rutkowski[1] (kr@iopan.pl), Wojciech Paciura[3] (wpac@pgi.gov.pl)

[1]Institute of Oceanology Polish Academy of Sciences (Poland)
[2]National Marine Fisheries Research Institute (Poland)
[3]Polish Geological Institute National Research Institute (Poland)
[4]Maritime Institute Maritime University in Gdynia (Poland)

Establishment of the ODIS infrastructure to enhance internal collaboration and promote FAIR (Findable, Accessible, Interoperable, and Reusable) data principle

Polish organizations engaged in oceanic research and exploration are active participants in international initiatives aimed at developing oceanographic data resources. As custodians of a substantial volume of oceanographic data, these organizations seek to make it accessible to a broader audience, including researchers, the industrial sector, and interested citizens. To enhance Poland's contribution to pan-European and global oceanographic data resources, these organizations have established the Oceanographic Data and Information System (ODIS) infrastructure to foster international collaboration and make Polish scientific data compliant with FAIR principles. In Figure 1, we present the web portal of Polish ODIS developed by consortium of Polish scientific institutes and universities conducting oceanographic research.

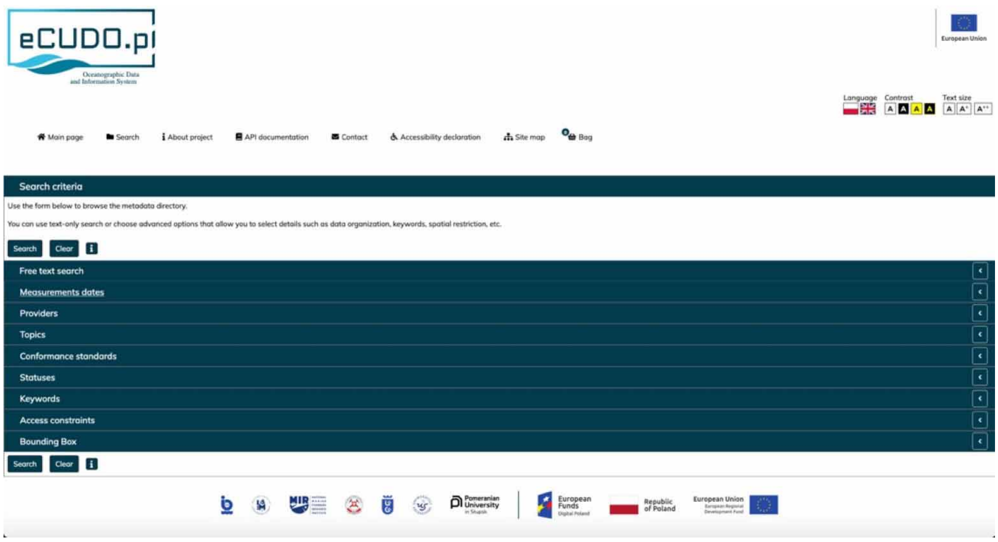


Figure 1 Web portal of Polish ODIS.

Consortium and National Laboratory

The consortium, consisting of the Institute of Oceanology Polish Academy of Sciences, Polish Geological Institute National Research Institute, National Marine Fisheries Research Institute, University of Gdańsk, Maritime Institute of Maritime University in Gdynia, Pomeranian Academy in Słupsk, and University of Szczecin, collectively immersed in marine research for an extended period, has consolidated efforts to make Polish oceanographic scientific data accessible to the public through a national repository – ODIS for Polish Scientific Data, available online at https://odis.ecudo.pl. Indexing over 7 million datasets derived from Polish scientific activity in the Baltic Sea and Polar Regions, this initiative is actively working towards deploying an operational state of the system to provide demanded oceanographic data and products to users, aiming to establish the Polish Oceanographic Data Committee. The dataset content, as well as the data format is variable according to domain specific definitions and practices. ODIS is capable to index documents in range containing single observation, through casts containing hundreds of measured values, to complex raster structures containing over milion of points.

eCUDO Platform Services

The project embraces a commitment to openness, rated at 5 stars according to the "5 Star Open Data" scale. As part of this commitment, a system will be developed to grant users access not only to current, open, and searchable data but also to connect with other open collections. The data will be described with metadata compliant with geospatial data standards, including ISO 19115-1:2014. Each dataset will be tagged with unique URIs, and metadata will contain appropriate links to other resources on the web. The eCUDO platform offers users various services through interfaces implemented on the platform, including:

- CSW (Catalog Service for the Web) - a metadata catalog service that allows you to download records that properly describe the data and search them based on a given criterion, compliant with the OGC (Open Geospatial Consortium) standard.
- WMS (Web Map Service) - a service that allows you to share maps in raster form, compliant with the OGC standard.
- WFS (WebFeatureService) - a service that allows you to share geographic features, compliant with the OGC standard.
- ExtendedSearch - a service that allows you to search metadata based on specific attributes.
- DataManagement - a service that allows you to enter, delete and update metadata.
- UserManagement - a service that allows you to manage users, allows you to add and remove users in federated systems.
- LinkedData - a service that allows you to share data and their description in the LinkedData format.
- AuthService - a service that allows authentication of the platform user.

The platform relies on the robust infrastructure provided by the Jupyter Hub environment, a versatile and interactive computing framework. This environment enables users to engage in thorough data exploration, leveraging the capabilities of the Python, Julia, and R programming languages. The incorporation of these languages offers a diverse and powerful toolkit, fostering advanced analytics, statistical modeling, and computational research within the platform. Users can harness the flexibility and efficiency of these programming languages to conduct a wide range of data-driven tasks, thereby enhancing the depth and breadth of their analytical endeavors on the platform.
Harmonization of environmental data collection in line with INSPIRE requirements and SeaDataNet standards.
Present activities focus on harmonizing environmental data collection and preserving it in line with INSPIRE requirements and SeaDataNet standards. This approach enable possibilities of continuous integration of these data with the European SeaDataNet network by which the Polish data will also become available for EMODnet and other European and international initiatives. Additionally, efforts are directed towards securing resources for data management and stewardship, along with the digitalization of hardcopy data archives. Anticipated outcomes include improved data discovery, findability, accessibility, interoperability, and a higher potential for the reuse of data collected over years of research activity.

The system's design aligns with program requirements, emphasizing user orientation and the development of services based on user demand. ODIS is open to all stakeholders and is prepared to accommodate other organizations and data resources.

Shaping an Ocean of Possibilities for science-industry collaboration – The data management behind an innovation platform

Linda Baldewein[1] (linda.baldewein@hereon.de), Carsten Schirnick[2] (cschirnick@geomar.de), Angela Schäfer[3] (Angela.Schaefer@awi.de), Sylvia Reißmann[2] (sreissmann@geomar.de), Marc Adolf[2] (madolf@geomar.de), Ulrike Kleeberg[1] (ulrike.kleeberg@hereon.de)

[1]Helmholtz-Zentrum Hereon (Germany)
[2]GEOMAR Helmholtz Centre for Ocean Research Kiel (Germany)
[3]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research (Germany)

The Helmholtz funded innovation platform SOOP aims to create an ocean of possibilities for science-industry collaborations building up sustainable structures for transfer and innovation in ocean observation by creating a paradigm shift towards a multi-sector integrated observing system. In SOOP, small modular plug and play sensor systems targeted at non-scientific sea-going vessels to be operated by laypersons are co-developed with industry partners. Access to ubiquitous ocean data is enhanced by equipping for example small private sailing boats with sensor systems, facilitating their data stream on shore, collection, and provisioning of acquired raw data and merging the results in public dashboards to foster the understanding and safeguarding of our oceans and seas.

Within the SOOP platform, the data management team develops and implements a concept for enabling their research infrastructures to host non-scientific partners, such as citizens, companies, and public authorities, regarding storage capacity, mechanisms to assess and control data quality and FAIRness of the resulting data services. We follow international standards and best practices where available or contribute to their development.

The newly developed small sensor systems are categorized and their metadata pre-registered via templates in the sensor REGISTRY. When an end user buys a system, its initialisation is carried out via automated workflows for metadata registration (e.g., scanning QR-code or NFC). A universal unique identifier complementing a not-so-universal serial number is registered, optional metadata information such as the name and owner of the vessel may be added while indispensable technical metadata of the sensor system is already merged from a matching template.

Once the device is operational, it can be monitored and controlled through BELUGA, e.g., in regional campaigns, and necessary software or operational parameter updates can be carried out. Data transfer in real-time or delayed mode is warranted using the secure pathways via Virtual Private Network using the HELMI solutions or retrieved from secure service providers, such as LoRaWAN networks with trusted data brokers implementing HELMI as a concept.

The data are processed and may optionally be automatically quality controlled according to best practice Standard Operation Procedures. Afterwards, near real-time data are ingested based on predefined structures into time-series databases. Curated datasets designed for scientific applications are citable published via DOI in performant, reliable data and metadata storage systems. All data are accessible via the SensorThings API. Using this Internet of Things Open Geospatial Consortium standard, data streams from different sources rely on common vocabularies and can be merged e.g., to be visualized in dashboard solutions: Specialized for the owner of a sensor system for real-time monitoring as well as general dashboards for the civil society and special interests. Data streams are included in the central data portal of participating research centers and can be integrated in international data portals.

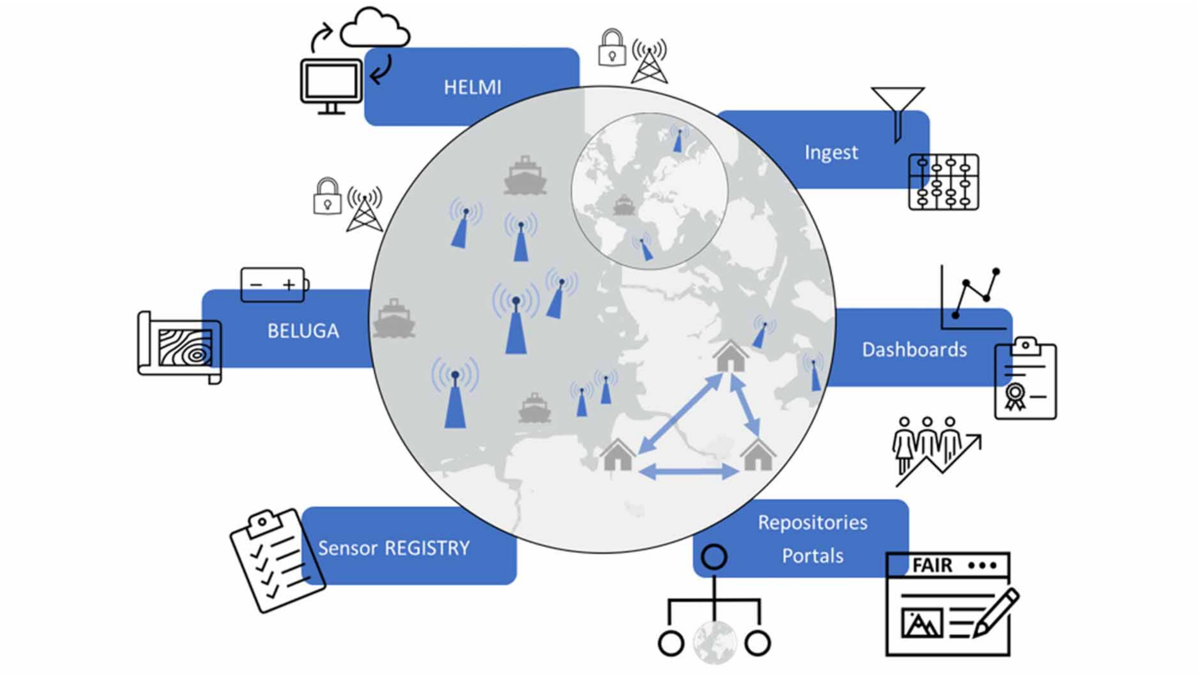


Figure 1 Data management overview diagram of SOOP innovation platform – New sensor installations are registered in the REGISTRY, can be monitored, and controlled via BELUGA, use HELMI as a concept for trusted data transfer on shore, and seamless ingest into databases, dissemination, and consumption for any purpose.

The whole workflow will be generalized and published as a concept of an open-source data platform. It shall be used as a blue-print for companies, public authorities and other research centers on how to manage the data workflow from decentralized sensor platforms. A proof-of-concept implementation within the Helmholtz MareHub and German Marine Research Alliance marine data portal is envisioned at the end of the initial phase in 2025.

A first a roadmap for data management solutions of low-cost ocean observing technology

Patrick Gorringe[1] (Patrick.Gorringe@smhi.se) Antonio Novellino[2] (antonio.novellino@grupposcai.it), Peter Thijsse[3] (Peter@maris.nl)

[1]Swedish Meteorological and Hydrological Institute - SMHI (Sweden)
[2]ETT Solutions S.p.A. (Italy)
[3]MARIS (The Netherlands)

Whether you're a researcher looking for alternatives to expensive scientific equipment, a citizen scientist group interested in building a marine monitoring program, or an ocean enthusiast seeking new ways to interact with the sea, this presentation is for you.

The technical tools necessary to study, explore, and understand the ocean are often inaccessible to the vast majority of ocean users. Financial and technical barriers severely limit access to observing the oceans. Many countries still face major challenges when it comes to observing and monitoring the ocean and accessing equipment and information. This often means knowledge of the seas and oceans depends on the amount of available financial resources. Low-cost sensor technology for marine observation is becoming more and more available, but still lacks standard operating procedures or best practices, as well as the widespread belief that high-end equipment and facilities are necessary. This blocks the opportunity for widespread uptake. At the same time, in more developed countries, it is necessary that our scientific communities, especially those specializing in instrumental research, question their resource consumption and environmental impacts associated with new technological developments. "Low-cost", cost-effective, cheaper technical sensor solutions, coupled with building capacity in communities of ocean stakeholders, can substantially broaden participation in ocean observation and research. Thus, we want to overcome the indicated barriers by nurturing a community of open-source hardware and software developers, scientists, and ocean stakeholders. Not only by creating a platform for technical concept developers to register themselves, but also developing the community which will work on best practices for deployment as well as data management (from the source to the end-users at EU/global level). Because even though there is a dramatic rise in interest and development of cheap ocean sensor technology, it still lacks data pipeline solutions to get these potentially transformative tools and the data into the hands of the people who need them most. Data collected by these new technologies are often not yet able to be submitted to national marine data centers or ingested to European data aggregators such as the European Marine Observation and Data Network (EMODnet) with insufficient metadata, resulting in a lack of data availability, loss of data and reduced data availability.

This presentation is geared towards providing an overview of a selection of cheap, low-cost, cost-effective, sensor technologies. We will propose a roadmap of low-cost technology goals for sustained ocean observation, including data management solutions and data services. Data and meta-data standards are key for facilitating data usage and interoperability where we aim to maximize the data value from low-cost sensors, establish best practices around deployment, data collection and management, from the sensor to data flow, data formats etc. in order to achieve FAIR (Findable, Accessible, Interoperable, and Reusable), along with CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics), when applicable, data and facilitate observation uptake.

A roadmap that will be supported by work in the recent JERICO-S3 and LandSeaLot EU-projects, where the data flow together with best practices from low-cost sensor technology will be developed and demonstrated, as well as EMODnet and with global energy from UN Decade of

Ocean Science programs and projects. The work done towards building the framework for an interactive, searchable directory of low-cost technologies for the greater ocean community will be presented with the overall goal to connect the ocean community and set forth an executable, long-term vision for cheap ocean observing technology.

Sharing how data and evidence underpins the UK's leading role in offshore wind to enable net zero, energy security and nature recovery

John Mitchell (john.mitchell@thecrownestate.co.uk), Chelsea Bradbury (Chelsea.bradbury@thecrownestate.co.uk)

The Crown Estate (UK)

The UK seabed holds some of the richest resources in the world, underpinning the UK's position as a global leader in offshore wind. The Crown Estate is responsible for sustainable development of critical sectors in the waters around England, Wales and Northern Ireland. This places The Crown Estate in a unique position to look at the bigger picture across a breadth of offshore sectors, and data is key to unlocking key consenting and planning challenges; ultimately enabling net zero, energy security and nature recovery.

The Crown Estate have been investing in data and evidence for over 20 years to help to balance competing demands on the seabed, but we know that we need to achieve more in the energy transition in the next ten years than we have in the last 20. The seabed and waters around the UK will get busier, and it is critical that development decisions are based on data and evidence to maximise growth in secure, affordable green energy, whilst also protecting and restoring the UK's thriving marine ecosystems. This is a collective challenge and one that cannot be achieved in isolation. We need to ensure that we're collaborating on a scale never seen before to share knowledge, data and insights.

Access to data is under the spotlight across the globe and The Crown Estate want to ensure that we are building upon the lessons learnt of the UK who have a world-leading model for industry data sharing.

The Crown Estate have required offshore wind developers to share data since 2003 by including a data clause in all projects around the UK. As a result of working with industry to understand barriers to data sharing, the Marine Data Exchange (MDE) was developed in 2013 as a bespoke platform and the first of its kind to manage multidisciplinary offshore survey data collected throughout the lifetime of offshore wind farms, from feasibility studies through to decommissioning or repower. As we celebrate over 10 years of the MDE, its data holding has developed significantly to having almost 300TB of survey data from over 50 offshore projects, and 60% of that data is freely available for others to access and use, with the rest of the data holding published as projects move out of sensitive phases such as consenting.

Whilst the MDE seeks to promote the re-use of survey data to inform evidence based decisions, it also strives to make it simple for offshore industries to provide survey data without processes and data requirements becoming overly onerous. Over the years The Crown Estate have worked with the offshore industry to define and to develop the right balance in terms of data requirements and the adoption of data management standards.

Through the MDE, the work of The Crown Estate and the willingness of industry to share data, barriers to industry sharing data have been identified and largely removed in the UK. Like the industry though, our data requirements and approach to managing the confidentiality of data, continue to evolve and adapt.

Whilst we have the trust of industry and a platform that is enabling industry to share data, we can't afford to slow down. Technology is advancing, the challenges facing industries are changing and the data and evidence that we hold on the MDE can help to accelerate the sustainable development of the seabed, whilst also answering some of the wider research questions that can help us to protect the marine environment.

The focus is now on what we do with industry data and how we enable others to use it to drive forward innovation.

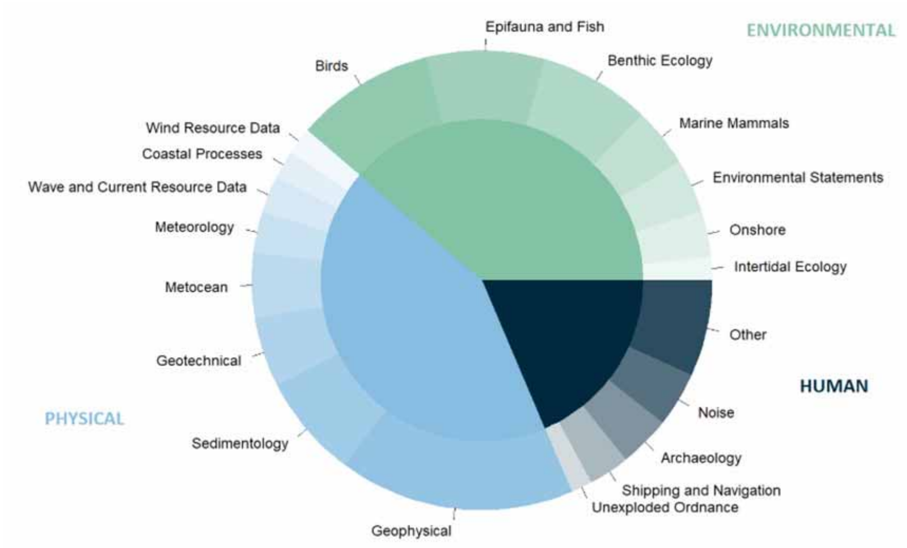


Figure 1 Survey data themes held on the Marine Data Exchange.

SESSION INFRA POSTERS

From deep sea to European servers: toward an integrated data flow to ensure traceability of biological resources

Clémence Cotten (clemence.cotten@ifremer.fr), Lénaick Menot (lenaick.menot@ifremer.fr), Brendan Hennebaut (brendan.hennebaut@ifremer.fr), Jérôme Detoc (jerome.detoc@ifremer.fr), Sylvie Van Iseghem (sylvie.van.Iseghem@ifremer.fr)

Ifremer (France)

The deep sea has suffered for a long time from a lack of knowledge due to its difficult accessibility. Today, thanks to increasingly sophisticated underwater systems, it is possible to operate at depths of several thousand meters and bring up samples to better understand these ecosystems. Ships in the French oceanographic fleet are equipped to operate manned submarines (Nautile) and ROVs (Victor 6000, Ariane) capable of descending to depths of six thousands meters.
In a context where the deep sea is increasingly threatened by the exploitation of living, energy and mineral resources, improving our understanding of deep-sea ecosystems is crucial for their protection.
Deep sea ecosystems are among the most fascinating on the planet. New species, perfectly adapted to this hostile environment, are discovered on a regular basis. How are organisms able to live at such depths? Through what physiological adaptations? Thanks to the biological samples brought up during these cruises, scientists are trying to answer these questions, and in so doing, improve our understanding of life.
It is essential to ensure that the data acquired during these cruises are stored in a database in a sustainable way, as they are both costly to acquire and highly accurate. In addition, it is important to provide to national and international scientists an open and easy access to the data for their research. To ensure the long-term preservation of data acquired on board of French oceanographic fleet ships, and to enable their export to European platforms, Ifremer has equipped itself with several tools organized into a data flow.

On-board acquisition
During the cruises, the scientists use on-board tool and database called *Sealog* installed on a fixed workstation. This does not require an Internet connection and can therefore be used easily in any part of the world. Sealog is fed by electronic logbooks, which are synchronised with ship navigation (Casino) or submarine engine navigation (Mimosa/Adelie). This on-board tool centralises information on navigation, sampling operations (dives and on-board) and physical samples. Each sample taken on board is labelled with a unique code and entered into the database. The label could be printed before cruises and a pre-filled spreadsheet make easier to associate a physical sample to its numerical equivalent and associated metadata.

Sample identification and local storage
On board or on return to the laboratory, the biological samples are taxonomically identified. The information linked to these samples (unique label codes, taxonomic information, biological traits, etc) are integrated into a LIMS (Laboratory Information Management System). This tool, which is accessible to all members of a laboratory, from anywhere, online, enables them to integrate easily a large number of samples (sometimes more than thousand for scientists working on genomic data).
LIMS *LabCollector* is currently being deployed at Ifremer and is highly popular with biology laboratories, whatever their research area. It is a tool that is easy to use and adapts well to the needs of the teams. However, LabCollector is not intended to replace the central systems in

place. The flexibility and the autonomy of laboratories in terms of data management mean that it cannot be used to ensure that data are stored in a standardised way over the long term.

Figure 1 Victor 6000 Remotely Operated Vehicle (ROV) @Ifremer, ChEReef 2022.

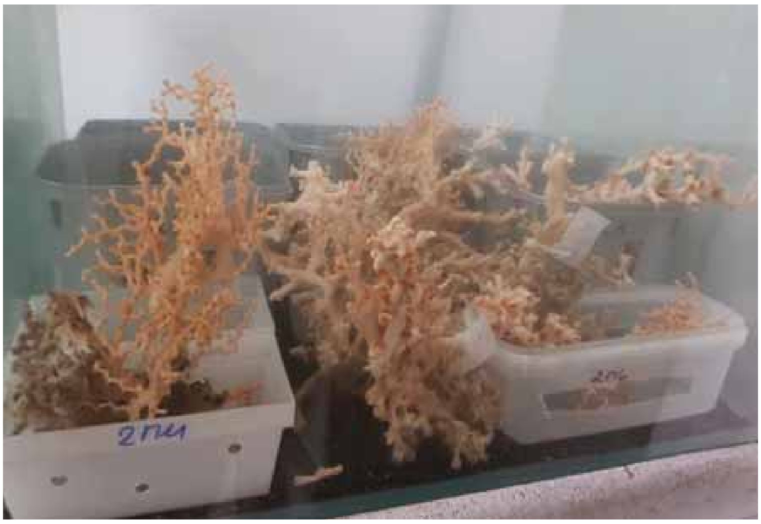


Figure 2 Coral samples from ChEReef cruise @Ifremer, ChEReef 2022.

Institutionnal storage of biological samples

On return from the cruises, the onboard database is updated to improve navigation information in particular. It is then integrated into Ifremer's sample database from oceanographic cruises, called BIGOOD. After a few months, once the taxonomic analyses have been carried out on the samples taken on board, Ifremer's team of data managers integrates the information from the

LIMS into the reference base French Oceanographic Campaign Samples BIGOOD using the unique code on the tags. It is then possible to trace the cruise detailled informations back to the taxon identified in the sample.

Export to International servers

Today, Biological data stored in BIGOOD cruise sample database are then manually sent to European infrastructures, in particular OBIS and GBIF. Works are ongoing first to ensure the exportation trough ISA standards (International Seabed Authority) and to use IPT (Integrated Publishing Toolkit) in order to improve the export trough International Data portals.

Example of ChEReef cruise

ChEReef 2022 is the second of six annual cruises The cruises are multidisciplinary, involving mooring and sampling operations both from the ship and from submarine engines. Some data from 2022 ChEReef cruise (*R/V Thalassa*) are presented as an illustration of the data flow.

Potential improvements

The data flow currently in place at Ifremer comprises several tools and requires numerous human operations on the part of the teams. In the future, we plan to develop a more integrated solution to facilitate the work of the scientists on board.

Moreover, the on-board database is quite time-consuming to use. There are also prospects of deploying the World Register of Marine Species (WoRMS) at each level of this data flow in order to facilitate export to European and international databases.

Next cruises, like Chereef 2023 will helps us improve our way of working.

Perspectives

Once the taxonomic analysis has been carried out, the samples are analysed in a variety of ways, including genomic bioinformatics analyses. The information previously acquired enables Ifremer's bioinformatics team to retrieve the metadata needed to make the genomic data they manage traceable and reusable, by ensuring a link between the sample and the genomic data.

In the current regulatory context, it is necessary to ensure that oceanographic campaigns (or projects) are in compliance with the differents regulations (APA, CITES, etc.). Improved management of data from biological samples and there traceability will enable us to improve Ifremer's responses to these international legislation. Our aim is also to enhance the reporting of data acquired on Ifremer mining permits to the ISA, using the BIGOOD database.

Copernicus Marine Service. Regional In Situ data production unit for the Arctic

Arnfinn Morvik (arnfinn.morvik@hi.no), Asuka Yamakawa (asuka.yamakawa@hi.no)

Institute of Marine Research (Norway)

Objective

The Copernicus Marine Service is based on a distributed model of service production centers, both for modelling and collection of in situ ocean observations. The service comprises of two kinds of production centres:

- In Situ Thematic Assembly Centers (INS TAC) – collection of ocean observations
- Monitoring and Forcasting Centers (MFC) – maintaining numerical models of the ocean

A Copernicus Marine Service INS TAC has two major deliveries:

- Near Real-time (NRT) data: Collect multi-source, multi-platform, heterogenous data, perform consistent automatic quality check in a common NetCDF format and deliver within 24 hours.
- Delayed mode data: Supply data with quality flags by human experts for the last 50 years.

Copernicus Marine Services' focus area is European waters including the North Atlantic, Mediterranean, Baltic, and the Black Sea. Therefore, INS TAC consists of 6 regional centers with overlapping areas and one center covers the global ocean. Figure 1 shows the geographical extent of the areas, and the Arctic region covers all data collected north of 62 degrees.
The Norwegian Marine Data center (NMD) at the Institute of Marine Research (IMR) in Bergen is one of the 7 regional components of the INS TAC and is responsible for delivering products from the Arctic region.

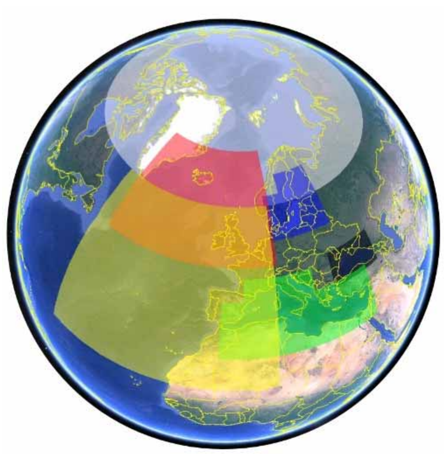


Figure 1

Data collection

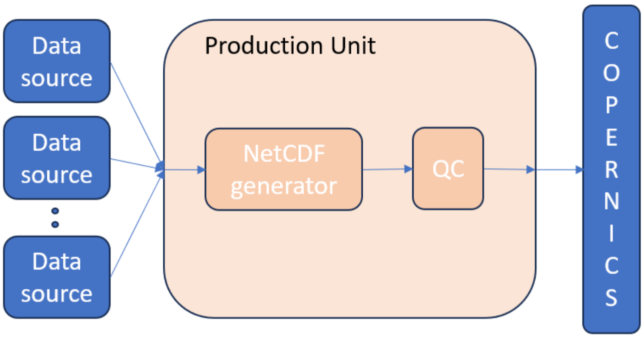
The Arctic is an area with rough weather conditions and big communication challenges. This reflects the availability of NRT data in the area. NRT data are collected from IMR research vessels, buoy data from various sources, and glider data from collaborated institutions.
In addition to NRT data, NMD seeks various data repositories publishing data from the Arctic region, both from Norwegian and international institutions. These data are reprocessed to the common NetCDF format and published as delayed mode data. NRT data are republished as delayed mode data when additional quality assurance has been performed at the originating institution.

Data production

NRT data arrives at the production unit from several different instruments and at different time intervals. This can vary from daily to several times each hour. The method for retrieving the data may also vary. Data can be pushed to the production unit, or the production unit downloads data from the data provider.

Figure 2 shows the data production chain at IMR. Data are retrieved from various sources in various formats and are fed into a NetCDF generator. This software generates files in NetCDF compliant to the Copernicus Marine In Situ NetCDF Format specification. Before the files are pushed to publication, NRT automatic QC is performed, and the data is flagged according to this procedure. To make sure that data is produced as frequently as possible, this procedure is repeated every hour.

Figure 2



Data publication

Data from Copernicus Marine Service INS TAC is collected from many different platforms/instrument types such as vessel CTD, gliders, moorings, etc. and the data available are grouped based on individual platforms. For example, CTD data from two different vessels are available in separate files.

NRT products are published per region in three different collections:

- Latest – daily files covering the last 30 days. Updated every hour
- Monthly – One file per month per platform
- History – One file per platform

The data provider may provide more quality assurance on the data. As an example, IMR do analyze water samples taken together with CTD. This might result in adjusting the salinity from the original values first produced in the NRT datafile. This will be reflected in the monthly and history files from the time when the adjusted data is available. The user of the data can verify this by checking the global attribute: "data_mode".

NRT data from the various production units can be accessed through OPeNDAP, ERDDAP, and FTP. Data can be visualized using WMS calls.

Collaboration

There are several international initiatives involved in the process of identifying and facilitating dissemination of data. An example is the European Marine Observation and Data Network (EMODnet). EMODnet provides European marine data on seven discipline-based themes, including physics and chemistry.

To avoid the situation where a possible data provider is approached by both Copernicus Marine and EMODNet, the Marine In Situ collaboration (MIC) working group has been established. The objective for the group is to reach and collaborate with new providers in a comprehensive and common way.

When data from the new data provider is available at Copernicus Marine, it will also be available at EMODnet, and vice versa. The data provider will be notified that the data will be available at both infrastructures.
This increases the availability of data at both EMODnet and Copernicus Marine Service, and the benefit for the provider is that she will only be approached by one data infrastructure.

Progress on increasing quality and transparency in ICES commercial fisheries data

David Currie[1] (david.currie@marine.ie), Edvin Fuglebakk[2] (edvin.fuglebakk@hi.no), Henrik Kjems-Nielsen[3] (Henrikkn@ices.dk), Colin Millar[3] (colin.millar@ices.dk), Sofie Nimmegeers[4] (sofie.nimmegeers@ilvo.vlaanderen.be)

[1]Marine Institute (Ireland)
[2]Institute of Marine Research (Norway)
[3]International Council for Exploration of the Sea – ICES (Denmark)
[4]ILVO (Belgium)

Quality assurance within ICES

The International Council for the Exploration of the Sea (ICES) is an intergovernmental marine science organisation meeting societal needs for impartial evidence on the state and sustainable use of our seas and oceans. Quality assurance is key to the ICES mission to advance and share scientific understanding of marine ecosystems and the services they provide. The ICES Quality Policy has been published recently and is founded on the 10 principles that guide the advisory process [ICES, 2023a]. The ICES Data Centre has attained CoreTrustSeal certification.

Regional Database and Estimation System, and the Transparent Assessment Framework

The ICES Regional Database and Estimation System (RDBES) stores detailed commercial fisheries biological sampling data, and aggregated effort and landings data from ICES member countries. Its aims include: 1) to ensure that quality controlled commercial fisheries data can be made available for the coordination of regional fisheries data sampling plans, 2) to provide a regional estimation system such that statistical estimates of quantities of interest can be produced from sample data; 3) to serve and facilitate the production of fisheries management advice and status reports; and 4) to increase the awareness of fisheries data collected by the users of the RDBES and the overall usage of these data.

The ICES Transparent Assessment Framework (TAF) is an online resource for yearly ICES fish stock and mixed fisheries assessments, as well as other types of assessments feeding into ICES advice on topics including contaminants, fishing impacts, survey indices, ecosystem and fisheries overviews.

The RDBES and TAF will replace the current ICES commercial fisheries data systems in the near future and will be used to provide inputs to fish stock assessments [ICES, 2023b]. The main improvements will be [Currie D. et al., 2023]:

1. Greater transparency and more detailed data. Commercial fisheries data is a key data source used in ICES fish stock assessments. The RDBES makes it possible to upload more detailed data in a structure that reflects how the biological sampling was actually conducted. This allows users to ensure they are processing the data correctly and to also gain more detailed information from it, thus increasing the potential of the collected data. TAF provides transparent estimation at each step from detailed data to total estimates for a fish stock. This allows calculations to be easily checked and verified – they can also be re-run if there are changes to the input data. The RDBES provides a wealth of information about data availability for new assessment models, and for dealing with problems like imputation of missing data.
2. Improved estimates from fisheries dependant data. The use of a variety of different estimation methods is enabled by the RDBES data model. In particular, unbiased design-based estimation methods are supported – these methods are based on assumptions

about our sampling design which can, in principle, be controlled. Estimates of variance can also be calculated and including these has the potential to improve the quality of stock assessment model outputs.

3. Sharing best practices. With the introduction of a common RDBES data model, the ICES community is empowered to work collaboratively on estimation functions and packages. This enables more effective and efficient code to be developed as compared to working in isolation, thus improving its quality.

Operationalising the RDBES and TAF

When replacing existing processes, it is important to verify that the new systems can, at the very least, reproduce current functions within an acceptable tolerance. To this end, a series of workshops have been held which have focused on data reproduction studies [ICES, 2023c; d]. Two main areas for these reproduction studies have been identified: national estimation, and stock estimation. National estimation is the process of creating an estimate of the length and age structure of fishing catches from a stock based on each country's national data – this requires the combination of biological sampling data with fishing effort and landings data. Stock estimation is the process of combining these national estimates into an estimate for the whole stock. Both of these processes may involve the imputation of missing data.

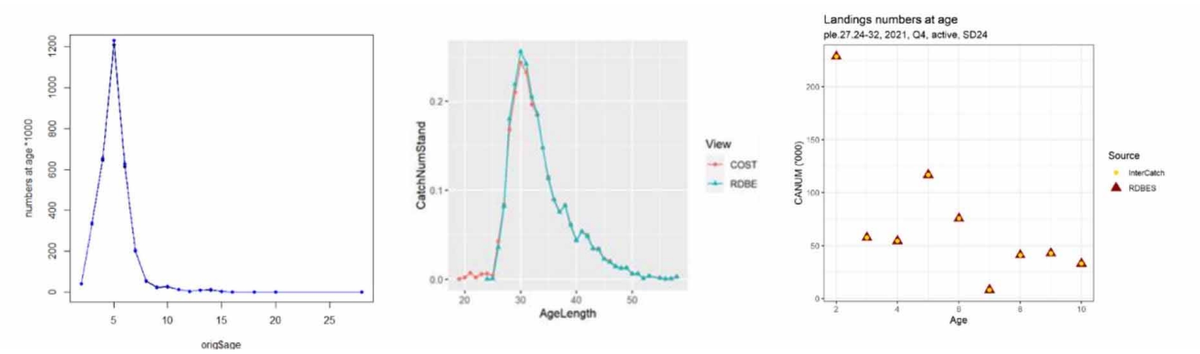


Figure 1 Selected national reproduction study results.

Good progress on both national and stock estimation reproduction has been made during the workshops - some selected national results are shown in Figure 1. Several countries have been able to satisfactorily reproduce their estimates using RDBES and have implemented their calculations in TAF, and at least one stock estimation task has been completely implemented in TAF. However, it should be noted that the aim of making an entire reproduction of stock estimation output from detailed national data submitted to RDBES has yet to be delivered - this will continue to be pursued.

These workshops have also served several other purposes for the RDBES development. They have: evaluated the use of TAF as a framework for estimation with RDBES data; encouraged adaptation of RDBES/TAF by data-providing institutions; served as a training arena for the ICES community to get acquainted with RDBES/TAF; developed recommendations for standardised output formats; elucidated roles and data access requirements for current and future estimation workflows; and produced potentially re-usable code for tasks shared between roles in the RDBES/TAF system.

References

Currie D., Fuglebakk E., Kjems-Nielsen H., Millar C., Nimmegeers S., (2023). *Improvements in stock assessment input data enabled by the Regional Database and Estimation System.*

ICES Annual Science Conference 2023, Bilbao, Spain. https://doi.org/10.17895/ices.pub.24305050
ICES (2023a). *ICES Quality Policy*. https://doi.org/10.17895/ices.pub.23864760
ICES (2023b). *Working Group on Governance of the Regional Database and Estimation System* (WGRDBESGOV; outputs from 2022). https://doi.org/10.17895/ices.pub.22786034.v1
ICES (2023c). *Workshop on Raising Data using the RDBES and TAF* (WKRDBESRaiseTAF; outputs from 2022). https://doi.org/10.17895/ices.pub.21995141.v1
ICES (2023d). Second Workshop on Raising Data using the RDBES and TAF (WKRDBES aise&TAF2). https://doi.org/10.17895/ices.pub.24648546.v1

A review of ERDDAP the established best practice in sharing gridded and tabular data from the Earth Sciences community

Kevin O'Brien[1,2] (kevin.m.o'brien@noaa.gov), Conor Delaney[1,3] (conor.delaney@emodnet.eu)

[1]Members of the ERDDAP Strategic Insight Group (USA)
[2]UW/CICOES, NOAA/PMEL (USA)
[3]EMODnet (Belgium)

ERDDAP is a data server that provides users with a simple and consistent way to download subsets of gridded and tabular scientific datasets into different file formats (Figure 1). It receives data queries and file format conversion commands via standard HTML URL calls, which users can share with others or integrate into their data flows. ERDDAP is Free and Open-Source Software (FOSS) originally developed at National Oceanic and Atmospheric Administration (NOAA) National Marine Fisheries Service (NMFS) Southwest Fisheries Science Center (SWFSC) Environmental Research Division (ERD).
Over the last decade usage of ERDDAP has increased and what was once a primarily NOAA driven and used software tool is now relied upon by over 100 organizations in at least 16 countries. ERDDAP is an open-source JAVA based web application, with online resources to support efficient interactions between ERDDAP developers, administrations, and user communities. These include:

- a GitHub Organization (https://github.com/erddap) that includes all ERDDAP source code.
- a GitHub-based issue/discussion (https://github.com/ERDDAP/erddap/discussions) process for technical input by developers and contributors.
- an active and community-supported user forum (https://groups.google.com/g/erddap) for support.
- examples of ERRDAP in action:
  - coastwatch.pfeg.noaa.gov/erddap/index.html
  - erddap.emodnet.eu/erddap/index.html

It is imperative to maintain this culture of openness and continue growing, supporting, and promoting ERDDAP. To help guide ERDDAP advancement and interaction with the global ERDDAP community, a Strategic Insight Group (ERDDAP-SIG) has been established by NOAA to provide support, oversight, and direction for the continued growth of ERDDAP. The ERDDAP developer and user communities are more vigorous than ever, and ERDDAP-SIG provides a clear opportunity to cultivate the momentum that ERDDAP has created in the data management community over the last decade. The global ERDDAP user community can be confident that ERDDAP continues to be strongly supported by NOAA (NMFS/ERD) as well as by an active and engaged international community.
In this poster we present and overview of the new advisory structure, and we will review the status ERDDAP data server and illustrate the multitude of benefits that the data platform provides. These benefits include: support for dozens of popular formats; standards-based metadata and data services and formats; support for both human and machine interactions; improved discovery of datasets; and many more. We will demonstrate how the use of ERDDAP, both from the data provider and data user perspectives, has resulted in the grass roots growth of a worldwide network of interoperable servers publishing scientific quality marine data.

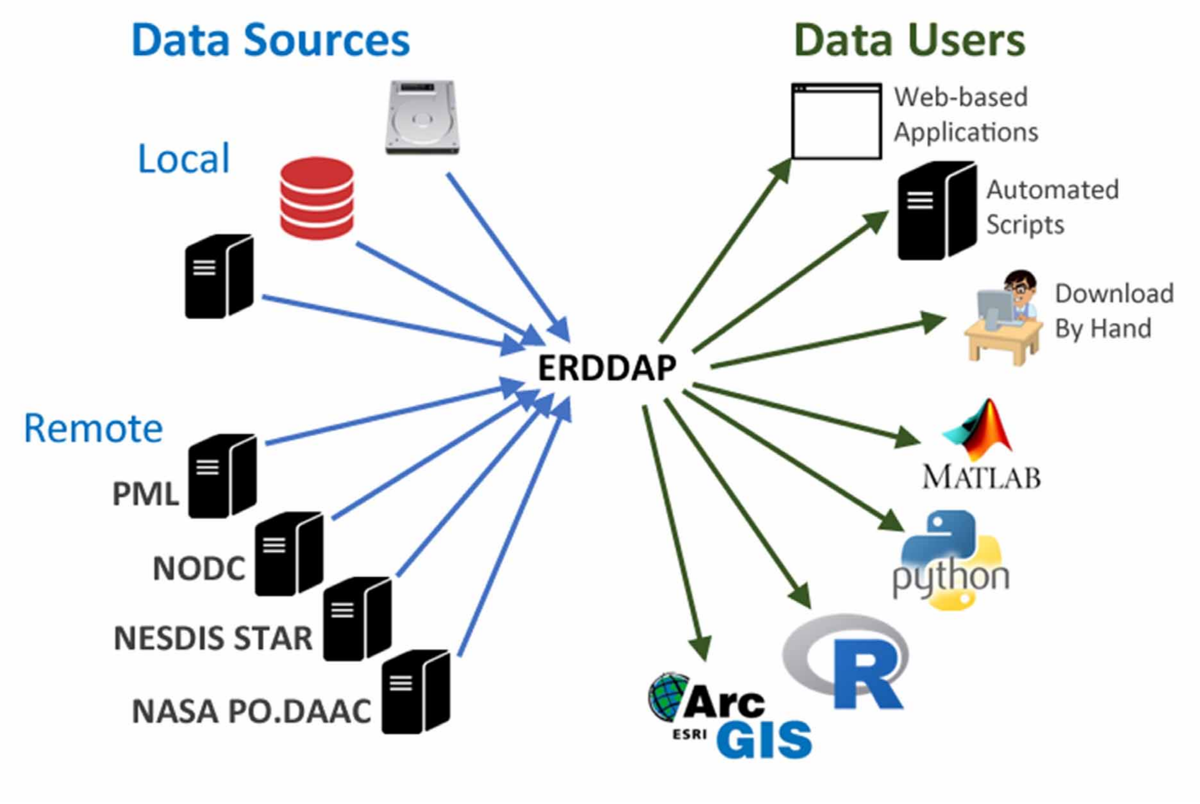


Figure 1 ERDDAP acts as a middleware between data servers and data users.

Poster session

The Marine Munition Data Compilation Germany

Marcus Krüger[1] (mkrueger@geomar.de), Robin Heß[2] (robin.hess@awi.de)

[1]Helmholtz Centre for Ocean Research Kiel - GEOMAR (Germany)
[2]Alfred-Wegner Institute, Helmholtz Center for Polar- and Marine Research (Germany)

Introduction

Worldwide coastal waters are contaminated by munitions from World War I and World War II. In the German part of the North Sea and Baltic Sea alone, there are about 1.6 million tons of munitions. The presence of ammunition in the sea is a latent threat to shipping and the offshore industry [Frey et al., 2022]. However, it is also an ecological problem, as explosive-related compounds (STVs such as TNT, ADTN, RDX, DNB, etc.) can dissolve in water as toxic and carcinogenic pollutants and can be found in both water and sediment as particles [Bünning et al., 2021].

Due to these environmental and health risks and the hazards to the marine industry there is increasing interest in the investigation and removal of underwater munitions in Germany. However, data on the topic Munitions in the Sea is widely spread over many public and private institutions. To collect and interconnect already existing datasets and for the holistic analysis of munitions-relevant data the *Marine Munition Data Compilation* (https://marine-data.de/?site=viewer&id=mmdc) (Figure 1) was established in the CONMAR project (https://conmar-munition.eu) at the Marine Data Portal (https://marine-data.de) of the German Marine Research Alliance (Deutsche Allianz Meeresforschung - DAM).

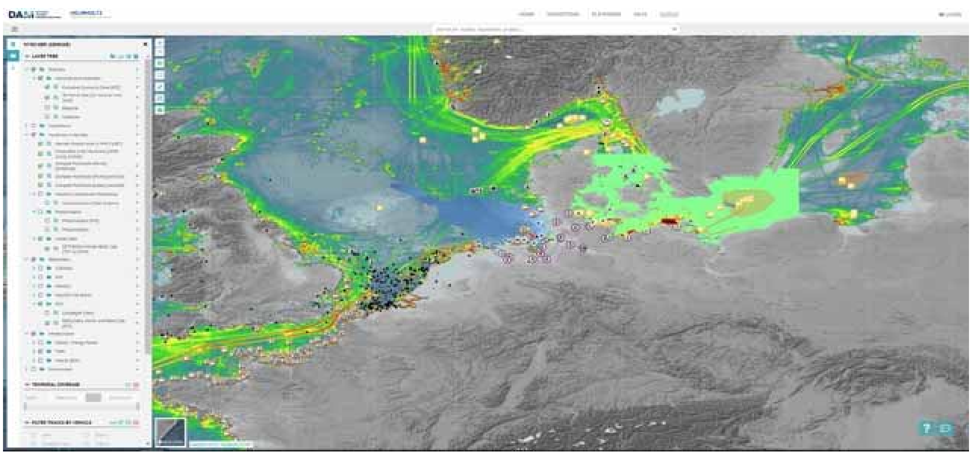


Figure 1 The Marine Munition Data Compilation.

A Federated Data Management Concept in the Light the Topic of Munitions in the Sea

Together with its member institutions DAM´s task is to develop an integrated, reliable and sustainable data management concept for the marine research environment. With the launch of the Marine Data Portal DAM introduced a cross-institutional platform for centralized access to decentral research data in line with FAIR principles [Wilkinson et al., 2019]. As part of the DAM research mission sustainMare the *Marine Munition Data Compilation* was implemented as a featured viewer within the DAM portal infrastructure. The contents of this data compilation are highly federated. The displayed data is therefore hosted at the data-producing locations (scientific centres, authorities, private sector, etc.) themselves and are only merged in the viewer.

Data Sources

Data from seafloor mapping by hydroacoustic, (electro)magnetic, or optical/imaging methods, as well as data from sampling for physical, chemical, and biological parameters in water, sediment, and biota, are partly held by federal or state agencies or research institutions. In addition, the German Navy maintains a database of munitions-related underwater objects. Furthermore, data collected by commercial companies or their contractors, for example, in the context of infrastructure projects exists.

Data Integration (Extract-Transform-Load)

As part of the federated data management concept in the CONMAR project, stakeholder data is extracted in various formats (Extract). Data that is already exchanged via OGC interfaces can be integrated directly into the portal via the service address (Load). However, non-service data must be either processed, aggregated or transformed first by CONMAR's data managers at GEOMAR (Transform). The Web-based open-source (geospatial) content management system GeoNode is used as a provisioning and catalogue component at GEOMAR. Registered users can upload and search for datasets, generate standardized metadata and grant access rights to other users. According to the data type (vector data/raster data), the data records are stored in a PostgreSQL database or in the file system. The actual provision of the data as an OGC service is performed by the GeoServer integrated into the system (Load).

Data Repositories

To sustainably store scientific research data under FAIR principles, long-term data repositories are already being used. PANGAEA - Data Publisher for Earth & Environmental Science is one of these digital library systems for data from earth system research and environmental science. Here, scientific data are linked to their metadata via an editorial system, which is then stored in a curated manner and made available on the Internet via standardized web services. The datasets are permanently referenced with a Digital Object Identifier (DOI) [Felden et al., 2023]. Quality-assured project data and datasets used for publications in CONMAR must be stored in PANGAEA eventually. Until then, data or metadata should be made available via federated web services from project partners. As soon as the data is published from the repository, the corresponding datasets are linked to the Marine Munition Data Compilation. However, security-relevant data cannot be stored in open repositories. Such data are transferred by GEOMAR to the Naval Command, Geo-Advisory Department / MWDC (Mine Warfare Data Center) in Rostock. Requests for this sensitive data are the responsibility of the Naval Command. However, GEOMAR provides metadata records for the data stock on the Internet.

References

Bünning T., Strehse J., Hollmann A., Bötticher T., Maser E., (2021). *A Tool-box for the Determination of Nitroaromatic Explosives in Marine Wa-ter, Sediment, and Biota Samples on Femtogram Levels by GC-MS/MS*. Toxics 9, 60, https://doi.org /10.3390/toxics9030060

Felden J., Möller L., Schindler U., Huber R., Schumacher S., Koppe R., Diepenbroek M., Glöckner F.O., (2023). *PANGAEA – Data Publisher for Earth & Environmental Science*. Scientific Data, 10(1), 347, https://doi.org/10.1038/s41597-023-02269-x

Torsten F., Seidel M., Koschinski S., Strehse J. S., Wehner D., Wichert U., Kampmeier M., Andresen C., Greinert J., (2023). *Kampfmittel im Meer – Der Umgang mit Belastungsschwerpunkten*. Handbuch Altlastensanierung und Flächenmanagement 93 (3).

Wilkinson M.D., Dumontier M., Jan Aalbersberg I. et al., (2019). *Addendum: The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 6, 6, https://doi.org/10.1038/s41597-019-0009-6

Poster session

Marine observations organized within EU "HORIZON 2020" Project "BRIDGE-BS", Georgia

Irine Baramidze[1] (irine.baramidze@gmail.com), Mustafa Yucel[2] (myucel@ims.metu.edu.tr), Kakhaber Bilashvili[1] (kakhaber.bilashvili@tsu.ge), Madona Varshanidze[1] (varshanidzem@yahoo.com), Eteri Mikashavidze[1] (eteri-mikashavidze@rambler.ru), Vazha Trapaidze[1] (vazha.trapaidze@gmail.com), Nino Machitadze[3] (ninomachitadze@gamma.ge), Nino Gelashvili[3] (ninogelashvili@gamma.ge)

[1]Iv Javakhishvili Tbilisi State University (Georgia)
[2]METU (Turkey)
[3]Scientific Research Firm GAMMA (Georgia)

The overall objective of EU "HORIZON 2020" Project "Advancing Black Sea Research and Innovation to Co-Develop Blue Growth within Resilient Ecosystems" — "BRIDGE-BS" is to advance the Black Sea's marine research and innovation to co-develop Blue Growth pathways under multi stressors for the sustainable utilization of the ecosystem services. The project is based on a novel 3 NODE structure (Service Dynamics, Blue Growth Incubators, and Empowered Citizens). The corresponding activities were planned in the WP5, that is situated within Node 2 – Blue Growth Indicators, but strongly linked to Node 1 of the project. One of the specific objectives of the WP is to execute basin wide process-oriented synoptic R/V expeditions to improve process understanding of stressor and service interactions in relation to physical and biogeochemical redox processes. Following the objective, samplings on the pre-identified pilot sites in the Black Sea, in particular, Pilot site 6, Batumi, Ajara region, Georgia has been organized (Figure1).

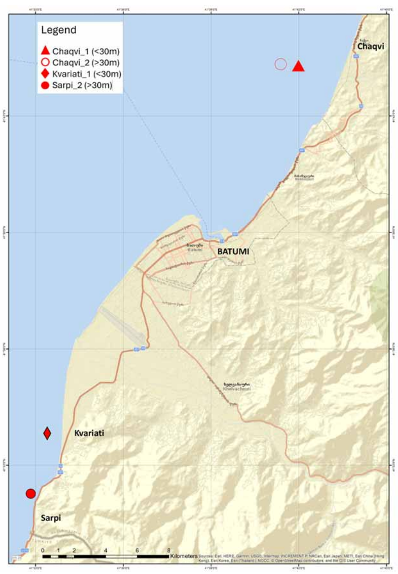


Figure 1 Batumi Pilot site 6, Ajara region, Georgia.

Introduction

Investigation on research stations have 4 main tasks:

- assessment of physical and chemical indicators of the quality of the marine environment;
- assessment of the level of eutrophication of the Adjarian Black Sea shelf and coastal waters;
- assessment of the level pollution of Adjara marine area;
- assessment of the state of marine biodiversity of the Adjara shelf and coastal waters.

Justification

Within the BRIDGE WP5 it has been identified research stations, sampling depths and biological and chemical indicators. Correspondingly, the chemical and biological observations will be conducted at 4 different stations on 2 transects.

Chemical indicators/parameters (including contaminants):

- Temperature, Transparency, Ph, TSS, Salinity, Dissolvedoxygen DO, Ammonia NH_4-N
- Nitrite NO_2-N, Nitrate NO_3-N
- Phosphate PO_4-P, TR, TN
- Heavy metals - Fe,Zn, Cu,Ni
- Polycyclic Aromatic Hydrocarbons, Pesticides.

Observations on Marine litter (macrolitter - riverine, beach and sea floating) will take place, as well.

Biological indicators:

- Phytoplancton. Total abundance cells/m^3, biomass mg/m^3, Species composition.
- Zooplankton. Total abundance cells/m^3, biomass mg/m^3, Species composition.
- Zoobenthos. Biodiversity, number ind/m^2), biomass (g/m^2).

During the conducted two sessions of the field works, the location of the stations became necessary to be changed as intercalibration of benthic organisms required the transfer of samples taken from 3 different benthic habitats (infrallitoral mud, Infralittoral sand, circalittoral mud, circalittoral mixed sediment - from the named 4 habitats we have the first, the second and fourth habitats).

#	Station name	Water area	Strata	Deapth (m)	Coordinate DD		
					N	E	
1	Kvariati_1	Kvariati	<30	28	41.56358	41.55633	Kvariati_1 (inf-mud)
2	Sarfi_2	Sarfi	>30	54	41.5375	41.54825	Sarfi_2 (circ-mix,54m)
3	Chaqvi_1	Chaqvi	<30	21	41.71992	41.69717	Chaqvi_1 (inf-sand,21m)
4	Chaqvi_2	Chaqvi	>30	36	41.7207	41.69177	Chaqvi_2 (sand,36m)

Important locations (for investigation) according to the proposed plan include: Sarpi-Kvariati: Off the coast, *Cystoseirabarbata* with seafloor communities are present. This is a pristine shore because of its water circulation and the fact that no sewage water is discharged into the river there. There are many fish species and dolphins. The existing seafloor communities are not protected, while the species' community has been listed in the Red Book of Marine Habitats. This area, proposed to be designated as Marine Protected Area, could, when linked, become one integrated site with the Chorokhi Delta Emerald Site. Protection of this area and the river mouth at national level is needed and the linkage to the marine environment is evident. All observed data are under formatting into CDI ODV files for further submission to SeaDataCloud Portal. Two more expeditions are planned to be conducted in the mentioned area within the project and, consequently, an obtained data will be provided to the portal as well.

Blue-Cloud 2026 – Federated Data Discovery & Access Service

Dick Schaap[1] (dick@maris.nl), Enrico Boldrini[2] (enrico.boldrini@cnr.it)

[1]MARIS (The Netherlands)
[2]Consiglio Nazionale delle Ricerche - IIA (Italy)

Blue-Cloud 2026 project

The EU pilot Blue-Cloud project combined developing a marine thematic European Open Science Cloud (EOSC) and serving the Blue Economy, Marine Environment and Marine Knowledge agendas. It deployed a versatile cyber platform with smart federation of multidisciplinary data repositories, analytical tools, and computing facilities in support of exploring and demonstrating the potential of cloud based open science for ocean sustainability. The successor Blue-Cloud 2026 project aims at a further evolution into a Federated European Ecosystem to deliver FAIR & Open data and analytical services, instrumental for deepening research of oceans, coastal & inland waters.

Blue-Cloud Data Discovery & Access service

The Blue-Cloud Data Discovery and Access Service (DD&AS) is one of the two main components of the Blue-Cloud technical framework, next to the Blue-Cloud Virtual Research Environment (VRE). It facilitates discovery and retrieval of marine data sets and data products for external users in stand-alone mode, also interoperable with the Blue-Cloud VRE. The Blue-Cloud data sets are managed in Blue Data Infrastructures (BDIs) that are connected to the DD&AS to serve federated discovery and access.

Figure 1 Homepage of the Blue-Cloud DD&AS.

Federating European Blue Data Infrastructures

In Europe there are several research infrastructures and data management services operating in the marine and ocean domains. These cover a multitude of marine research disciplines, and providing access to data sets, directly originating from observations, and to derived data products. A number are ocean observing networks, while others are data aggregation services. Moreover, there are major EU driven initiatives, such as EMODnet and Copernicus Marine, that focus on generating and giving access to derived data products. Together, these infrastructures constitute a diverse world, with different user interfaces. The Blue-Cloud DD&AS has been initiated to overcome this fragmentation and to provide a common interface for users by means of federation. In the current DD&AS version, federated discovery and access is given to metadata, data, and data products from the following Blue Data Infrastructures: SeaDataNet, EMODnet Chemistry, EurOBIS, EcoTaxa, ICOS-Marine, ICOS-SOCAT, ELIXIR-ENA, and EuroArgo – Argo. Together their offer comprises more than 10 million data sets from physics, chemistry, geology, bathymetry, biology, biodiversity, and genomics.

How does the DD&AS work?

A common interface is provided for discovery and retrieval of data sets and data products from each of the federated BDIs. The interface also includes facilities for mapping and viewing the locations of data sets, as this is part of the query dialogue. The interface has a shopping mechanism, facilitating users to compose and submit mixed shopping baskets with requests for data sets from multiple BDIs. The DD&AS is fully based and managed using web services and APIs, following protocols such as OGC CSW, OAI-PMH, ERDDAP, Swagger API, and others, as provided and maintained by the BDIs. These are used to deploy machine-to-machine interactions for harvesting metadata, submitting queries, and retrieving resulting metadata, data sets and data products.

The query mechanism has a two-step approach:

- The first step enables users to identify interesting data collections, with free search, geographic and temporal criteria as main query operators on a common catalogue including entries from all federated BDIs;
- The second step enables users to drill down per interesting BDI to get more specific data sets at granular level, using again free search, geographic and temporal criteria, but this time at granular level, and including additional search criteria, specific for a selected BDI;
- Finally, users are able to compose and submit shopping requests at the granule level and after processing by the Blue-Cloud data broker to retrieve the data sets by downloading from their MyBlueCloud dashboard.

Metadata and Data Brokerage Services and Shopping mechanism

The DD&AS works with brokerage services both at metadata and data level. For the metadata brokerage at the first level of data collections, use is made of a DAB metadata broker service as developed and managed by CNR-IIA. This middleware harvests metadata at collection level from each of the BDIs, using their existing web services or APIs. The DAB service transforms the harvested XML files from each of the BDIs into a common ISO Blue-Cloud collection profile, which is published by means of a Blue-Cloud OGC-CSW service with a common XML profile for each BDI.

The second step drills down within identified collections to get more specific data, using free search, geographic and temporal criteria, but this time at granular level, and including additional BDI-specific search criteria. The Blue-Cloud Data brokerage service, developed and operated by MARIS, performs the master role in the DD&AS. It retrieves and indexes the latest Blue-Cloud level 1 metadata into an Elastic database, and adds metadata from the web services of the BDIs at level 2, including associated data download links. Level 1 metadata includes sufficient metadata for users to identify which of the BDIs holds interesting data sets. For level 2, the common search profile is expanded for each BDI with extra specific criteria, facilitating discovery down-drilling. Finally, the Data brokerage service also contains a shopping mechanism with basket and ledger, by which users (external users and VRE) and BDIs can stay informed about shopping transactions and their status.ù

Further evolution

Developments are underway for expanding the DD&AS with more BDIs, namely: EMSO, SIOS, EMODnet Physics, and ELIXIR-MGnify. Furthermore, activities are undertaken for optimizing the DD&AS functionality, such as adding semantic interoperability, and making the services and content as provided by the BDIs more FAIR. The DD&AS is an important component of the Blue-Cloud infrastructure for streamlining the provision of multidisciplinary data sets as needed as input for the Blue-Cloud Virtual Labs and EOV Work Benches developments. The DD&AS is also a major service for researchers in the marine and ocean domains, and wider via the European Open Sciece Cloud (EOSC). It is also aiming for serving the Digital Twins of the Oceans (DTO's).

Blue-Cloud 2026 - Virtual Research Environment Service

Massimiliano Assante (massimiliano.assante@isti.cnr.it), Leonardo Candela (leonardo.candela@isti.cnr.it), Luca Frosini (luca.frosini@isti.cnr.it), Francesco Mangiacrapa (francesco.mangiacrapa@isti.cnr.it), Elisa Molinaro (elisa.molinaro@isti.cnr.it), Pasquale Pagano (pasquale.pagano@isti.cnr.it)

Consiglio Nazionale delle Ricerche - CNR (Italy)

Blue-Cloud 2026 project

The EU pilot Blue-Cloud project combined developing a marine thematic European Open Science Cloud (EOSC) and serving the Blue Economy, Marine Environment and Marine Knowledge agendas [Schaap et al., 2022]. It deployed a versatile cyber platform with smart federation of multidisciplinary data repositories, analytical tools, and computing facilities in support of exploring and demonstrating the potential of cloud based open science for ocean sustainability. The successor Blue-Cloud 2026 project aims at a further evolution of the Blue-Cloud platform into a *federated European ecosystem* to deliver FAIR open data and analytical services, instrumental for deepening research of oceans, coastal and inland waters.

Blue-Cloud Virtual Research Environment Service

The Blue-Cloud Virtual Research Environment (VRE) is one of the two main components of the Blue-Cloud technical framework, next to the Blue-Cloud Data Discovery and Access Service (DDAS). The Blue-Cloud VRE components are developed and operated by relying on the D4Science infrastructure [Assante et al., 2019; 2023; Candela et al., 2023] and range from services to promote the collaboration among its users to services supporting the execution of analytics tasks embedded in a distributed computing infrastructure, and to services enabling the co-creation of entire Virtual Laboratories (VLabs), also interoperable with the Blue-Cloud DDAS. The VRE services are instrumental in advancing Open Science practices within VLabs, empowering researchers to harness the advantages of state-of-the-art e-infrastructures. By leveraging these services, researchers can capitalise on the power of the Cloud and of e-infrastructures, driving scientific progress and enabling collaborative research efforts within the realm of Open Science.

VRE Service Overview

VRE services offer a comprehensive array of features, fostering collaboration, facilitating data analytics, enabling result dissemination, and ensuring seamless integration with external systems. In fact, the VRE caters to the entirety of the research lifecycle, providing diverse services.

Specifically, the *Collaborative Storage Framework* fosters collaboration among users by providing shared storage spaces where researchers can easily access and manage datasets collaboratively. This accessibility ensures that all team members have up-to-date information, which facilitates decision-making and accelerates the research process.

In terms of data analytics, the *Analytics Engine Framework* empowers VLabs with powerful tools and resources. The *DataMiner* and the *Cloud Computing Platform* enable data analysis, allowing researchers to extract valuable insights from complex datasets coming from Blue Data Infrastructures. Additionally, VRE services offer specialised solutions such as the *RStudio*-based, the *JupyterHub*-based, the *ShinyProxy*-based for ShinyApps, and the solution for Dockerised Applications. These tools cater to different analytical needs and preferences, ensuring VLab members to have the flexibility and diversity of options to conduct their analyses effectively.

The *Publishing Framework* within VRE services facilitates the dissemination of research outcomes by means of the *VRE Catalogue* and the *Spatial Data Catalogue*, that provide a means to organise and publish research results, making them accessible to the wider scientific community. This framework ensures transparency and the sharing of valuable knowledge generated within VLabs. Within a VLab, these services are conveniently organised and accessible through dedicated menu items displayed on the VLab landing page, as depicted in Figure 1.

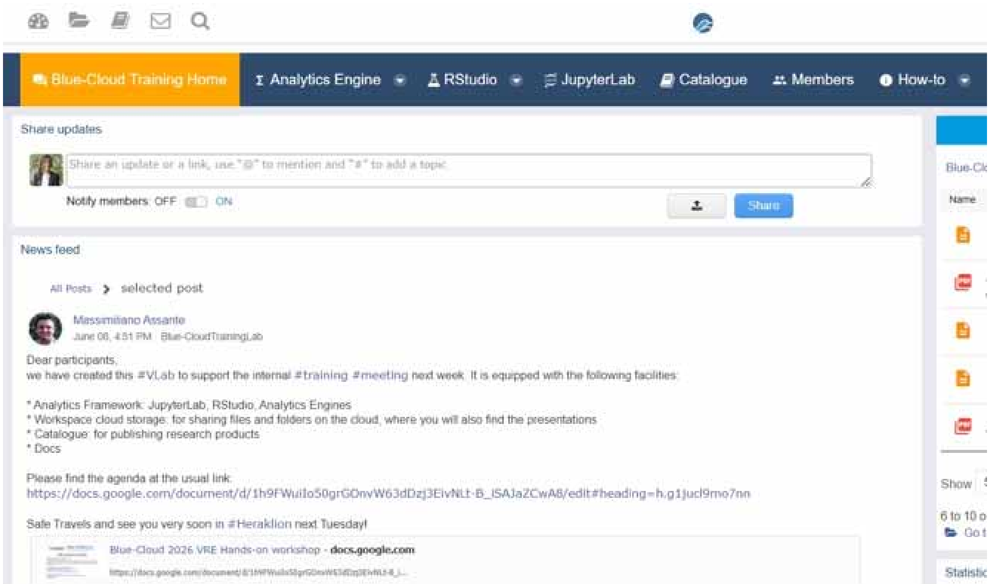


Figure 1 One of the Blue Cloud Virtual Lab set up on the VRE.

References

Schaap D., Assante M., Pagano P., and Candela L., (2022). *Blue-Cloud: Exploring and demonstrating the potential of Open Science for ocean sustainability*. IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea), Milazzo, Italy, pp. 198-202, https://doi.org/10.1109/MetroSea55331.2022.9950819

Assante M. et al., (2019). *Enacting open science by D4Science*. Future Gener. Comput. Syst. 101: 555-563, https://doi.org/10.1016/j.future.2019.05.063

Assante M. et al., (2023). *Virtual research environments co-creation: The D4Science experience*. Concurrency Computat Pract Exper. 2023; 35(18):e6925, https://doi.org/10.1002/cpe.6925

Candela L., Castelli D., and Pagano P., (2023). *The D4Science Experience on Virtual Research Environments Development*. In IEEE Computing in Science & Engineering, https://doi.org/10.1109/MCSE.2023.3290433

Digital Twin of Land and Sea: Federated cross-standard Marine SDI Use Cases

Sina Taghavikish[1] (staghavikish@ogc.org), Piotr Zaborowski[2] (pzaborowski@ogc.org)

[1]Open Geospatial Consortium (Canada)
[2]Open Geospatial Consortium (Poland)

Introduction
The Federated Marine SDI (FMSDI) 2023 pilot represents the fourth phase of the FMSDI pilot program in Opent Geospatial Consortium (OGC) with a focus on Digital Twins and the interface between land and sea. A primary goal of this pilot is to advance the FMSDI concept with a focus on coastal areas vulnerable to increasing threats of climate change. Coastlines around the world are undergoing transformations due to climate change, marine pollution, and population growth. This results in habitat loss, sea level rise, and other side effects such as an increase in microplastics. Stakeholders face the challenge of quantifying the rapid changes in the environment affecting dependent ecosystem services. FMSDI pilot 2023 had three threads focused on specific global regions: Singapore, the Arctic, and the Caribbean, each with its own challenges and requirements.

Digital Twin of Land and Sea Interface in Singapore
Singapore, with one of the world's busiest seaports, is particularly susceptible to sea level rise, storm surge, and major coastal weather events. The overall scenario for this thread involved a storm surge flooding event in Singapore. The objective was to simulate this event using a digital twin in order to aid future planning and decision-making processes. The methodology employed by the participants mainly involved bringing together land and sea data from separate agencies, assessing how well this data fit together, identifying any gaps, and finally integrating the data within a 3D/4D environment. Participants developed four solutions that independently addressed the digital twin challenge. They used restricted data from the local land and marine authorities, as well as open-source data sets.
The conclusion was that without authoritative data from the land and sea interface (i.e., intertidal zone), this gap could not be filled. This can be achieved using expansive LiDAR observations or state-of-the-art solutions such as Satellite-Derived Bathymetry (SDB). The challenge with SDB to be used by hydrographic offices is to verify, validate, and adjust it against corrected LiDAR bathymetry referenced to chart datum. This practice is already planned to be employed in Canadian Arctic regions and is a focus for Singapore.

Integrating Land & Sea for Various Use Cases – Caribbean
The Caribbean islands are vulnerable to the warming oceans that surround them. Their economies heavily depend on the services provided by their coastal zones, but coastal erosion poses a significant threat to their populations. The Caribbean thread consisted of five use cases that aimed to demonstrate alternative uses of navigational charts, which could prove helpful in fields such as disaster management and the blue economy. The titles included the concentration and distribution of micro-plastics before and after climate events, improving decision-making using enhanced data availability through standards during tropical storms and hurricanes, developing a framework for an Offshore Wind Farm Decision Support System to assist in site selection, creating a Hurricane Health Risk Index to identify the underlying health and medical needs of a population affected by hurricanes, and lastly, making use of navigational data for non-navigational purposes by building a geospatial information management foundation.

In conclusion, the Caribbean thread revealed a need for making datasets more openly available and highlighted the widespread use of OGC standards (e.g., WFS, WMS and WCS) and the IHO S-xx series. This underscores the solution's potential to consolidate existing data practices while expanding the use of standardized data formats and APIs for efficient data integration and scenario building. The future work will be to enhance cooperation with UN-GGIM and IGIF, particularly in the marine domain (IGIF-Hydro), to address data discovery and licensing issues, ensuring the inclusion of the marine domain in the larger geospatial information ecosystem.

Digital Arctic Connecting Land & Sea – Canadian Arctic

In the Arctic, the focus was on identifying vulnerable areas due to rapidly changing sea ice concentrations, anticipating the consequences of rising temperatures, changing habitats, and biodiversity loss, and the thaw rate of permafrost with its diminishing role in sequestering global greenhouse gases. Another major challenge discussed in this thread was to efficiently discover and present diverse data from numerous sources for various scenarios in the Arctic. One aspect of this challenge was to evaluate the best practices and standards aligned with the OGC suite of standards. This evaluation aimed to demonstrate an improved methodology for discovering and integrating land and marine geospatial information in support of analysis services from multiple sources.

The four participants identified three scenarios to showcase the aforementioned challenge in the context of climate change. The scenarios were the impact of sea ice loss and the migratory patterns of Arctic fauna (bow whale and killer whale), interoperability of geospatial information across land and sea, health risk indicators of microplastics in the environment and the effect on coastal communities, and satellite-derived essential climate variables with an application to large marine ecosystems and

The resulting catalogue provides a platform for discovering, organizing, and visualizing critical data related to climate change and coastal vulnerabilities. It leverages OGC and ISO standards for interoperability and efficient access to information. The architecture is built on OGC API standards, including Records, Processes, and Coverages, enabling effective management and dissemination of multi-dimensional geo data cubes (GDC).

Summary of the results

The FMSDI 2023 pilot was specifically designed to assess the advantages of a standards-based approach to data discovery and application. The evaluation aimed to support stakeholders involved in the changing coastal environments of the Canadian Arctic, the Caribbean islands, and the Republic of Singapore. The coastal environment is represented as a seamless transition from the ocean floor to the land surface through the integration of distinct data products. This transitional realm has its unique organization and functions as a host to critical habitats and essential ecosystem services that serve the needs of coastal communities.

One of the significant challenges of the project was to overcome the disparity between marine and terrestrial data systems and create a digital twin representation of the coastal environment to improve the 'time to decision' for stakeholders. Another challenge was to identify use cases for the non-navigational application of navigational charts and build capacity for the blue economy. The project also aimed to address data discovery issues, particularly in areas like the Canadian Arctic and the Caribbean, where data availability levels varied. Lastly, there was a lack of quality control and validation for online datasets that needed to be addressed by data providers and users.

The FMSDI pilot 2023 yielded promising results, and a detailed open-access report will be published soon, highlighting new insights and challenges. One of the key findings was the effectiveness of the GeoDatacube architecture in addressing scale and positioning accuracy issues. Another significant observation was the difficulty in correlating data based on existing local vertical datums, such as the Singapore Vertical Datum, with global vertical datums like WGS84. The use of Vectors of flow to visualize meteorological effects and migratory patterns in the Arctic environment was also discussed, with extended models showcasing the usefulness of the OGC Features and OGC Moving Features specifications to model movement. The pilot project also identified the crucial role of core OGC standards in coastal observation workflows and their application in more strategic standards. Moreover, the interdependencies with external standards bodies such as the IHO and UN-GGIM were also recognized.

One Ocean Expedition

Arnfinn Morvik (arnfinn.morvik@hi.no), Marte Strømme (marte.stromme@hi.no)

Institute of Marine Research (Norway)

Background

During 2020-2023 One Ocean Expedition, the Norwegian tall ship Statsraad Lehmkuhl circumnavigated the globe, sparking attention and sharing knowledge about the ocean for a sustainable development in a global perspective. One Ocean expedition is a recognized part of the UN Decade of Ocean Science for Sustainable Development.

The voyage started from Arendal in the southern part of Norway in August 2021 and ended in Bergen in April 2023. 37 ports and 24 countries were visited during the 60000 nautical miles voyage. The ship served as a floating university and training vessel combined, bringing students, scientists, trainees and professionals together on different legs. During port visits, the ship was used for conferences, diplomacy, high level meetings and cooperate hospitality.

Instrumentation

The vessel was equipped with state-of-the art research equipment continuously collecting data throughout the voyage. Sensors include EK80 fish finder echosounder, water sampling ferrybox system, pCO2 sensor, weather station, light sensor, wave radar and hydrophone. Data from the instruments were automatically stored on disk.

Data collection

The expedition was divided into ten legs. For each leg, two students served as research assistants in charge of data collection. They supervised the instruments for continuous data collection and handled the manual instruments.

Water samples were taken several times per week, to perform eDNA analysis and to collect micro plastic. When the weather conditions were acceptable, the ship was stopped to enable the use of a CTD sensor, plankton net and secchi disk. Fishing and whale watching was conducted daily, following scientific procedures.

Data storage/transmission

The One Ocean Scientific Committee, led by the Norwegian Institute of Marine Research (IMR),

has developed methods for dissemination of the observations to scientists and the public.
Throughout the expedition, four Network Attached Storage (NAS) systems were used for data storage. During each leg, data was stored on two separate NAS racks. When the leg was finished, one NAS rack was sent to the Norwegian Marine Datacenter at the IMR in Bergen, the other one kept on the vessel as backup. After copying data from the NAS, the disk was sent back to the vessel. Following this procedure, three NAS racks were always on the vessel and the fourth NAS in transit to or from Bergen. The data copied from the disk is stored at IMR for long time preservation.
In addition to the data stored on disk, data from the ferry box, pCO2 sensor and weather station were sent to IMR over satellite when the satellite link was making this possible. An Application Programming Interface (API) for uploading data was implemented at IMR, with data from the vessel uploading every 10 minutes. This data was publicly available immediately after uploading.
Figure 1 shows the One Ocean Expedition data flow. Data from the various instruments were collected on a central server and stored to disk.

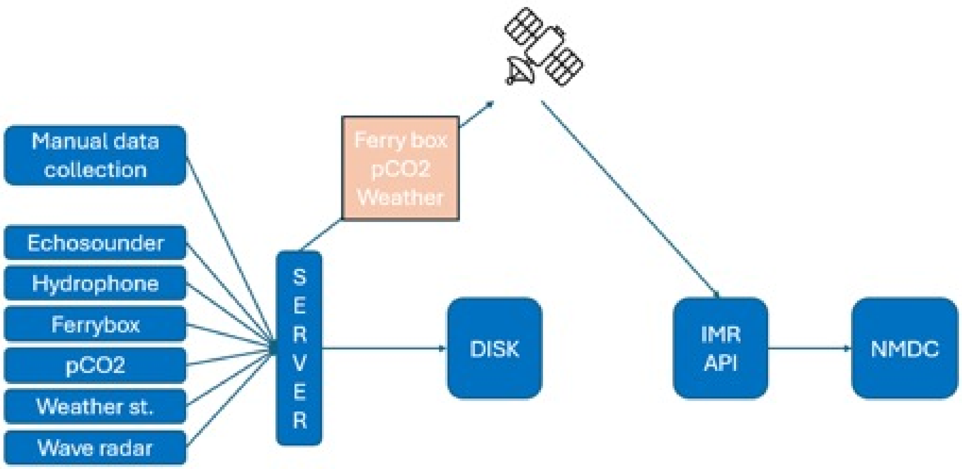


Figure 1 Data storage/transmission.

Data Publication
Data collected at the One Ocean Expedition are publicly available on NMDC, a Norwegian national infrastructure for marine data hosted by IMR.
During each leg of the expedition, API links to the automatically uploaded data where available for anyone to download. The data available on the API was used to report weather conditions and position on the homepage of the expedition. It was also used in a showroom for the expedition set up at the Bergen Aquarium.
All the raw data are published on NMDC. Some of the data needs further processing before publishing, and this work is still ongoing. These datasets will be published as soon as they are available.
The data from the expedition are available at this url: https://doi.org/10.21335/NMDC-1572929066

Iliad Digital Twins of the Ocean: An Innovative Approach to Marine Data Systems

Ute Brönner[1] (ute.broenner@sintef.no), Raymond Nepstad[1] (raymond.nepstad@sintef.no), Harald Warmelink[2] (harald.warmelink@buas.nl), Jacopo Fabrini[2] (fabrini.j@buas.nl), Emlyn Davies[1] (emlyn.davies@sintef.no)

[1]SINTEF Ocean (Norway)
[2]Breda University of Applied Science (The Netherlands)

Introduction

The OceanLab Observatory (*www. oceanlabobservatory.no*) is a full-scale ocean space laboratory that can be used by national and international partners, academia and industry and was established in Trondheim, Norway, as part of a state-of-the-art national research infrastructure in collaboration with SINTEF and NTNU, funded by the Norwegian Research Council. This innovative facility is designed to cater to the needs of education, research, and innovation in the marine and maritime sectors.

As a dynamic field laboratory, the Observatory offers unique opportunities for marine environmental research and sensor technology development. It encourages cross-disciplinary collaboration in technology development, autonomous operations, and ocean ecosystem monitoring, including the development of Digital Twins of the Ocean. The facility aligns with the United Nations sustainable development goals, contributing to societal benefits through cutting-edge research, education, and science communication initiatives.

Figure 1 One of the Blue Cloud Virtual Lab set up on the VRE.

One of the Observatory's primary objectives is to support the Blue Economy and the Green shift of the ocean industries, by providing a platform for the rapid development and prototyping of sensors and digital marine technologies, offering easy access to test facilities, telemetry-coupled digital infrastructure, and deep-water testing environments as means to foster sustainable marine and maritime practices.

Equipped with buoys and sensors, the Observatory facilitates a range of experiments and tests, collecting time-series data critical for marine research. These resources are available for the global research community, emphasizing the Observatory's commitment to open access and collaborative innovation.

A digital twin for environmental monitoring and pollution detection

In our case, OceanLab also allows us to advance data availability and sharing through developmet of a Digital Twin (DTO), targeting environmental monitoring and pollution detection using advanced particle observation and classification with SilCam and PyOPIA (*www.sintef.no/silcam*, *pyopia.readthedocs.io/*).

The Trondheim fjord Observatory exemplifies cutting-edge data acquisition for Digital Twins of marine environments. Through advanced sensing and monitoring technologies, essential ocean variables and additional data are continuously collected to form the backbone of the DTO.

Data Infrastructure

Central to the DTO is a dynamic, adaptive digital infrastructure. It processes and stores real-time data, ensuring the digital twin accurately reflects the fjord's marine ecosystem. This infrastructure is pivotal for the effective management and interpretation of complex marine data. The digital twin is developed as one of the pilots in the Digital Twins of the Ocean project Iliad (*www.ocean-twin.eu*). It will be based on an interoperable system of systems architecture, i.e., the twin can easily be transferred and enhanced by data and models from other data infrastructures. This is called a federated approach.

Digital Twin Core

The digital twin core consists of frequent particle transport modelling of the particles observed in the observatory, combined with particle classification through our softwware PyOPIA: Python Ocean Particle Image Analysis. PyOPIA started in Feb. 2022 as a 'spin-off' from elements of PySilCam (*github.com/SINTEF/PySilCam/wiki*) which had been developed to distinguish oil droplets and bubbles from sediment particles and zooplankton like fish eggs and copepods.

The DTO is not a static digital replica of the ford area of the OceanLab observatory but an evolving model that integrates data with themed modelling. This Digital Twin Core is crucial for maintaining an up-to-date, interactive representation of the ocean environment of the Observatory, facilitating real-time analysis and decision-making. When the Digital Twin detects an unusual event through significant deviations in the data it will immediately issue a warning. At the same time, the frequency of the particle transport simulations is increased.

Twin Visualisation and Gaming Technology

One of the key innovations in developing the digital twin involves the application of gaming technology (Unity, *www.unity.com/*) for data visualisation. The challenge with observing small particles in the ocean is that while these particles play a crucial role in the ecosystem, they are too small for the human eye. To make the invisible visible, we employ immersive 2D, 3D, and 4D representations, to provide an intuitive understanding of these observation and marine data. While these visualisation techniques might suit stakeholder engagement, scientist would rather interact directly with the data. The Iliad project has defined standard APIs for sharing ocean observations, model output and other sources of marine data to ensure easy interoperability between Digital Twin building blocks like Twin Core and Twin Visualisation. These APIs will serve the data to other platforms where they can be inspected through notebooks and dashboards.

Interoperability and FAIR Data Sharing

The system's design prioritises high interoperability and adherence to FAIR data principles. Well-defined APIs ensure that the essential ocean variables are accessible, usable, and meaningful, fostering collaboration within the scientific community. Future projects that collect data within the OceanLab Observatory should be able to plug into the Digital Twin and leverage the underlyig technology to access and share their data easily.

The authors received funding through the Iliad project, funded by the European Commission's Horizon 2020 Research and Innovation programme under grant agreement No 101037643.

Sextant, a marine spatial data infrastructure: implementation of ogc protocols for the dissemination of marine data

Mickael Treguer (Mickael.Treguer@ifremer.fr), Julien Meillon (Julien.Meillon@ifremer.fr), Amandine Thomas (Amandine.Thomas@ifremer.fr)

Ifremer (France)

At national and European levels, in various projects, data products are developed to provide end-users and stakeholders with homogeneously qualified observation compilation or analysis. Ifremer has developed a spatial data infrastructure for marine environment, called Sextant, to manage, share and retrieve these products for its partners and the general public. Thanks to the OGC and ISO standards and INSPIRE compliance, the infrastructure provides a unique framework to federate homogeneous descriptions and access to marine data products processed in various contexts, at national level or European level. For example EMODnet (Bathymetry and Chemistry), Copernicus Marine Service, SeaDataNet.
The discovery service of Sextant is based on the metadata catalogue. The data description is normalized according to the ISO 191XX series of standards and to the Inspire recommendations. Access to the catalogue is provided by the standard OGC service, Catalogue Service for the Web (CSW 2.0.2).

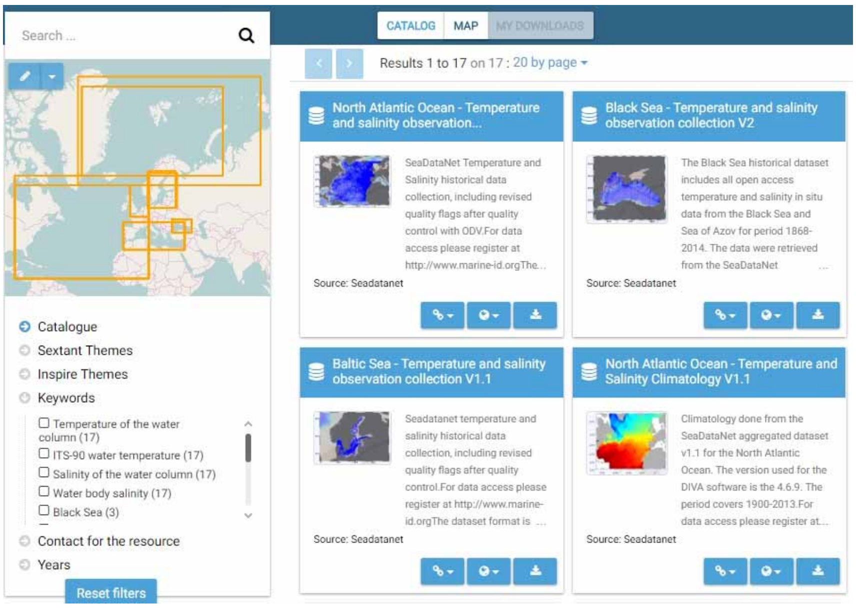


Figure 1 SeaDataNet content catalogue.

Data visualization and data downloading are available through standard OGC services, Web Map Services (WMS) and Web Feature Services (WFS). Several OGC services are provided within Sextant, according to marine themes, regions and projects. Depending on the file format, WMTS services are used for large images, such as hyperspectral images, or NcWMS services for gridded data, such as climatology models.
New functions are developed to improve the visualization, analysis and access to data. For example: data filtering, online spatial processing with WPS services and access to sensor data with SensorThings API.

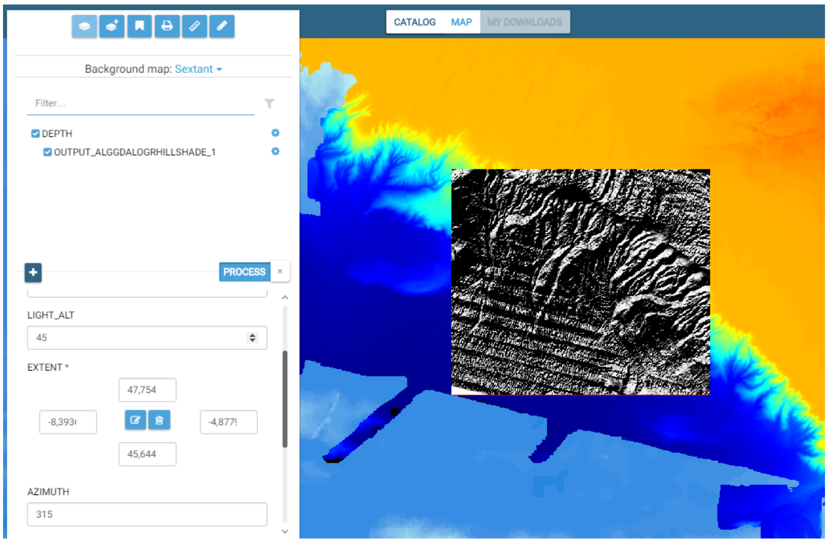


Figure 2 Online Process output for hillshade on DEM OGC WPS.

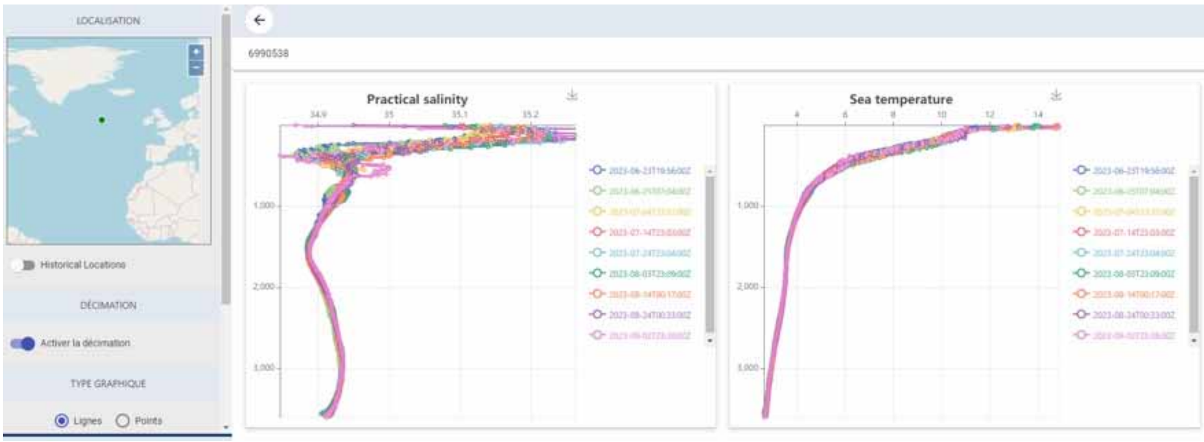


Figure 3 Argo profile from OGC SensorThings API.

The New Croatian Met-ocean Moored Buoy Network and Data Service Operating by DHMZ

Dijana Klarić (dijana.klaric@cirus.dhz.hr), Tvrtko Horvat (tvrtko.horvat@cirus.dhz.hr), Eugen Hukić (eugen.hukic@cirus.dhz.hr), Stjepan Ivatek-Šahdan (stjepan.ivatek-sahdan@cirus.dhz.hr)

Croatian Meteorological and Hydrological Service -DHMZ (Croatia)

DHMZ met-ocean moored buoy network.

The Croatian Meteorological and Hydrological Service (DHMZ) is the state authority for marine meteo services at the Adriatic Sea. The complex structure of the Eastern Adriatic coastal zone is an area of fast changing local meteorological and oceanographic regimes that requires dense observations of met-ocean phenomena. DHMZ has developed the network of met-ocean moored buoys at the transition between inner waters and the open sea. The network is a first-rank observing system with up to 22 sensors per buoy. The main goal of the network is to provide real time data for safety on the sea (public, marine transport) and to improve Adriatic Sea monitoring and warning systems and is listed as part of the Croatian waterways safety system. Taking into account the huge multi-sensors data-streams from fixed buoy platforms, the data will be an important contribution to meteorological and oceanographic forecasting models and will provide validations for satellite products over the Adriatic Sea. The network should operate for at least 10 years and produce interdisciplinary data for the new Adriatic Sea climatology. The DHMZ met-ocean network has become standard for some similar platforms deployed from Croatian marine institutes in other parts of Adriatic Sea.

Figure 1 Croatian met-ocean data buoys at East Adriatic; positions and mooring depths.

The DHMZ met-ocean buoy network has been deployed in summer 2022. Buoys were tested during the winter-spring season, sensors were tuned. Severel updates were applied to the buoy data stream collection and pre-processing procedures. Futher challenges are the establishment of advanced first-error checks at the buoy data pre-processing system, as data are also disseminated via radio broadcasting to the nearby ships. The network has been running operationally from autumn 2023.

Met-ocean buoys data set.

- Meteo system – double sensors at 3m height
- Sea waves – directional sea waves
- Biochemical – dissolved oxygen, chlorophyll –a
- Optical – turbidity
- Sea currents – ADCP at buoy hull; profiling range 8 m - 80 m depths
- CTDO profiling – fix sensors; 4-6 per buoy; range 1.2m - 150m depths

Met-ocean buoys data hub and processing.

Each buoy has 2 telecommunication systems (GSM, Iridium satellite) to ensure real-time dissemination. Up to 10 minutes' interval data are sent to the data cloud and transmitted to the data-hub server. The stored CSV files undergo a preparation phase on the Linux server. Validation processes are applied to identify and handle duplicates within the data. Data types may be adjusted or transformed to ensure compatibility with the intended PostgreSQL database schema. A FastAPI web framework is used for the backend. The API returns data in a standardized format, making it easy for clients to parse and work with the information. The frontend solution is built using Next.js, React, and D3.js. Next.js facilitates server-side rendering for React applications. React is used for building the user interface, providing a dynamic and efficient rendering of components. D3.js is employed for data visualization, enabling the creation of interactive graphs that can be dynamically updated as new data arrives. The data is made accessible through a FastAPI backend, and a frontend solution enables users to perform data analytics through responsive graphs. This end-to-end pipeline ensures the availability and analysis of near real-time data from the buoy. Data are monitored by operators and internal users – marine forecasters.

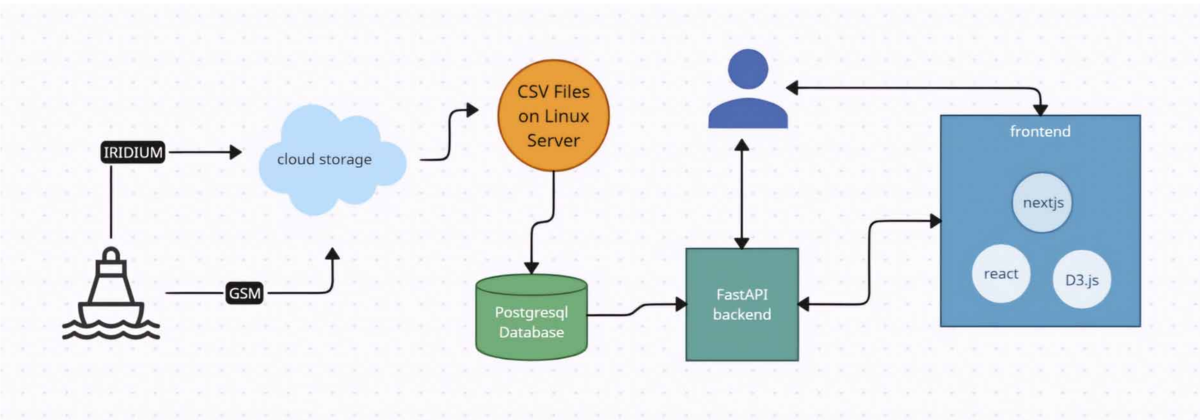


Figure 2 Buoys data processing.

Central DHMZ data integration platform.

The Met-ocean data buoy network is part of the new DHMZ meteorological network developed under the METMONIC project (Modernisation of the National Weather Observation Network in Croatia). As part of the project DHMZ established 390 new meteo stations, 6 meteo radars, 5 met-ocean open-sea buoys and 28 coastal sea-temperature buoys. All systems disseminate data in real time, at least every half hour. Data are collected at dedicated hubs and data bases. The public access to METMONIC data and to the other DHMZ historical meteo data are provided by CIP – Central Integrated Platform (https://dhmz-cip.igea.hr/). For the purpose of marine and general forecasters, graphics for 3-days data streams were made.

Sea wave observing system to support blue economy

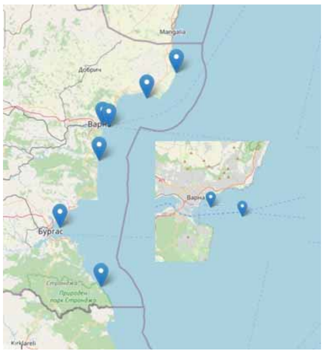
Veselka Marinova (marinova@io-bas.bg), Violeta Slabakova (v.slabakova@io-bas.bg), Atanas Palazov (palazov@io-bas.bg)

Institute of Oceanology (Bulgaria)

Introduction

The Bulgarian Sea Wave Observing System is jointly managed by the Institute of Oceanology of the Bulgarian Academy of Sciences (IO-BAS) and the National Institute of Meteorology and Hydrology (NIMH). The data provided by this system allows for real-time wave monitoring and helps improve marine forecasts. In addition to its scientific research contributions, the system is designed to support the sustainable development of the Bulgarian Black Sea coast and the blue economy. The system ensures operational information for various sectors, including marine energy, maritime transport, safety and rescue operations, fisheries, offshore facilities, coastal structures, tourism, and other coastal zone services, addressing the needs of all stakeholders.

Figure 1 Location of wave buoys on the Bulgarian.



System

The development of the system began in 2020 with the deployment of six moored wave buoys, three by IO-BAS and three by NIMH. The buoys were of two types: Spotter and Anteia. The following year, NIMH deployed an additional three Spotter buoys. In 2023, two new buoys were delivered by IO-BAS with the support of the DOORS EC project. The deployment positions were chosen to provide optimal coverage of the Bulgarian Black Sea coast (Figure 1). Some of the buoys were destroyed by passing vessels and were replaced. In some cases, vessels damaged the mooring system, causing the buoys to drift. Immediate action enabled their recovery and return to their original location. Based on our estimations, 6-7 buoys are sufficient to provide a comprehensive picture of sea waves along the Bulgarian coastal zone. Their position depends on user needs and service capabilities. The main challenge in maintaining the system is the intensive biofouling in the Black Sea, especially during spring and summer. Buoys and moorings need to be cleaned four to five times per year.

In situ wave measurement data

The wave buoys collect data on 3D displacement, sea surface temperature, and barometric pressure at the sea surface. Bulk wave parameters (significant wave height, wave period, wave direction, and directional spread) are computed over 30-minute intervals onboard the buoy. Additionally, wind speed and direction estimated from the measured wave spectra are provided. Data and bulk parameters are transmitted via GPRS or satellite connection every hour and stored in databases at both the Bulgarian National Oceanographic Data Center (BGODC) and the NIMH data center through API integration. Real-time wave data are accessible to stakeholders through websites hosted by IO-BAS and NIMH as shown in Figure 2. These portals allow users to directly monitor up-to-date information on wave conditions.

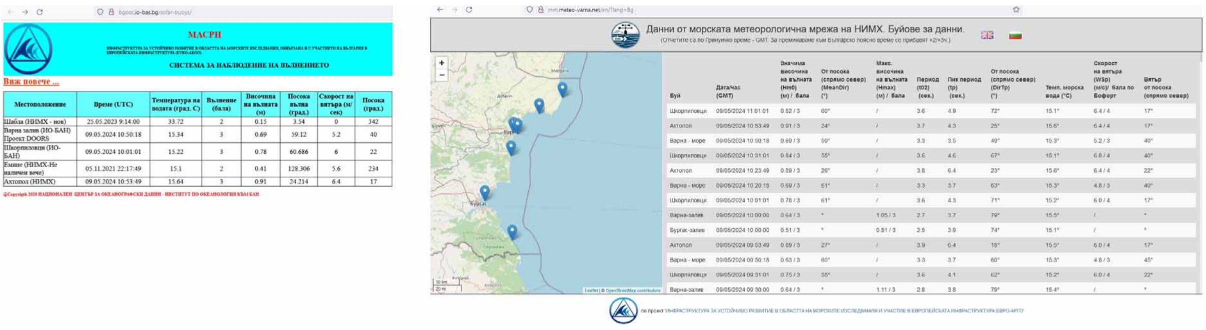


Figure 2 IO-BAS and NIMH websites.

The wave monitoring system has been operational for over three years, during which a substantial amount of data have been collected. Throughout this period, several high-energy events have occurred. Figure 3a shows the progression of the most recent significant storm, named Frederico (Linus), which made landfall on the Bulgarian Black Sea coast on November 18-19, 2023. With winds reaching almost hurricane strength of 90 km/h with gusts up to 144 km/h along the northern Bulgarian coast, this storm generated sea waves with a significant height exceeding 4.5 m. The storm resulted in extensive damage and human casualties along the Bulgarian Black Sea coast, attributed to the extraordinary wind and waves, compounded by heavy rainfall and snowfall.

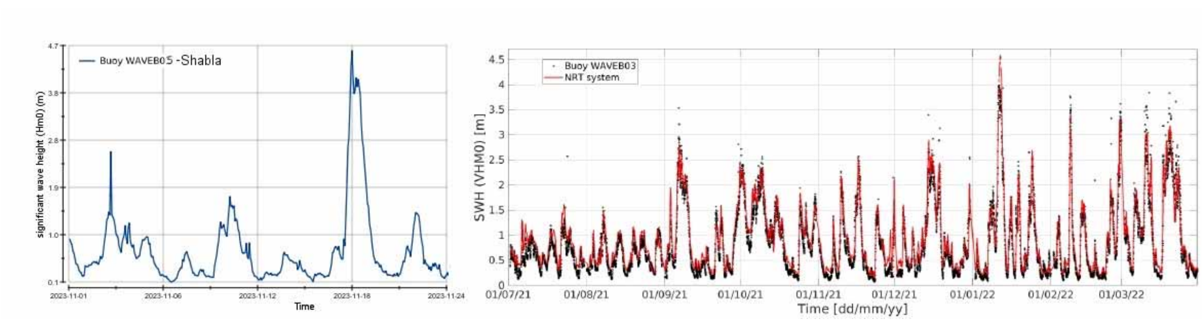


Figure 3 a) Storm Frederico registed by buoy WAVE05; b) wave model results validated by buoy WAVE03.

Data collected by these buoys are also distributed through the CMEMS In Situ products and are used to validate Black Sea wave model results in Copernicus Marine Services (Figure 3b, taken from CMEMS-BS-QUID-007-003, 2022).

Conclusions

The wave observing system stands as a unique source of in situ wave data in the Bulgarian Black Sea waters, offering both real-time wave data and long-term datasets crucial for scientific research and marine industry applications. During the period of operation, valuable experience has been accumulated in maintaining the system to ensure the provision of reliable sea wave data. Biofouling and vandalism are identified as the main factors impacting system performance.

Acknowledgements
Sea wave observing system is a part of Bulgarian National Operational Marine Observing System – NOMOS, which development and maintaining is supported by MASRI - Infrastructure for Sustainable Development of Marine Research and Participation in the European Infrastructure Euro-Argo, a project of the National roadmap for scientific Infrastructure (2020 – 2027) of Republic of Bulgaria. The building of the system was partly supported by European project DOORS - Developing Optimal and Open Research Support for the Black Sea.

References

CMEMS-BS-QUID-007-003, 2022
https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-BS-QUID-007-003.pdf

A Marine Spatial Data Infrastructure (MSDI) for the Italian National Future Biodiversity Centre (NBFC)

Valentina Grande[1] (valentina.grande@cnr.it), Federica Foglini[1] (federica.foglini@cnr.it), Mariacristina Prampolini[1,2] (mariacristina.prampolini@cnr.it), Giorgio Castellan[1,2] (giorgio.castellan@cnr.it), Simonetta Fraschetti[2,3] (simonetta.fraschetti@unina.it)

[1]Consiglio Nazionale delle Ricerche - Istituto di Scienze Marine- CNR - ISMAR (Italy)
[2]National Biodiversity Future Centre (Italy)
[3]University of Naples Federico II (Italy)

One of the main objectives of the National Biodiversity Future Centre (NBFC) is the integration and harmonization of the knowledge about the Italian marine biodiversity, considering also environmental variables, and human pressures. Data from institutions, national and international programs, literature, public marine data portals will be collected. The NBFC has also the ambition to review marine monitoring activities in Italy, to support the promotion and adoption of the most cost/effective new/emerging monitoring methodologies and technologies. This knowledge will be introduced into the National (Marine) Biodiversity Observatory System (NMBOS) and integrated in the Biodiversity Gateway, a tool to transform the scientific research in shared knowledge, fostering a better integration between science, policy and general public with the final aim to improve conservation, restoration, and valorization of marine biodiversity.
In this framework, a solid and shared data management becomes crucial. This study describes the architecture and the implementation of a Marine Spatial Data Infrastructure (MSDI) designed to collect, homogenize, integrate, share and access in an interoperable way: biodiversity data, environmental variables, and human pressures, data coming from the pilot studies, information about biological samples for biodiversity monitoring at the species/habitat level and at ecosystem/seascape level. The MSDI allows scientists to store and access data in a systematic and standardized way, to capitalize on this information to reach biodiversity targets and to implement maritime spatial plans in Italy. The architecture is designed to fit all standards and best practices for data collection, integration, availability, and shareability, following the FAIR principles (Findability, Accessibility, Interoperability, Reusability).
The MSDI consists of five main components:

- Cloud system
- Database Management System (DBMS)
- Geoportal
- Metadata catalogue
- Web services

Cloud system
The cloud system consists of a NAS-QNAP with QTS as the operating system and myQNAPcloud for remote access. It enables data providers to store digital resources in an organized file system reachable by the web using credentials and, to associate a URL and password for sharing them.

Database Management System (DBMS)
The RDBMS is based on an Oracle spatial database managed through ArcSDE and ArcGIS desktop. The structure of the relational database follows a specific data model based on the INSPIRE Directive data specifications and customized in the framework of several research projects dealing with conservation and maritime spatial planning.

Geoportal

The Geoportal integrates biodiversity data, environmental variables, and human pressures; data coming from the pilot studies; information about biological samples for biodiversity monitoring at the population/species level, and at ecosystem/seascape level. The data stored in the DBMS are transformed into thematic layers with specific portrayal rules, cataloged using the Moka Content Management System (CMS) (© Semenda s.r.l.), organized in thematic groups (such as Protected sites, Habitats and Biotopes, and Human activities), and published with ArcGIS Server in the Geoportal.

Metadata catalogue

Layers, digital data, and physical samples are described in the GeoNetwork metadata catalogue using the standard ISO19115 for spatial datasets, ISO 19115-2 for Imagery and gridded data, ISO19119 for web services, Dublin Core for a non-spatial resource (such as samples, figures, and tables). The catalogue is the main interface for search, find, share, preserve and harvest metadata, containing information such as abstract, points of contact, distribution format, data quality, data history, use and access policies, Digital Persistent Identifiers (PID), link for the data download.

Web services

Interoperability between systems is guaranteed by the publication of Web Map Services (WMS) permitting the integration of the geodata in other platforms. The Catalogue Service for the Web (CSW) allows the harvesting of the metadata in other metadata catalogues.

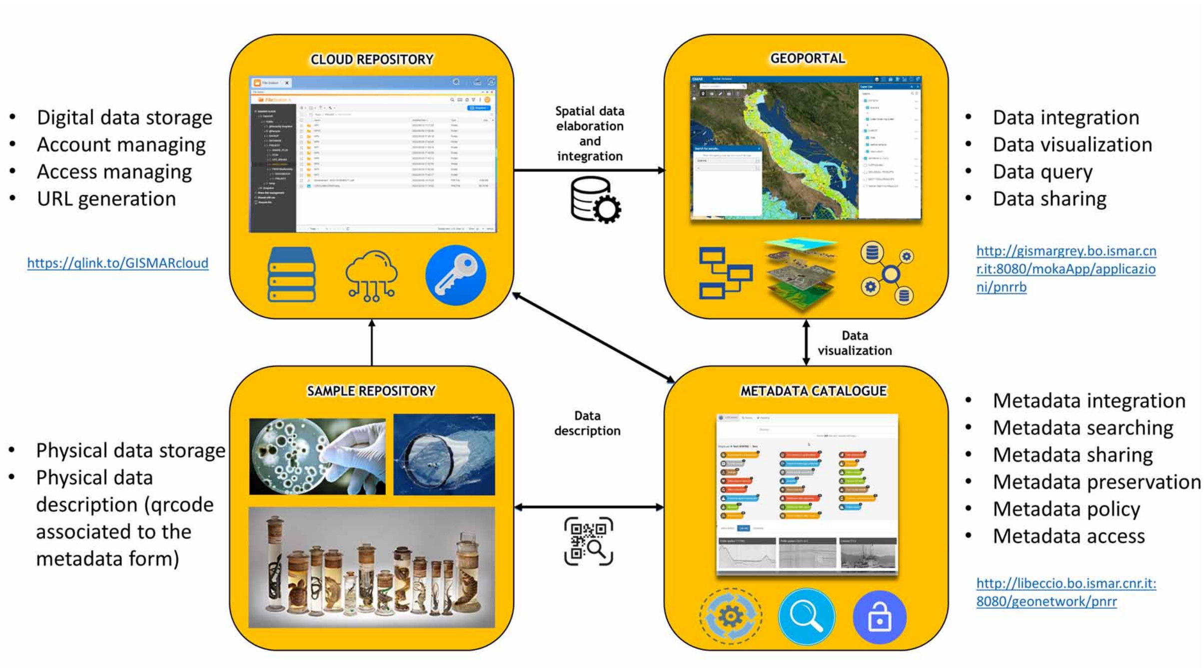


Figure 3 Sketch of the MSDI implemented for the Italian NBFC.

Multidisciplinary environmental marine portal in the southeastern Bay of Biscay (e-begi)

Lohitzune Solabarrieta (lsolabarrieta@azti.es), Andrea del Campo (adelcampo@azti.es), Ane Iriondo (airiondo@azti.es), Asier Nieto (anieto@azti.es), Beatriz Sobradillo (bsobradillo@azti.es), Irati Epelde (iepelde@azti.es), Irene Ruiz (iruiz@azti.es), Isabel García Barón (igarcia@azti.es), Ivan Manso-Narvarte (imanso@azti.es), Izaskun Zorita (izorita@azti.es), Javier Franco (jafranco@azti.es), Julien Mader (jmader@azti.es), Manuel Gonzalez (mgonzalez@azti.es), Maria Santos (msantos@azti.es), Pedro Liria (pliria@azti.es), Anna Rubio (arubio@azti.es)

AZTI (Spain)

Introduction

In 2022, AZTI initiated the e-begi project, with the aim of optimizing communication and coordination in the development of a common scientific and technological strategy between different observation disciplines towards the establishment of a multidisciplinary observatory. One of e-begi project main expected outcomes is to enhance the provision of Marine Ecosystem Data (both historical and real-time) that respond to the needs associated with ecosystem-based management and other demands in the fields of conservation and recovery of Biodiversity and Habitats, the challenges of Climate and Global Changes and the implementation of Policies and Directives on the management of the marine environment.

The first development achieved within the framework of the e-begi project, has been the development of the e-begi dashboard (https://aztidata.es/ebegi/) (Figure 1), which integrates all the metadata of the existing measurement systems and activities in the area.

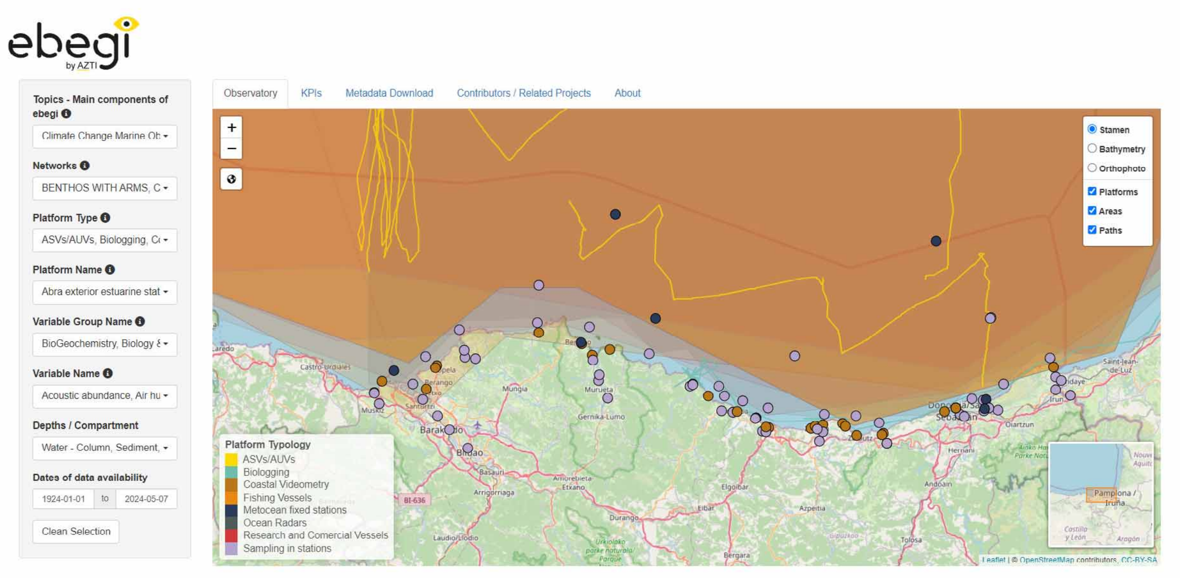


Figure 1 Screen captures of e-begi dashboard: available at: aztidata.es/ebegi.

This online application is a first step towards identifying additional measurement needs and integration opportunities to build an optimized observatory. It also provides visibility to the available information collected by AZTI for decades in its oceanographic surveys and observation activities, information that is key for the actors involved in the management of the marine environment. E-begi dashboard alows the exploration, selection and download of the metadata of the multidisciplinary and multiplatform observational datasets at the core of the e-begi project. Furthermore, e-begi dashboard offers the possibility to create plots of the main KPI, based on the selection of the user, who can filter by area, time, variables, topics, depths, ... (Figure 2)

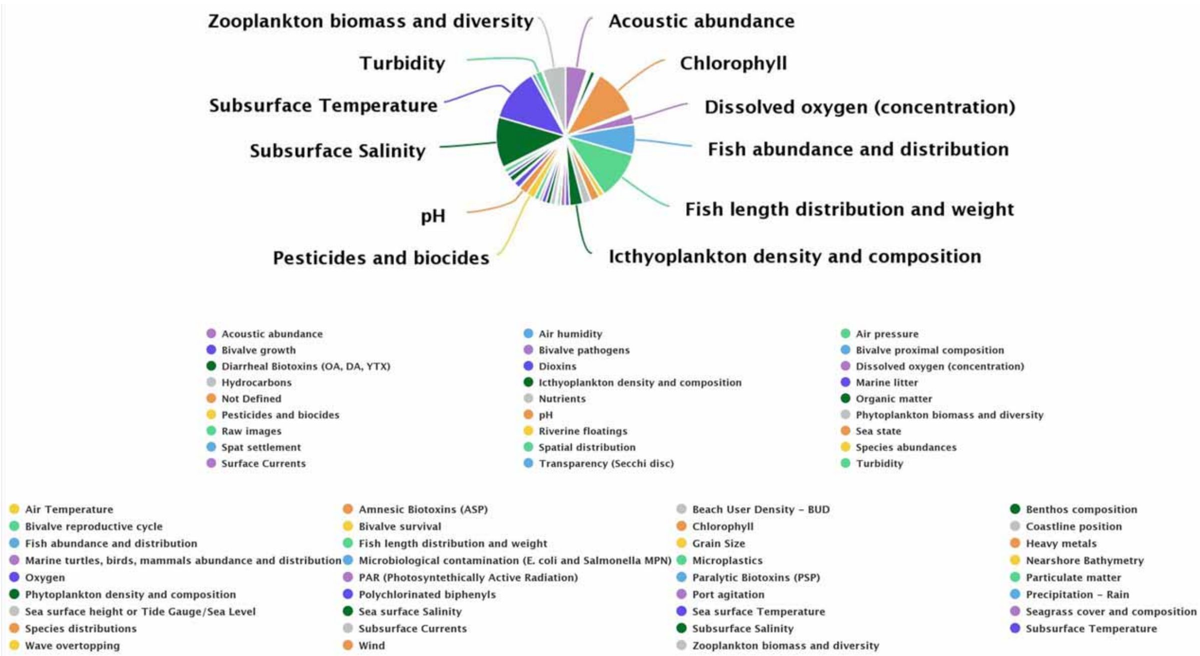


Figure 2 Example of a plot for some of the variables included in the ebegi database. The most common ones (by number of measuring stations/surveys or locations of continuous monitoring) are highlighted in the upper panel.

e-begi Strategy

The e-begi project strategy is based on the following principles:

- Promotion of open data publication and the use of DOIs for e-begi components and outputs (data and data products, software, documentation).
- Co-design and development of new data products (mainly oriented to Marine Strategy Framework Directive MSFD and fisheries).
- Reinforce a multiplatform and multidisciplinary sampling strategy, ensuring a continuous and sustained observations over time.

The next steps in the development of the e-begi dashboard are: (i) to optimise FAIRness of the dashboard, adopting global standards and vocabularies, (ii) to share metadata with the European data management networks and (iii) evolve into a machine accessible metadata model and interface.

EMODnet Chemistry, the data infrastructure for seawater quality: key services and latest results

Alessandra Giorgetti[1] (agiorgetti@ogs.it), Chiara Altobelli[1] (caltobelli@ogs.it), Dick Schaap[2] (dick@maris.nl), EMODnet Chemistry Consortium

[1]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[2]MARIS (The Netherlands)

Introduction

Launched by the European Commission in 2009, the European Marine Observation and Data Network (EMODnet) is a network of organisations working to gather, process the data according to international standards and make them freely available. EMODnet offers a single access point for submission and access to the data and products managed by the EMODnet Ingestion Service and the 7 thematic groups: Bathymetry, Biology, Chemistry, Geology, Human Activities, Physics and Seabed Habitats. EMODnet Chemistry mission is providing easy access to marine chemical data that support the implementation of the EU policy. The EMODnet Chemistry network consists of 66 data centres in 32 countries that collect and manage data to meet the FAIR principles, which require findable, accessible, interoperable and reusable data. The partnership collect and validate data on eutrophication and ocean acidification, in the water column, contaminants, including marine litter, in water, biota, and sediment and then make them freely accessible and interoperable. Secondly, EMODnet Chemistry generates and publishes standardised, harmonised and quality-controlled data collections and derived maps.

Main services

The EMODnet Products Catalogue offers functions for searching and downloading the metadata of chemical data collections, the associated maps and the data used to create them.
The Map Viewer Service allows users to search, view, and query EMODnet maps.
There are also specific EMODnet Chemistry services. These include the Data Discovery and Access Service and the webODV Data Explorer and Extractor tool. The former is the SeaDataNet interface for searching for measurement data, narrowing down queries and downloading selected chemical data. The webODV Data Explorer and Extractor allows users to create subsets, perform scientific analysis, create various graphs and download datasets in different formats from EMODnet Chemistry's validated, aggregated data collections.

Latest achievements

EMODnet Chemistry fully integrated in the EMODnet central portal

EMODnet partnership has developed a unique web portal under the domain of the European Commission. Launched in January 2023, the portal provides a more efficient, integrated and fit-for-purpose services for industry, policy makers and scientific data users.

The quantity and quality of data has steadily increased

The number of data records has risen from around 1,000,000 to 1,267,321 (as of March 2024). These data have been harmonised, standardised and validated to create data collections for many groups of variables, e.g., fertilisers, chlorophyll, acidity, heavy metals, hydrocarbons, marine litter, per each European sea basin. Data collections for beach litter, seafloor litter and micro-litter in the sea surface and sediment was also carried out at European level.
As for quality control, a new questionnaire for contaminant data has been added to improve

transparency of data quality, reliability and comparability. The questionnaire contains a record of field and laboratory information for chemical parameters submitted by a data producer. Forty-two questionnaires were received from 21 institutes and 18 countries.

The number and variety of maps has grown.
Twenty-five new maps for contaminants were created on the basis of the data collected (Figure 1, left). The new maps show the spatial and temporal variability of the sampling stations for twelve pollutants and their concentrations measured in the water column, in biota and in sediment. Five new maps were created using the webODV Explorer and Extractor Service (Figure 1, right). These are based on the latest data collections on eutrophication and are published as an OGC WMS-WFS service in the EMODnet Map viewer.

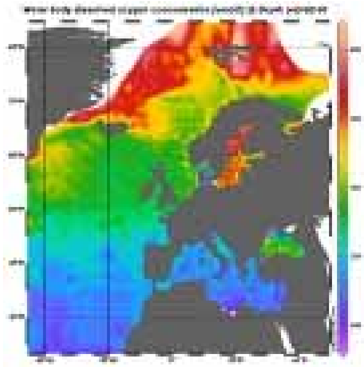


Figure 1 Left - Benzo(a)pyrene concentration in water, sediment and biota (median values of the last 6 years, using the percentile median ranges). Right - webODV map of water body dissolved oxygen concentration at 20 meter b.s.l.

Finally, 42 marine litter maps were also created (Figure 2). These include for the first time 16 maps for floating microlitter in all European sea basins and beyond and three new maps for seafloor litter.

Figure 2 Density of floating microlitter from research and monitoring protocols.

Conclusion
In 2023, the present EMODnet Chemistry partnership received a rewarding confirmation from the European Commission for the next two years and is currently discussing the long-term vision. The network is also strengthening cooperation with experts from the Regional Seas

Conventions, the European Environment Agency and the Joint Implementation Strategy for the MSFD. Finally, the partnership is working towards full interoperability with the services of other EMODnet thematic groups and with other data infrastructures, e.g. Copernicus, the Marine Environment Database of the International Council for the Exploration of the Sea and the World Ocean Database.

SESSION SERVICES ORAL PRESENTATIONS

Ireland's Digital Ocean enhancing Ocean Sustainability and Informed Decision-Making through Ocean Digital Twins and Effective Data Management

Eoin O'Grady (eoin.ogrady@marine.ie), Tara Keena (tara.keena@marine.ie), Andrew Conway (andrew.conway@marine.ie)

Marine Institute (Ireland)

Introduction

As offshore activities continue to grow, coastal cities and ports expand, and changes to climate affect the environment, the need to better understand our marine and coastal areas is becoming increasingly evident. The capability for responsive evidence-based decision-making is critical. Legislative and strategic drivers such as the European Union's Maritime Spatial Planning Directive, Ireland's new national strategy for offshore renewable energy the Offshore Renewal Energy Development Plan 2, and the digitalisation of maritime transport and ports, require the development of an integrated and evidence-based approach to manage marine areas effectively. To meet these requirements in Ireland, the Marine Institute is developing advanced digital systems, such as Ocean Digital Twins. Underpinned by established data acquisition programmes and effective data management, Irish marine data and domain expertise can be leveraged to support the delivery of information used for informed decision making. This strategic initiative is happening in the wider context of the European Digital Roadmap and Ireland's national Harnessing Digital Framework and the United Nations Decade of Ocean Science for Sustainable Development.

Understanding Digital Twins and their role in Informed Decision-Making

An ocean digital twin is a virtual representation of an ocean environment integrating a range of data to simulate, model, and represent various aspects of the ocean. It can be used for scientific research, environmental monitoring, climate modelling, and sustainable resource management, supporting use cases including for marine spatial planning, for aquaculture and food safety, for climate adaptation and offshore renewable energy development, and for transport and logistics, among others. Overall, ocean digital twins can provide valuable answers to "*what-is*" and "*what-if*" questions in a way that supports informed decision making of marine-focused stakeholders.

Challenges: Data Quest

Challenges associated with Digital Twins range from the quest of suitable data, data quality and interoperability [Tzachor et al., 2023], to a clear definition of what a Digital Twin is [Jones et al., 2020]. Other challenges include cross-disciplinary collaborations, for example science, technology, and policy communities. In Ireland, there are a range of organisations acquiring marine data for specific research, public sector and commercial purposes. These data are often captured for a single purpose and not reused. The current fragmented landscape hinders decision making. There is growing coordination at an EU level, with a focus on regional models and services such as through Destination Earth and EMODNET. However, programmes need to build on national capabilities to provide high resolution data for local decision making. Marine-related data from non-traditional sources, including NGOs and industry, can enrich the view of what is happening but are underutilised. Therefore, a coordinated national approach to marine data and digital services is needed to add significant value to marine research, innovation and the development of new marine-related public services.

"Repurposing the Wheel, Not Reinventing": A Development Pathway for Ireland's Ocean Digital Twins

The development of Ireland's Ocean Digital Twins will leverage existing capabilities, frameworks and data. This includes the use of existing tools and services such as *Ireland's Marine Atlas*, the *Digital Ocean Portal*, the *Marine Institute Data Catalogue and* related federated data services, as well as the Marine Institute's *IODE-accredited Data Management Quality Management Framework (DM-QMF).*

Existing collaborative processes will connect various stakeholders and sectors enabling innovation with an ecosystem of internal and external partners. The accuracy, coverage, and capabilities of Ocean Digital Twins and related services will be developed through thematic use cases and with nationally and internationally collaboration.

Data Integration and Interoperability is required to ensure that data from various sources can be integrated. Data acquisition programmes from a range of organisations contribute data, including sensor observations from data buoys and tide gauges, data from a range of oceanographic models, and archive data including underpinning seabed data, from multiple organisations. The integration and application of these diverse inputs requires the use of developing new digital technologies.

Ireland's existing Digital Ocean Programme aims to provide easy access to marine information in ways that end-users can understand and engage with. Together, the Digital Ocean Programme with developing Ocean Digital Twin capabilities, can help to bridge the science user divide, making marine information more readily available and directly useful to policy makers, businesses, scientists and citizens through visually engaging and interactive digital representations of our marine spaces.

Effective data management practices are also vital in improving the reliability and accuracy of the data that underpins a digital twin. Data Governance, reuse and provenance within the Marine Institute is managed by the IODE Accredited Data Management Quality Management Framework and related systems and practices.

Conclusion

The development by Ireland's Digital Ocean programme of new digital twin capabilities represents a significant step towards better management of Ireland's marine and coastal regions. This approach provides new insights into the current state of our marine environment and enables scenario modelling for future planning, helping to design ways to plan for future uses, to restore marine and coastal habitats, and to support the development of a sustainable blue economy. While acknowledging many challenges lie ahead, it is important to recognise that fundamental building blocks and efforts already exist that can be built upon. Leveraging existing data acquisition chains, data management practises, digital capabilities, and scientific and marine domain expert co-creation and collaborative processes, can provide a strong foundation for delivering on the promise of Ocean Digital Twins.

References

Jones D., Snider C., Nassehi A., Yon J., Hicks B., (2020). *Characterising the Digital Twin: A systematic literature review*. CIRP Journal of Manufacturing Science and Technology. 29. https://doi.org/10.1016/j.cirpj.2020.02.002.

Tzachor A., Hendel O., Richards C.E., (2023). *Digital twins: a stepping stone to achieve ocean sustainability?* npj Ocean Sustain 2, 16, https://doi.org/10.1038/s44183-023-00023-9

A Comprehensive Approach to Marine Biodiversity Data Validation: The Biocheck Tool

Ruben Perez Perez (ruben.perez@vliz.be), Bart Vanhoorne (bart.vanhoorne@vliz.be), Lynn Delgat (lynn.delgat@vliz.be), Joana Beja (joana.beja@vliz.be), Leen Vandepitte (leen.vandepitte@vliz.be)

VLIZ, Flanders Marine Institute (Belgium)

Rationale

Precise and detailed marine biodiversity occurrence data plays a vital role in comprehending and forecasting species population trends. This information serves as the foundation for ecological and conservation research. EMODnet Biology (European Marine Observation and Data Network - Biology lot), in collaboration with other international initiatives such as the Ocean Biodiversity Information System (OBIS) and the Global Biodiversity Information Facility (GBIF), is dedicated to gathering and publishing high-quality biodiversity and sample related data. Maintaining data quality requires careful curation and validation, which can be challenging due to scattered and non-standardized quality check tools tailored to specific user requirements.

The Biocheck tool

The Biocheck tool is a web application that brings the capabilities of the EMODnetBiocheck R package to the wider public, regardless of whether they have knowledge of the R programming language or not. Both the web application and the R package expand on and are built upon the obistools R package developed by OBIS and are a result of the collaborative work of the LifeWatch Belgium and EMODnet Biology projects. The primary objective of this tool is to conduct a thorough exploration and comprehensive quality control of marine biodiversity occurrence (meta)data, aligning with the widely recognized Darwin Core biodiversity data standard. Consequently, it serves as the default quality control tool for all types of marine biodiversity occurrence data.

The Biocheck tool performs a series of essential quality checks, including:

- assessment of (meta)data format and integrity;
- taxonomic verification, using the World Register of Marine Species (WoRMS) as reference;
- evaluation of marine and coastal biogeography;
- scrutiny of parameters linked to occurrences, encompassing organism quantifications and facts, biometrics data and environmental and sampling-related data.

Conclusion

The tool generates exploratory maps and graphs tailored to each dataset, providing valuable insights into the data. Additionally, it compiles and highlights a comprehensive list of records that fail the quality checks, not meeting the required quality standards. The Biocheck tool thus allows its users to identify and fix potentially problematic records and to increase the overall quality of their datasets before or after publication. Therefore, holding a crucial position in enhancing the accuracy and trustworthiness of marine biodiversity occurrence data. This, in turn, aids to our comprehension of species populations and supports well-informed decisions in conservation and ecological matters.

Figure 1 Illustration on the Biocheck tool procedures from Perez Perez et al. [2023].

References

OBIS. https://obis.org
GBIF. https://www.gbif.org
Biocheck tool. https://rshiny.lifewatch.be/BioCheck
EMODnetBiocheck R package. https://github.com/EMODnet/EMODnetBiocheck
Darwin Core. https://dwc.tdwg.org
LifeWatch Belgium. https://www.lifewatch.be
EMODnet Biology. https://emodnet.ec.europa.eu/en/biology
Perez Perez et al., (2023). *The Biocheck tool: Enhancing data quality in marine biodiversity research.* Poster in EMODnet Open Conference 2023.

BEACON - High performance data access supporting the marine data lake

Peter Thijsse (peter@maris.nl), Robin Kooyman (robin@maris.nl), Tjerk Krijger (tjerk@maris.nl)

MARIS (The Netherlands)

In many of the societal and scientific challenges, such as Digital Twins of the Oceans and virtual research environments, fast access to a large number of multidisciplinary data resources is key. However, achieving performance is a major challenge as original data is in many cases organised in millions of observation files which makes it hard to achieve fast responses. Next to this, data from different domains are stored in a large variety of data infrastructures, each with their own data-access mechanisms, which causes researchers to spend much time on trying to access relevant data. In a perfect world, users should be able to retrieve data in a uniform way from different data infrastructures following their selection criteria, including for example spatial or temporal boundaries, parameter types, depth ranges and other filters. Therefore, as part of the EOSC Future and Blue-Cloud projects, MARIS has started developing a software system called "BEACON" with a unique indexing system that can, on the fly with incredible performance, extract specific data based on the user's request from millions of observational data files containing multiple parameters in diverse units.
The BEACON system has a core written in RUST (low-level coding language) and its indexed data can be accessed via a REST API that is exposed by BEACON itself meaning clients can query data via a simple JSON request. The system is built in a way that it returns one single harmonised file as output, regardless of whether the input contains many different data types or dimensions. It also allows for converting the units of the original data if parameters are measured in different types of units (for this it e.g., makes use of the NERC Vocabulary Server (NVS) and I-Adopt framework). It is important to mention that the system can be applied to different data infrastructures and is not tailor made for one specific type of database.
Showcasing the performance and usability of BEACON, as part of the EOSC-Future Environmental Dashboard, the BEACON system is applied to the SeaDataNet CDI database, Euro-ARGO and the ERA5 dataset from the Climate Data Store.

SeaDataNet CDI example
For CDI, an example workflow using BEACON has enabled obtaining specific data on-the-fly based on a set of filter options (parameter selection, bounded box with minimum and maximum longitudes and latitudes, depth range and time period). For the parameter selection and aggregation, the P01 vocabulary from the NERC Vocabulary Server is used, as the data within the SeaDataNet CDI is mapped with these elements that are identifiers of physical parameters such as Temperature or Oxygen. The data can be retrieved in formats such as one single flattened NetCDF or JSON with dimensions LONGITUDE, LATITUDE, PARAMETER VALUE, TIME and DEPTH. In this example, this data is used as input towards the Marine Data Viewer, which is elaborated on below.

Marine Data Viewer
The user interface of the Marine Data Viewer (https://eosc-future.maris.nl/ see Figure 1) is designed to allow (citizen) scientists to interact with the large data collections and retrieve parameter values from observation data. Enabled by the performance of BEACON, the user can filter the data on-the-fly using sliders for date, time and depth. At present, the ocean variables concern temperature, oxygen, nutrients and pH measurements, from Euro-Argo and SeaDataNet. The in-situ values are overlayed at the same time and space with product layers from Copernicus Marine, based upon

modelling and satellite data. The related in-situ data sets concern single observations such as profiles, tracks, analysed water samples, etc which are not available as continuous time series. Moreover, the spatial coverage varies in time, density, and numbers. The in-situ observations can be plotted on the map by selecting an EOV, time period (years or months) and depth range (e.g., [0, 5] m). The user can also select multiple EOVs at once, creating an aggregated data collection.

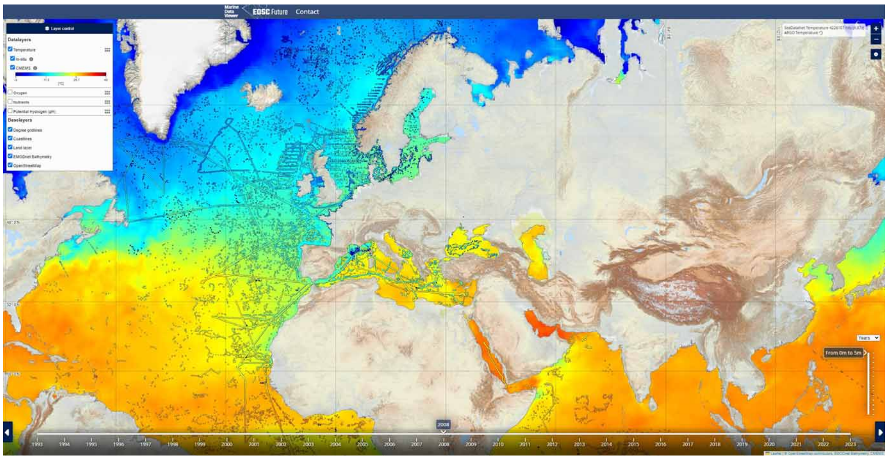


Figure 1 Marine data viewer.

A colorbar is displayed with a clear colour scheme, depicting the unit of the EOV and the corresponding values. Depending on the EOV a linear or log scale is used for visual purposes. There is a layer system in place where the user can choose the desired hierarchy that determines which data should be shown if different parameters are at the same location.
After the user has selected the parameters, the number of measurements retrieved from the SeaDataNet and Euro-Argo databases can be seen in the top right corner. Here, users are also able to click on the information icon in the top right and then click on the map near measurements to get more detailed information. The information is stored in a table, and when the measurements concern Argo floats, the user is able to click on the profile info to get the information of the whole depth profile. Or when it concerns SeaDataNet measurements the user is redirected to the corresponding CDI landing page.
The viewer also includes background layers from CMEMS that include modelling or satellite products. This is done by using OGC Web Map Service (WMS), which is a standard protocol providing a simple HTTPs interface for requesting geo-registered map images. The in-situ data can be plotted on top of these products as seen in the figure above, such that a comparison can be made.

IMDIS Session
During the session at IMDIS the presenter will update the audience on the mentioned developments, including the latest results and how the BEACON software could be accessed and used by developers.

References
Vermeulen A., Schaap D., Adamaki A., Krijger T., Bardaji R., Forno, A., Moncoiffé G., (2023). *Dashboard for the State of the Environment*. ARPHA Preprints. https://doi.org/10.3897/arphapreprints.e107578

EMODnet - Ingestion and safekeeping of marine data sets

Anastasia Iona[1] (sissy@hnodc.hcmr.gr), Dick Schaap[2] (dick@maris.nl)

[1]HCMR (Greece)
[2]MARIS (The Netherlands)

Introduction

Many data collected by public authorities, researchers and private operators of coastal or offshore facilities still do not arrive to national or regional repositories and are thus unavailable to potential users. This creates additional costs for those working on marine issues who will have the choice of accepting lower confidence in their analysis than would otherwise be the case, or being compelled to needlessly repeat observations. There is therefore the need to streamline the data ingestion process so that data holders from public and private sectors can easily release their data for safekeeping and subsequent distribution through EMODnet or other means.
EMODnet Ingestion is a key pillar of the European Marine Observation and Data network (EMODnet). It seeks to identify and reach out to organisations from research, public, and private sectors who are holding marine datasets and who are not yet connected to the European marine data exchange and contributing to EMODnet. EMODnet Ingestion motivates and supports those potential data providers for releasing their datasets for safekeeping and subsequent free distribution and publication through EMODnet. This can concern historic data sets that can become part of large European archives that might be of use for many applications. This can also concern operational oceanography data streams from monitoring platforms that can become part of the European operational oceanography data exchange for feeding forecasting models and supporting various operations.

Network of EMODnet ambassadors

EMODnet Ingestion activities are undertaken by a European network of 43 organisations (marine research institutes, governmental agencies, and Small or Medium-sized Enterprises (SME's)) from 27 coastal countries. Geographically, the network has nodes in the countries around all European marine basins and it covers all EMODnet data themes. Most partners are data centres and are qualified as National Oceanographic Data Centres (NODC), National Geological Surveys, Biology Institutes, or as National Hydrographic Agencies. Moreover, the consortium includes coordinators of all EMODnet thematic groups, who involve also their networks. EMODnet Ingestion has many contacts at national level for identifying, encouraging, and supporting potential data providers. In addition, it has many experts in marine data management for EMODnet thematic disciplines, and it has working relations with the leading European marine data management infrastructures, that are pillars under EMODnet. Promotion is made in various ways, through presentations at conferences, articles in publications, and webinars for specific target groups, like the offshore renewable energy sector, marine aqua culture sector, various European marine Research and Technological Development (RTD) projects, citizen science groups, and by personal contacts.

Data Ingestion processes

EMODnet Ingestion operates services for streamlining the data ingestion process from data providers to the leading European marine data management infrastructures that are feeding EMODnet. These comprise Pan-European infrastructure for ocean & marine data management (SeaDataNet), European node of the Ocean Biodiversity Information System (EurOBIS), Copernicus Marine, European Global Ocean Observing System (EuroGOOS), International

Council for the Exploration of the Sea (ICES), and some others. EMODnet Ingestion provides opportunities for data providers for contributing to EMODnet data products, which are popular by many users, whereby data providers will be acknowledged. Also, it offers long term safekeeping of their data by inclusion in national and European infrastructures and subsequent elaboration of their data into standard formats and quality validation. A core service is the Data Submission service which facilitates data providers to submit their delayed mode data sets. A low threshold is offered by splitting the completion of the submission form in 2 parts, whereby a data submitter only completes a part of the metadata together with the uploading of a data package. Each data submission is then assigned to a competent data centre for completing the metadata of the submission. As a next step, assigned data centres elaborate selected submissions further to make (subsets of) the data fit for population into national, regional, European and EMODnet thematic portals. This depends on data centres assessing the added-value of the submitted data and the efforts needed for elaborating the data to common formats, if anyhow possible. Elaboration includes activities like review, validation, conversions to standard formats, and further population to the relevant European infrastructures.

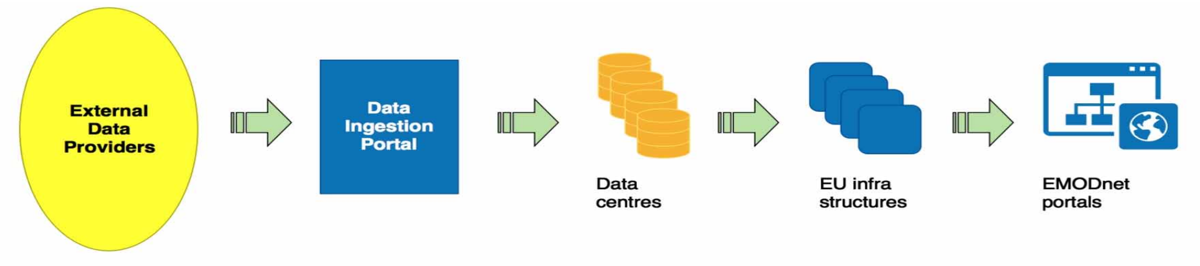


Figure 1 Workflow from submission to elaborating and processing for publishing in EMODnet.

Another aspect, is the facilitation of Near Real Time operational data streams, establishing direct data exchange connections with (networks of) monitoring instruments. This is approached by developing and establishing machine-to-machine transfers between monitoring instruments and repositories and by working out a scenario whereby monitoring data remain at the source repository and are made available to EMODnet by services. The operational oceanography exchange is done by EMODnet Ingestion together with EMODnet Physics.

Key Performance Indicators

Since the establishment of EMODnet Ingestion in 2016, 1385 data submissions have been received, processed, and published as part of the Ingestion Viewing service. Of those, so far **633** have been elaborated and included in the European data management infrastructures that feed into EMODnet. These submissions originate from circa 200 Organizations from Research, Government, Business, and NGOs. The majority (circa 60%) is from the Academic and Research Institutes, while there is great potential for growth from Non-Governmental Organization (NGOs) through the collaboration of EMODnet Ingestion with these communities, in particular for marine litter data.

Overall, there is great momentum and also direct machine-to-machine exchanges are being deployed with other Ingestion services, at national level (such as the Marine Data Exchange service (UK)) and European level (such as the SeaDataNet SEA scieNtific Open data Edition (SEANOE) Data Citing and Publishing service). This leads to waking up ever more data and data providers, that find their way to the European marine data exchange.

The data conversion process for the M-VRE webODV

Ingrid Linck Rosenhaim (ingrid.linckrosenhaim@awi.de), Sebastian Mieruch-Schnülle (sebastian.mieruch@awi.de), Reiner Schlitzer (reiner.schlitzer@awi.de)

Alfred Wegener Institute (Germany)

Introduction

webODV is the online version of Ocean Data View (ODV, https://odv.awi.de), an established software for analyzing and visualizing oceanographic and georeferenced data, and, therefore, suitable for the visualization and analysis of the Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) Expedition data. The MOSAiC Expedition had the German icebreaker Polarstern frozen in the Arctic Sea ice for a year (2019-2020) following the Transpolar Drift. During MOSAiC, scientists from different countries and research disciplines collected an incredible amount of measurements now archived in Pangaea (https://pangaea.de) and available as open-source data. webODV (https://webodv.awi.de) is an interactive and powerful tool accessible via the browser in a user-friendly virtual environment, where MOSAiC data is already uploaded, and analysis and visualizations can be directly performed (https://mvre.webodv.cloud.awi.de).

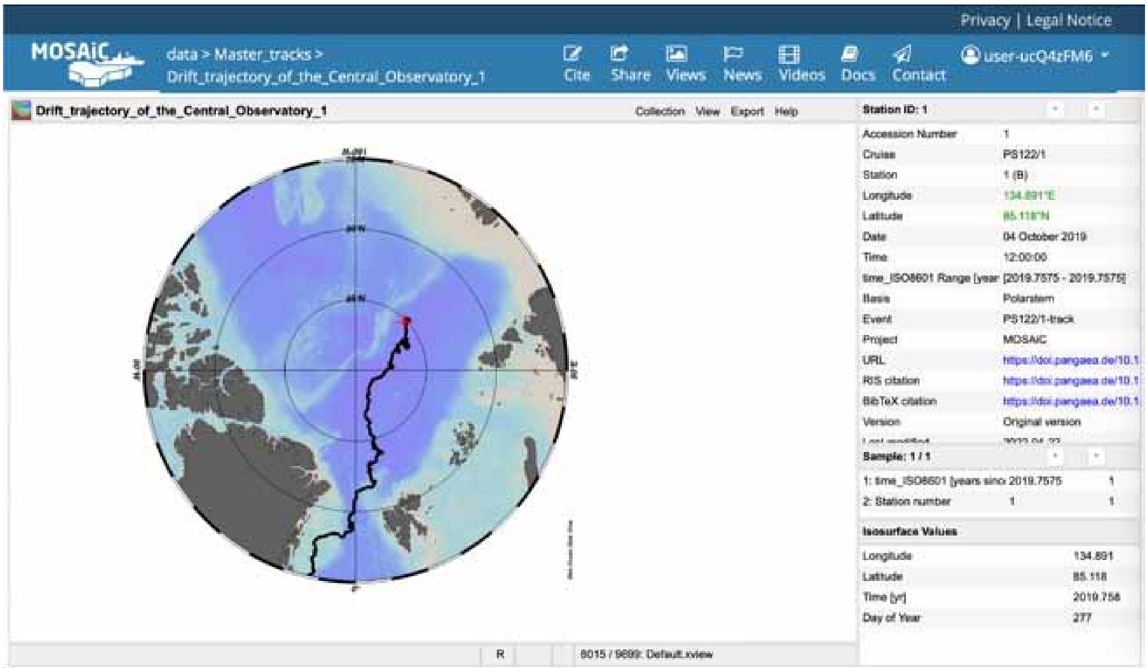


Figure 1 Default view of the drift of Polarstern in the MOSAiC Central Observatory 1 [Nicolaus et al., 2021] in webODV.

webODV and FAIR data

webODV follows the concept of FAIR (Findable, Accessible, Interoperable, and Reusable) data. Findable: every data set inside a collection comes with a direct link to itself in the Pangaea archive. Accessible: data collections are similar measurements combined into one file. However, these data can be analyzed and visualized individually by filtering the data. The collections are held up-to-date with the data, metadata, and data references in Pangaea. Interoperable: Data collections are provided in webODV for download in three consistent formats: text, ODV, and netCDF. Reusable: copies of analysis and visualizations of collections created in webODV are easily shared with co-authors and colleagues via links. These are easily replicated without changing the analysis and visualizations of the sender. All metadata and references to Pangaea are provided within the data collections in webODV to ensure transparency and traceability of the data source and authors. Hence, after analyzing the data collection, the visualizations created in webODV are ready to be used in publications, and the citations to the data in Pangaea, can be directly downloaded from webODV. Thus, webODV contributes to increasing the visibility and access to MOSAiC Expedition measurements.

MOSAiC measurements in webODV

The MOSAiC measurements in webODV are presented as data collections and separated into categories like Atmosphere, Ocean, Sea ice, and others. For the exploration, visualization, and analyses of data in webODV, the data are converted from the tab format used by Pangaea in an ODV readable format. An automatic process queries, filters, and downloads the measurements from the Pangaea archive onto our server, where they are compiled and aggregated into collections before being converted to the ODV format.

Data conversion to ODV format

An advanced search of MOSAiC measurement entries is performed in the Pangaea archive using the Python class panquery. The measurements are then scanned and filtered, and only the ones consistent with ODV are selected for download. During the download process, the Python code pangaea2odv[37] generates a text file with the contents of the Pangaea tab and xml files with a header containing the data source details, a short description of the measurements, MetaVariables, and DataVariables, as ODV requires, and the table with all meta- and data variables values. The text file created by the pangaea2odv process contains the following MetaVariables: Basis, Cruise, Event, Station, Project, URL, RIS and BibTeX citations, Version, Last modified, Scientists, Main scientist, Contact, Method, Bottom Depth [m], Original file URL, Longitude and Latitude, and all DataVariables from the original tab file. Similar measurements are gathered in collections whose names correspond to the title given in Pangaea.

Most of the downloaded MOSAiC data, after the procedure described above, are ready to be aggregated in collections and converted to ODV format. However, the MOSAiC data are not only diverse in scientific topics and different measurements, but they are also different in format and structure. Hence, some data sets require an extra conversion, like measurements stored in netCDF, tar, and gz files or others. These data receive a unique script that downloads the files, reads the data, and writes a text file containing all meta- and data variables. Some data, due to its structure, requires a unique conversion. For instance, when a variable is measured at different depths, and each depth is a column, these data are transformed and presented in two columns: one has the depth values and the other the value of the measurement in the respective depth. This type of conversion is necessary due to how ODV works with the data. Data containing statistical analysis results, e.g., standard deviation values, also receive a different conversion. However, only the variable label changes, so ODV knows this variable is a statistical analysis value. The MOSAiC data in webODV must be similar, if not identical, to the original data published in Pangaea. No data values are modified in the conversion process, only the data format. After the conversion from tab to text format, a script performs the data aggregation of each collection into one single text file, making these collections ready for conversion to the ODV format. This step generates the .odv file and Data folder for each data collection, which, after thorough technical verification, are uploaded into webODV.

[37] Pangaea2odv is a Python code written by Dr. Roland Koppe, Alfred Wegener Institute, Bremerhaven.

References

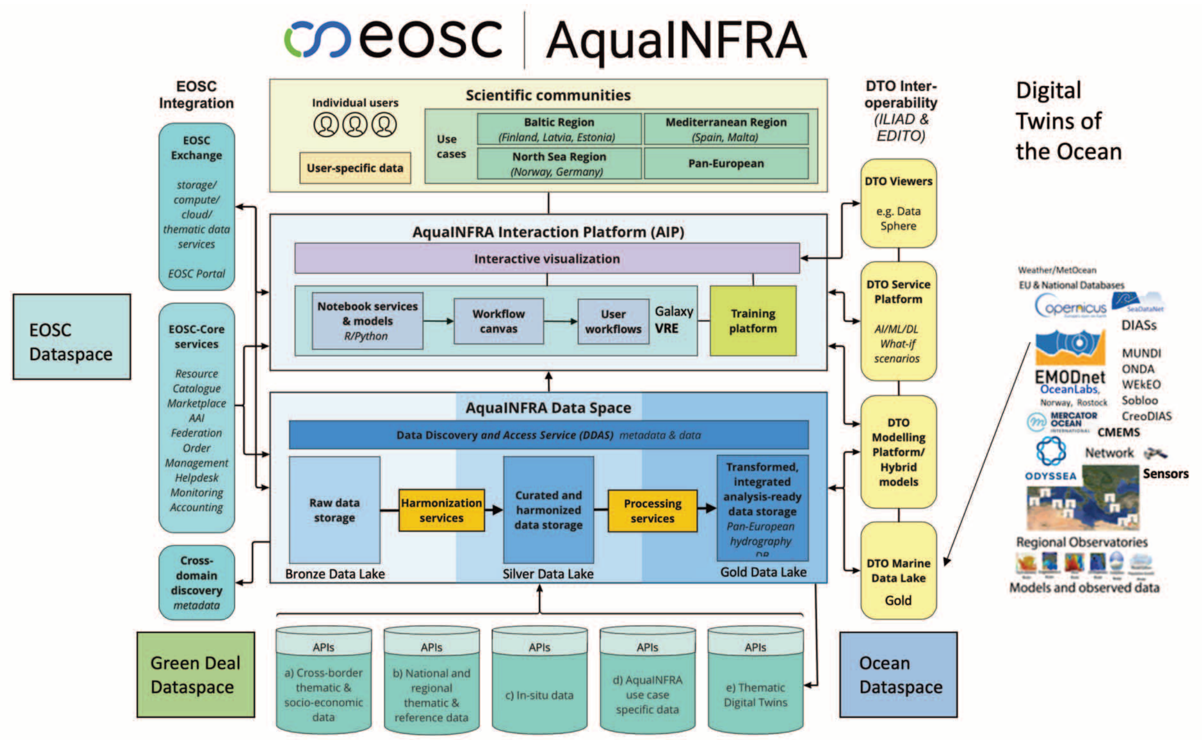
Nicolaus M., Rieman-Campe K., Bliss A., Hutchings J.K., Granskog M.A., Haas C., Hoppman M., Kanzow T., Krishfield R.A., Lei R., Rex M., Li T., Rabe B., (2021). *Drift trajectory of the Central Observatory 1 (CO1) of the Distributed Network of MOSAiC 2019/2020*. Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, PANGAEA, https://doi.org/10.1594/PANGAEA.937184

The AquaINFRA Interaction Platform: Access to FAIR Multi-disciplinary Freshwater and Marine Data and Services to Restore our Ocean and Waters by 2030

Henning Sten[1] (hsh@plan.aau.dk), Keeble Simon[2] (simon@bluelobster.co.uk), Keeble Kathryn[2] (kathryn@bluelobster.co.uk)

[1]Hansen, Aalborg University Copenhagen (Denmark)
[2]Blue Lobster IT Limited (UK)

The AquaINFRA project is a transformative initiative aimed at restoring healthy inland waters, coasts, seas and oceans, addressing the urgent need for high-quality, multi-disciplinary data. A virtual environment will equip freshwater and marine scientists and stakeholders with FAIR (Findable, Accessible, Interoperable, and Reusable) data services, fostering collaboration across research infrastructures and disciplines. Leveraging on the European Open Science Cloud (EOSC) and other operational dataspaces, AquaINFRA aspires to transcend national borders, merging freshwater and marine research communities to work conjointly. The project will provide an EOSC-compliant research infrastructure, augmenting a cross-domain search and discovery mechanism alongside tools for spatio-temporal analysis and modelling through a Virtual Research Environment (VRE).



With strategic use cases across various European waters, AquaINFRA is positioned to co-design and test its services. Its ultimate contribution to the EU's Mission of restoring marine and freshwater ecosystems aligns with the call to reduce human pressure on these environments. AquaINFRA will deliver a suite of resources, including the intelligent Data Discovery and Access Service (DDAS) utilising hydrological connectivity as the backbone of the search engine, harmonisation and processing services, and a robust interactive platform, all consistent with the EOSC interoperability framework.

AquaINFRA encompasses integrating existing domain services with EOSC core services, providing an enhanced user experience, and facilitating the discovery of datasets in a FAIR manner via the EOSC-Portal and marketplace. Key outputs involve a catalogue of integrated components, the establishment of a seamless pan-European hydrography for environmental analysis, and the demonstration of improved research practices through the integration of diverse data types.
The technical methodology of the AquaINFRA project is a multi-faceted approach to achieving interoperability, usability, and service integration within a virtual environment aimed at enhancing the research capacity for marine and freshwater sciences.

Data Harmonisation and Integration
One of the cornerstones of AquaINFRA is data harmonisation, essential for integrating diverse datasets from disparate sources, ensuring consistency and compatibility across different systems. The process will involve standardising data formats, metadata, and vocabularies to align with FAIR principles. The harmonised data, once processed, will feed into the AquaINFRA Data Space. A subsequent level of integration will attach this harmonised data to a high-resolution pan-European hydrography database. This will ensure that the resulting dataset is not only of high quality but also reflects the spatial-temporal connectivity and lag-effects inherent in aquatic systems.

Virtual Research Environment (VRE)
The AquaINFRA Virtual Research Environment (VRE) will provide researchers with computational tools and services necessary for sophisticated data analysis and modelling. The VRE will support a range of programming languages, such as R and Python, and will include notebook services that offer customisable workflows. These workflows will facilitate the analysis of raw or harmonised data, allowing researchers to create tailored processes for their specific research needs. The VRE will be intuitive, with a workflow canvas that visually represents analytical processes, enhancing transparency and understanding of the data analysis pathways.

Interoperable Architecture and Standards
A key aspect of the technical methodology is to adhere to and promote interoperability standards. The AquaINFRA architecture will utilise international standards such as those from ISO TC211, the Open Geospatial Consortium (OGC), and the EU INSPIRE Directive, ensuring that data are handled in a way that promotes broad usability and integration into global systems. Moreover, AquaINFRA will contribute to the development of emerging standards such as Environmental Data Retrieval API and Cloud-Native Geospatial, indicating the project's commitment to staying at the forefront of technological advancements in data handling.

High-Performance Computing Integration
To manage the extensive computational demands of data processing, harmonisation, and analysis, AquaINFRA will leverage the capacities of the AAU High-Performance Computing Centre (CLAAUDIA). The integration of AquaINFRA services with high-performance computing resources is vital to ensure that complex models and large datasets can be processed efficiently. This will enable real-time data analysis and facilitate the handling of large-scale hydrological models, which is particularly critical when dealing with high-resolution, pan-European datasets.

Interdisciplinary Cross-Domain Discovery Services
The technical framework is designed to enhance cross-domain research, enabling seamless discovery and access to marine and freshwater data via interdisciplinary discovery services such as EUDAT B2FIND, OpenAIRE EXPLORE, and others. This will be enabled through the DDAS, which will act as a bridge connecting diverse data catalogues, APIs, and Digital Twins, allowing

users to navigate through a comprehensive data landscape without the barriers typically posed by domain-specific silos.

Training and Capacity Building

AquaINFRA will develop a suite of educational materials and resources, which will be disseminated through the project's training platform. This will ensure that users not only have access to the data and tools they need but also possess the skills and knowledge necessary to utilise these resources effectively, promoting a sustainable culture of open science within the marine and freshwater research communities.

The combination of these technical methodologies underscores AquaINFRA's commitment to providing a robust, user-friendly, and innovative environment that enables and propels forward the research in marine and freshwater sciences. Through leveraging high-performance computing, adhering to and advancing standards, and promoting open science and FAIR data principles, AquaINFRA aims to be a transformative force in aquatic research infrastructures.

Blue-Cloud 2026 - Federating European FAIR and Open Research Ecosystems for Marine data

Rita Meneses[1] (r.meneses@trust-itservices.com), Sara Pittonet[1] (s.pittonet@trust-itservices.com), Dick Schaap[2] (dick@maris.nl)

[1]Trust-IT Services (Italy)
[2]MARIS (The Netherlands)

Ocean is a key component of climate change. The Ocean provides 50% of the oxygen we breathe, is absorbing 90% of excess heat, and it is absorbing 1/3 of CO_2 emissions. If we want to understand the relationship between the ocean and climate change, we need data, long time series data, and we need to be sure we understand them in cross-disciplinary contexts.
Data collection, analysis & application are key enablers for Ocean data innovation. Integration is needed, open science is one of those transformative changes that the ocean needs. Research infrastructures (RI) are key elements to implement open science and are fundamental fuel of data to the nascent Digital Twin of the Ocean (DTO) and Digital Twins of the Oceans (DITTO). Distributed RI work as symbiotic ecosystems for effective implementation of open science. RIs provide data to different sectors of society, from coastal research to beach safety to sustainable fisheries to marine heatwaves, from carbon monitoring to citizen science applications. Within RI "ocean integration" can actually take place for enhanced science and responding to society challenges.

Advancing ocean research through cloud-based data integration and Open Science
The Horizon Europe Blue-Cloud initiative developed a collaborative web-based environment that enables open and simplified access to an unprecedented wealth of marine data resources and interoperable tools. By federating different "blue" Research Infrastructures (EurOBIS, Euro-Argo, ELIXIR-ENA, SOCAT, EcoTaxa, ICOS-Ocean, EMSO, SIOS, and ELIXIR-MGnify), data infrastructures (SeaDataNet), major European data initiatives (EMODnet) and e-infrastructures such WEkEO (CMEMS DIAS), Blue-Cloud is contributing to increasing the available data volume and to an easier discovery and access of blue data. This is open to researchers, industry and society from Europe and beyond, to address issues related to climate change, food, biodiversity conservation, sustainable ocean economy, pollution and natural hazards.
The Blue-Cloud services improve the quality of data available thanks to harmonisation, validation, qualification and assessed interoperability of the various sources as provided by each federated Blue Data Infrastructures (BDIs). It benefits researchers working at research infrastructures worldwide, managers of ocean data repositories and research performing organisations, data providers and data scientists active in ocean science; policy makers at ocean science research infrastructures worldwide.
Blue-Cloud 2026 "de facto" federation approach taking place at 3 levels: the data, the computing resources and the analytical services, is already recognised as a success story at European level, as mentioned by representatives of the European Commission and European Open Science (EOSC) policy actors. It's important to highlight that the work completed under Blue-Cloud on harmonisation of data discovery and retrieval services across its infrastructures bring key data sets to direct availability for the data lake of the European DTO and to feed European data sets into the global DITTO initiative. Furthermore, Blue-Cloud 2026 has also taken a pivotal step by pledging its support to the EU Mission "Restore our Ocean and Waters" charter.

High standard suite of Marine Data Services for open science Blue-Collaboration
The current Blue-Cloud technical framework is extensible and open by design, constantly

evolving according to the needs of the community, facilitating collaborative research and the uptake of Open Science principles, through a distinguished set of marine data services.

- **Data Discovery & Access Service (DD&AS)**: an easy and FAIR service for discovering and retrieving multi-disciplinary data sets and products managed and provided by Blue Data Infrastructures. The federation facilitates sharing of datasets as input for analytical and visualisation services and applications, that are hosted and further developed in the Blue-Cloud Virtual Research Environment.
- **Virtual Research Environment (VRE)**: an Open Science platform for collaborative marine research, using a wide variety of datasets and analytical tools, complemented by generic services, such as sub-setting, pre-processing, harmonising, publishing and visualisation. The VRE hosts different Virtual Labs and is going to include thematic Workbenches, which users can access with existing credentials in EOSC, the European Open Science Cloud. Multi-disciplinary datasets retrieved from the Blue-Cloud DD&AS can be exploited in the VRE. All methods and services in the Catalogue are exchanged with the EOSC Portal Catalogue & Marketplace.
- **Thematic Virtual Labs (VLabs)**: where researchers work closely together with the Blue-Cloud 2026 technical team to describe Virtual Lab workflows and technical requirements, in order to implement them in the Blue-Cloud VRE and further test its capabilities on specific topics. The following Vlabs are available: Carbon-Plankton Dynamics, Global Fisheries Atlas, Coastal currents from observations, Integration of coastal ocean observations along Europe, Plankton Genomics, Marine Environmental Indicators, Aquaculture Monitor, Fish, a matter of scales and Zoo and Phytoplankton EOV products.
- **Workbenches for Essential Ocean Variables (EOVs)**: a number of data-intensive Workbenches for selected EOVs are being where ocean and data scientists can implement efficient workflows that allow them to harmonise, validate and qualify large and various in situ data sources, exploiting the blue analytical services available in the Blue-Cloud VRE. The following EOVs are available: Ecosystem-level EOVs, Eutrophication: chlorophyll, nutrients, oxygen and Physics: temperature & salinity.

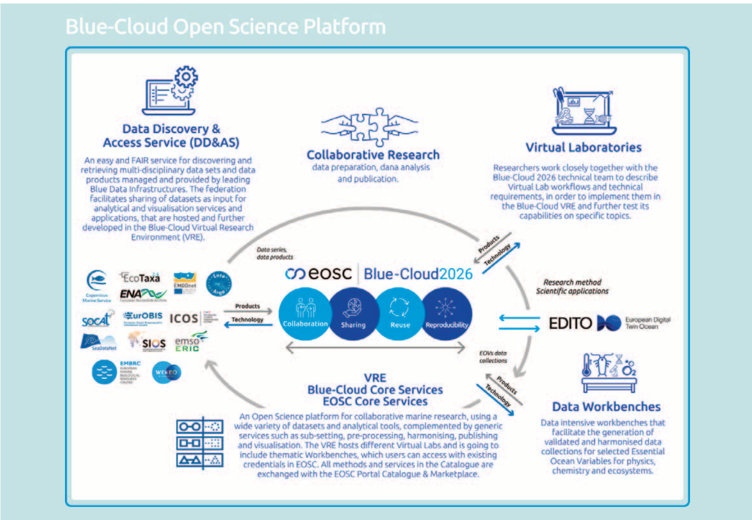


Figure 1 Blue-Cloud 2026 Vlabs & EOVs.

To support the marine data community in exploiting to the fullest all Blue-Cloud 2026 services, users can count on the Blue-Cloud Training Academy, a space with lessons and materials to guide on how to utilize Blue-Cloud services for Open Science in marine research. Dedicated webinars to debate around the uptake of Open Science practices and FAIR data principles in the marine domain are also organised.

The Galaxy platform as a ground breaking FAIR tool for Earth System's analytical workflows

Marie Jossé[1] (marie.josse@ifremer.fr), Jérôme Detoc[2] (jerome.detoc@ifremer.fr), Erwan Bodéré[2] (erwan.bodere@ifremer.fr)

[1]CNRS (France)
[2]Ifremer (France)

FAIR-EASE, an EOSC project

In today's interconnected world, collaborative efforts among scientists across diverse fields are increasingly urgent. The FAIR-EASE project (https://fairease.eu/) addresses the challenges of data accessibility and fosters integrated uses by opening gateways for earth and environmental sciences. The project's primary objective is to enable communication, integration, and data processing for multidisciplinary use cases applications across domains. It focuses on different domains of study providing a comprehensive assessment of the Earth System.
The Earth System is a complex and dynamic system that encompasses the interactions between the atmosphere, oceans, land, and biosphere. Understanding and analyzing data from the Earth System Model is essential, for example to predict and mitigate the impacts of climate change. The ES use cases that the project tries to implement on a cloud-based environement include coastal water dynamics, ocean bio-geochemical in-situ data, marine omics observations in the marine domain but also volcano activities and land degradation.

Galaxy and analytical workflows

Thereby, FAIR-EASE took a particular interest into the Galaxy platform. It's an open-source platform and can be used as an IT toolkit to meet the multidisciplinary data visualisation, analyse and processing needs of the different domain chosen for their diversity in terms of fields studied. By design, Galaxy manage data by sharing and publishing results, workflows, and visualizations, ensuring reproducibility by capturing the necessary information to repeat and understand data analyses. Therefore, with FAIR-EASE and Galaxy Earth System we aim at directing users toward standardized tools that can be plugged into cross-domains workflows (see an example of a worklow in Figure 1).

Galaxy Training Network

The Galaxy Training Network significantly contributes to enhancing the accessibility and reusability of tools and workflows. The Galaxy Training platform (available at training.galaxyproject.org) hosts an extensive collection of tutorials authored by administrators, developers, and contributors. These tutorials serve as valuable resources for individuals seeking to learn how to navigate Galaxy, employ specific functionalities like Interactive Tools or how to execute workflows for specific analyses. By mixing trainings and tools in a same friendly user webapp, Galaxy is a tool perfectly suited for open science.

Galaxy Earth System

A first step in FAIR-EASE was the creation of a Galaxy declination for Earth System studies (earth-system.usegalaxy.eu) with dedicated models, data, tools and data visualisation. It will make Earth System analysing more accessible to researchers in different fields.

- Within the FAIR-EASE use cases, 3 pilots are related to marine and ocean studies:
- Coastal Water Dynamics: focuses on the coastal marine environment near river estuaries, where important processes take place.

- Ocean Bio-Geochemical Observations: addresses questions regarding the health of marine ecosystems (e.g., ocean acidification, ...) and needs for ocean resource management.
- Marine Omics: analyses of spatial- and time-comparable marine microbial metagenomics data sets for the exploration of biodiversity and its correlations with environmental quality.

To address these ocean and marine studies, multiple tools have been, are, and will be implemented into earth-system.usegalaxy.eu. The aim is to create full operationnal workflows and answering users'needs. Some of these tools are:

- DIVAnd (Data-Interpolating Variational Analysis in n dimensions) performs an n-dimensional variational analysis/gridding of arbitrarily located observations (Barth et al. [2014], https://doi.org/10.5194/gmd-7-225-2014).
- QGIS is a geographic information system (GIS, https://qgis.org/en/site/) to visualize, manage, edit, analyse spatial files, and compose printable maps.
- Scoop (Detoc et al. [2021], SEANOE, https://doi.org/10.17882/48531): visual quality control of a series of Argo floats NetCDF files.
- Canyon-B (Bittig et al. [2018], https://doi.org/10.3389/fmars.2018.00328): alternative to static climatologies provides robust estimation of open ocean CO_2 variables and nutrient concentrations from T, S and O_2 data using Bayesian neural networks.
- Ocean Data View (Schlitzer R. [2023], ODV, https://odv.awi.de/): interactive exploration, analysis and visualization of oceanographic and other geo-referenced profile, time-series, trajectory or sequence data.

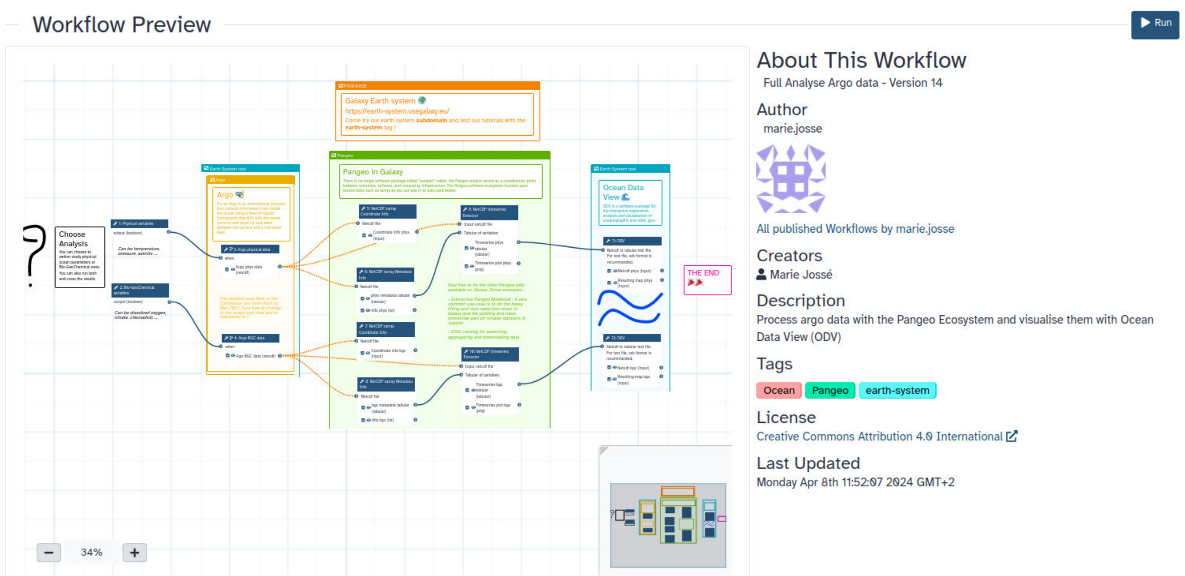


Figure 1 A Galaxy workflow to retrieve, analyse, and visualise oceanographic data.

In conclusion, during this presentation we will introduce the main principles of Galaxy and then do a focus on how to profit of galaxy to analyse oceanographic data with recently implemented workflows (one example in Figure 1). This talk aims at discovering Galaxy from a user perspective, without diving into the complexity of development and deployement.

D4Science: Advancing Ocean Science Through Collaborative Data Analysis

Massimiliano Assante (massimiliano.assante@isti.cnr.it), Leonardo Candela (leonardo.candela@isti.cnr.it), Luca Frosini (luca.frosini@isti.cnr.it), Francesco Mangiacrapa (francesco.mangiacraoa@isti.cnr.it), Elisa Molinaro (elisa.molinaro@isti.cnr.i), Pasquale Pagano (pasquale.pagano@isti.cnr.it)

Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche - CNR-ISTI (Italy)

Introduction

In the realm of ocean science, addressing intricate challenges necessitates collaborative analysis of extensive datasets. This underscores the significance of infrastructures that facilitate multidisciplinary collaboration, effective communication, and timely data sharing. D4Science [Assante et al., 2019], an operational infrastructure initiated 18 years ago with European Commission funding, has evolved into an efficient solution. Utilizing the "as a Service" paradigm, D4Science provides web-accessible Virtual Laboratories [Assante et al., 2023; Candela et al., 2023] (VLabs) that proved to be also suitable for ocean science collaboration [Schaap et al., 2022]. These VLabs simplify access to marine datasets, concealing underlying complexities. Key functionalities include a cloud-based Workspace for file organization, a platform for large-scale data analysis on a distributed computing infrastructure, a catalog for publishing research results, and a communication system based on social network practices.

D4Science has been actively supporting diverse marine and ocean science Virtual Laboratories (VLabs), adapting to evolving research needs. Notable initiatives include contributions to the European Open Science Cloud (EOSC), starting with the 'Blue-Cloud' project in 2020 and its subsequent extension, 'Blue-Cloud2026.' In 2015, D4Science played a pivotal role in the BlueBRIDGE Horizon 2020 Project, which aimed to provide user-friendly data services and tools for the aquaculture, fisheries, and environmental sectors. Additionally, in 2013, D4Science contributed to the iMarine FP7 Project, which has since evolved into the current iMarine initiative. This ongoing effort is dedicated to establishing and operating an e-infrastructure that aligns with the principles of the ecosystem approach to fisheries management and the conservation of marine living resources, further supporting the Food and Agriculture Organization's (FAO) Blue Growth Initiative.

D4Science is currently supporting over 20 scientific communities and over 150 VLabs, and pioneers Open Science in ocean research. It fosters collaboration, offers user-friendly environments, and provides service for accessing, sharing, analyzing, and publishing oceanographic data. A detailed description of these services is given in the following.

D4Science services overview for Ocean Science Virtual Laboratories

The D4Science services are instrumental in advancing Open Science practices within VLabs, empowering researchers to harness the advantages of state-of-the-art e-infrastructures. By leveraging these services, ocean science researchers can capitalise on the power of the Cloud and of e-infrastructures, driving scientific progress and enabling collaborative research efforts within the realm of Open Science.

The D4Science services offer a comprehensive array of features, fostering collaboration, facilitating data analytics, enabling result dissemination, and ensuring seamless integration with external systems. In fact, they cater to the entirety of the research lifecycle, providing diverse services. Specifically, *(i)* the *Collaborative Storage Framework* fosters collaboration among VLab users. The Workspace provides a platform for VLab members to collaborate, share resources, and work together on projects. This collaborative environment enhances the efficiency and effectiveness of research activities within VLabs. In terms of data analytics, *(ii)* the *Analytics Engine Framework* empowers VLabs with powerful tools and resources, *(iii)* the *Publishing Framework* within VLabs facilitates the dissemination of research outcomes by means of the Metadata Catalogue and the Spatial Data Catalogue, that provide a means to organise and publish research results, making them accessible to the wider scientific community. This framework ensures transparency, reproducibility, and the sharing of valuable knowledge generated within VLabs.

These services are made available either by default or through specific requests. Every VLab can be equipped with:

- **Communication area** for collaborative and open discussions on any topic and disseminating information of interest for the community, for example, the availability of a research outcome;
- **Administration area.** User and groups Management dashboard for managing membership and roles;
- **Analytics Computing Framework area:**
  - The D4Science Analytics Engine, specifically developed to leverage advancements in IT and software engineering over the past decade where analytical methods can be integrated by means of the Container technology. The engine provides multiple execution infrastructures selectable by users, uses OGC API Processes (JSON) as a standard protocol, integrates with various CVS systems, and supports parallel executions. It offers many out-of-the-box methods, with a focus on flexibility and ease, provides JupyterLab and RStudio integrations, a web component-based front-end technology, and streamlined code generation tools.
  - **RStudio** allows users to perform online statistical analyses. Rstudio is no longer shared and it is now persistent. It offers a predefined list of R packages, and each VLab can define its RStudio servers configuration, Standard (4 cores/8GB RAM) and Large (8 cores/32GB RAM).
  - **JupyterLab**, a web-based interactive development environment for Jupyter notebooks, code, and data. It allows users to configure and arrange the user interface to support a wide range of workflows in data science, scientific computing, and machine learning.
- **Workspace cloud storage** for storing, organising and sharing items, such as datasets, scripts or outputs. This cloud-based solution provides two storage options: the Workspace and the Dataspace. Both can be accessed through RStudio and JupyterHub. The computations run in the analytical services can take inputs from the Workspace.
- **Spatial Data Infrastructure (SDI) services** provides users with the capability to store, discover, access, and manage vectoral and raster georeferenced datasets. The SDI exploits the following technologies: GeoServer equipped with PostgreSQL and PostGIS, GeoNetwork, Thredds.
- **Catalogue Framework** to document and publish any generated research product. Its primary component is the VRE Data Catalogue. The VRE Data Catalogue service is a catalogue service built on open-source technology for data catalogues (CKAN ckan.org). Via this Catalogue users can also search and browse data, products, and resources (posters, deliverables, etc) of interest from the Ocean Science community.

References

Assante M. et al., (2019). *Enacting open science by D4Science*. Future Gener. Comput. Syst. 101: 555-563, https://doi.org/10.1016/j.future.2019.05.063

Assante M. et al., (2023). *Virtual research environments co-creation: The D4Science experience*. Concurrency Computat Pract Exper. 2023; 35(18):e6925, https://doi.org/10.1002/cpe.6925

Candela L., Castelli D., and Pagano P., (2023). *The D4Science Experience on Virtual Research Environments Development*. In IEEE Computing in Science & Engineering, https://doi.org/10.1109/MCSE.2023.3290433

Schaap D., Assante M., Pagano P., and Candela L., (2022). *Blue-Cloud: Exploring and demonstrating the potential of Open Science for ocean sustainability*. IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea), Milazzo, Italy, pp. 198-202, https://doi.org/10.1109/MetroSea55331.2022.9950819

Eurofleets+ Virtual playground

Paolo Diviacco (pdiviacco@ogs.it), Alessandro Busato (busato@ogs.it)

Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)

Introduction

Modern scientific research is increasingly complex and cross-disciplinary. No scientist, anymore, can cover all activities and topics alone. Collaborative research is invoked then to overcome such limitations. Scientific collaborative research is a wide area that spans several thematic fields and that, due to the introduction of Information Technologies (IT) and the Internet, is changing radically towards e-research: a perspective where the life of researchers is augmented by on-line interactions, high computing power, communities and knowledge.

Virtual Research Environments

Current trends in collaborative science support focus mostly on the technical issues related to the consolidation of data in a remote facility, where it can be processed using top computing resources and from where data can be retrieved many times and by multiple researchers. With some specific extensions this is the perspective commonly known as Virtual Research Environments (VRE).

We propose to extend the notion of VRE towards that of a Virtual Playground (VP) aiming to pave the road to a new vision of collaborative science that aims to bridge the gaps between different scientific domains and ways of thinking while at the same time preserving peculiarities and backgrounds. In this perspective, since, by definition, "collaboration" targets a common goal, a common cognitive space (the VP) must be created, maintained and referred to by all partners.

Boundary objects

Promoting and providing a common space for cross-domain collaboration bring several difficulties that originate in the different semantic and pragmatic backgrounds of participants. To address such issues contemporary Sociology of Science advocate the use of artifacts called "boundary objects". These can provide a common path to share meaning and practice which could, eventually, allow researchers to collaborate even when coming from different perspectives. Our work proposes to instantiate boundary objects as conceptual maps that using nodes and their relations can represent meaning. Nodes can be considered repositories for data (files) and information (messages exchanged among team members). In this perspective the system with its map together with data, information and knowledge that populate each node, becomes the current status of knowledge of a project or of the designated team/community.

Eurofleets+ Virtual Playground

The EU Horizon 2020 EurofleetsPlus project aims to facilitate open access to an integrated and advanced research vessel fleet, designed to meet the evolving and challenging needs of the user community. Among EF+ and Joint Research Activities, OGS extended the already existing web based collaborative tool it developed, named COLLA [Diviacco, 2018; Diviacco et al., 2015; 2014] to obtain a fully functional VP with a particular focus on Geophysics, but that can be applied to any other scientific field. The EF+ VP offers the possibility to create projects where Principal Investigators are enabled to upload an ontology that is visualized as a concept map acting as the boundary object of the collaborative project (Figure 1 left). The VP offers tools to work together interacting with the data (viewing, analysing, annotating) and communicating and discussing with the other members of the team (messaging, forums, posting, Figure 1 right).

Concept map/ontologies are currently created externally using Protégé, while a specific internal tool will be made available soon. Navigating the conceptual map users can identify and access the project's data that can be then processed using a dedicated script based notebook facility based on Apache Zeppelin. This tool allows to use several programming languages and script. As a testbed for the framework we have developed a specific set of tools to process seismic data directly from within EF+ VP, which provided good results and have been used also during EF+ training workshops.

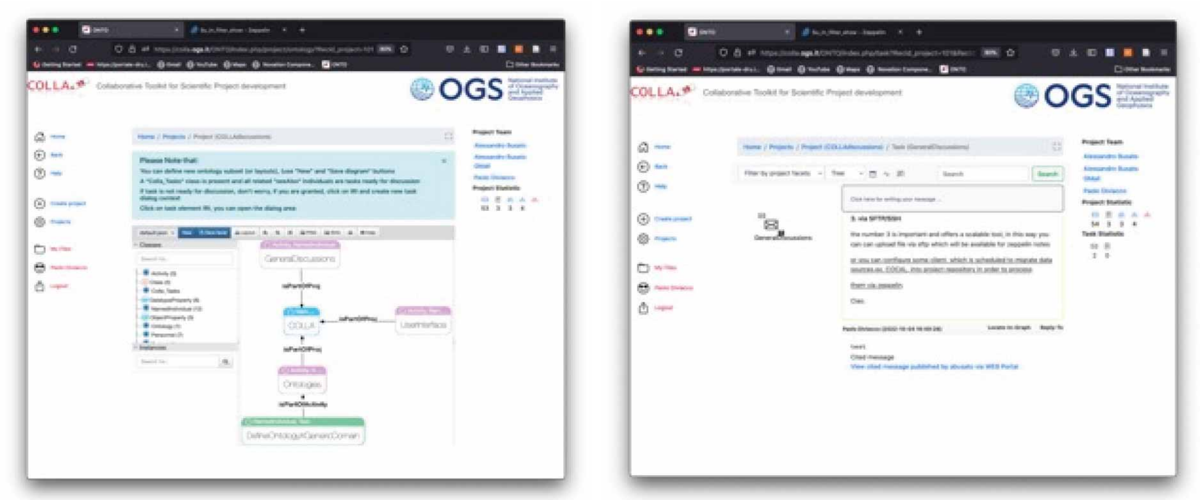


Figure 1 EF+ Virtual Playground Concept map (left) and messaging tool (right).

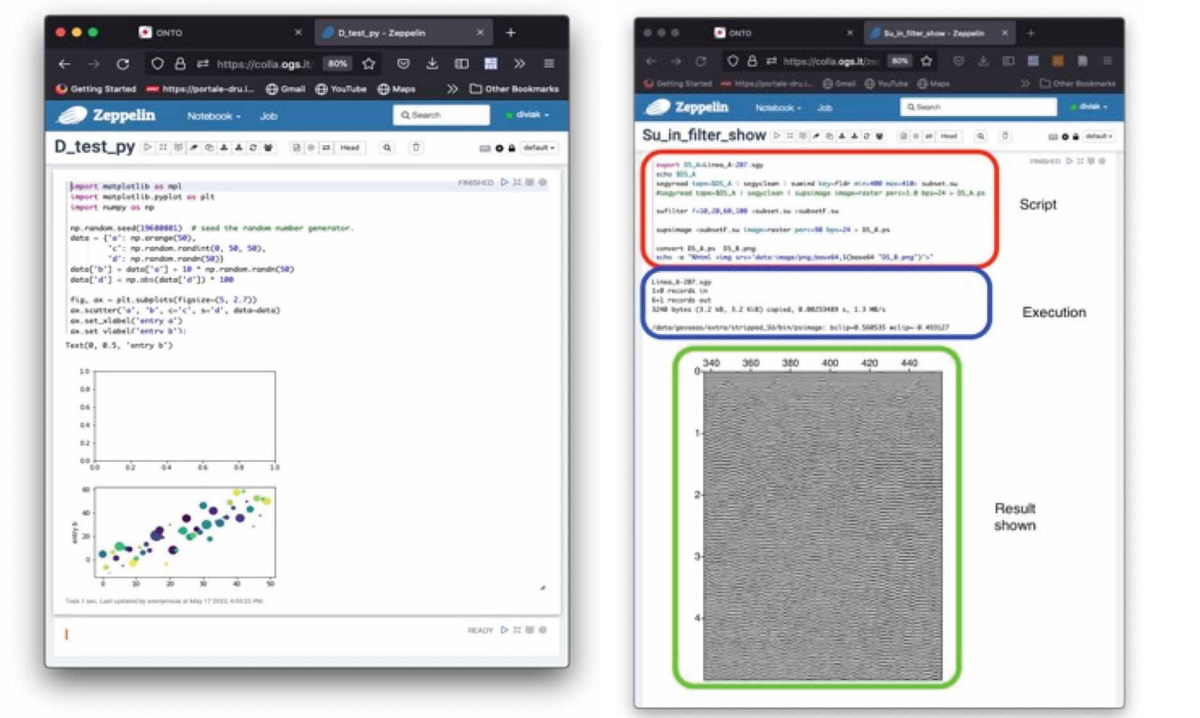


Figure 2 EF+ Virtual Playground data access and processing.

References

Diviacco P., Busato A., (2014). *Maps, graphs, and annotations as boundary objects in knowledge networks, distributed cognition, and collaborative e-research*. Collaborative Knowledge in

Scientific Research Networks, pp. 387- 408. https://doi.org/10.4018/978-1-4666-6567-5.ch019

Diviacco P., Pshenichny C., Carniel R., Khrabrykh Z., Shterkhun V., Mouromtsev D., Guzmán S., Pascolo P., (2015). *Organization of a geophysical information space by using an event-bush- based collaborative tool*. Earth Science Informatics, 8 (3), pp. 677-695. https://doi.org/10.1007/s12145-014-0182-2

Diviacco P., (2018). *On formalization and representation in collaborative research*. Dynamic Knowledge Representation in Scientific Domains, pp. 89-97. https://doi.org/10.4018/978-1-5225-5261-1.ch004

NERSC ARctic VirtuAl Laboratory (NARVAL)

Adrien Perrin (adrien.perrin@nersc.no), Johnny A. Johannessen (johnny.johannessen@nersc.no), Antonio Bonaduce (antonio.bonaduce@nersc.no)

NERSC (Norway)

NARVAL is a web visualization portal for satellite, model and in-situ Earth data

The growing amount of Earth Observation (EO) data and continuous improvement of the numerical models offer outstanding possibilities for environmental monitoring, as successfully demonstrated by the Copernicus European Programme. In this, a versatile and simple way of exploring multiple datasets is the key to ensure FAIR access to environmental monitoring data. With this in mind, we are building on the synergy between satellite sensors, in-situ records, and models to advance the data discovery and visualization capabilities with a highly interactive, flexible, and intuitive web visualization portal.

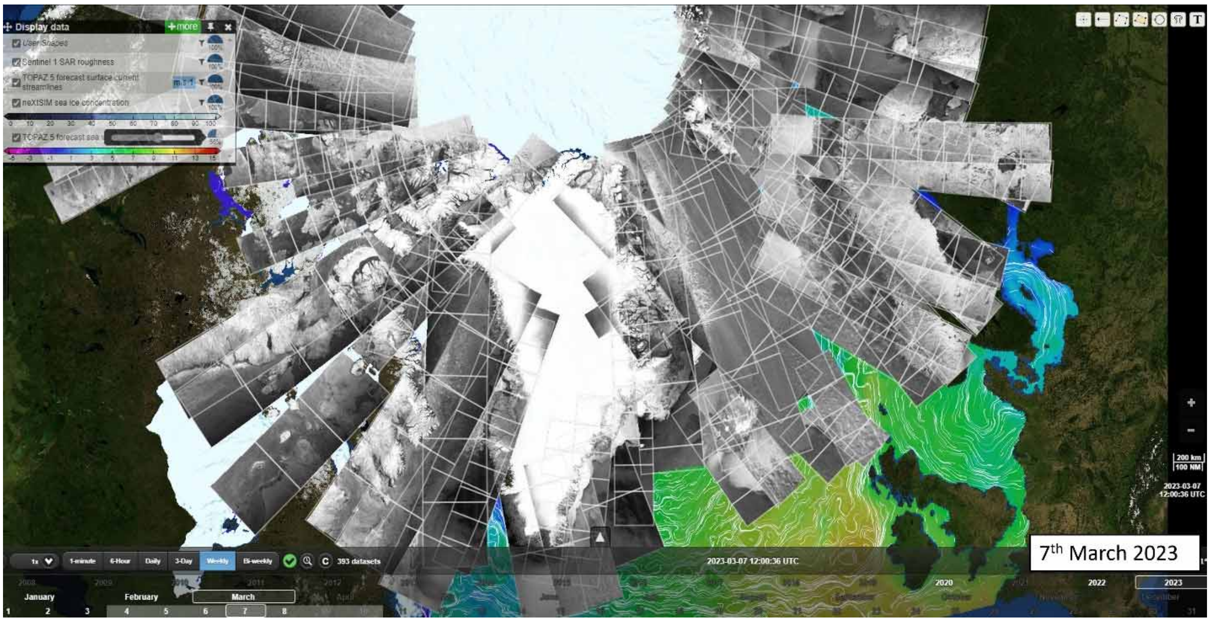


Figure 1 Screenshot of the visualization portal showing Sentinel 1 SAR images, NeXtSIM sea ice extent and TOPAZ open ocean data.

NARVAL is a web visualization interface coupled with a data catalog, which results in a web portal automatically populated with fresh data every day.
We use Syntool for the web interface. It is an open-source earth data visualization web interface developed by our partner OceanDataLab. It provides an intuitive and interactive user experience. As it displays data from satellite, models and in-situ instruments, it is a very convenient tool to explore data relevant to scientific studies.
We use GeoSPaaS to find and process the data displayed in the portal. It is an open-source geospatial data management service developed at NERSC. It allows us to easily find data from the numerous available providers (ESA, NASA, etc.), download it and convert it to the right format so that it can be displayed in Syntool.

The resulting system is a web interface displaying multiple data products in polar stereographic projection, automatically updated every day. This provides a convenient way to explore complex environmental monitoring and forecasting information and exploit synergies between multiple sensors and models. The portal is currently in beta version, offering a number of data products. New ones will be added gradually.

NARVAL offers multiple useful applications, from scientific work to communication and outreach. As stated earlier, NARVAL offers a convenient way to explore the growing volume of environmental data available to scientists, making it easy to identify interesting cases at a glance. Moreover, the ability to simultaneously display data from multiple sensors or models can highlight synergies that would be difficult to identify otherwise.

Finally, NARVAL has great potential for communication and outreach, offering easy viewing of scientific data to stakeholders as well as fellow scientists. It can also be used to great effect for education and training both at universities and summer schools.

NARVAL will be available starting December 1, 2023 at the following URL:
https://narval.nersc.no/

Modernised Cruise Summary Report directory: enhanced access to and management of marine cruises

Marine Vernet (marine.vernet@ifremer.fr), Vanessa Tosello (vanessa.tosello@ifremer.fr), Nolwenn Danioux (nolwenn.danioux@ifremer.fr), Fred Merceur (fred.merceur@ifremer.fr)

Ifremer (France)

Introduction

Oceanographic research cruises play a pivotal role in advancing our understanding of the oceans, generating a large number of marine observations. Cruise Summary Reports are a mean to report on these marine cruises and in-situ experiments. Traditionally sent by Chief scientists 2 months after the cruise, these reports provide a first level inventory of measurements and samples collected at sea.

The Cruise Summary Report catalogue inventories more than 67500 cruises from 1500 research vessels and 41 countries, conducted in European Seas and global ocean between 1873 and nowdays. Completely integrated into SeaDataNet infrastructure, the CSR catalogue provides an interoperable and swift access to rich and standardized cruises' information, from the cruise name to the data description.

Initially managed by the BSH (German Federal Maritime and Hydrographic Agency) and subsequently transferred to IFREMER (French Research Institute for the Exploitation of the Sea) in 2021, the CSR catalogue now provides dedicated interfaces for users, providers and machines. The catalogue serves as a valuable resource for the scientific community but also for stakeholders and decision-makers that are in need of rapid and accessible information about oceanographic cruises, i.e., to collect statistics about national cruises, improve cruises' cost effectiveness or identify unexplored areas. Its information and interface are also being reused by international projects and associations such as POGO (Partnership for Observation of Global Oceans) and ARICE (Arctic Research Icebreaker Consortium).

A comprehensive search interface

The CSR search interface developed by IFREMER offers multiple search facets to select cruises by year, platform, data collate centre, but also data type or research discipline. Free text search and additional filters are also available to refine the search. Moreover, diverse geographical filters can be applied, to select the cruises conducted in a specific area through standard marine area terms, geographical coordinates or directly on a map. The shopping basket deals with unique or multiple CSR export, in text, XML or CSV format. A machine-to-machine interface following W3C rules has also been developed to query the content of the CSR database through a dedicated SPARQL endpoint.

Figure 1 Description of RHOSOS cruise (2008, Le Suroît) on the CSR search interface.

Interconnected and standardized cruises' description

To ensure CSR standardized description, the CSR are described using a specific metadata format based on ISO19139 Inspire-compliant standard. Most metadata tags are supported by SeaDataNet FAIR vocabularies exposed on the NERC Vocabulary Server, enhancing interoperability.
The CSR description includes references to cruise's persistent identifier when available and is also connected to cruise data (SeaDataNet Common Data Index - your data published on SeaDataNet), projects (EDMERP - European Directory of Marine Environmental Research Projects), organisation (EDMO – European Directory of Marine Organisations) and with different infrastructure and data flows such as EMODNet Ingestion.

A dedicated back office interface

The CSR backoffice interface, accessible through MarineID authentication, empowers chief scientists or national data centres to easily create, update and manage their CSRs through XML file uploads or online forms. To avoid duplication, the system performs several checks and provides warnings and a preview page. IFREMER's validation and curation team ensures quality of CSR content via a dedicated administration interface.

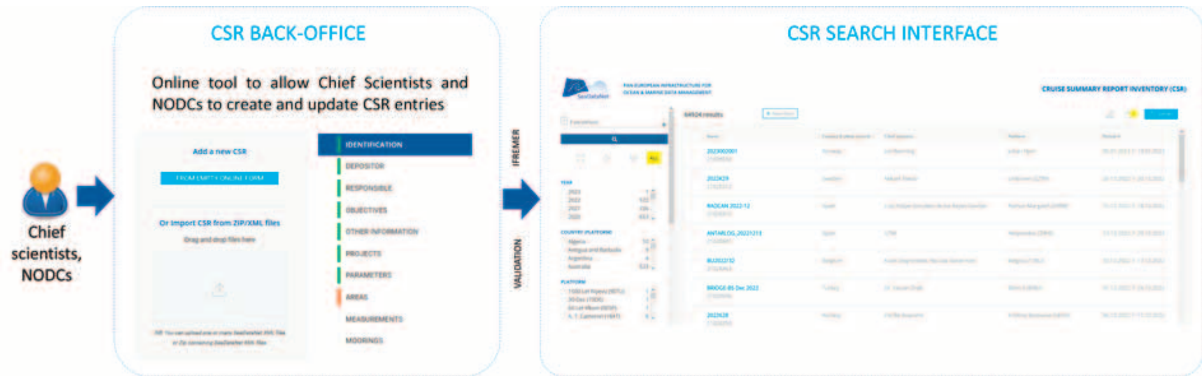


Figure 2 CSR workflow.

CSR technical architecture

The CSR system relies on a robust architecture composed of an ORACLE database to store cruise metadata and an ElasticSearch index synchronised nightly using a Java batch. The CSR interface queries the ElasticSearch index using a REST API and share functionalities through an Angular library.

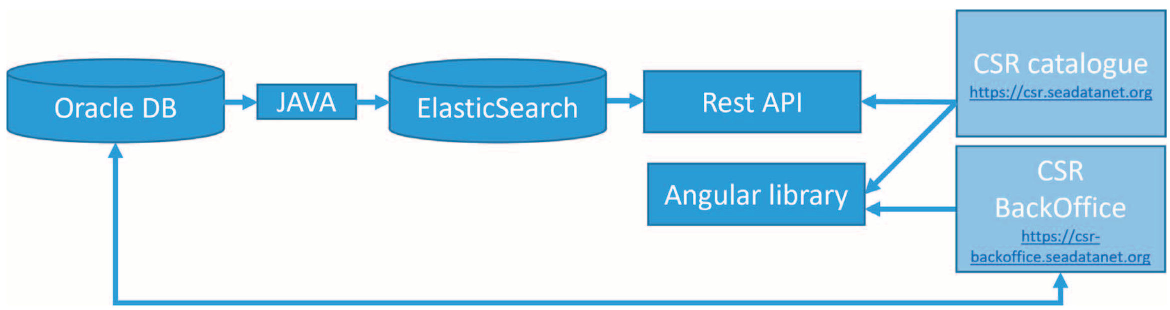


Figure 3 CSR technical architecture.

New perspectives

Future perspectives include enhancing user experience for both data providers and data users. Several developments are planned, such as facilitating management at national data centre's level, refining the geoviewer for interactive geographical searches, diversifying export types for easy integration into cruises reports and data management plans, and proposing a DOI attribution service to facilitate cruises'citation and valorisation. With a growing number of users and data, these developments aim to provide a higher quality of service and further support oceanographic research endeavors.

Optimizing Data Discovery and preservation: A FAIR-Compliant Metadata Management System

Asuka Yamakawa (asuka.yamakawa@hi.no), Cristian Muñoz Mas (cristian.munoz.mas@hi.no), Arnfinn Morvik (arnfinn.morvik@hi.no), Helge Sagen (helge.sagen@hi.no)

IMR (Norway)

Introduction

The Institute of Marine Research (IMR) is a national research institute affiliated to the Ministry of Trade, Industry and Fisheries. Since its establishment in 1900, IMR has been producing a significant amount of marine environment, physical and fisheries data. The Norwegian Marine Data Center (NMD) at IMR was established as a national Data Center for professional management and publication of marine data. NMD maintains the largest collection of marine environmental and fisheries data in Norway. NMD is responsible for long-term preservation of all data collected and generated by IMR.

NMD is starting a digital transformation process that will increase the efficiency, variety and quality of products and services that generate value to our stakeholders. Operationalizing the FAIR principles in our data strategy is a priority, and metadata is a key aspect of this process. As defined by Danish National Forum for Research Data Management, *"from a FAIR (Findability, Accessibility, Interoperability, and Reuse of digital assets) perspective, metadata are more important than data, because metadata would always be openly available, and they link research data and publications on the Internet of FAIR Data and Services".*

The adoption of standards for metadata collection is necessary for the succeeding in rescue and preservation initiatives [Hills D., 2015]. NMD receives near real time (NRT) data from research vessels via satellite every day. It is required to automate the NRT (meta)data processing system to timely publish it with a standard format as soon as the (meta)data is delivered to the Data Center. In addition, there are many challenges associated to the recovery of historical metadata such as varying and inconsistent format of the metadata sources, or the lack of digitalization and/or machine-readability capability.

The NND Metadata Management System and its operation

One of the main objectives for NMD is to increase internal coordination mechanisms to improve the overall metadata lifecycle management between different stakeholders and enhance the quality assurance and consistency across metadata bases with the purpose of unlocking the value contained in all data generated through IMR.

A metadata management system (NMD-MMS) has been established for collecting, processing, storing, preserving and distributing metadata. This system is the core for providing centralized and interlinked access to associated resources (data, publications, methods, code repositories, training materials, etc). It was designed with a scalable architecture approach with particular focus on (1) assuring the coupling of the different subsystems that provide metadata to maintain consistency in the content between the subsystems, (2) allowing asynchronous processing and reducing dependencies between subsystems, (3) leveraging on existing open-source infrastructure, and (4) complying with community-adopted best practices and standards (Figure 1).

The system is operated by data curators responsible for the metadata lifecycle management. They are divided in two categories: (1) experts for each data thematic domain, project, or research group at IMR who are responsible for the accuracy and completeness, and (2) experts in data management from the NMD who safeguard the entire lifecycle management process and ensure consistency and validity. Metadata curation occurs in different stages of the metadata lifecycle implemented in the NMD-MMS.

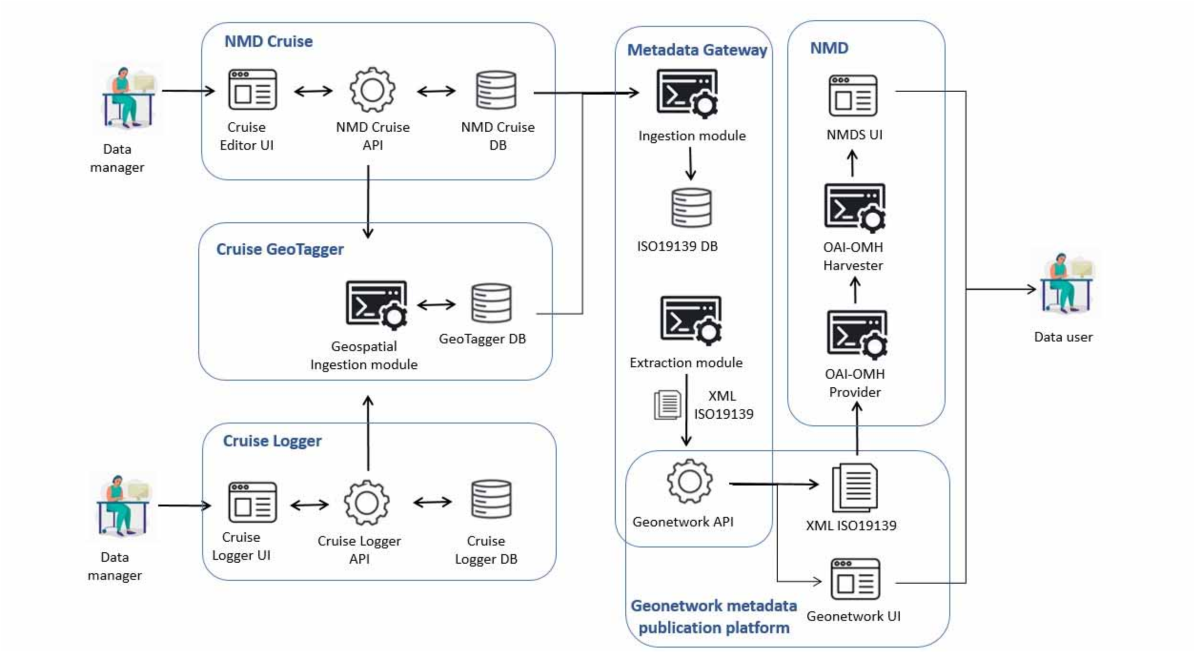


Figure 1 NMD's Metadata Management System architecture.

NMD's Metadata Rescue Program (MRP)

NMD tackles preserving all data at risk of being lost due to deterioration of the medium and digitizing the data. The NMD's Metadata Rescue Program (MRP) is implemented as a part of the development of the NMD-MMS. It establishes mechanisms for transitioning the historical metadata into a format and digital infrastructure that will allow setting the cornerstone for making all data generated findable. This process includes a thorough exercise of recovering metadata from non-reusable and non-consistent formats to a state that can be catalogued using community-agreed standards such as ISO19115/19139.

Future work

NMD is improving both the way it publishes NRT data and historical data complying with the FAIR principles, and the way it links up with external data systems to promote the use of marine data and information to generate knowledge. NMD-MMS is driving such improvement. However, assuring quality in the metadata content is a must. Text mining and machine learning will support the future process of automated curation of metadata content.
NMD-MMS will also become a key component for collecting and distributing the metadata of the future Norwegian Marine Data Space (NMDS) towards the global digital ecosystem provided by the Ocean Data and Information System - Ocean InfoHub (ODIS-OIH).
Finally, NMD-MMS will support capacity development in metadata management with partner National Oceanographic Data Centers (NODCs) and Associate Data Units (ADUs) from the International Oceanographic Data and Information Exchange (IODE) Network, where NMD is already an accredited NODC.

Making methods available through a globally distributed system: Introducing the Ocean Practices Federated Network

Cristian Muñoz Mas[1] (cristian.munoz.mas@hi.no), Marc Taconet[2] (marc.taconet@fao.org), Neil Holdsworth[3] (NeilH@ices.dk), Peter Pissierssens[4] (p.pissierssens@unesco.org), Ruth Anderson[3] (ruth.anderson@ices.dk), Gercende Courtois[5] (gtricor@hotmail.com), Pauline Simpson[4] (p.simpson@unesco.org), Pier Luigi Buttigieg[6] (pier.buttigieg@awi.de), Jay Pearlman[7] (jay.pearlman@ieee.orgj), Tamsin Vicary[2] (tamsin.vicary@fao.org), Ffion Bell[3] (ffion.bell@ices.dk), Anton Ellenbroek[2] (anton.ellenbroek@fao.org), Tiziano Di Condina[2] (tiziano.dicondina@fao.org), Aureliano Gentile[2] (aureliano.gentile@fao.org), Kiran Viparthi[2] (kiran.viparthi@fao.org)

[1]IMR (Norway)
[2]FAO (International)
[3]International Council for Exploration of the Sea – ICES (International)
[4]UNESCO-IOC (International)
[5]ULPGC (Spain)
[6]AWI (Germany)
[7]IEEE (France)

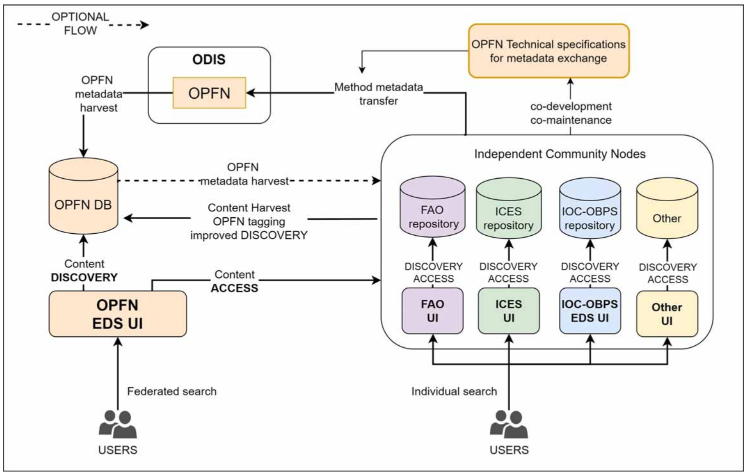
Introduction

The general paradigm for data and information exchange is shifting towards distributed rather than centralized systems. Such a model allows organizations to maintain autonomy and control over their digital assets, choosing how and what to share with specific partners or the open web. In this context, organizations may wish to form digital federations, where they forge agreements on how to share (meta)data to accomplish common goals. This contribution outlines such a federation, where sharing metadata about methodological records is furthering cross-indexing and discovery across repositories (methodology management systems; MMSs) under an initiative of the UNESCO Intergovernmental Oceanographic Commission Ocean Best Practices System (OBPS) project and the Ocean Data and Information System (ODIS), the Food and Agriculture Organization (FAO), and the International Council for the Exploration of the Sea (ICES).

The Ocean Practices Federated Network (OPFN)

The OPFN will facilitate a coordinated approach across MMSs residing in individual organizations. This will enable siloed MMSs to offer federated content discovery with semantic indexing capabilities as well as advanced metrics for methodology uptake. The OPFN is being developed under the UN Ocean Decade Programmes OceanPractices for the Decade and OceanData2030, both of which aim to support other Ocean Decade Actions in better-aligned implementation of the FAIR principles. The OPFN will advance the exchange of methods and information across its partners, and enhance linkages to other methods and data. Methods themselves (in whichever format they are stored) remain with each node, under the appropriate access control, but users will more efficiently and comprehensively discover and access the metadata describing them and links to their archival locations provided by the OPFN nodes (Figure 1). The OPFN, as an Ocean Data and Information System (ODIS) thematic subsystem, will leverage and benefit from the ODIS architecture. The OPFN will use the ODIS-Arch sub-pattern for documents and methods for establishing and sustaining interoperability in the exchanges of (meta)data between its nodes. The OPFN will have a dedicated portal with enhanced discovery capabilities. Leveraging the ODIS interoperability framework (which in turn reuses generic web architectural patterns and structured data on the web practices) will promote interoperability across and beyond the ocean community.

Figure 1 Conceptual architecture of the OPFN.



Sustainably and robustly sharing content among partners requires the creation of a data governance framework to manage data sharing, access control, data ownership declarations, and information promoting compliance with legal and regulatory requirements, aligned with the policies of its participating organizations. The partners within the OPFN are drafting more methods-focused governance policies to ensure their diverse requirements and responsibilities to their individual stakeholders are fulfilled by OPFN membership. For example, ownership of the document content remains with the copyright owner who may provide access through terms of use that may include standard licenses. Discovery through OPFN will transfer the user to the methods provider for access to full text where metrics of the traffic of users will be maintained through their own infrastructures as well as the OPFN. The OBPS, as a key component of the OPFN, will provide an MMS to those organizations and communities who are unable to support their own methodological repositories.

The OPFN will allow users to have greater opportunities to discover relevant methods across the partner holdings. It will also increase the generation of value to users through interlinking of methodologies - this will support the development of improved practices by facilitating collaboration, cross-validation, broader community review and assessment, convergence, and endorsement by users and organizations. While the ODIS interoperability architecture provides a compelling starting point, the process of creating deeper interoperability conventions (e.g. shared/interoperable semantics, mapping, and content negotiation) between OPFN partners is needed to make the OPFN discovery from different nodes transparent to users. This deeper interoperability can be used to improve the description and cross-indexing of records shared over the OPFN to enhance search and discovery services. As new OPFN nodes come online, or new challenges and needs are encountered, the ODIS-Arch patterns used to exchange (meta)data about methods will be updated. While backward compatibility should be the norm, partners will have the opportunity to leverage updates by aligning their own (meta)data catalogs to improve the visibility of their records in downstream tooling.

Implementation of a Pilot Use Case

Following the foundational discussions that have formed the prototypical OPFN during 2023 (currently between UNESCO-IOC OBPS, ICES Library service, and FAO OpenASFA system) and through the OBPS Workshop VII, a working group is creating a pilot demonstration focused on the aquaculture domain. Stakeholders in the aquaculture community will co-design and validate the capabilities for an operating capability. We aim to expand the OPFN to include MMS from other domains, upgrading the generic interoperability conventions and downstream technologies by the end of 2024.

Building on an interoperable, federated system to make it easier to find marine data

Clare Postlethwaite (cfpo@noc.ac.uk), Colm Walsh (colm.walsh@noc.ac.uk), Gaynor Evan (gaynor.evans@noc.ac.uk), Charlotte Miskin-Hymas (charmisk@noc.ac.uk)

Marine Environmental Data and Information Network - MEDIN (UK)

This paper describes work done by the Marine Environmental Data and Information Network (MEDIN) to enable further federation of an interoperable, federated marine data network

The Marine Environmental Data and Information Network (MEDIN) is the hub for UK marine data and has been promoting good data management practices since its inception in 2008. Providing a network of specialist and accredited Data Archive Centres (DACs), a marine discovery metadata standard, marine data guidelines, tools and an online portal, MEDIN aims to make UK marine data Findable, Accessible, Interoperable and Re-usable (FAIR). The MEDIN Portal contains over 17,500 marine environmental datasets from around the world, owned or managed by UK organisations. Over the past year, MEDIN has enabled a series of developments to further enhance the findability of UK marine data, thereby increasing the potential reuse value of the UK's marine data holdings. The first of these developments was to make the catalogue of metadata in the MEDIN portal available using an Application Programming Interface (API). The second development was to enhance the metadata formats available to our users, which allowed us to expose UK marine data holdings to the UN Ocean Decade digital infrastructure, Ocean InfoHub. Both of these enabled one of MEDIN's key objectives: to support the UK marine sector to implement globally and cross domain interoperable marine data services.

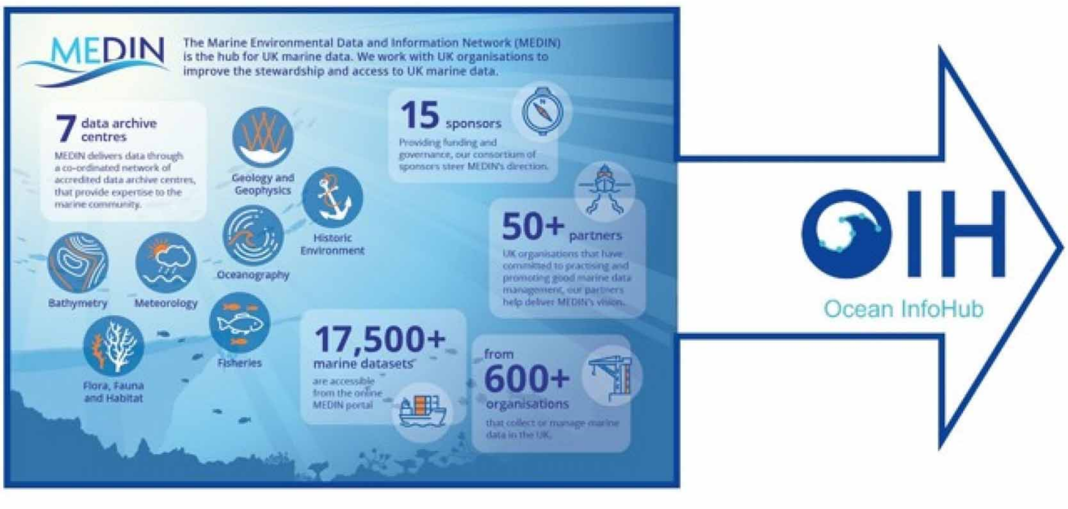


Figure 1 Infographic showing some of the components of the Marine Environmental Data and Information Network (MEDIN) and recent work to make UK marine data available via Ocean InfoHub.

The MEDIN Portal is underpinned by the MEDIN Discovery Metadata Standard

The MEDIN Discovery Metadata Standard contains the list of information required to accompany a dataset to allow other people to find out what the dataset contains, where it was collected and how they can access it. The MEDIN Discovery Metadata Standard is a marine profile of the UK Government Standard GEMINI2 and complies with other international conventions such as INSPIRE and ISO19115. Version 3.1.2 was published in 2022 and sets out a specific format to record details of a dataset, so that, in the future, other people can easily discover marine data that may be of use to them. MEDIN provide a suite of tools, training and helpdesk support to enable widespread uptake of this standard by the UK marine community. Metadata complying with the MEDIN Discovery Metadata Standard feed into the MEDIN Portal. Roughly 85% of the metadata in the MEDIN portal come directly from the MEDIN DACs (British Oceanographic Data Centre; UK Hydrographic Office; British Geological Survey; DASSH; Met Office; Centre for Environment, Fisheries and Aquaculture Science; Marine Directorate; Archaeology Data Service; Historic Environment Scotland; Royal Commission for the Ancient and Historic Monuments of Wales) with the remaining metadata coming from a mix of public and private UK organisations. Ultimately, the MEDIN portal is a federated system to help users find UK marine data.

MEDIN Application Programming Interface

Recognising the role that MEDIN can play to support other federated data systems access UK marine data, MEDIN has recently exposed its metadata catalogue using an API. This allows all the metadata that MEDIN has already collated from hundreds of sources to easily be included in other portals or catalogues. This API is available for anyone to use, but the first application that we know of is the inclusion of MEDIN metadata into the Offshore Wind Evidence and Knowledge Hub (OWEKH), which will present marine data from MEDIN alongside a range of other datatypes valuable for the offshore energy sector.

Exposing metadata in formats to enhance interoperability

The Ocean InfoHub project facilitates discovery and interoperability of existing information systems through the Ocean Data and Information System (ODIS) infrastructure, with the ultimate goal of coordinating action and capacity to improve access to ocean data and knowledge. MEDIN has been contributing to the development of Ocean InfoHub by sharing its catalogue of dataset discovery metadata records with the ODIS Catalogue where they are now searchable in the Ocean InfoHub web portal. Ocean InfoHub and the ODIS Catalogue use the schema.org vocabulary and JavaScript Object Notation for Linked Data (JSON-LD) mark-up language to publish metadata resources. MEDIN have already exposed schema.org semantics using JSON-LD to make discovery metadata available in the Google Datasets search engine. This made it possible for MEDIN to add 17,460 dataset records to the Ocean InfoHub web portal with only the need for some minor modification to its pre-existing JSON-LD code. Ocean InfoHub contributes to the Ocean Data 2030 Programme of the United Nations Decade of Ocean Science for Sustainable Development.

Summary

MEDIN's longstanding commitment and experience making UK marine environmental data FAIR has enhanced access to the evidence base for decision-making and research across commercial, policy, academic and conservation sectors. Exposing metadata using an API and being an early contributor to Ocean InfoHub enables a broader, international audience to benefit from MEDIN's 15 years' experience improving access to valuable marine data resources. Exposing the MEDIN catalogue of datasets in these ways increases the findability and potential reuse value of those datasets and supports MEDIN's ethos of "measure once, use many times".

A new online app for ocean temperature quality control - An artificial intelligence approach

Sebastian Mieruch (sebastian.mieruch@awi.de), Mohamed Chouai (mohamed.chouai@awi.de)

Alfred Wegener Institute (Germany)

Introduction

Marine data quality control (QC) is crucial to provide robust data products for climate analyses, monitoring, process- and model studies and much more. However, the QC of marine measurements of e.g. temperature, salinity, nutrients (phosphate, nitrate, ...), oxygen etc. is challenging. Measurements are prone to errors due to external forcing (sun, wind, currents, ...), internal variability (e.g., extremes), biogeochemical processes, instrument errors or failures and more. Ocean data QC is an international effort and large marine infrastructures, like SeaDataNet (https://www.seadatanet.org/), EMODnet Chemistry (https://emodnet.ec.europa.eu/en/chemistry), Argo (http://www.argodatamgt.org/) or IQuOD (https://www.iquod.org/) have created sophisticated QC processing schemes. Typically, ocean data QC is a semi-automatic process, whereas the ocean experts use algorithms to identify potentially "bad" data, which are accordingly often visually inspected to make a final decision and to give the data sample a quality flag, i.e., an indicator such as "good", "bad", "probably bad", etc. One widely used tool for the QC is the Ocean Data View (ODV, https://odv.awi.de) software, which is also available as the online version webODV (https://webodv.awi.de). Because of the diverse nature of errors in the data, fully automated QC without expert visual checks is still less skillful and yields to too many misclassifications. However, visual QC is highly time demanding and skillful algorithmic support is needed.

Approach

In this study we trained a deep neural network with the knowledge of an ocean QC data expert to mimic the human visual QC. Behrendt et al., 2018 created a large arctic temperature and salinity dataset, called UDASH (Unified Database for Arctic and Subarctic Hydrography), which is based on 288.532 profiles and more than 74 million single measurements (samples) for the time 1980-2015, north of 65° N. For the QC of this dataset, the authors used classical algorithms to identify "bad" data. These algorithms have been too sensitive, i.e., they misclassified too many "good" data samples as "bad". Therefore, Behrendt et al. [2018] visually inspected all potentially "bad" data and changed their flags back to "good" in the case of misclassifications. We used this "labeled" dataset to train a deep neural network (an MLP, multilayer perceptron) to "learn" this human QC, i.e., turning apparently "bad" data back to "good". Thus, we have developed an AI version of a human QC expert and we name this algorithm *SalaciaML-arctic*.

Results

As a first try we concentrated only on temperature data and only on two types of errors in the data, i.e., "Spikes" and "Suspect Gradients", which are anomalous excursions from the expectations. We tested our algorithm *SalaciaML-arctic* on 1.234.730 data samples, where the classical algorithms detected 5.038 samples as "bad". Behrendt et al. [2018] turned 2.443 of those back to "good" and *SalaciaML-arctic* turned 2.082 back to "good", without introducing any other errors. Thus the skill of *SalaciaML-arctic* is ca. 85 % of that what Behrendt et al. [2018] achieved.

Conclusion and App

We have developed an AI algorithm based on "learning" human QC skills, which we name *SalaciaML-arctic*. The algorithm can help QC experts, researchers, data managers and others to

remove “bad” measurements from datasets. At the moment it can only be applied to arctic temperature profile data, which was the basis for training. It is superior to classical algorithms by reducing the number of misclassifications, i.e., “good” data wrongly labeled “bad”. Therefore *SalaciaML-arctic* saves important data, also in terms of variability and monetary.
We recommend the use of *SalaciaML-arctic*:

1. Fully automated: if a study does not allow for or does not have the resources for visual QC.
2. Semi-automated: using the flags from *SalaciaML-arctic* as a first guess for a following visual QC. The first guess helps the experts in taking a decision and can reduce the workload.

Our algorithm *SalaciaML-arctic* is written in Python and easy to use. It will be made available via GitHub soon. In addition we have developed an online app at https://mvre.autoqc.cloud.awi.de/, where own data can be processed easily. The app provides detailed documentation and processed data are exported as simple .csv files or ODV Spreadsheet, which can be used directly in ODV or webODV. Figure 1 shows a screenshot of the app, where the user has to transform the data into a simple column based structure and upload the data. The processing is accomplished in the background and as soon as the QC is finished, the user is informed via email and the final data can be downloaded.

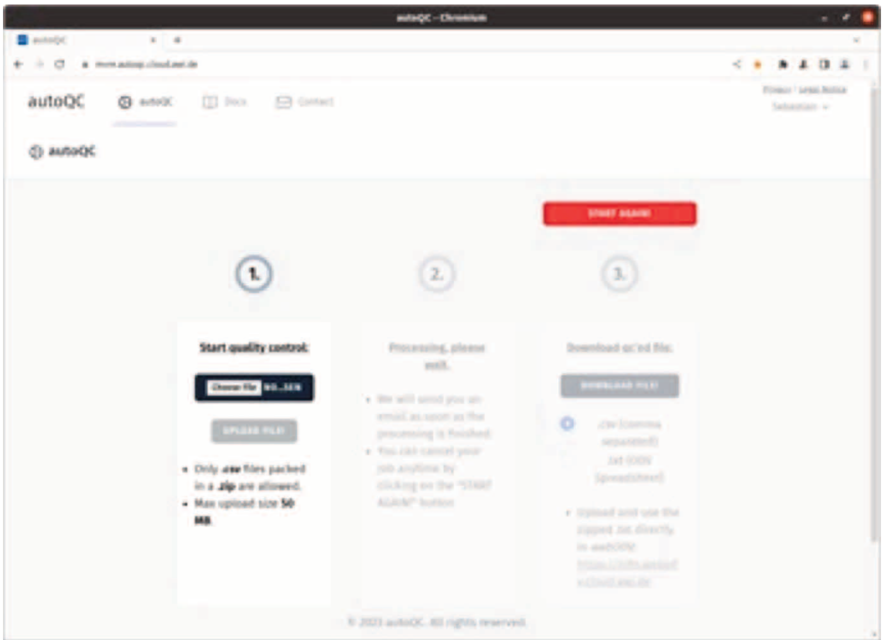


Figure 1 The autoQC app at https://mvre.autoqc.cloud.awi.de/

References

Behrendt A., Sumata H., Rabe B., and Schauer U., (2018). *Udash – unified database for arctic and subarctic hydrography*. Earth System Science Data 10, 1119–1138. https://doi.org/10.5194/essd-10-1119-2018

iMagine, AI-supported imaging data and services for ocean and marine science

Gergely Sipos[1] (gergely.sipos@egi.eu), Dick Schaap[2] (dick@maris.nl),
Alvaro Lopez Garcia[3] (aloga@ifca.unican.es), Valentin Kozlov[4] (valentin.kozlov@kit.edu)

[1]EGI (The Netherlands)
[2]MARIS (The Netherlands)
[3]CSIC-IFCA (Spain)
[4]KIT-SCC (Germany)

Introduction

Aquatic bodies, such as oceans, seas, and coastal and inland waters are home to diverse ecosystems and habitats, strongly regulate climate, and offer many resources for economic opportunities. However, the combination of long-term global change and multiple local stressors affects ecosystems in unpredictable ways to a point of no return with significant socio-economic consequences. Better understanding of the dynamics and complex geochemical interactions is key to allowing sustainable mitigation, and/or restoration plans for these essential ecosystems. Increasingly, the potentials of processing image data is being explored and successfully exploited by researchers and Research Infrastructures that develop and operate high-performance image analysis tools, including tools using Artificial Intelligence (AI) techniques. In the aquatic domain, the application areas stretch broadly from remote sensing images, observed by aerial platforms such as satellites, planes, and drones, footage and images observed by cameras mounted at beaches, rivers, and lakes, image footage and sound recordings from cameras and hydrophones, mounted underwater at environmental monitoring installations, to images, acquired from microscopes, screening water samples. Overall, there is great potential, both already proven and underexploited, for multiple types of image data and derived types of observation parameters, which could complement more traditional data acquisition. More adoption, developments, and open provision of image processing services and AI should be stimulated and pushed in the aquatic sector for a wider uptake and more significant impact. This has motivated the EU iMagine project to deploy, operate, validate, and promote a dedicated iMagine AI analaytics framework and platform connected to the European Open Science Cloud (EOSC), giving researchers in aquatic sciences open access to a diverse portfolio of AI-based image analysis services and image repositories, to facilitate gaining knowledge on and of relevance for healthy oceans, seas, coastal and inland waters.

iMagine Platform for development, testing and operation

The iMagine AI Platform has been established as a scalable, shared IT platform that delivers a generic framework for AI model development, training, and deployment, which researchers can adopt researchers for image analysis in marine and freshwater research. It uses the DEEP-Hybrid-DataCloud services, and its successor, the AI4OS software as underlying technology. The platform covers the whole machine learning development cycle, ranging from the model creation, training, validation and testing to the models serving as a service, sharing and publication. The framework allows transparent access to existing e-Infrastructures, effectively exploiting distributed resources for the most compute-intensive tasks coming from the machine learning development cycle. Moreover, it provides scientists with a set of Cloud-oriented services to make their models publicly available, by adopting a serverless architecture and a DevOps approach, allowing an easy share, publishing and deploying of the developed models. The compute layer of the iMagine AI Platform is powered by four cloud providers (based in in Spain, Portugal, Turkey, and Ireland) from the pan-

European EGI federation infrastructure, collectively offering a federated infrastructure and over 132,000 GPU-hours, 6,000,000 CPU-hours and 1500 TB-month for image hosting and processing.

Portfolio of Use Cases

iMagine has five use cases that are based on existing image analytics services. These use cases are further developed and made fit for operational deployment at the iMagine platform, after which these services will be opened with Virtual Access to external researchers. The development work focuses on labelling training images, migrating AI-powered image classification models into the iMagine AI Platform, training the model with the images, validating the prediction accuracy, and improving the models and images. This development process is facilitated by the iMagine Competence Centre, a virtual team that provides guidance and support to the use case developers. The use cases so far developed five models in the iMagine AI platform and also rely on third-party models, for example, YoloV8.

Table 1 Overview of mature use cases.

Aquatic Litter Drones: Aquatic Litter monitoring system using drones	
EcoTaxa pipeline: Taxonomic identification of zooplankton using Zooscan	
Ecosystem monitoring at EMSO sites by video imagery	
Oil Spill Detection: Oil spill detection from satellite images	
Taxonomic identification of phytoplankton using Flowcam images	

In addition, the project supports three further use cases that are new in AI-based image processing. These use cases are being made familiar with the concepts of machine learning, and learn to use the iMagine AI Platform as a tool for development, testing, and implementation. Furthermore, a recent Open Call resulted in the selection of two external use cases, and these will be also developed and deployed on the iMagine platform.

Best Practices

Best practices for the development and operational deployment of AI image analytical services are being captured, and will provide instrumental information for researchers, image data providers, and image analysis service providers. The best practices are derived from the iMagine AI platform providers and from the use case developers. The synergies between aquatic use cases will lead to common solutions in data management, quality control, performance, integration, provenance, and FAIRness, contributing to harmonisation across Research Infrastructures, which will also be included in the iMagine Best Practices results. Dissemination is performed through guidelines, presentations, and wider distribution of documents and slides using ZENODO, in particular connecting with the European science community, and the IODE Ocean Best Practices System (OBPS), reaching out to a large international ocean research community.

EOSC

The iMagine AI platform can also be accessed through the EOSC Marketplace. This way, the iMagine cutting-edge environment designed to facilitate the advancement of Artificial Intelligence, Machine Learning, and Deep Learning in the field of aquatic sciences, is promoted and made available for the wider EOSC research community.

Introducing adviceXplorer – a new app that makes ICES fisheries management advice interactive and accessible

Luca Lamoni (luca.lamoni@ices.dk), Sarah Louise Milla (sarah-louise.millar@ices.dk)

International Council for Exploration of the Sea – ICES (Denmark)

Introduction

From data collection to knowledge production, fisheries management advice takes thousands of hours to produce, and draws on many data measurements. The end result of this is typically distilled into a few pages of text and graphs entombed in a PDF. AdviceXplorer is the next generation of management advice that is connected to the data sources, interactive and downloadable.

The International Council for the Exploration of the Sea (ICES) is an intergovernmental marine science organization, meeting societal needs for impartial evidence on the state and sustainable use of our seas and oceans. ICES goal is to generate state-of-the-art advice for meeting conservation, management, and sustainability goals. ICES receives requests from public authorities for scientific advice on a range of issues relating to fisheries, marine policies, and management. The most established and downloaded ICES product is the advice on fishing opportunities, previously available to the general user only in PDF. The purpose of the advice is to present the level of precautionary removals from each stock assessed by ICES (as total catch or as landings and discards) following the ICES framework for providing catch advice. This is supported by standard figures and tables that explain how the headline value was obtained, alongside some history and background information relevant to the current fishery.

To meet the growing demand for interactive and accessible data and to move away from the standard PDF product, ICES has launched a new web-app - adviceXplorer - to access and visualise ICES advice on fishing opportunities. The app is fully integrated with ICES databases and services and allows users to quickly find, visualize and access the data used for advice.

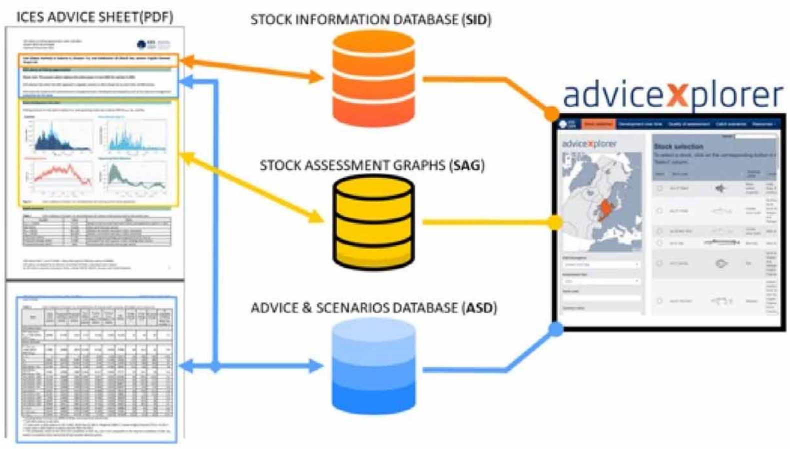


Figure 1 Infographic representing how the three ICES databases are a key part of the advice process and represent the foundation upon which adviceXplorer is functioning.

How does it work?

All current users of ICES advice - stock assessors, scientists, managers, and decision makers – are now able to search for fisheries advice by using an interactive map of ICES ecoregions, as

well as additional filtering options such as assessment year, stock code, and species. This filtering system is designed to accommodate different types of users; for example, a stock assessor interested in a specific stock could directly use the stock code field, while a manager focused on one or more species could use the species field to get an overview of several stocks at once. Once a fish stock is selected, trends in catch and biological parameters over time, headline advice values, and catch scenarios are displayed through interactive plots and tables. The data displayed in adviceXplorer can be easily downloaded and direct links to fishing opportunity advice sheets (pdf format) published in ICES library are readily available.
The data infrastructure behind this application includes three key ICES databases:

- the Stock Assessment Graphs (SAG)
- the Stock Information Database (SID)
- the Advice and Scenarios Database (ASD)
- AdviceXplorer retrieves data from these databases through publicly available ICES webservices. This approach has the advantage of delivering data to the user securely and making the application lighter and more responsive, which results in a seamless user experience. The development of this web-app followed the recently published guidelines on ShinyApps hosting and publication by ICES (ICES, 2023), ensuring that FAIR principles of scientific data manangement were upheld.
- Currently, adviceXplorer displays the last six years (2018 - 2023) of ICES advice on fishing opportunities for the majority of stocks. The handling of the diversity and complexity of ICES 200+ annual fish stock assessments is a challenge that will be addressed in future releases of adviceXplorer. Furthermore, future development will focus on issues of automation, reliability and transparency. This approach is part of a general ICES strategy that strives to improve transparency in both advice processes and data management using the Transparent Assessment Framework (TAF).

References

ICES (2023). *ShinyApps hosting and publication: Context and process*. ICES General and Science Guidelines. Report. https://doi.org/10.17895/ices.pub.22695898.v1

Assessment of Online Translation Tools for Indigenous Languages in the Ocean Domain in the Arctic

Sigmund Kluckner[1] (sigmund@kluckner.eu), Maria van Veldhuizen[2] (mariavveldhuizen@gmail.com), Jay Pearlman[3] (jay.pearlman@fourbridges.org)

[1]Independent consultant (AT)
[2]Maria van Veldhuizen Consulting (The Netherlands)
[3]Four Bridges (USA)

Introduction & Background

The vast array of knowledge communities dedicated to Arctic environments, maritime studies, and ocean sciences generate a wealth of valuable information, including reports, best practices, standards, and white papers. While this information holds immense potential to benefit indigenous communities in the Arctic and coastal regions globally, a significant barrier exists: much of it is currently accessible only in major world languages, usually either English or French. As part of the CAPARDUS project[38], the authors investigated the capacities of Neural Machine Translation (NMT) systems to translate ocean/Arctic-related knowledge artifacts into several other major languages, some of which are relevant to the Arctic (Danish), and performed a quality check on the outputs. Additionally, the option to use these tools to translate knowledge to and from indigenous languages in the Arctic, like Kalaallisut, Inuktitut or Iñupiaq ('low resource languages') were investigated. Translation into local languages is important, as the indigenous peoples in the Arctic share geographic and economic ties, but their languages, while sharing the same roots, are not always mutually intelligible. Besides sharing the latest research findings with indigenous communities, automatic translation would also be valuable to safeguard and share the knowledge held by them. The overall goal was to ascertain whether NMT could, in principle, be used to translate Ocean-related knowledge artifacts into Arctic indigenous languages if these were supported more effectively.

Methodology - three distinct steps

1. **Desk research on existing tools**: the authors catalogued translation tools with a focus on NMT systems. These were then assessed based on set parameters, including the languages they support (a sample of 15 relevant languages was used; including, for comparison, two non-Arctic, coastal indigenous languages), or their features for translating plain text, images, files or websites.
2. **Assessing the quality** of translation tools overall: a test document with excerpts from a variety of ocean- and Arctic-related papers and publications was compiled and run through six translation tools, each translating into German, Dutch, French and Danish (where available). These languages were chosen based on the team's language competencies. The outputs were then quality checked and rated by native speakers, focusing on overall accuracy, readability, and performance on various challenges: the test included segments like special symbols and calculations ("NH_4^+ ions"), diagrams and challenging sentence structures.
3. **Translation capacity and quality assessment** for Arctic indigenous languages: using excerpts from multi-lingual documents, the authors ran tests and compared computer-translated outputs with human translations.

[38] CAPARDUS - Capacity-building in arctic standardization development; Coordination and Support Action under EC Horizon 2020 - Grant Agreement no. 869673 https://capardus.nersc.no/

Outcomes – following the steps from above

Translation tools: In total, 43 online tools were catalogued and assessed. In terms of languages, there is widespread support for English, German and French; less support is available for Danish, Norwegian, Swedish, Finnish and Dutch. Barely any support was found for Kalaallisut (Greenland), Inuktitut (Canada), and no support at all for Iñupiaq (Alaska) or Saami (Scandinavia).

Quality of tools for major languages: Overall, the quality of translations was high – not faultless, but certainly sufficient for most uses. DeepL (highest score) and Google Translate (runner up) returned the most useful outputs. Reverso and Systran Pro scored midfield, while results from Yandex and Alibaba were poor to very poor. It is worth mentioning that Google Translate provides free translations also of entire documents, and as such is probably most useful for most people. Tools struggled the most with **consistency** (same terms translated in various ways throughout a document), **symbols** (often rendered incorrectly) and **diagrams with labels** (due to little context, none of the tools was able to provide a useful translation). Grave errors (such as wrongly translated instructions or the use of offensive language) were rare but present in the outputs of all translation tools; a quick check by a human translator is therefore still required in all cases.

Translation for indigenous languages: Due to the lack of support for most languages of the Arctic region, it was only possible to run limited tests. For example, only Microsoft offered translations to and from Inuktitut, while Alibaba claimed to support Greenlandic/Kalaallisut to English. The translations, as can be seen clearly in the examples below, may not be usable by themselves, but can paint a coarse picture of what is being said and what the contents of the original text allude to (with human interpretation, for example climate change and weather change below).

Excerpt "People of the Ice Bridge"[39]	MS Bing Translation	Human translation (in the report)
ᑕᐃᔅᓯᒪᓂᓕ, ᑭᓯᑐᐃᓐᓇᐃᑦ ᐊᓯᔾᔨᕐᓂᐊᖅᑐᕐᔫᓚᐅᖅᓯᒪᙱᑦᑐᑦ. ᑭᓯᐊᓂᓕ ᐅᓪᓗᒥ ᓯᓚ ᐊᓯᔾᔨᖅᑐᖅ ᐊᒻᒪᓗ ᖃᐅᔨᒪᔪᖓ ᑲᓛᓖᑦ ᓄᓈᖓᓂᒥᐅᑦ ᐃᓅᖃᑎᕗᑦ ᐃᓄᒃHᐅᐃᑦ ᑕᒪᐅᙵ ᑎᑭᐸᒃᑯᓐᓃᖅᑐᑦ. (ᓕᐅᕆ ᐊᐅᓪᓚᓗᒃ, ᐊᐅᔨᐃᑦᑐᖅ).	At that time, things seemed to change. But today the weather is changing and I know that Greenland is not the people who come here anymore. (Larry Audlaluk, Grise Fiord).	Back then, it seems like nothing would change. But today there is climate change and I am aware of that since the Greenlandic people who are our fellow-Inuit Inughuit do not travel here anymore. [Larry Audlaluk, Aujuittuq (Grise Fiord)].

Excerpt "User Guide, Greenland Self-Rule"[40]	Alibaba Translation	Human translation (in the report)
Pinngortitalerinermi apeqqutaasinnaasut, assersuutigalugu: 1. Uumasut amerliartussappat imaluunniit ikiliartussappat? 2. Aalisakkat pisarineqartartut alliartussappat imaluunniit milliartussappat? 3. Uumasut takkuttarnerat pisarnermiit kingusinnerulissava imaluunniit siusinnerussappat?	For example: 1. Will the animals be multiplied or decreased? 2. Should the fish grow, or should they grow up? 3. Animals arrive later. Be ahead?	Biological questions, e.g.,: 1. Are there more or fewer animals? 2. Are the fish caught bigger or smaller? 3. Are the animals arriving/leaving later or earlier than 'normal'?

Conclusion

Our assessment of machine translation tools shows that they can generally provide high quality, useful outputs for ocean domain knowledge. Unfortunately, high-quality models are not available for indigenous languages, mostly due to the limited amount of model training data. As translation systems improve, these challenges could soon be solved: for example, Meta's "No Language Left Behind" project already supports over 200 languages with plans on expanding to many more.

[39] People Of The Ice Bridge: The Future Of The Pikialasorsuaq, Report of the Pikialasorsuaq Commission, Nov. 2017

[40] Local documentation and management of living resources: User Guide, Greenland Self-Rule, Jan. 2014

Using NMT, but also other new generative AI tools, for translations will bring challenges for the users, but with some foresight and care, there are huge benefits to be gained. Paying attention to writing original documents in a more machine-translation friendly style (e.g. less ambiguity) can improve the translation quality of the tools - and increase readability in both the original and target language.

Finally, using these tools to automate the translation of ocean related publications for the use by many in the Arctic region would represent a tremendous improvement in knowledge access and availability, with the potential to reach a broader audience.

SESSION SERVICES POSTERS

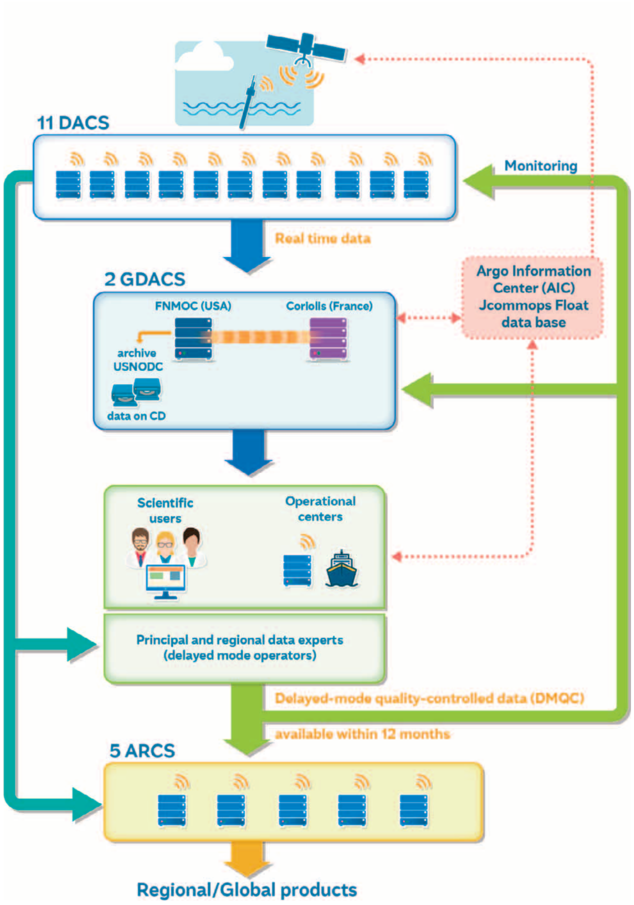
Quality control on Argo data: from the automatic tests to the MinMax method

Christine Coatanoan[1] (christine.coatanoan@ifremer.fr), Jérôme Gourrion[2] (jerome.gourrion@pokapok.org)

[1]Ifremer (France)
[2]POKaPOK (France)

Two Global Data Assembly Centres (GDACs), located at Coriolis/France and FNMOC/USA, are in charge of collecting the processed Argo data from the 11 national Data Assembly Centres (DACs) and to provide users with access to the best version of an Argo profile (Figure 1). DACs receive the data from the satellite operators, decode and quality control the data according to a set of 21 real time automatic tests agreed by the international Argo program [Wong et al., 2024]. Erroneous data are flagged, corrected where possible and then passed to the two GDACs in NetCDF format and to the WMO GTS in the BUFR format. Several Argo Regional Centres (ARCs) provide wide expertise on specific geographical ocean regions in order to provide the most comprehensive data sets (including non-Argo data) of the highest quality.

Figure 1 Argo data system management (https://www.euro-argo.eu/Activities/Data-Management/Argo-Data-System).



Due to the requirement of delivering data to users within 12 - 24 hours of the floats reaching the sea surface, the real-time quality control test procedures on Argo data are automatic but limited. When a file is detected to have been uploaded/updated on the GDAC, it is also uploaded/updated on the Coriolis Oracle database. These files then receive supplementary tests. The resulting alerts are checked visually by using the Scoop software [Detoc et al., 2021] and the QC flags are modified in the Coriolis database where appropriate. These alerts are based on the MinMax test [Gourrion et al., 2020].

This test proposes to discard the Argo profiles directly from minimum and maximum observed values instead of a classical approach that estimates validity bounds from first- and second-order moments of the climatological parameter distribution (mean, variance). The minimum and maximum values are computed for bins of longitude, latitude and pressure in hexagonal cells (110-km distance between two opposite vertices) (Figure 2) and are closer to the known local variability than a classical climatology. The MinMax test can detect contaminated portions of a profile, as well as anomalies caused by sensor drift. Whenever a quality code is changed in the Coriolis database, an automatic message is generated and sent to the corresponding DAC. Then the DAC has all the information to correct the QC on the data and upload the corrected files on the GDACs.

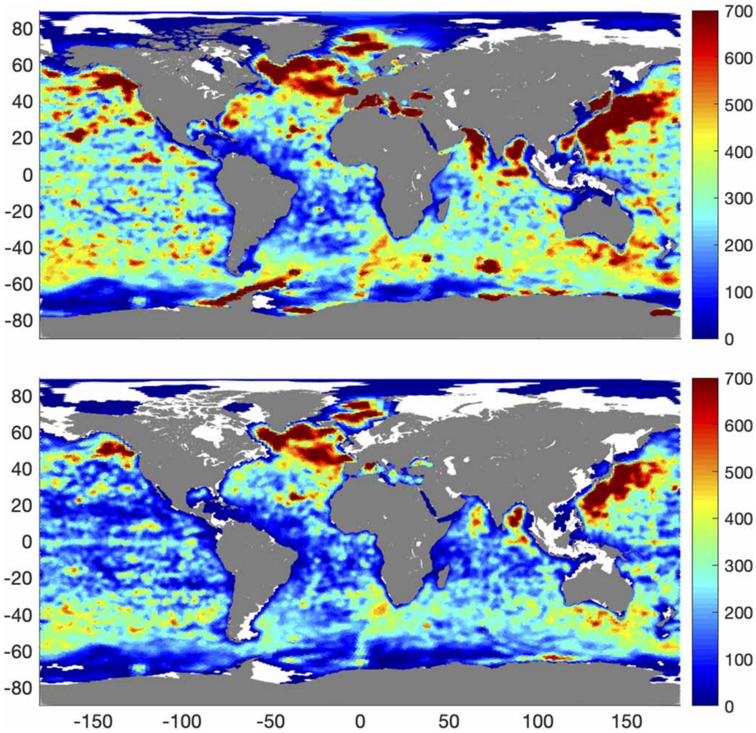


Figure 2 Number of profiles per grid cell: (top) the 0 – 20-m layer and (bottom) the 1000 – 1020-m layer.

This work also forms part of the Blue-Cloud 2026 project, and in particular of the tests carried out in the Workbenches (WBs) for physical variables.

Reference

Detoc J., Thepault B., Carval T., Mahoudo P., Garo M., (2021). *Scoop-Argo: visual quality control for Argo NetCDF data files*. SEANOE. https://doi.org/10.17882/48531

Gourrion J., Szekely T., Killick R., Owens B., Reverdin G., & Chapron B., (2020). *Improved Statistical Method for Quality Control of Hydrographic Observations*. Journal of Atmospheric and Oceanic Technology, 37(5), 789-806. https://doi.org/10.1175/JTECH-D-18-0244.1

Wong A., Keeley R., Carval T., Argo Data Management Team, (2024). *Argo Quality Control Manual for CTD and Trajectory Data*. https://doi.org/10.13155/33951

Supporting Ocean Digital Twins with FAIR marine image and video annotation in BIIGLE 2.0

Martin Zurowietz (martin@cebitec.uni-bielefeld.de), Daniel Langenkämper (dlangenk@cebitec.uni-bielefeld.de), Tim W. Nattkemper (tim.nattkemper@uni-bielefeld.de)

Bielefeld University (Germany)

Underwater visual data (digital images and video) collected with different platforms (e.g., ROV, AUV, towed or stationary observatories) play a key role in the analysis of marine ecosystem dynamics, the assessment and monitoring of biodiversity and the inspection of marine infrastructure (offshore platforms, windfarms, ships etc). With the advent of marine mining operations or the observed worldwide decrease of biodiversity, the pressure on the development of methods to manage and analyse the growing volumes of visual data increases. This pressure challenges the available working time capacities of marine biologists or engineers regarding the visual inspection and interpretation of the data. Nowadays, Ocean Digital Twins (ODT) are proposed as next generation digital platforms supporting monitoring, modelling, and management in marine operations, investigations, and research. However, the development and operation of ODT requires the integration of the marine visual data which also increases the pressure on the analysis of the data, i.e., the detection and classification of regions of interest. As an example, ROV videos showing marine infrastructure must be screened for potential cracks or damages. Another example task relevant for academia, authorities and marine industry is biodiversity assessment, where marine species must be quantified and classified according to specified taxonomic guidelines in a marine habitat. One very important aspect in the analysis of marine visual data is maintaining a high level of consistency in the quality of the data extracted from images when data are collected over extended periods, using varying equipment, and analysed by different individuals [Langenkämper et al., 2020]. To provide a solid data basis for ODT development, this kind of data extracted from digital visual data must a) be consistent in classification and detection and b) adhere to FAIR data principles (Findable, Accessible, Interoperable, Reusable). Since 2017 the web-based annotation tool BIIGLE 2.0 [Langenkämper et al., 2017] is available open source and free for academic use (https://biigle.de). It has become one of the most prominent online tools for marine visual data annotation worldwide with more than 2800 registered users (November 2023) from 59 countries and a growing number of on-site software instances in institutes such as Ifremer, Senckenberg or Helmholtz [Zurowietz et al., 2021]. This may qualify BIIGLE to be considered as a reference regarding software and methodology in protocols for marine visual data annotation in the context of monitoring marine environments and infrastructures. Here we present recently implemented BIIGLE 2.0 features that help users to achieve the goals a) and b) above.

Methods and Results

In the following, we will consider the case that several images or video frames have been inspected by some users and the data needs to be prepared for a next analytical step, which can, for instance, be a multivariate statistical analysis or feeding the data into an ODT. To achieve consistency in image annotation, BIIGLE offers two tools to evaluate the quality of the classification labels and of the location information that define the annotations.

In a first step, the consistency in the classification labels must be evaluated and approved. To this end, a new feature of BIIGLE's LARGO (LAbel Review Grid Overview) tool is now implemented to enable a quicker review of image annotations in a grid of image patches showing instances from the same label category (see Figure 1 upper right).

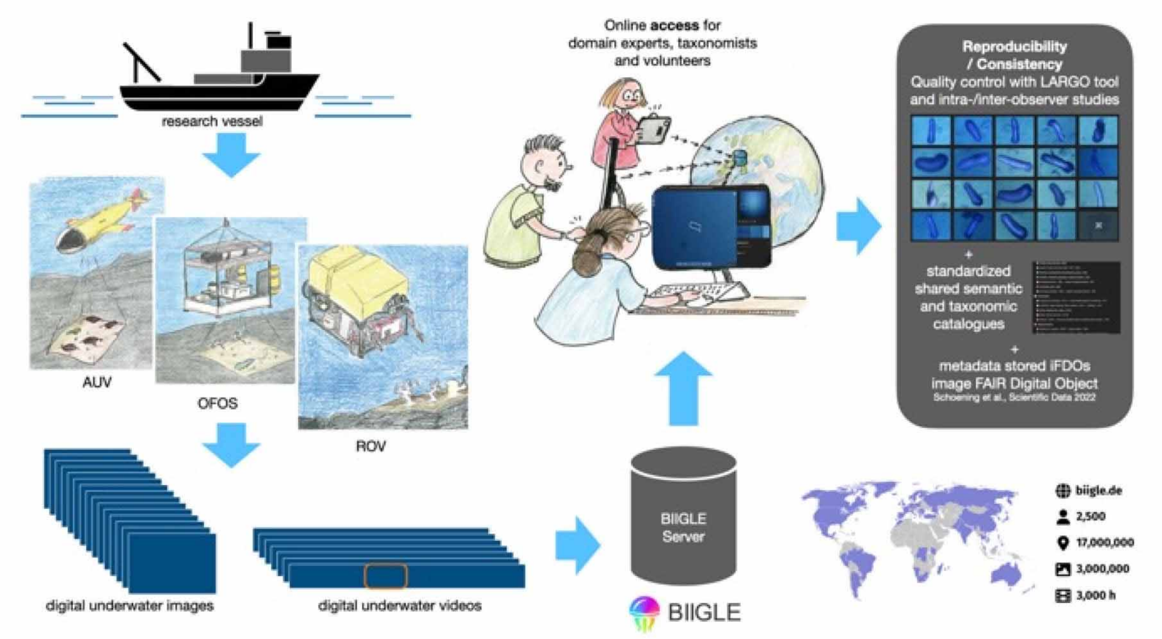


Figure 1 The overview for the proposed workflow starting after the underwater image / video acquisition and ending with the storage of standardized metadata using image FAIR digital objects.

The image patches can now be sorted according to the morphological similarities between objects in the patches. To this end, image features are computed for each patch with the pre-trained deep neural network DINOv2 [Oquab et al., 2023, arXiv preprint arXiv:2304.07193] and these features are compared using a cosine similarity function. When the LARGO image patches are sorted by inverse similarity to a reference patch of a given label category, the identification of annotation mistakes requires much less time, as these appear first in the grid.
In the second step, the consistency of the locations of the annotations (i.e., the detection performance of the users) can be analyzed using the BIIGLE 2.0 feature for "annotation sessions". Here, a collection of marine images can be selected for a session, and users annotate the images blind from each other. In the resulting annotation data from this kind of study, for each object (e.g., a structural damage or a species) it can be determined, how many users have marked this object, which allows for the calculation of an inter-observer agreement factor that quantifies the consistency in the detection. If one user annotates the same data in such a study twice (usually with a few days delay) the intra-observer agreement factor can be calculated.
In the third step, the principles of FAIR data are supported in BIIGLE 2.0 making use of the recently proposed iFDO [Schoening et al., 2022]. The idea behind iFDO is that digital marine image collections require a set of standardized metadata such as the date, deployment location, dive ID, mission ID, and hardware parameters (camera, ROV etc.). Information on ownership and authorization should also be registered, together with annotation results. The iFDOs standard aims to build a bridge between (and closing gaps) in other metadata standards (such as DublinCore, Audubon, SmartarID, PDS4). BIIGLE 2.0 supports the import of iFDO files to create new image or video collections and the export of iFDO files that are enriched with annotation data created with BIIGLE. Metadata combined with annotation data from exported iFDO files are an important resource for the development of ODTs.

References

Langenkämper D. et al., (2017). *BIIGLE 2.0 - Browsing and Annotating Large Marine Image Collections*. Front. Mar. Sci., 4, 2017, 83. https://doi.org/10.3389/fmars.2017.00083
Langenkämper D. et al., (2020). *Gear-Induced Concept Drift in Marine Images and Its Effect on Deep*

Learning Classification. Front. Mar. Sci., 7, 2020, 506. https://doi.org/10.3389/fmars.2020.00506
Schoening T., Durden J.M., Faber C. et al., (2022). *Making marine image data FAIR*. Sci Data 9, 414. https://doi.org/10.1038/s41597-022-01491-3
Zurowietz M. et al., (2021). *Current trends and future directions of large-scale image and video annotation: Observations from four years of BIIGLE 2.0*. Front. Mar. Sci., Sec. Ocean Observation. Volume 8 – 2021. https://doi.org/10.3389/fmars.2021.760036

A Mesoscale Events Classifier for Sea Surface Temperature Data

Marco Reggiannini (marco.reggiannini@isti.cnr.it), Oscar Papini (oscar.papini@isti.cnr.it), Gabriele Pieri (gabriele.pieri@isti.cnr.it)

Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche - CNR-ISTI (Italy)

The identification of mesoscale phenomena, such as upwelling, countercurrents and filaments, is an important task for oceanographers. Indeed, the occurrence of such processes involves variations in the density of nutrients which, in turn, influences the biological parameters of the habitat. In this work, we describe a novel method for an automatic classification system, the Mesoscale Events Classifier (MEC), dedicated to recognising marine mesoscale events. MEC is devoted to the study of these phenomena through the analysis of Sea Surface Temperature (SST) images captured by satellite missions.
Figure 1 shows the steps of the MEC algorithm and a classification example. MEC is able to detect different upwelling regimes; in this paper it is shown the case study of the southwestern part of the Iberian Peninsula, where four main types of patterns can be recognised, hereby called E1–E4. Two sources of SST data have been used in this analysis: the satellites of EUMETSAT's Metop programme and the satellite Aqua of NASA. In particular, SST data has been downloaded from the respective data centres in form of NetCDF files, and a preliminary selection has been performed in order to discard files without a significant amount of data located in our area of interest (between 35° and 40° latitude N, and 12° and 6° longitude W).

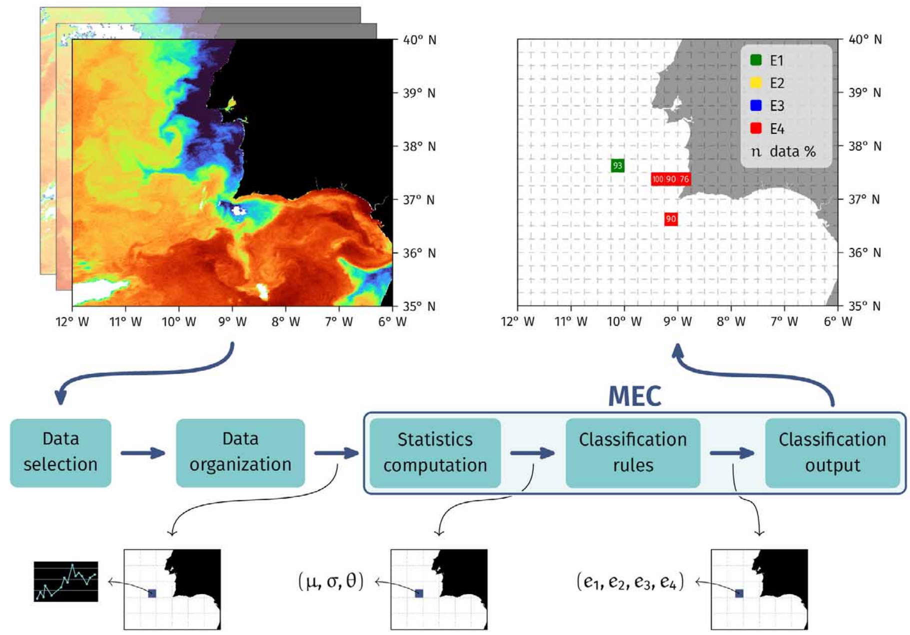


Figure 1 Block description of the MEC algorithm. Top left: map representing the SST at a certain time, as extracted from satellite data (blue: lower SST, red: higher SST); top right: map resulting from the application of the algorithm, where different types of patterns are located and classified.

A custom tool has been developed to extract the relevant information from the files and prepare it to be fed to the main part of the algorithm. The area of interest is first subdivided in squares with side length of 0.25°; for each square a time series is computed by averaging the SSTs acquired in the 15 days preceding the final observation. The multiple time series obtained by repeating such operation within a given region can be plotted in a common reference system to obtain a spaghetti plot that can be visually analysed to recognise patterns in the different temperature trends of the squares and associate them with a corresponding mesoscale event.

Once all the time series have been collected, the main part of MEC can begin. The first step consists of the computation of three statistics for each series, namely the temporal mean (μ) of the SST, the standard deviation (σ) and the slope of the straight line that better interpolates the data (θ). Next, these values are passed to a set of custom-crafted classification rules, such that for each square a vector of four scores (e_1, e_2, e_3, e_4) is computed depending on the values of μ, σ and θ in the square and in its neighbourhood. Each score e_i is a real number between 0 and 1 that represents a belief index for an event of type E*i* to have occurred in the square in the considered time period.

Finally, the last step determines a label ideally for every square in the map depending on its scores, without considering that the occurrence of such phenomena is usually observed only in quite defined geographical domains. The introduction of geographical constraints allows to filter out physically unrealistic classifications. This way, a label is assigned only to squares that fulfil the following conditions: considering the maximum score of the square (suppose that it is e_k), if it is higher than a fixed threshold and the square belongs to a geographical zone in which an event of type E*k* may occur, then the square is given the label "E*k*". Moreover, the percentage of the actual amount of data with respect to the expected one (computed using the spatial resolution and the temporal frequency of the satellite images) is reported for each square. This provides immediate feedback about the reliability of the classification. The application of MEC to remote sensing data has been presented in previous works [Pieri et al., 2023; Reggiannini et al., 2023]. In particular, a promising alignment between the classifier output and the ground truth has been clearly observed and described.

It is worth pointing out that a geometrical model for a mesoscale pattern cannot be defined with a sufficiently general criterion. This fact, which is due to the variability of the spatial configurations of mesoscale phenomena, actually prevents the problem from being tackled with a template matching approach. The extension from the analysis of the mere spatial properties of the SST data, considered at fixed time, to the analysis of the data within an extended spatial/temporal domain enables to counterbalance the lack of information about the geometry of the pattern with the information related to the history of the pattern formation. Hence, the search for a specific temperature spatial pattern is replaced by the detection of a specific temperature behaviour over time, within a specific spatial domain. To the best of the authors' knowledge this represents the main novelty introduced by MEC to the detection and classification of mesoscale phenomena. Moreover, under a classical approach a given SST map is associated, at most, with a single mesoscale class only. As mentioned, MEC integrates within a single frame the joint analysis of the temperature spatial and temporal properties. Thanks to this approach, it may occur that fragments of the life cycles of multiple patterns fall within the same observation window considered by MEC. These patterns, concurrently forming and developing, may then be detected thanks to the MEC enriched point of view.

The tools described in this paper have been implemented in a Python framework and are available on GitHub (https://github.com/ospapini/nautilos-T8.5-sst).

The authors thank Prof. Flávio Martins and Dr João Janeiro, from the University of Algarve, for their support. This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000825 (NAUTILOS).

References

Pieri G., Janeiro J., Martins F., Papini O., and Reggiannini M., (2023). *MEC: A Mesoscale Events Classifier for Oceanographic Imagery*. Applied Sciences 13 (3), 1565. https://doi.org/10.3390/app13031565

Reggiannini M., Papini O., and Pieri G., (2023). *Evaluation of a Mesoscale Event Classifier*. IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW). https://doi.org/10.1109/ICASSPW59220.2023.10193234

SNAPGlobeHub - Geophysical data discovery and access on a global scale

Paolo Diviacco (pdiviacco@ogs.it), Alessandro Busato (busato@ogs.it), Mihai Burca (mburca@ogs.it), Alberto Viola (aviola@ogs.it), Nikolas Potleca (npotleca@ogs.it)

Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)

Introduction

Geophysics is an exploration method that can provide insights into the earth's crust down to a depth of several kilometers. This approach has been used in many areas of the Earth, collecting large data sets that allow to improve the knowledge of the Earth's dynamics, and that can be used for environment protection, energy and engineering projects.
Access to current and legacy datasets for a broad scientific community is of paramount importance to promote collaborative research among scientists. In this context, the ability to find, view and potentially work on data directly on the Internet is of great importance.
Initiatives aiming at such goals often focus on specific geographic areas, such as the European seas or Antarctica, which are far apart and have different needs in terms of projections or bounding boxes. It is not easy to find a one-size-fits-all solution for web mapping and data access to georeferenced data in such different environments. At the same time, the analysis and comparison of data from such different and distant regions is extremely important, for example, for climate studies.

SNAP

Notwithstanding the fact that the goal of openness is of paramount importance in a perspective of an increasingly interconnected world of scientific research and of Open Access publications, it is also important to consider that, a certain reluctance in data sharing exists in the field of geophysics. In this sense any web based geophysical data management initiative needs to implement a flexible balance between protection and openness.To address such problems, OGS has developed a specific web-based geophysical data management system called SNAP (Seismic data Network Access Point) (Diviacco et al., 2019; 2016; 2015) that can handle the previously mentioned issues using a controlled data access workflow, while being, at the same time, compliant to the FAIR paradigm, handling DOIs and following international standards such as those developed within SeaDataNet, Inspire, ISO, OGC and SEG.
SNAP is used as the main web based data portal for geophysical data acquired by OGS and can be accessed at the following web address: https://snap.ogs.trieste.it. It contains mainly recent and vintage geophisical data from the Mediterranean and the Black Sea areas. As a web framework, SNAP is used also within several international data dissemination initiatives such as EMODnet, SeaDataNet, the SCAR Antarctic Seismic Data Library (SDLS), PNRA-NADC and others.
All these initiatives live generally isolated from each other, although the SNAP engine is common to all of them. This is due to organizational and funding reasons but also to difficulties in having a common and easy to be used geographic mapping tool to employ during data discovery and access. Areas that are far from each other have different needs in terms of extension and projections that generally is solved using separated views when not completely different portals. For example the SDLS (Antartica) cannot simply use the Mercator web projection (EPSG:3857) that is commonly used by most of the online map providers such as OpenStreetMap. In this case, instead, an Antarctic Polar Stereographic projection (EPSG:3031) should be used in order to reduce projection distortions and to be able to follow surveys across the International Date Line.

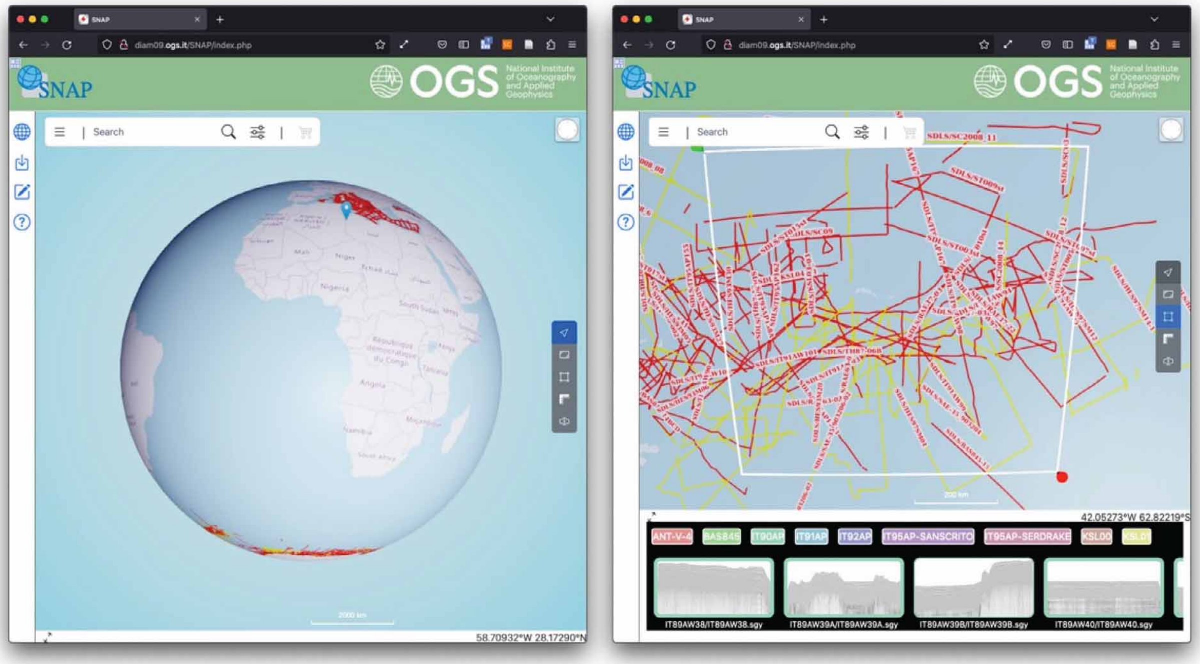


Figure 1 Globe based data discovery and data selection of data acquired by OGS in the Mediterranean and Antarctic area

SNAPGlobeHub

To overcome these problems we developed a new version of SNAP called SNAPGlobeHub, that avoids using local projections being based on an interactive 3D digital globe where all data positioning can be rendered simultaneously. This way distortions are avoided, all surveys can be easilly identified and followed, and comparisons between different and distant areas can be easilly done.

SnapGlobeHub is an HTML5 + CSS + Javascript + PHP Web Application that offers the possibility to display on a model of the planet (the globe), WHERE the geo-refereenced datasets are and HOW to interact with them exploiting the WebGL computer graphics extension offered by most of the popular web browsers. The graphic part of SnapGlobeHub, the mathematical libraries and the interactivity functions are provided by the OpenGlobus library (https://openglobus.org).

SNAPGlobeHub take advantage of the possibilities offered by OGC services like WMS (Web Map Service) and WFS (Web Feature Service) so that multiple initiatives (such as the already mentione SDLS and main SNAP portal but also others) can integrate their public datasets catalog in a single data space relying on a WMS service to display geometries on the globe texture and a WFS service to query the catalog upon either geographic bounding-boxses or textual queries using the CQL language. Queries results are a list of dataset features like the canonical name and description, the URL of the thumbnail that identifies the dataset, the URL to a service that offers a page with all the information and metadata that characterize the dataset(for example. GeoNetwork) and if available an URL to the viewer that allows users to interact with a graphical representation of the dataset or to eventually download the data.

References

Diviacco P., Firetto Carlino M., Busato A., (2019). Enhancing the value of public vintage seismic data in the Italian offshore. Geoscience Data Journal, 6 (1), pp. 6-15. https://doi.org/10.1002/gdj3.58

Diviacco P., Munoz R.C., Sorribas J., Loubrieu T., (2016). A metadata model for cross-domain marine data management, the SeaDataNet/Geo-Seas experience. Bollettino di Geofisica Teorica ed Applicata, 57 (1), pp. 59-69. https://doi.org/10.4430/bgta0163

Diviacco P., Wardell N., Forlin E., Sauli C., Burca M., Busato A., Centonze J., Pelos C., (2015). Data rescue to extend the value of vintage seismic data: The OGS-SNAP experience. GeoResJ, 6, pp. 44-52. https://doi.org/10.1016/j.grj.2015.01.006

STREAMing on-demand satellite data

Aldo Drago[1] (aldo.drago@mcast.edu.mt), David Debono[1] (david.debono@mcast.edu.mt), Eugenio Busuttil[1] (eugenio.busuttil@mcast.edu.mt), Francis Delicata[1] (francis.delicata@mcast.edu.mt), Roberto Drago[2] (roberto.drago@mst.com.mt), Ryan Vella[2] (ryanvella96@gmail.com), Bjorn Azzopardi[3] (bjorn@think.mt), Nick De Gabriele[3] (nick@think.mt), Glyn Mercieca[3] (glyn@think.mt), Vlatko Roland[4] (vlatko@prehnit.hr), Ghada El Serafy[5] (ghada.elserafy@deltares.nl), Lorinc Meszaros[5] (lorinc.meszaros@deltares.nl), Anna Spinosa[5] (anna.spinosa@deltares.nl), Robyn Gwee[5] (robyn.gwee@deltares.nl), Maciej Markowski[6] (maciej.markowski@ug.edu.pl)

[1]MCAST (Malta)
[2]MST AudioVisual Ltd (Malta)
[3]THINK DESIGN Ltd (Malta)
[4]Prehnit Ltd (Croatia)
[5]DELTARES (The Netherlands)
[6]GIS Laboratory, University of Gdansk (Poland)

Aim

Innovative downstream services are evolving to popularise the uptake of data, catering to the demand of a knowledge-based society seeking faster and selective access to information. These services are easily accessible and allow for the retrieval of data via popular smart mass media without requiring data science skills. STREAM builds on these aspects to provide at fingertips, on-demand satellite earth observations to general users with access over smartphones and tablets. The work done in this project can be followed on www.stream-srf.com.

STREAM enhances the level of access and application of space data with local users, including the citizen level, to facilitate its integration into daily endeavours by operators and responsible entities, for decision-making, higher efficiency, and performance at reduced risks, delivering tailor-made and added-value services needed at different scales, disciplines, and levels of society. The project provides an intermediate platform dedicated to data fields covering the sea areas around the Maltese Islands, and more broadly the Malta shelf area up to the southern Sicilian coast, easily channelling data customised to user needs from different satellite data sources.

Results

The project is funded by the Space Research Fund of the Malta Councuil for Science & Technology. The Institute of Engineering and Transport of the Malta College of Arts, Science and Technology leads its scientific coordination. The projectis implemented jointly with two Maltese SMEs, MST AudioVisual Ltd and THINK Design Ltd, together with specialised staff including the scientific excellence of DELTARES on subcontracting basis. This collaboration brings into the project expertise for system design, setup, and operation, as well as for the user interfaces and the branded service delivery on electronic hand-held devices.

STREAM focuses on dedicated services that aim to enhance and expand the local capability in utilizing full resolution operational satellite data merged to numerical modelling data. The service streams raw data upon user-defined requests, and additionally delivers specialised derived products in the form of indicators, including those for heat wave alerting and water quality. The STREAM service is easy to use through a mobile-enabled interface that allows the user to make specific individual requests, selected and compiled on multi-functional input dashboards. Although currently intended as a proof of concept, the service targets different users, professional and non-professional, for research, education and awareness, monitoring and expert assessments, decision-making and planning. The services also support third party initiatives boosting private enterprise in many sectors. The area of interest is the Malta Shelf area covering the stretch of sea between Malta and Sicily, zooming on the sea around the Maltese Islands.

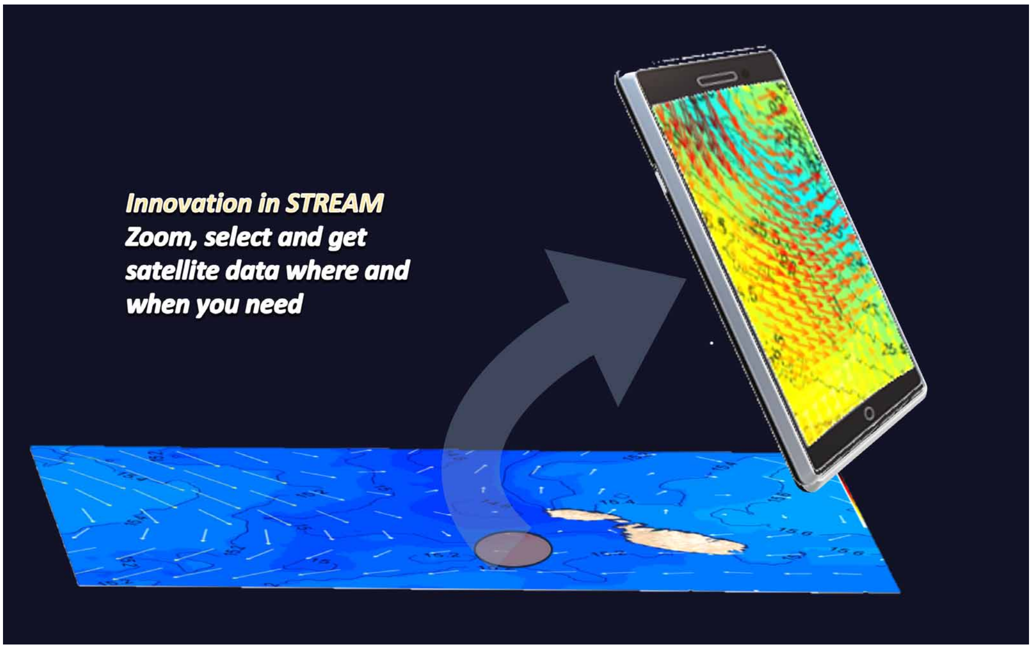


Figure 1 Concept of the STREAM service for on-demand satellite data delivered on mobile phones.

Research impact

The main project challenge lies in the various technical aspects that need to be brought seamlessly to work together in a multi-functional process chain, encompassing tasks from data sourcing to operational scripts that automate data harvesting, automation of data processing and elaboration, to storage on a dedicated server that handles and organises spatial data, backend processes that channel data and respond to requests from a dedicated user interface that is intended to deliver peer services on a mobile application. These are the requisites of the STREAM digital platform that follows a modular design, using the latest available technologies, providing a versatile system architecture to accommodate future evolutions. Hardware components consist of multiple redundant servers running VMware as the hypervisor, together with a centralised Network Attached Storage (NAS) interconnected over 10G connections to the core switch. A number of linux virtual machines are also used to house all the operational activity, providing SSL termination, firewalling and VPN connectivity. The GeoServer is the main software component through which end users will access data on the platform. Data is stored as a combination of GeoTiff and NetCDF files, alongside a PostgreSQL database with PostGIS extensions for data preparation to front-end via mapping and tiling Application Programming Interfaces (APIs) for delivery to users through OpenLayers and KML integration. Other software components comprise the JAVA Spring framework which is used as the main interface between internal components of the platform and the database. This same framework is also used to expose a restricted set of REST APIs to the front-end application for metadata sharing. Python is the primary programming language for automation, scheduling, data processing, and value addition. The Featured Manipulation Engine (FME) is also used as the main Extract Transform Load (ETL) tool for modular data elaboration and integration, simplifying data workflows through in-built extract, transform and processing functions, with data being subsequently loaded into

the central database and file repository. The FME's graphical user interface speeds up the overall design project and favours the performance of the system at execution.
Several innovations are implemented in STREAM:

1. EO data are fused to other data fields such as from models to enable data inter-comparisons and forecasting;
2. besides some web-based applications, the main service is made available on smartphones and tablets for easier on the spot accessibility to common users;
3. visualisation will allow superposition of fields with user-friendly functionality;
4. features include the viewing of time series at selected points, transects or areas of interest, together with the elaboration of simple statistics;
5. real-time data can be compared to climatological values;
6. data selection and downloading is available from the same service to provide one system for viewing, selecting and access to data. Moreover, the data fields will be restricted on the limited area of interest to enable higher resolution data fields at faster performance.

Citizen Science, data quality and beyond

Paolo Diviacco (pdiviacco@ogs.it), Rodrigo Carbajales (rcarbajales@ogs.it), Gilda Savonitto (gsavonitto@ogs.it), Alberto Viola (aviola@ogs.it), Riccardo Gerin (rgerin@ogs.it), Massimiliano Iurcev (miurcev@ogs.it), Cinzia Comici (ccomici@ogs.it), Lorenzo Grio (lgrio@ogs.it), Nikolas Potleca (npotleca@ogs.it), Stefano Kuchler (skuchler@ogs.it)

Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)

Introduction

Citizen science (CS) is a term that designate scientific research where some activities are outsourced to laypeople. These activities are generally limited to data acquisition, measurements and observations. This approach has many advantages such as the reduction of costs in personnel, the increase in geographic and temporal resolution and coverage, and, very importantly, the possibility to spread awareness of scientific topics among the general public. The downsides of this paradigm are related mainly to the quality of the data that are acquired by volunteers. Often these latter lack the skills to perform optimal measurements. Often, when a large number of volunteers is required to cover large areas, in order to keep the overall costs of the projects under control, low cost (LC) sensors need to be used. QC/QA is then an important step in CS.

The Gena infrastructure

OGS developed and maintains a large CS based infrastructure (Gena) that follows the complete path from acquisition, transmission, storage, integration to real time web visualization and mapping of mobile CS data acquired in several research fields from air quality [Diviacco et al., 2022], noise in urban areas to marine environmental parameters such as pH, temperature, dissolved oxygen and salinity. [Diviacco et al., 2021]. The acquisition system is based on a hub box to which various LC sensors can be connected depending on the aim of the survey. All boxes embed a GPS device that allows geolocation and timestamping. Transmission is granted by a specific unit that allows to opt between GSM, WiFi or LoRaWAN technology depending on the available coverage. Data are collected in an InfluxDB database, which allows easy integration with TheThingsNetwork for LoRaWAN network management and directly with GSM and WiFi connections. Data are processed and rendered in a web GIS using standard OGC services and following a strict FAIR approach.

QC /QA and CS within the Gena Infrastructure

In the marine case, considering for example pH sensors, following Yang et al. [2014] the accuracy of different types of technologies, corresponding to different ranges of costs, can scale down from ±0.001 to ±0.1. Okazaki et al. [2017] maintained that at sea, where temperature, gas exchange, pressure, and biological processes simultaneously affect seawater, precision and accuracy is not easy to be controlled, so that in these conditions, the gap between LC and high-end probes is moved from measurement accuracy and precision more towards repeatability and stability. Several LC temperature and conductivity sensors were thoroughly evaluated and compared with laboratory-level secondary reference standards at the OGS Oceanographic Calibration and Metrology Center (CTMO) in a temperature- and humidity-controlled environment. The temperature sensors showed a good performance while the salinity sensors showed limited accuracy. The issue of accuracy of LC sensors is consistent with what already demonstrated in other environments by Diviacco et al. [2022]. This suggests that they can be used to reconstruct a 'relative' map of distribution of a parameter but cannot provide absolute' values. After an accurate initial calibration, within the Gena Infrastructure each measured value is QA/QC in order to identify outliers that can be generated by GPS errors or electronic glitches. Later, measurements are gridded and averaged within each cell. Cell size is kept by default at 200m by 200m but this setting can be changed depending on the rate of sampling, the number of participants and their speed. It is possible, then, to map geographically these value obtaining the distribution of the parameter in an area.

Reference correction

If the accuracy of the distribution obtained from CS measurements is disputable, envisioning it as a surface, its shape, can be instead generally considered representative of the actual physical phenomena. When an accurate source of measurements is available within the surveyed area the CS surface can be scaled in order to minimize the difference between the CS distribution and the reference source of measurements. In the marine case this can be obtained using reference sensorized buoys. In the Gulf of Trieste this is obtained using the OGS Mambo buoy site that sends in realtime measurements that can be immediately compared with CS data. This approach allows to extend geographically high quality measurements that otherwise would remain confined to the location where the reference sensors are located. The corrected distribution will then allow to highlight trends, hotspots and anomalies in a larger area with a reasonable precision and accuracy.

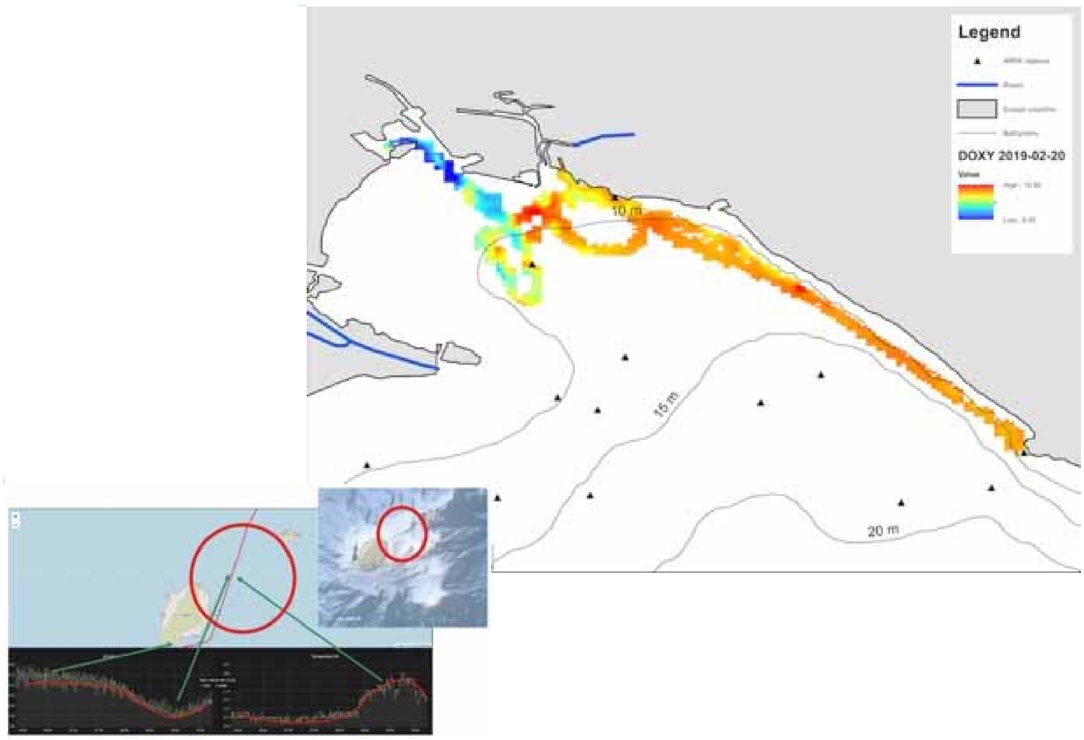


Figure 1 Example of CS based marine data surveys. Left box: a survey in the vicinity of the Panarea island (Italy) characterized by calc-alkaline volcanism and related hydrothermal systems. The red circle in the map highlights an area where pH values decrease (bottom left) while temperatures rise (bottom right). Right box: an area of the Gulf of Trieste where multiple boats acquire data simultaneously

References

Diviacco P., Iurcev M., Carbajales R.J., Potleca N., Viola A., Burca M., Busato A., (2022). *Monitoring Air Quality in Urban Areas Using a Vehicle Sensor Network (VSN).* Crowdsensing Paradigm

Remote Sensing, 14 (21), art. no. 5576, https://doi.org/10.3390/rs14215576
Diviacco P., Iurcev M., Carbajales R.J., Potleca N., (2022). *First Results of the Application of a Citizen Science-Based Mobile Monitoring System to the Study of Household Heating Emissions*. Atmosphere, 13 (10), art. no. 1689, https://doi.org/10.3390/atmos13101689
Diviacco P., Nadali A., Iurcev M., Carbajales R., Busato A., Pavan A., Burca M., Grio L., Nolich M., Molinaro A., Malfatti F., (2021). *MaDCrow, a Citizen Science Infrastructure to Monitor Water Quality in the Gulf of Trieste (North Adriatic Sea)*. Frontiers in Marine Science, 8, art. no. 619898, https://doi.org/10.3389/fmars.2021.619898
Okazaki R., Sutton A., Feely Andrew R., Dickson G., Simone R., Alin L., et al., (2017). *Evaluation of marine pH sensors under controlled and natural conditions for the Wendy Schmidt Ocean Health XPRIZE*. Limnol. Oceanogr. Methods 15:2017, https://doi.org/10.1002/lom3.10189
Yang B., Patsavas M.C., Byrne R.H., and Ma J., (2014). *Seawater pH measurements in the field: a DIY photometer with 0.01 unit pH accuracy*. Mar. Chem. 160, 75–81, https://doi.org/10.1016/j.marchem.2014.01.005

SmartDots, from age determination to fish family planning with eggs, larvae and maturity determination

Carlos Pinto[1] (carlos@ices.dk), Julie Olivia Davies[2] (joco@aqua.dtu.dk), Karen Bekaert[3], (Karen.Bekaert@ilvo.vlaanderen.be), Joana Ribeiro[1] (joana.ribeiro@ices.dk), Kevin De Coster[3] (Kevin.DeCoster@ilvo.vlaanderen.be), Wim Allegaert[3] (wim.allegaert@ilvo.vlaanderen.be)

[1]International Council for Exploration of the Sea – ICES (Denmark)
[2]DTU (Denmark)
[3]ILVO (Belgium)

Introduction

SmartDots is a platform for quality assurance of biological parameters as input for stock assessment. It was launched in 2018 and has since been developed by The ICES Working Group on SmartDots Governance (WGSMART) who oversee all improvements and ensure all developments are in line with the ICES quality assurance framework (QAF). SmartDots has become a core tool in calibration and training of technicians across national laboratories, supporting the standardisation of procedures and data output. There are a number of modules which are interlinked; the software where images are annotated, a database where all images and associated data are stored, a web application which communicates between the two and serves as a management site for users, and a reporting module which produces a report template based on a standardised statistical analysis run from an r-script. It is hosted at ICES and currently has 862 users registered from 45 countries.

SmartDots platform description (for age reading)

The SmartDots software supports the age reading module which facilitates age reading based on otolith images. The software enables the age readers to participate in ICES age reading workshops and exchanges by annotating a set of otolith images within an event, annotations and final age estimations can be compared against other readers in other laboratories, either remotely or in a workshop setting. User friendly functions within the age reading software include image sorting, image adjustment, measurement, identification of specific features for discussion and assignment of a scale for quality control purposes. The web application enables the workshop or exchange manager to manage all users and meta data related to the SmartDots events. All registered data are available in the connected reporting environment where r-scripts produce results from a standardised statistical analysis which can be directly included in the stock assessment model runs.

Request to integrate maturity, egg and larvae modules

ICES was requested and financed by EU and UK to develop additional SmartDots software modules to host a maturity staging, a fish larvae and an egg identification module and to provide user training for these modules. There was the need for an API and software specific features for each of the modules. These were developed individually but relied on the work done for the age reading module. Testing and documentation of the API and software needed to be developed. The age reading module served as a template for how to further develop the maturity, egg and larvae modules, both in a technical sense but also in a wider sense where the benefits of the improvements accredited to SmartDots are extended into management and advice processes.

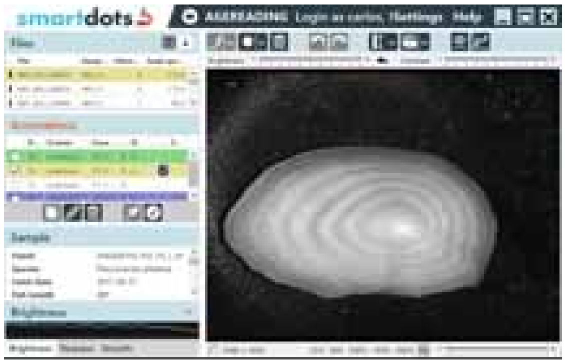


Figure 1 Age reading module.

Figure 2 Maturity staging module.

Figure 3 Larvae identification module.

Figure 4 Egg identification module.

Technical description software, WebAPI, reporting and webinterface

The SmartDots platform is an open-source solution (licensed under GPL v3.0) with a client-side Graphical User Interface, developed as a Microsoft WPF application by the Flanders Research Institute for Agriculture, Fisheries and Food (ILVO). This software facilitates the exchange of annotation data via a dedicated Web API, developed by ICES. Within the SmartDots system, events are not only created but also managed via the user-friendly web application, a network of Smartusers fosters a sense of community and expertise mantained by the country coordinators. Participation in these events requires a username, while public events can be accessed using a convenient guest token. SmartDots simplifies the process of engagement, enabling users to harness the power of data annotation and collaboration with ease.

Conclusion

These three new modules—maturity staging, fish larvae, and egg identification—represent significant advancements in the SmartDots platform. They are expected to streamline the quality control and calibration processes by enabling a greater number of events of these types without the logistical challenges of gathering in one location. SmartDots is set to facilitate the way data is collected, analyzed and quality assured, making collaboration more accessible and efficient.

Future

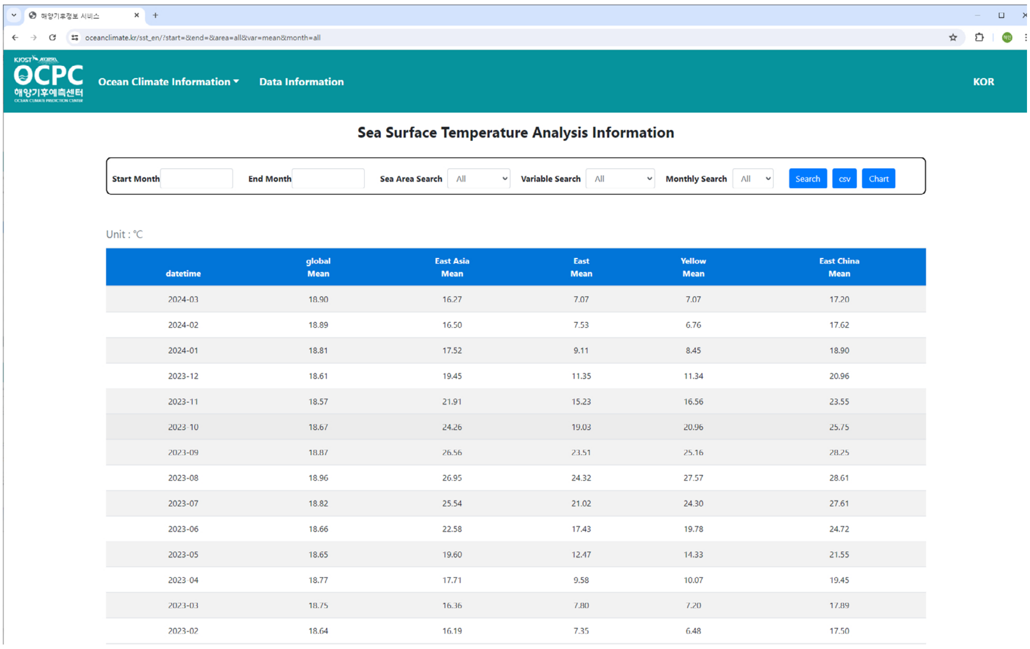
Up next in the development pipeline are “reference collections” and training modules. These modules will empower SmartDots to store community built reference collections, facilitating reader training while tracking changes in users interpretations overtime Furthermore, they open doors AI algorithm evaluation, ensuring a well-developed platform.

Data Service System for Ocean Climate Analysis Information

Hyuk-Min Park (hyukmin@kiost.ac.kr), Min-Ho Kwon (mhkwon@kiost.ac.kr),
Hyoun-Woo Kang (hwkang@kiost.ac.kr), Jiyun Shin (shinjy@kiost.ac.kr), Ho-Jeong Ju (jussi7434@kiost.ac.kr), Mi-Jin Kim (mijin1902@kiost.ac.kr)

Korea Institute of Ocean Science and Technology (Rep. of Korea)

Climate change is increasingly becoming a global issue, and the seas around Korea are significantly affected by these changes. In response, we are conducting thorough analyses of marine and atmospheric variables in the World Oceans and the East Asia Seas. The East Asia Seas area, spanning from 20°N to 53.5°N and 110°E to 150°E, is further dissected into specific regions such as the East Sea, Yellow Sea, and East China Sea for more detailed examination. To make our analysis results easily accessible not only to researchers but also to the general public, we have developed a marine climate analysis information system.
Data collected from various organizations, such as the NOAA National Centers for Environmental Information (NCEI), Copernicus Marine Environment Monitoring Service (CMEMS), European Centre for Medium-Range Weather Forecasts (ECMWF), NOAA National Snow and Ice Data Center (NSIDC), and Polar Science Center (PSC), are used to analyze the monthly conditions and variations of nine oceanic variables (Sea Surface Temperature, Absolute Dynamic Topography, Significant Wave Height, Mean Wave Periods, Sea Ice Extent, Thickness, El Niño/La Niña, Indian Ocean Dipole, River Discharge) and four atmospheric variables (2m Air Temperature, Sea Level Pressure, Accumulated Precipitation, 10m Wind Speed) in the World Ocean, East Asia Sea, East Sea, Yellow Sea, and East China Sea. Additionally, monthly anomalies are analyzed based on the standard deviations and average conditions over the past 30 years (1991 - 2020, with Absolute Dynamic Topography using 1993 - 2022 as its reference period). These monthly analyses of ocean and atmospheric variables are updated around the 15th of each month in the data service system.
This system is constructed using the Python web framework Django, employing the Model-Template-View (MTV) pattern, and utilizes Nginx as the Web Application Server (WAS). Gunicorn serves as the communication interface between Python and WAS. The database management system used was MariaDB, and the Highchart library was used for the chart creation function. The analyzed data (monthly average, monthly anomaly, base period average, standard deviation) are stored in a database, allowing users to search for specific seas, time periods, and variables of interest through this system (Figure 1). Users also have the option to download these results in CSV format and to create interactive charts for each sea area and variable (Figure 2).
Additionally, we plan to provide a spatial distribution visualization service to display the geographical patterns of the analyzed data. Access to this system is available at https://oceanclimate.kr.



datetime	global Mean	East Asia Mean	East Mean	Yellow Mean	East China Mean
2024-03	18.90	16.27	7.07	7.07	17.20
2024-02	18.89	16.50	7.53	6.76	17.62
2024-01	18.81	17.52	9.11	8.45	18.90
2023-12	18.61	19.45	11.35	11.34	20.96
2023-11	18.57	21.91	15.23	16.56	23.55
2023-10	18.67	24.26	19.03	20.96	25.75
2023-09	18.87	26.56	23.51	25.16	28.25
2023-08	18.96	26.95	24.32	27.57	28.61
2023-07	18.82	25.54	21.02	24.30	27.61
2023-06	18.66	22.58	17.43	19.78	24.72
2023-05	18.65	19.60	12.47	14.33	21.55
2023-04	18.77	17.71	9.58	10.07	19.45
2023-03	18.75	16.36	7.80	7.20	17.89
2023-02	18.64	16.19	7.35	6.48	17.50

Figure 1 Database Query Service (oceanclimate.kr).

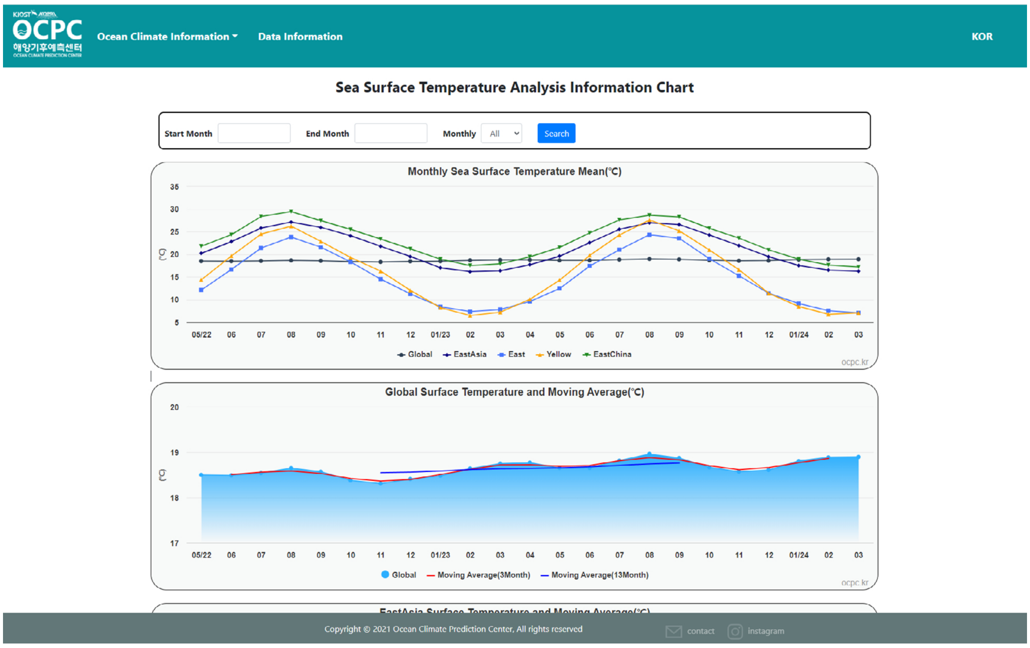


Figure 2 Chart Service (oceanclimate.kr).

HELMI data management - Secure sensor data transfer and publication

Linda Baldewein[1] (linda.baldewein@hereon.de), Max Böcke[1] (max.boecke@hereon.de), Housam Dibeh[1] (housam.dibeh@hereon.de), Tim Leefmann[2] (tim.leefmann@gv.hamburg.de), Joost Hemmen[3] (joost.hemmen@ufz.de), Rehan Chaudhary[1] (rehan.chaudhary@hereon.de), Christian Jacobsen[1] (christian.jacobsen@hereon.de), Jörn Plewka[1] (joern.plewka@hereon.de), Ulrike Kleeberg[1] (ulrike.kleeberg@hereon.de)

[1]Helmholtz-Zentrum Hereon (Germany)
[2]Freie und Hansestadt Hamburg, Landesbetrieb Geoinformation und Vermessung (Germany)
[3]Helmholtz-Centre for Environmental Research (Germany)

The collection and usage of sensor data are crucial in science, enabling the evaluation of experiments and validation of numerical simulations. With the rise of cyber security attacks on research centers in the last years, a secure data transfer from sensor systems, that are oftentimes located outside of the campus of a research center and thus are not connected to the same network as the rest of the IT infrastructure, is paramount. The HEreon Layer for Managing Incoming data (HELMI) management workflow establishes a secure pathway from external moving or stationary sensor platforms via a secure gateway using a VPN to data storage. It is followed by modules for quality control, visualization, and publication features.

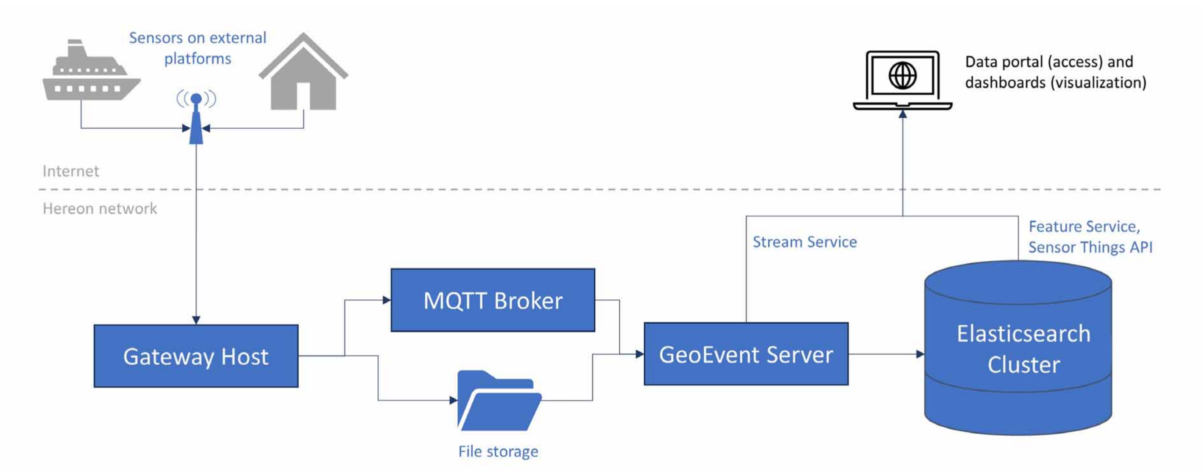


Figure 1 The HEreon Layer for Managing Incoming data (HELMI) management workflow, depicting data from sensor platforms entering the Hereon network through a gateway, being stored, and processed by an MQTT broker. It is followed by rerouting of data for publication as stream, SensorThings API and feature services.

To guarantee a secure data transfer, a gateway host ensures the secure Virtual Private Network communication between the sensor and the Hereon network. Only once the secure handshake between the external sensor system and the gateway host is successful, a connection for data transfer is established.

HELMI uses two different data transfer and processing pathways depending on the data generated by the sensor systems, either real-time data or data files. For real-time data, a fast transport to the centralized data infrastructure for display in for example web applications is critical. They are transferred using the MQTT protocol, an extremely lightweight messaging protocol which can work with low bandwidth or unreliable connections. The protocol is often used for the internet of things (IoT) and is an international standard (OASIS). Instead of using the common client-server pattern, MQTT uses the publisher-subscriber pattern, with topic-based message filtering. This means that MQTT messages published to a topic are only received by authorized subscribers listening to that topic, ensuring security. Incoming real-time data are automatically quality checked according to EuroGOOS and immediately processed by a real-time ingestion server called GeoEvent and published as stream services for integration in live dashboards.

The data files contain all measured variables of the sensor systems, including many maintenance variables. While transferring these files, it is important to ensure data integrity while at the same time deleting the files after a successful transfer to reduce the required storage space on the external system. Rsync, a network protocol based on TCP/IP and a program for the synchronization of data, is used to transfer the output files of the sensor systems in the external system. In contrast to the real-time data transfer, timing is not so critical, as these data are not displayed in real-time. Due to bandwidth considerations this allows for the complete dataset, rather than a subset used in real-time dashboards, to be transmitted whenever a better Internet connectivity is available, e.g., once a research vessel is in a harbor. The data files contain a lot more information necessary for delayed quality control than what is possible via MQTT. The data are automatically quality controlled and ingested into a spatial-temporal big data store in form of an Elasticsearch cluster. In a delayed mode quality control user interface build upon this database cluster, the scientists can apply various quality control routines as well as manually flag data as good or bad data. Once the scientists approve of the publication of the quality-controlled sensor data, they are published via feature services and the SensorThings API.

HELMI allows for two secure distinct data transfer pathways from external sensor systems to research centers and directly connects the data to quality control, storage, visualization, and publication tools. It can be used as a blueprint to modernize existing sensor data transfer workflows to newer, more secure web technologies. HELMI has been successfully established at various research vessels, ships of opportunity and stationary sensor platforms with difficult network connectivity. Some examples are the Hereon research vessel Ludwig Prandtl for transmission of the meteorological data, the ship of opportunity Magnolia and the container platform Tesperhude. In the future, data encryption will be implemented.

A Novel Approach in Oil Spill Detection, Identification, and Classification via Multisource Technologies and Artificial Intelligence

Tom Avikasis Cohen (tomchu7@gmail.com), Anna Brook (abrook@geo.haifa.ac.il), Dror Angel (dangel@univ.haifa.ac.il)

[1]University of Haifa (Israel)

Introduction

The Mediterranean Sea has a substantial volume of maritime traffic, including many tankers ferrying oil from eastern sources to western refineries. This critical maritime front, vital for trade and connectivity, also poses a significant risk of oil spills due to these busy shipping routes. Despite a decline in the number of large and medium oil spills since the 1970s, the environmental and economic consequences of oil pollution continue to be devastating. To safeguard the marine environment, early detection and response is imperative to minimize the impact of oil spills. This paper explores the use of remote sensing technology, specifically satellite-based systems, to detect oil spills as anomalies in the marine environment.

The conventional methods for early oil spill detection have encountered numerous challenges, primarily due to the complex and variable nature of spill events. This study focuses on the use of satellites for comprehensive coverage of large maritime areas. However, identifying oil spills at sea is a complex task due to factors like sun glint, mixed pixels, different water types, and limited spectral resolution. These factors make it difficult to develop a universal oil detection formula or algorithm applicable in all scenarios. The diversity of sensors employed on various remote-sensing systems further complicate the pursuit of uniformity in oil spill detection. Oil exhibits several distinct spectral characteristics across different parts of the electromagnetic spectrum, making it challenging to rely on a single sensor. Synthetic Aperture Radar (SAR) systems have traditionally been the most reliable method for detecting oil spills under variable weather conditions and oil types. However, their low frequency of coverage limits their utility.

This paper advocates for an anomaly-based approach, considering oil spills as anomalies in the environment rather than trying to identify the oil itself. By comparing the oil-contaminated area using a baseline of different water parameters, we can effectively detect and monitor the sea for oil spills. A pilot study conducted on the February 2021 Israeli oil spill successfully identified this spill as an environmental anomaly by utilizing a combination of indices indicating there was oil pollution in the water. The results from this study hold promise for developing tools that enable users worldwide to detect and monitor oil spills, contributing to early risk mitigation and preventing future ecological and environmental damage.

Methodology

This approach focuses on treating oil spills as anomalies in the marine environment. To achieve this, a multi-faceted approach is employed, integrating remote sensing technology, satellite-based systems, and a baseline for water parameters. The use of satellites provides a wide view of the region, enabling comprehensive monitoring of potential oil spills. In this study we worked with data from two different Sentinal products (Sentinel 1 and Sentinel 2), and are currently working on integrating data from Sentinel 3 as well.

The proposed methodology follows the concept of anomaly-based detection, where oil spills are identified as anomalies in the environment, rather than focusing on the direct identification of the oil itself. The study establishes a baseline for water parameters, encompassing various environmental factors. Deviations from this baseline are considered anomalies and potential indicators of oil pollution. After establishing a *clean* water baseline in our case study region, we established a number of indices relevant to the anomaly detection process. The detection solution is intended to be global in its application but involves method calibration as the baseline water properties change from region to region. In addition, we are using machine learning in the detection process for both validation and to test our results and identification, as well as for inhancement of our data from different sources.

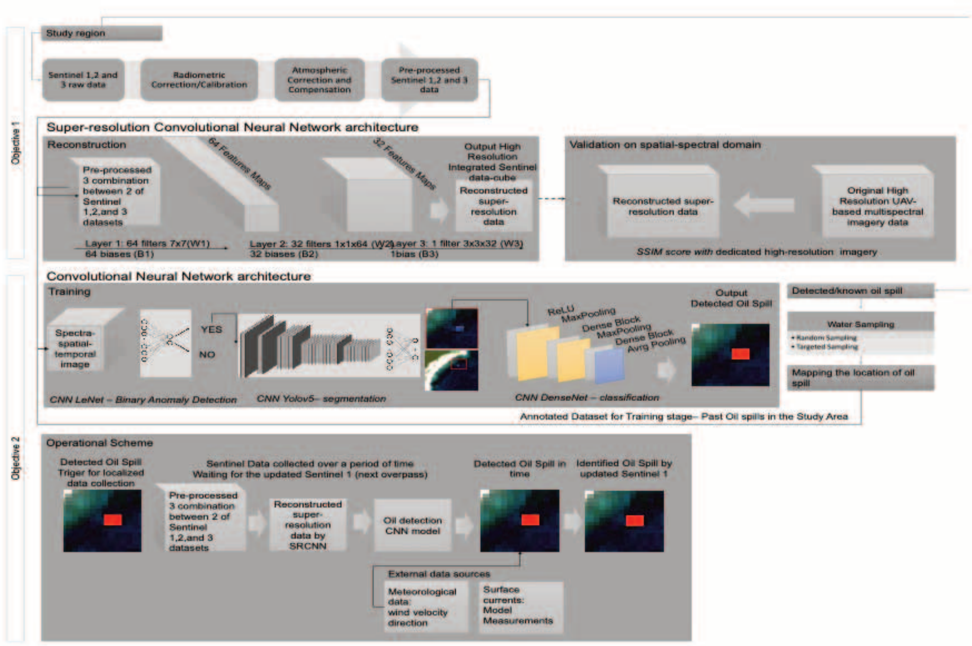


Figure 1 General Concept of Methodology Using Machine Learning in the Detection Process.

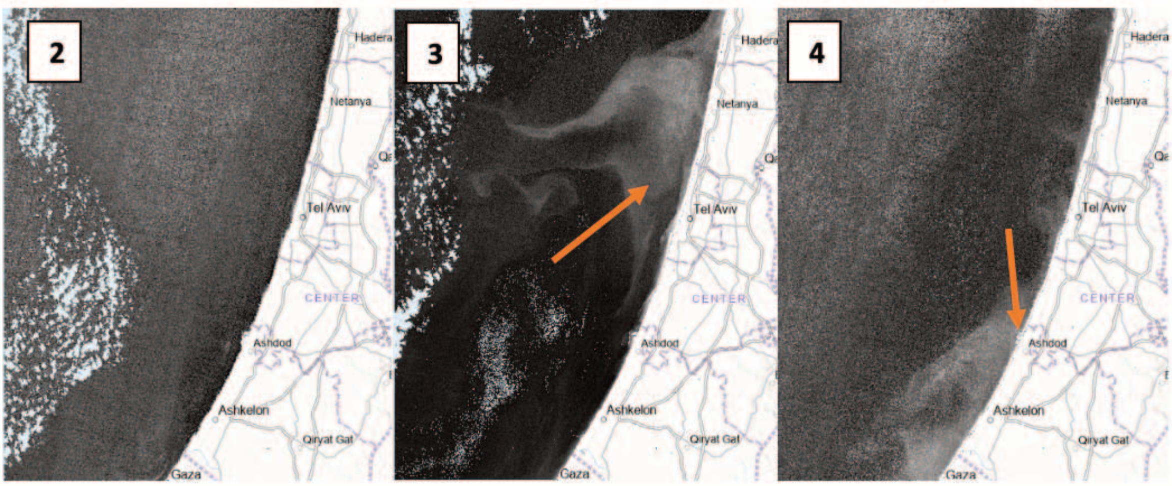


Figure 2 - 4 Identification Result on 3 Dates: 2) 14/02/21 - Prior to Storm; 3) 24/02/21 - Days After the Storm; 4) 06/03/21 - Weeks After the Storm.

Results

Our research efforts have culminated in a novel approach to detect oil spills. The visual representation of our findings in Figures 2 - 4 vividly demonstrates the regions of brightness, each indicative of the presence of an oil spill. Figure 2, captured prior to a storm in our case

study area, shows no discernible areas of brightness corresponding to oil spills. In Figure 3, we successfully pinpoint the oil spill in the central coastal areas. Figure 4 reveals that weeks after the storm, the oil had drifted south, aligning with the prevalent currents in the region.
The accuracy and reliability of our detection method were rigorously validated using a variety of sources, reinforcing the effectiveness of our innovative approach.

Unraveling the Technological Framework of MSFD Platform

Orestis Moresis[1] (o.moresis@hcmr.gr), Angelos Lykiardopoulos[1] (angelo@hcmr.gr), Konstantinos Kalkavouras[1] (ckalkav@hcmr.gr), Anastasia Iona[1,2] (sissy@hnodc.hcmr.gr), Paraskevi Drakopoulou[1] (vivi@hcmr.gr), Kalliopi Pagkou[1] (popi@hcmr.gr)

[1]Hellenic Centre for Marine Research (Greece)
[2]HNODC, Hellenic National Oceanographic Data Centre (Greece)

One of the main purposes of the HCMR, is to support policy and decision makers at national, regional and EU level regarding marine policies and regulations, especially considering societal and economic issues. The HCMR provides support to the Ministry of Environment and Energy as is the main public research organization implementing the monitoring programmes for both the Water Framework (WFD, 2000/60/EC) and the Marine Strategy Framework (MSFD, 2008/56/EC) Directives in Greece. The operational monitoring networks of the EU directives such as MSFD (2018 - 2023) are implemented in 6-year cycles. In MSFD are monitored several stations and platforms to cover the official Hellenic Republic marine waters ~480000 km^2. Monitoring frequency and techniques are strictly depending on the elements (biological, physicochemical, hydrological and pressures components i.e., contaminants, noise, plastics) monitored and following the protocols and frequencies established by Greece.

Aim

MSFD is a European collaboration that aims to unify the data collection and distribution regarding the ocean environment. In order to follow a holistic approach that fullfills that aim to the fullest and at the same time be customizable to any future modification or requirement, rethinking and remodelling of a new system had to be implemented. The paper describes the technology used to complete this task as well as the reasons behind the selection of its building components and the final design. Moreover, the paper points out problems and issues that the team had to overcome in the process of the making mainly trying to simplify the final result for the final user as much as for the maintainer of the project.

Software Architecture

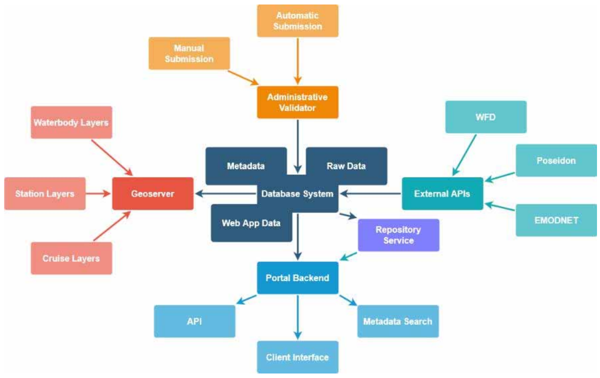
The architecture can be divided in several distinct components that interoperate between each other as well as with external services. The architecture is database centric, which means that it some cases acts as a middleware in the communication between different system nodes e.g., (repository service – main platform logic). In order to achieve asynchronous data delivery, a repository service has been implemented to be responsible for the data preparation, restriction checks and download provision.

Database

Database has been implemented into three distinct schemas to achieve separation of concerns. Metadata, raw data and web app. The approach in the metadata section has taken into consideration the universal vocabularies structure of descriptor, criteria and parameter as well as the connection with aliases for localization. An extra reason for following simplistic approach was to achieve maintenability and manual editing/input.

The database has geospatial capabilities as well as custom views and functions to achieve its purpose. It has been organized in such a way to accept multi-disciplinatry data in order to not depend to nature of the data submitted and achieve maximum flexibility.

Figure 1 Platform Logic.



Submission and sources

The data are gathered from a submission system (actively) as well as external APIs (passively). In both cases, the content is checked and converted to be compliant with international standards (e.g., INSPIRE, OGC). Furthermore, to meet the standards and improve the discovery results, the metadata are organized in metadata catalogues (geonetwork) and connected to the corresponding database entries.

The platform includes data from sources outside MSFD as well such as Greece Water Framework Directive, Hellenic National Ocecanographic Data Centre (HNODC), Poseidon etc. which also provide important information and aggregate it into one portal.

For the forementioned reasons, a special platform has been developed to fullfil the submission demands because the sources vary in structure and subject thus auto transformation and checking has been considered. The procedure of submission is transparent and both submitters and administrators have full control of it.

Web application

The client portal has been developed and customized with ease of use in mind but by approaching commercial geospatial software. The interface is fully capable providing download services, user roles, guest login, administrator panel and statistics. Data are prepared after request and delivered according to the state of restriction that accompanies them. There is also free data that is available for the public. The code is fully opensource as well as the libraries and tools that were used to implement it.

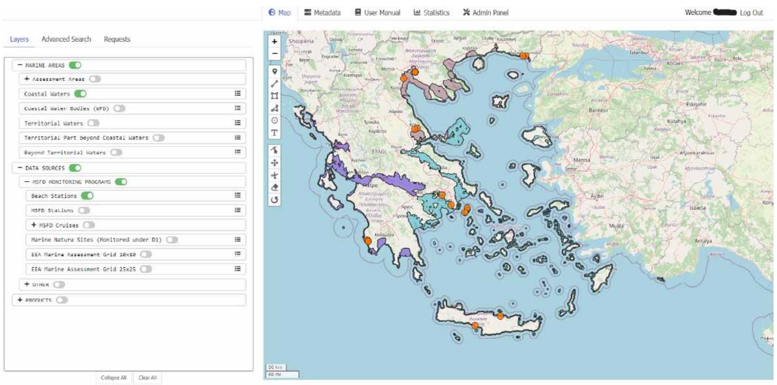


Figure 2 Portal Interface.

ResourceCODE: A Big Data toolbox to accelerate the development of wave and tidal energy

Nicolas Raillar[1] (Nicolas.Raillard@Ifremer.fr), Charlie Andre[1] (Charlie.Andre@Ifremer.fr), Christelle Riou[2] (Christelle.Riou@capgemini.com), Mickael Accensi[1] (Mickael.accensi@ifremer.fr)

[1]Ifremer (France)
[2]CAPGEMINI (France)

Introduction

ResourceCODE[41] is a project funded within the OCEAN ERA-NET Cofund call, that aims to provide the wave and tidal energy industry with the tools and data it needs to grow and thrive. The ResourceCODE toolbox is a Python toolbox that helps technology developers and supply chain companies to make informed technical and commercial decisions, improve engineering designs, and optimize operations. The project will help identify the sites most suitable for wave and tidal energies exploitation.

Hindcast database

The main outcome from the ResourceCODE project consists in the release of a state-of-the-art hindcast database of sea-state. This database that powered the toolbox, has been released early 2022, providing 27 years of model data, creating a high-resolution wave model in North West Europe. The database provides validated, transparent and reliable marine renewable energy waves and current resource data.
More specifically, the 5TB (compressed) FIELD data, refers to sea-state physical parameters (currents, winds and waves) obtained from a numerical wave model (WWIII). It is composed of 84 physical parameters. For each physical parameter, a value is defined at 320k nodes, and each hour of the 27 years covered period. More details about the hindcast database can be found in [Accensi et al., 2021].

Motivations

The hindcast database is stored in NetCDF files, with a general-purpose partitioning, like a file per month. The main challenge is that NetCDF is not optimized for cloud data access, or cloud computing workflows. Access performance for a specific use case will be indeed highly dependent on the data storage partitioning and chunking methodology, and what the access patterns are. The usual use case is to access the timeseries corresponding to a specific position quickly, from online API or services.

Methodology

Our approach is to look at the needed access patterns, then chunk the data against them, to ensure maximum performance at query time. In an offline way, on our HPC (Datarmor), we start to build small chunks of data with appropriate partitioning, which can be ingested in online tools. Regarding the offline part (see Figure 1), the first step converts NetCDF files into Parquet which is a popular file format for storing and processing large amounts of data. Then we use Spark to better embrace the industry Big Data ecosystem. Indeed, we serialize our data and some metadata with Protocol Buffers, that give performance and flexibility, and we can use actively maintained Spark connectors to ingest data to online tools.

[41] http://resourcecode.ifremer.fr

Regarding the online part, we use Spark on a permanent cluster, that allows us to ingest chunks in Cassandra, and pointer to those chunks in Elastic. Elastic and Cassandra are online tools; the first is relevant to provide high-availability and fast data access noticeably from scalable web services, while the second allows for quick reversed indices spatio-temporal searches. We use Airflow to orchestrate our different ETL (Extract, Transform, Load) jobs. For some ETL, we use Jobard. which is a Dask-based tool that allows for submitting job arrays on different backends ranging from HPC to K8s. It contains tools to structure, distribute and monitor the processing of a large number of files simultaneously. Especially, we used a special custom operator called JobardOperator, to make the link between Airflow and the final backend for processing.

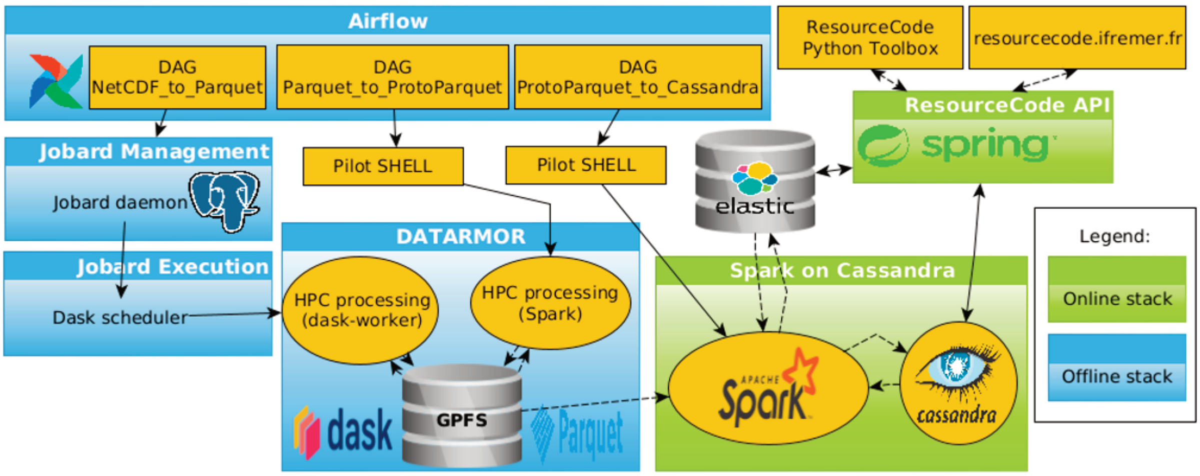


Figure 1 technical architecture.

Results

Regarding the processing part, we are able to process the dataset in a distributed way, from the archival NetCDF files, to the Cassandra ingestion, within 48 hours. Note that this result highly depends on the number of nodes implied in each calculation. What is important is that our ETL is scalable, so the more nodes you have, the less time it takes, up to a certain extent.

For the final user, the data can be accessed from the Python package [Raillard et al., 2023]: the user selects a particular position, and the package allows the corresponding time series to be returned in a few seconds, in a Pandas dataframe. The user can also use the online web platform, which provides node selection and plots within a few seconds, without the need of Python being installed. Most of the time is due to Notebook resource allocation setup cost.

About the usage, we have studied the nginx logs. We have seen 55425 unique users in 2022. The point with the most-consulted time series is the one near the Fécamp wind farm in France. In terms of physical parameters, significant wave height (Hs) is the most popular parameter, followed by peak period (Tp), peak direction, and wind data.

Conclusion

ResourceCODE is a significant step forward in the development of wave and tidal energy. By providing a unified data infrastructure and a suite of tools and services, ResourceCODE will help to accelerate the development of this emerging industry.

In this article, we discussed a technical stack that combines modern Big Data tools coming both from Scientific community, and industry community, to create production-grade high-performance data pipelines that powered web services and API.

For the future, we plan to process other parts of the reference database like spectrum data, with the same methodology, from studying access patterns, to then implement suitable chunking. The stack may also evolve to include streaming and ML libraries that take advantage of data distribution.

References

Accensi M. et al., (2021). *ResourceCODE framework: A high-resolution wave parameter dataset for the European Shelf and analysis toolbox*. Proceedings of the 14th EWTEC 5-9th Sept 2021, Plymouth, UK. https://archimer.ifremer.fr/doc/00736/84812

Raillard N. et al., (2023). *RESOURCECODE: A Python package for statistical analysis of sea-state hindcast data*. Journal of Open Source Software. https://doi.org/10.21105/joss.04366

NAUTILOS Data Management Infrastructure

Gabriele Pieri[1] (Gabriele.Pieri@isti.cnr.it), Marco Tampucci[1] (Marco.Tampucci@isti.cnr.it), Federico Volpini[1] (federico.volpini@isti.cnr.it), Marco Alba[2] (marco.alba@ettsolutions.com), Francesco Misurale[2] (francesco.misurale@ettsolutions.com), Antonio Novellino[2] (antonio.novellino@ettsolutions.com)

[1]Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche - CNR-ISTI (Italy)
[2]ETT Solutions S.p.A. (Italy)

Introduction and problem description

NAUTILOS Project (New Approach to Underwater Technologies for Innovative, Low-cost Ocean observation) is an H2020 project devoted to fill-in marine observation and modelling gaps for chemical, biological and deep ocean physics variables through the development of a new generation of cost-effective sensors and samplers. It brings together many partners with multidisciplinary expertise ranging from ocean instrumentation development and integration, ocean sensing and sampling instrumentation, data processing, modelling and control, operational oceanography and biology and ecosystems and biogeochemistry such, water and climate change science, technological marine applications, and research infrastructures. In this context, the data generated within the NAUTILOS project are meant to be highly complementary to the existing observing systems and thus with high impact and value for any NAUTILOS stakeholder.
One specific goal of the project is to facilitate interoperability towards stakeholders (e.g. National Oceanographic Data Centres and thematic and international data assembly repositories, EMODnet, SeaDataNet, Copernicus Marine Environmental Monitoring Services), hence the project has organized data management according to a workflow that makes its data as ready as possible. This includes the adoption and application of standardized and harmonized standards and formats.

Data management Back-end

NAUTILOS infrastructure is organized with a data layer designed to manage data and data products, a service layer to organize them to offer the services, and an application layer i.e., the end-user interface (and features) to access and use the developed and provided services. The data layer is in charge of ingesting datasets and is able to make available data in different formats, like csv, txt, netcdf, etc.

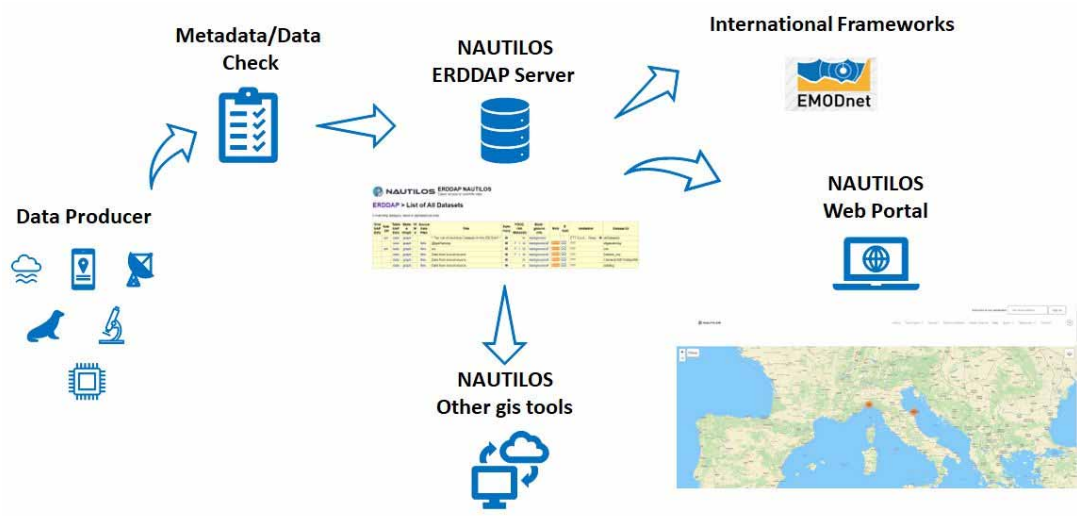


Figure 1 NAUTILOS Data flow and infrastructure.

Figure 1 shows NAUTILOS data publication workflow that processes data towards a multi-step tier approach where a first round is visible to internal users and permit to validate the new flow with the proper application of harmonized standards and quality checks/quality flags, etc. The next round makes this workflow operational, hence there is a direct streamless flow from the deployed platform towards the NAUTILOS backend. The level of discovery is still limited to NAUTILOS users for a second level of checks. The last step is the full publication of the data (as well as the organization of the data into collection that can be easily consumed by stakeholders). A single dataset source may be integrated in more products and the service layer organizes data within the NAUTILOS datasets and data products publication services and catalogues (e.g., ERDDAP). While implementing these three tiers, NAUTILOS applies special measures to improve and ensure data interoperability: data format, metadata and data services. As an example, to improve data provenience visibility and NAUTILOS data citation, it adopts digital object identifier (doi). Notably a DOI is also important when datasets are used from large databases that are subjected to changes, e.g., annual reprocessing, versioning, etc.

Data management Front-end

The development of the NAUTILOS Front-end is based on a web portal[42] providing an end-to-end data and information management, search, discovery and access system. The portal implements a WebGIS viewer with a dynamic map that provides the user with zooming and panning features, as well as with features to include and select data and data-products. Each dataset is enriched with metadata, and tools are integrated to pre-view these metadata contents. The front-end also provides several functionalities to filter the data and change their rendering behaviour. For each dataset, the front-end provides a set of dedicated options that can be used to filter data to be shown. Moreover, selecting the “time series” option changes the data representation providing a view of the data that evolves over depending on registered time. In order to prevent a performance, drop the front-end is able, for each dataset, to automatically group near points of interest and explode them when zoom-in. For specific datasets, users can change the rendering colour depending on stored data, i.e. for the “mini-drifter” datasets, users can change the polyline segments depending on drifter speed or measured temperature (Figure 2). Finally, displayed data can be downloaded as reports containing all the activated layer data and relative diagrams.

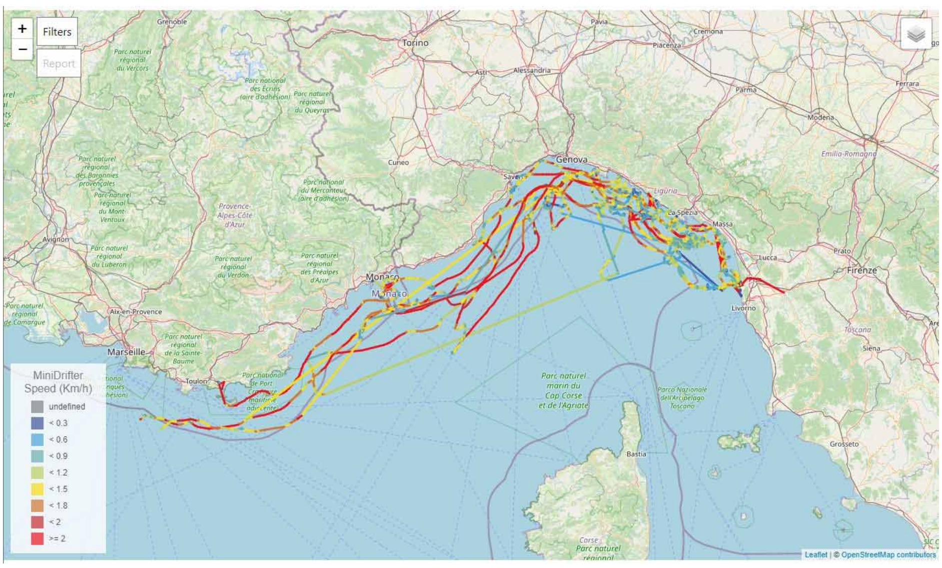


Figure 2 NAUTILOS front-end example.

[42] https://nautilos-h2020.eu/data-portal/

Conclusions

NAUTILOS is developing new tools to unlock measurement and sharing of chemical, biological and deep ocean physics variables with a focus on places and parameters that are still missing in the standard "networks" and programs. The NAUTILOS data management system is adopting the latest development from the international community and is developing updates and new workflows to advance these standards further. This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000825 (NAUTILOS).

Observations towards a modular digital twin architecture to support multi-community research and decision making

Thomas Mansfield (tma@pml.ac.uk), Claire Szostek (csz@pml.ac.uk), Benjamin O'Driscoll (bod@pml.ac.uk), Stephen Watson (stw@pml.ac.uk)

Plymouth Marine Laboratory (UK)

Introduction; modular digital ecosystems

Digital twins provide a virtual representation of physical objects and systems. Pioneered by engineering communities, digital twins have been proven to combine data to enable what-if simulations, bring together multi-disciplinary communities and improve the understanding of physical systems [West et.al., 2018]. Plymouth Marine Laboratory (PML) is incrementally developing the foundations of future, multi-community digital twins that encourage data sharing and collaboration among research communities [Tzacho et al., 2023; Bennett et al., 2023]. To identify areas of interest and trial new ways of working, this extended abstract presents the implementation and test of modular and interoperable data ecosystems as a first step towards the design of future multi-community digital twins.

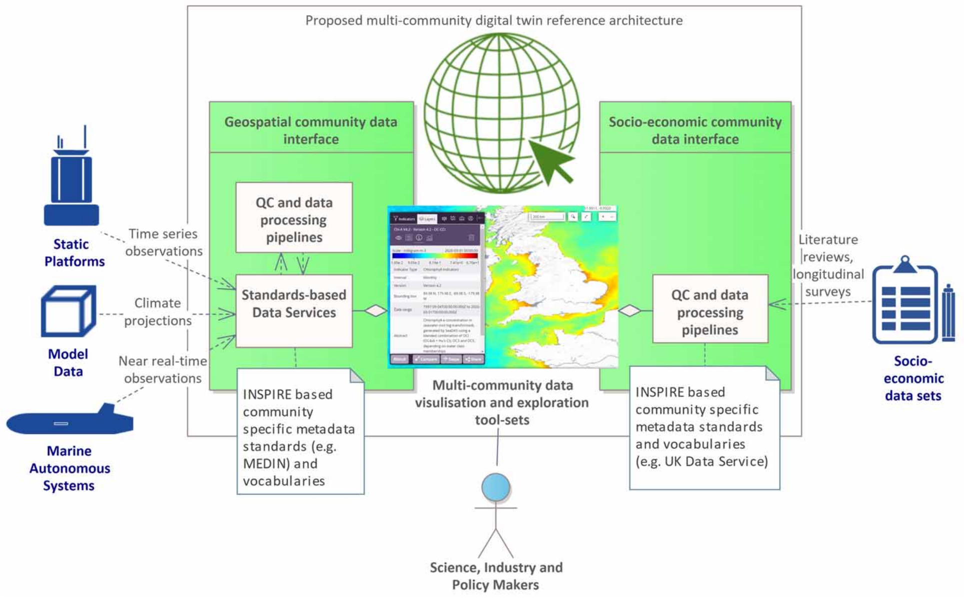
Combined geospatial and non-geospatial data; A proposed reference architecture

Sharing data among research communities is a foundation for cross discipline working. One key barrier to sharing data has been identified when communities need to combine geospatial and non-geospatial data [Walsh C., 2023]. Geospatial data communities include those that generate physical or biological timeseries data from static platforms at fixed sites, collect ad-hoc observations from ships or unmanned systems, remote sensing datasets and model outputs that generate forcasts and long-term projections for envionmental conditions. Data providers in marine geo-spatial communities often generate large quantities of raw data and have developed domain-specific foundations for data management cultures, tools and processes for handling their data. Work is currently underway to align these domain-specific appraoches with more general data service frameworks [DAMA International, 2017; Mansfield et al., 2024], with a significant focus on the implementation of the MEDIN discovery metadata standard as an extension of the INSPIRE metadata standards in addition to the use of common vocabulary servces [MEDIN Standards Working Group, 2022; European Commission 2018; BODC, 2023]. Non-geospatial data services are being developed by communities such as the UK Data Service [UK Data Service, 2023], with metadata elements that place a higher focus of on the reusability of a wide range of data types, with an additional emphasis placed on licensing, ethics and embargo periods.

To align each of these existing data communities, a reference architecture is proposed that enables existing data foundations within each community to be federated though the development of two separate digital ecosystems. To enable the combination of these digital ecosystems, two key areas have been considered;

1. The use of INSPIRE metadata standards in both ecosystems to develop broadly compatible geospatial and non-geospatial data sets.
2. The development of commonly structured data visuliation and, where required, human-centric data Quality Control (QC) toolsets. An overview of the proposed arechitecture of encouraging collaboration among geospatial and non-geospatial communities is provided in Figure 1.

Figure 1 Proposed multi-community digital twin reference architecture.



Development focus area; integrated portals for data visualisation and exploration

This paper illustrates key elements within the proposed reference architecture by presenting developments to support the joint discovery of geospatial and non-geospatial data through innovative data visualisation techniques. As a patiular area of innovation, a primary tier of data discovery toolsets are closely coupled to tools that encourage the exploration, analysis and user-driven QC of non-geospatial socio-economic data assessing topics such as the impact of windfarm developments within the context of other geospatial marine datasets. Each of these toolsets have been tested and trialed in multi-stakeholder workshops with elicited feedback being used to improve services, mature data architectures and align with standardisation.

Conclusions and next steps

This extended abstrract reports on recent progress to design, develop and test of a reusable and modular reference architecture that aims to realise the long-term benefits of using digital twins to support multi-community research. This federated and multi-domain approach suports the inclusion of data and methodologies from a range of marine research communities into an imporved and data-driven scientific toolset.

References

West T.D., et al., (2018). https://doi.org/10.1016/j.heliyon.2024.e26503
Tzachor A., et al., (2023). https://doi.org/10.1038/s44183-023-00023-9
Bennett H., et al., (2023). 10.5281/zenodo.7840266
Walsh C., (2023). *Non-spatial Discovery Metadata*, MEDIN, Southampton, UK.
DAMA International, (2017). *Data Management Body of Knowledge (DAMA BoK)*.
Mansfield T., et al., (2024). *Digital twins of the ocean: Comparing approaches from the defence nd envionmental research communities*” in, Undersea Defence Technologies, London, UK.
MEDIN Standards Working Group, (2022). https://dx.doi.org/10.25607/OBP-1625
European Commission (EC), (2018). *Technical Guidance for the implementation of INSPIRE Discovery Services*, EC, Brussels.
BODC, (2023). *NERC Vocabulary Server (NVS)*, BODC.
UK Data Service, (2023). *Offer a new dataset*. University of Essex and University of Manchester.

NAUTILOS Citizen Science App

Marco Tampucci[1] (Marco.Tampucci@isti.cnr.it), Eva Chatzinikolaou[2] (evachatz@hcmr.gr), Kleoniki Keklikoglou[2] (keklikoglou@hcmr.gr), Gabriele Pieri[1] (Gabriele.Pieri@isti.cnr.it)

[1]Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche - CNR-ISTI (Italy)
[2]Hellenic Centre for Marine Research (HCMR), Institute of Marine Biology, Biotechnology and Aquaculture (IMBBC), Heraklion, Crete, (Greece)

Citizen Science projects including crowd-sourcing of plastic litter data have been implemented successfully in many countries and within several long-term monitoring projects. The involvement of local communities in the collection of data allows them the opportunity to actively participate in the problem's solution and has the potential to raise environmental awareness. The engagement of citizens in everyday science may ultimately inspire public and even political change [Duckett et al., 2015]. Citizen science can be a useful approach to increase the available information on environmental problems such as the investigation of marine litter sources, distribution and ecological impacts [Hidalgo-Ruz et al., 2015].
Within the NAUTILOS Project, many different partners are involved in Citizen Science (CS) campaigns. Each one does not necessarily perform the same type of activities or focus on the same kind of citizens or citizen organisations. The primary goal of this activity is to develop an App which can support various CS activities arranged by different partners in the NAUTILOS project and with different focuses. Still, the goal is not to provide general tools that anyone could use to increase and collect information. The heterogeneity of CS activities in NAUTILOS entails the need to deliver a broader instrument that allows to acquire data more homogeneously and export them to the web platform, which also provides for the sharing of data collections in a standardised way. Aiming to guarantee more straightforward access to all scientific data collected inside the project, the data resulting from the CS activities will be compliant with the ERDDAP data format.Four main scenarios, following from the project's activities, have been identified which are covered by the developed App: CS plastics-related campaigns; divers' campaigns; crowdsourcing for visual marine image annotations; and, algal bloom-related campaign. In this paper we will present the design and structure of the CS App, along with a specific use case of the application concerning the plastic collection campaign scenario.

NAUTILOS Application and Use Case

In order to allow crowdsourcing campaigns as requested from some scenarios such as the plastic/litter collection campaigns, the application has been designed and developed to be accessible from the majority of devices, and people. Thus, the React Native framework has been adopted to develop the application that allows to develop an application that can run both on Android and iOS devices without any code adjustment or rewriting. The application is able to send the reports, resulting from the campaigns, to the NAUTILOS ERDDAP data server through a set of dedicated HTTP services. Indeed, the application provides two main typologies of usage: (i) storing, (ii) visualisation.
Storing usage is focused on the generation of the reports for the identified scenarios that will be stored in the ERDDAP data server; due to the heterogeneity of the scenarios and the differences that occur between them, the application manages them separately. For each scenario, the application provides the user with a set of dedicated textual fields, dropdown menus and three special fields (i) coordinates, (ii) date and (iii) picture. The coordinates are, by default, taken from the GPS sensor of the device but users can select a different pair by selecting from a map or directly editing the values; the date field, by default, takes the actual date and

offers a calendar view to change the date; finally, a picture can be chosen from the device gallery or, otherwise, directly taken from the device camera (Figure 1, right).
On the other hand, the visualisation modality retrieves the coordinates from the GPS sensors, is used to center the map on the actual location, and, directly querying ERDDAP data server, retrieves all the reports sent by the user and displaces them on the map. Users can have a brief view of the sent reports and, from here, edit or delete them.

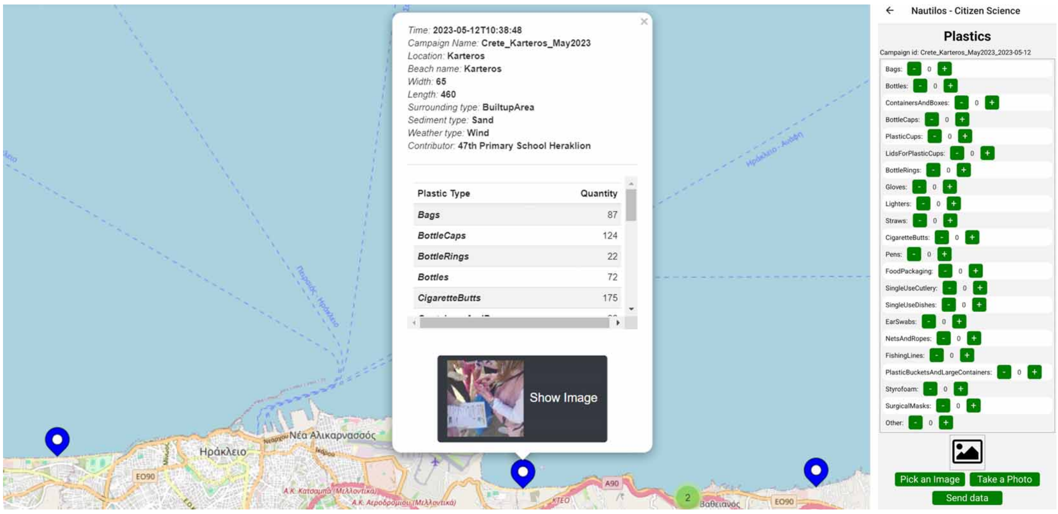


Figure 1 Map of Citizen Science plastic litter campaigns and format of data display (left). Appearance of the CS App (right).

Within the framework of the NAUTILOS project 12 Citizen Science field campaigns have been performed which involved the collection of plastic litter from 6 beaches in Heraklion (Greece) during the period May 2022 - May 2023, and they are still ongoing. A total of 231 primary and secondary school students of all grades (ages 7 - 18) were involved in the campaigns and recorded the plastic litter they collected on the beach on forms with predefined categories while working on the field. The students categorised the plastic litter into 22 types according to the J code list of the official EMODnet Chemistry Thematic Lot n°4 vocabulary. These categories included mainly single-use plastics such as bags, bottles, containers, and cutlery, remains from fishing equipment and small plastic items (lighters, pens, straws, cotton bud sticks), while also large quantities of unidentified plastic fragments were found. Other data collected during the campaigns include contributor, date, location name, coordinates, width and length of the sampled area, surrounding type, sediment type and weather conditions. While on the field or upon return to their classroom the students can add the data and relevant photos to the NAUTILOS Data Portal using the Citizen Science App of Nautilos. The data are mapped and their graphical representation is possible, while a data report can be also produced (Figure 1, left).

Conclusion

The NAUTILOS Citizen Science application is a new tool to measure and share information about chemical, biological and deep ocean physics variables with a focus on places and parameters that are still missing in the standard “networks” and programs. To enhance data distribution and information accessibility, plastic reports are also ingested to the European Marine Observation and Data Network (EMODnet) under the Chemistry section; indeed, the mandatoriness of the fields and the values of the dropdown menus follow the EMODnet guidelines.

This work is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 101000825 (NAUTILOS).

References

Duckett P.E., & Repaci V., (2015). *Marine plastic pollution: using community science to address a global problem*. Marine and Freshwater Research, 66(8), 665-673.

Hidalgo-Ruz V., & Thiel M., (2015). *The contribution of citizen scientists to the monitoring of marine litter*. Marine anthropogenic litter, 16, 429-447.

Marine Biogeochemical data quality control in EOSC Fair-Ease cloud

Catherine Schmechtig[1] (catherine.schmechtig@imev-mer.fr), Virginie Racapé[7] (virginie.racape@pokapok.org), Erwan Bodéré[2] (Erwan.Bodere@ifremer.fr), Thierry Carval[2] (Thierry.Carval@ifremer.fr), Fabio Conversano[3] (fabio.conversano@szn.it), Jérôme Detoc[2] (Jerome.Detoc@ifremer.fr), Delphine Dobler[4] (delphine.dobler@euro-argo.eu), Alessandra Giorgetti[5] (agiorgetti@ogs.it), Claire Gourcuff[4] (claire.gourcuff@euro-argo.eu), Marie Jossé[6] (Marie.Josse@data-terra.org), Marco Miralto[3] (marco.miralto@szn.it), Catalina Reyes[5] (nreyessuarez@ogs.it), Raphaëlle Sauzède[6] (raphaelle.sauzede@imev-mer.fr), Reiner Schlitzer[7] (reiner.schlitzer@awi.de), Alban Sizun[8] (Alban.Sizun@pokapok.org), Clément Weber[8] (Clement.Weber@pokapok.org)

[1]Sorbonne Université (France)
[2]Ifremer (France)
[3]SZN-Stazione Zoologica Anton Dohrn (Italy)
[4]Euro-Argo ERIC (Europe)
[5]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[6]CNRS Data-Terra (France)
[7]Alfred Wegener Institute (Germany)
[8]POKaPOK (France)

The observation of marine biogeochemical (BGC) properties is fundamental to address scientific processes regarding the health of marine ecosystems (e.g., ocean acidification, oxygen minimum zone, biological carbon pump, phytoplankton communities, etc.) and needs for ocean resource management. BGC sensors have been deployed through various autonomous platforms (floats, gliders, sea mammals, moorings, etc.) by observing networks under GOOS-OCG international coordination (Global Ocean Observing System - Observations Coordination Group), leading to a dramatic increase of BGC observations at global scale for last decades. There are currently more than 900,000 BGC profiles measured either by a BGC-Argo float, a glider or a sea-mammal throughout the global ocean, including European marginal seas (source: Copernicus Marine Service - https://doi.org/10.48670/moi-00036, Dec. 2nd, 2022): 18% are already in delayed mode status (usable by scientists in total confidence), 10% are automatically adjusted in real time whereas the remainder needs to be adjusted before it can be used. Thus, BGC data adjustment and validation represent a major challenge to significantly increase the volume of high quality BGC data available for the scientific community.

At present, the BGC-Argo science team is a major contributor on an international level to calibrate, validate and trigger alerts on in situ BGC data at global level. In recent years, BGC-Argo sensors have diversified (oxygen, nitrate, chlorophyll, suspended particles, pH) and methods of data quality assessment and control, validation and adjustment have become more complex. Most of the methods are available as open source tools, available on various public github repositories. These methods require an efficient access to external data: gridded climatologies, model outputs (meteorologic, oceanographic), discrete in-situ data, satellite data, etc. Moreover, applying these methods requires combining the data in space and in time using extraction and colocation. Thus, a massive, high-performance, distributed data infrastructure that would combine in situ, satellite and models data would definitely help the data scientist community.

Today, softwares development and metadata standards are specific to the Argo format. However, methods are essentially sensor-dependent and not platform-dependent, meaning that it is applicable to BGC sensors deployed on gliders, sea-mammals or other platforms.

As part of FAIR-EASE, a Horizon Europe project whose objective is to customize and operate distributed and integrated services for observation and modeling of the Earth system, a demonstrator is being developed for contributing to the overall improvement of BGC data quality, through software standardization and easy cloud development. The aim of the demonstrator presented here is to provide a single and efficient access to three services: qualification/calibration/validation of BGC data through a web portal. Tools deployed for the calibration focus on ocean BGC observations essentially measured by sensors deployed on BGC-Argo floats, gliders or sea mammals.

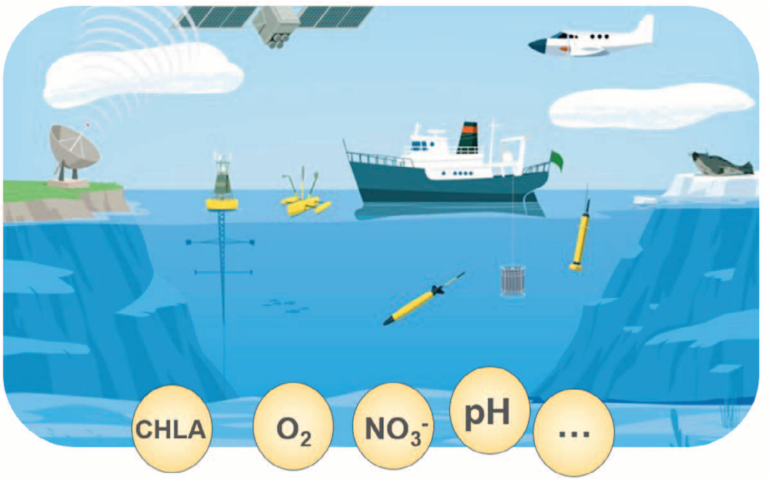


Figure 1 BGC in situ ocean observations.

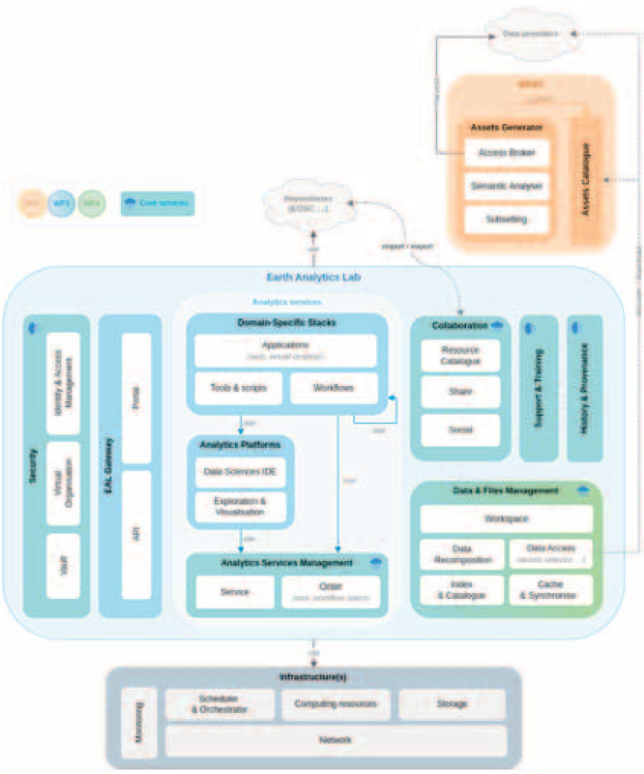


Figure 2 EOSC FAIR-EASE infra for BGC data scientists.

We shall present the scientific challenge to quality control, calibrate and validate in situ BGC Ocean data, the solutions and the technical infrastructure for European data scientists. The infrastructure is part of the European Open Science cloud, developed within the European project FAIR-EASE.

Poster session

Discovery, visualization and dissemination of marine spatial data products from the Spanish Institute of Oceanography

M. Olvido Tello-Antón (olvido.tello@ieo.csic.es), Gerardo Bruque Carmona (gerardo.bruque@ieo.csic.es), Luis Miguel Agudo Bravo (luismi.agudo@ieo.csic.es), Dulce Mata Chacón (dulce.mata@ieo.csic.es), Paula Gil Cuenca (paula.gil@ieo.csic.es)

Spanish Institute of Oceanography (Spain)

Spanish Institute of Oceanography (IEO-CSIC) www.ieo.es

The Spanish Institute of Oceanography (IEO) is a public research organization, established by the Royal Decree of April 17, 1914 and recently integrated into the Spanish National Research Council (CSIC). The IEO-CSIC is committed to marine science research, particularly concerning the scientific knowledge of the oceans, sustainability of fishing resources, and marine environment preservation. The IEO-CSIC depends on the Ministry of Science and Innovation.

The IEO-CSIC has a central office in Madrid, nine coastal oceanographic centres, five marine culture experimentation plants, twelve gauge stations, and a satellite image reception station. Its oceanographic fleet consists of five vessels.

Researchers of IEO-CSIC carry out multidisciplinary studies of the marine environment, utilizing various methodologies and data acquisition techniques are used depending on the study. The IEO acquires and generates a large amount of marine data, which are organized into different databases based on subject matter and the represented variables (e.g., geology, fisheries, aquaculture, pollution, habitats, species, microalgae, marine litter, and oceanographic variables such as temperature, salinity, oxygen, etc.).

The raw data adquired are integrated into the databases of the Oceanographic Data Center (CEDO) of IEO. From CEDO, the IEO shares data with the international data network SeaDataNet previously standardized according to international standards.

Moreover, data products resulting from the processing, treatment and analysis of raw data, whether stored in CEDO or provided directly by resarchers, are incorporated into the Institutional Geographical Information System of IEO, called "IEO MARINE GIS".

With the aim of enabling the discovery, visualization and dissemination of marine data products from IEO-CSIC, the Spatial Data Infrastructure of IEO (IDEO) was developed in 2010. IDEO is based on open standards mainly ISO (OGC) and INSPIRE standards.

IDEO provides access to the IEO marine data integrated in the MARINE GIS through interoperable cartographic web services, and interactive viewers that allow the visualization, querying, and analysis. Metadata for both, data and services, are accessible through the IEO Data Catalogue. The Spatial Data Infrastructure of the IEO provides a technological and organizational framework for the managing, using and disseminating geospatial information in the field of IEO.

Key components of IDEO include:

- Marine data, integrated into IEO MARINE GIS
- Web Map Services (WMS) and Web Features Services (WFS)
- A data catalogue developed with Geonetwork software
- A multidisciplinary marine data viewer http://www.ideo-base.ieo.es/Home
- Six additional specialized data viewers from special topics

Brief explanation of each of them

IEO MARINE GIS

The IEO MARINE GIS offers the operational capabilities of a GIS, with integrated data harmonized and standardized according to international norms (INSPIRE, OGC, etc.), which are subject to exhaustive quality controls. GIS tool is essential for decision-making related to marine environment protection, resource sustainable management, natural hazard mitigation, marine spatial planning, and more. The data include in the database of IEO MARINE GIS are de basis for the development of WMS and WFS services of IDEO.

Web Map Services (WMS) and Web Feature Services (WFS)
More than 400 web cartographic services (WMS & WFS) has been developed with the objetive of providing access to data for viewing, consulting and downloading. These services supply data to the IEO data viewers.

IEO Data Catalogue
It is a Geonetwork catalogue created according to the needs of IEO. It contains metadata for all datasets included in GIS-IEO, as well as containing metadata on oceanographic surveys and samplers such as CTD, etc. This tool allows finding and discovering the IEO spatial data. www.datos.ieo.es.

Multidisciplinary Marine Data Viewer
It is a web application enabling the visualisation, querying, downloading and dissemination of marine spatial data products, proceeding of relevant projects carried out by IEO researchers and technicians. Its offers an intutitive and simple interface for accessing marine information on habitats, species, geology, fisheries, pollution, marine litter, noise, bathymetry, seabed substrate, and other relevant topics. http://www.ideo-base.ieo.es/Home
This viewer utilizes Web Map Services and Web Feature Services developed based on GIS-IEO data. The technical specifications are as follows:

- The management of the viewer generation template for the IEO is a project based on REACTJS development technology. The development environment follows the principles established by this technology, including the use of the Webpack development web server.
- For the client map library, version 4 of the ESRI javascript API (version 4.7) has been chosen. When working with this API, some of its functionalities.
- The permissions management is controlled by database (MSSQL Server) through a Microsoft MVC .net project.

Six specialized data viewers from special topics
These data viewers cover topics such as the El Hierro volcano; Cabrera Natural Park (marine-terrestrial); Walvis Ridge; TPEA project; MarSp project; CCLME Eco-GIS Viewer. http://wiki.ieo.es/books/ideo-%E2%80%93-sig-marino/page/visores-ideo.

Applications and future of IDEO
The spatial data products from IEO relevant projects such as the implementation in Spain of the Marine Strategy Framework Directive 2008/56/EC (MSFD); LIFE-IP-INTEMARES project; Seagrass Atlas of Spain; Radiales program, etc., are accessible for viewing, consultation and dissemination from the Main Data Viewer of IDEO.
Looking ahead, our challenges include expanding available data through map viewers, facilitating data downloads of as much data as possible, in order to continue applying the FAIR Data Principles, improving selection options, 3D visualization, mapping capabilities, and developing online geoprocessing.

The DeepSea Monitoring Data Portal

Marcus Krüger (mkrueger@geomar.de), Valentin Buck (vbuck@geomar.de)

GEOMAR Helmholtz Centre for Ocean Research Kiel (Germany)

Background

The DeepSea Monitoring (DSM) working group at GEOMAR Helmholtz Centre for Ocean Research Kiel consists of 25 people. Natural scientists from variuos scientific fields), computer scientists and technicians (in geochemical laboratory analysis, electronics and precision engineering) collaborate on technology development, data generation and analysis to answer our scientific questions. In addition to application-oriented development research we address the following scientific issues:

- Detection, quantification and monitoring of methane seeps from the sea floor into the water column and further into the atmosphere.
- Application of high-resolution hydroacoustic and optical seafloor mapping for quantitative assessment of sea floor resources or habitat maps.
- Identification and annotation of benthic organisms from optical data.
- Environmental monitoring before, during and after the delaboration of munitions in the sea.

Using hydroacoustic (singlebeam and multibeam systems, sub-bottom profilers), electromagnetic (magnetometers) and optical (cameras, fluorometers, lasers) methods the group produces vast amounts of raw data and data products. Providing sustainable access to our research data is the key to prevent loss of knowledge not only at working group level, but also at institute level. A data catalogue is needed to enable researchers to discover and access these datasets.With the ocean science information system (OSIS), GEOMAR provides an institutional data access platform which is a central information hub with links to publications, reports, maps, data- and code repositories and stores metadata on GEOMAR´s ship-based research expeditions. However, the basic function of uploading and saving geodata sets, i.e., the possibility of having our research data catalogued, is currently lacking.

GeoNode - A (geospatial) Content Management System

The Deep Sea Monitoring Data Portal (Figure 1) is an implementation of the Web-based open-source (geospatial) content management system GeoNode (https://geonode.org). GeoNode is developed as an Open Source Geospatial Foundation (OSGeo) core project andorchestrates mature and stable software (Figure 2) to set up a geoinfrastructure as a data portal. Users are able to ingest, share, discover and visualize datasets and map collections and generate standardized metadata.

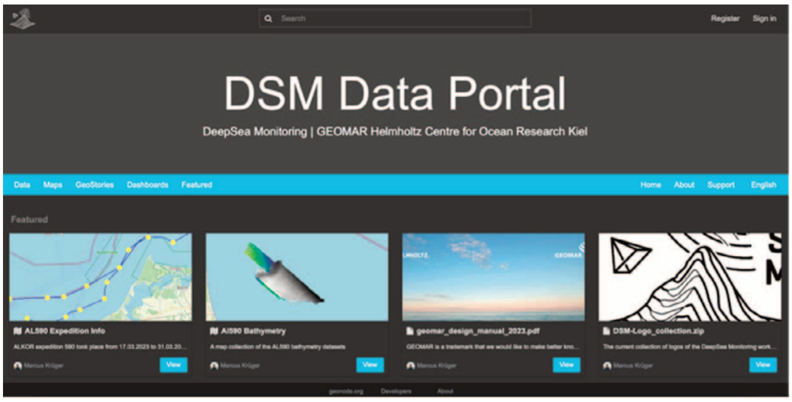


Figure 1 The landing page of the DeepSea Monitoring (DSM) Data Portal.

Figure 2 GeoNode Software Stack.

Since mid of 2023 the system runs on a virtual machine inside a docker environment at GEOMAR servers and is mainly used as a data catalogue and a platform to share datasets with users from GEOMAR or other institutes. It therefore amends GEOMAR´s institutional data management on the mission towards FAIR research data on working group level [Wilkinson et al., 2019].

Data Ingestion and Storage

The catalogue distinguishes between five kinds of resources: (i) datasets which are further subdivided into vector, raster, time series data and remote services, (ii) documents, (iii) maps, (iv) geostories and (v) dashboards. The latter three are configured inside the portal only whereas datasets and documents can be uploaded / ingested into the portal. Datasets can be uploaded using a large variety of common geospatial formats (ESRI Shapefile, GeoTIFF, CSV, GeoPackage, GeoJson, …) which then are stored in a PostgreSQL database or in the file system according to the data type (vector data or raster data). At the same time the data records are registered at the GeoServer backend and are provided as Open Geospatial Consortium (OGC) services. Documents can also be uploaded by users in various formats and are stored in a file system. Additionally, datasets and documents can be linked in the system which offers the opportunity to e.g., provide a dataset together with the corresponding standard operating procedure (SOP) giving information on how the dataset was created or processed.
The DSM Data Portal is mounted into the centralized storage system hosted by the Information, Data and Computing Centre at GEOMAR which also allows for data integration from storage side via the GeoServer backend. Therefore all resources of the portal are not only stored but also backed up on institutional servers.

Providing Research Data as a Service

When uploading a geospatial dataset into the system it is included in the integrated GeoServer and made available via standard OGC protocols, such as Web-Map-Services (WMS), Web-Feature-Services (WFS) and Web-Coverage-Services (WCS). Metadata records for all those resources can be retrieved via a Catalogue Service for the Web (CSW). Classified data access authorisations must be granted for some of the data obtained by the DSM working group, which GeoNode facilitates. Based on the rights and role concept given by the users in the frontend the access restrictions are also assigned to the GeoServer backend and the corresponding OGC services.

Conclusion and Outlook

The institutional data management at GEOMAR provides infrastructure, multiple tools and services to make data findable, accessible and reusable. The DSM data portal closes the gap in the cataloguing of research data and thus complements GEOMAR’s institutional research data software towards FAIR research data. In any institutional data management interoperability remains the main challenge. Due to its standardized protocols and expandable open-source software stack GeoNode has the potential to overcome these challenges. Currently in the fourth generation GeoNode is a mature software to build flexible geoinfrastructures that has been on the market for more than a decade now.

References

Wilkinson M.D., Dumontier M., Jan Aalbersberg I. et al., (2019). *Addendum: The FAIR Guiding Principles for scientific data management and stewardship*. Sci Data 6, 6, https://doi.org/10.1038/s41597-019-0009-6.

Iliad: Forging the Future of Marine Sustainability through Digital Twin Ocean Systems

Keeble Simon[1] (simon@bluelobster.co.uk), Keeble Kathryn[1] (kathryn@bluelobster.co.uk), Berre Arne[2] (Arne.J.Berre@sintef.no)

[1]Blue Lobster IT Limited (UK)
[2]SINTEF Digital (Norway)

The Integrated Large-scale Innovative and Ambitious Digital Ocean (Iliad) project represents a ground-breaking endeavour, funded by the European Union under the €1 billion European Green Deal, to pioneer the digital transformation of oceanic understanding and stewardship. With a consortium of 56 partners across 18 countries from Europe, the Middle East, and North Africa, Iliad is at the forefront of creating state-of-the-art Digital Twins of the Ocean (DTOs), a composite virtual representation of marine systems. These DTOs serve as a cornerstone for realising the vision of a more sustainable and well-preserved marine environment by enhancing real-time, interactive access to integrated data, tools, and information.
At the core of Iliad lies the ambition to bolster the planning and decision-making processes in marine conservation, as well as environmental and socio-economic domains. By enabling complex analysis, simulations, and predictive management scenarios through DTOs, stakeholders and end-user groups are empowered with unprecedented foresight and adaptability in addressing oceanic challenges.

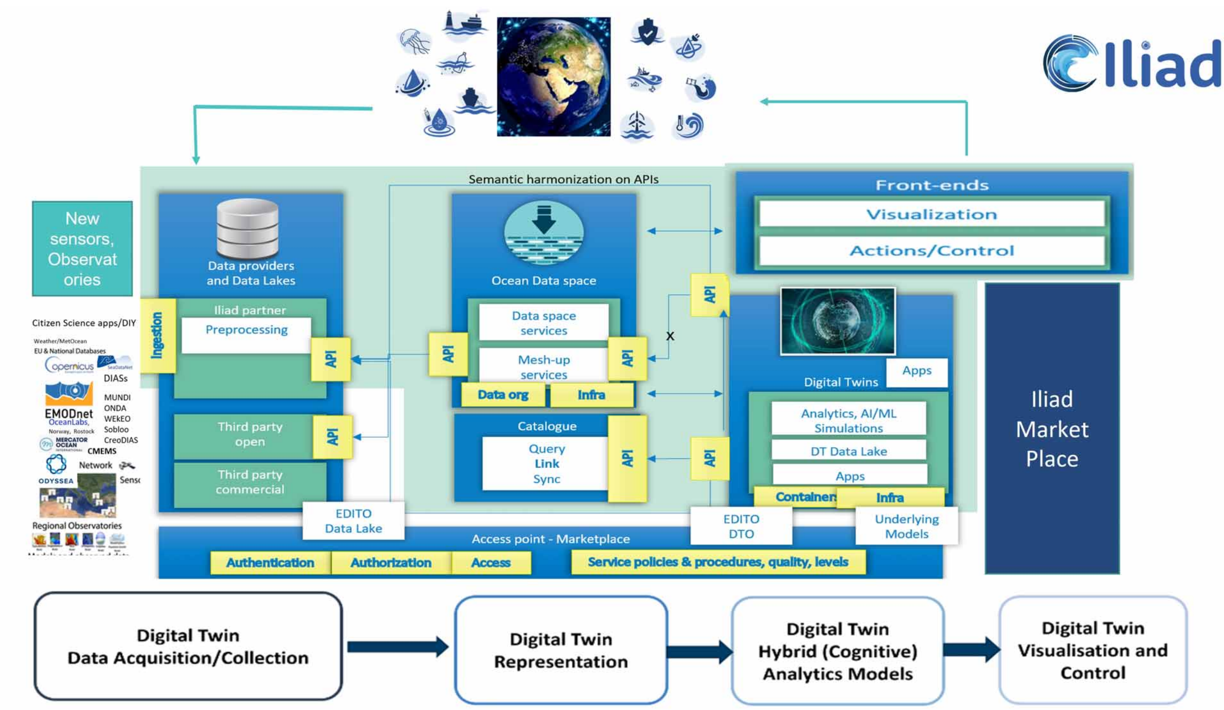


Figure 1 Iliad architecture.

The Iliad marketplace, set for launch in February 2024, will act as a central hub, facilitating the discovery, access, and utilisation of pilot DTs (and their component parts) in alignment with the FAIR Data principles. The marketplace will be open to all creators of DTO's to provide a central platform for DTO discovery and knowledge sharing. This initiative is complemented by the advent of TwinLab, an innovative space fostering the Collaborative Development and Co-design of DTOs. TwinLab serves as a crucible for sharing components of nascent DTOs, promoting a community-driven approach where pilot projects can evolve into mature services, subsequently integrated into the Marketplace.

The pilot DTO's are being co-created directly with stakeholders across energy, fisheries, aquaculture, marine traffic, environmental/pollution monitoring, biodiversity, insurance, harbour safety, plastics, sediment transportation, oil spill, jellyfish, swarms, among many others.

The relationship between Iliad, the European Open Science Cloud (EOSC), and the Environmental Data and Information Transfer Operational (EDITO) framework is pivotal. Iliad's DTOs are envisaged to be synergistic extensions, enriching the data landscape and analytical capabilities of both EOSC and EDITO. Through a federated system approach, Iliad aims to seamlessly integrate with these existing frameworks, allowing for a harmonious and comprehensive data ecosystem that facilitates broader reach and utility.

In its quest to develop a comprehensive network of Digital Twins of the Ocean (DTOs), this in turn supports the European Union's efforts as part of the EDITO (Environmental Data and Information Transfer Operational) projects towards developing a cohesive European DTO. EDITO's mission to streamline environmental data workflows and information transfer processes within the European context lays a vital foundation upon which Iliad's objectives are furthered. By harnessing the EDITO framework, Iliad enhances its capability to assimilate and orchestrate vast arrays of oceanographic data, driving forward the creation of interoperable digital marine ecosystems. This symbiotic relationship ensures that the intricate digital fabric woven by Iliad's DTOs not only aligns with but actively advances EDITO's strategic focus on fostering data accessibility and seamless information exchange across European environmental data spaces. Iliad's endeavour also complements EDITO's vision by expanding the digital horizon through the TwinLab initiative, fostering a collaborative environment where ocean data and models are co-designed. Such collaboration is necessary to set benchmarks for best practices in digital twin development, catalysing the adoption of these practices across European DTO efforts. This confluence of expertise and resources effectively accelerates the continent's march towards a digitally integrated marine knowledge system. By sharing DTO components and piloting novel DT applications, Iliad ensures that the cumulative knowledge gained becomes a cornerstone for the EDITO projects, enhancing Europe's ability to make data-driven decisions in ocean management and policy formulation.

Moreover, The Iliad project is uniquely positioned to significantly bolster the overarching objectives of the EU Mission Ocean initiative. The Mission Ocean initiative seeks to protect and restore the health of our oceans and waters by 2030, a noble and ambitious goal necessitating innovative approaches to environmental stewardship and resource management. Iliad's cutting-edge DTOs provide an essential tool for achieving this, offering a dynamic platform for real-time monitoring, simulation, and predictive analysis of marine ecosystems. By providing stakeholders with advanced tools for decision-making and planning, Iliad facilitates proactive measures in conservation efforts and sustainable use of ocean resources, aligning closely with Mission Ocean's vision of achieving good environmental status of all marine waters.

The interactive and integrative nature of Iliad's DTOs also support Mission Ocean's aim to prevent and eliminate pollution, make the blue economy carbon-neutral and circular, and preserve and restore biodiversity and ecosystems. Through the predictive capabilities and scenario simulations that DTOs offer, policy-makers, scientists, and conservationists can anticipate environmental shifts, design and implement effective marine policies, and engage in restoration activities with enhanced foresight and precision. The Iliad project, by extension, becomes an instrumental driver for Mission Ocean's goals, transforming the EU's approach to ocean governance into one that is not only informed by the best available data but also adaptable to the rapidly changing marine environment.

The concept behind the Iliad project is not merely a significant advancement in digital technology; it is a beacon of collaborative progress towards the sustainable management of oceanic resources. As the project moves towards its marketplace launch, the continual development and integration of DTOs into the wider data infrastructure signify a transformative step in marine science and management. The academic and scientific communities, along with environmental and socio-economic stakeholders, stand at the edge of a new era in ocean stewardship, propelled by the innovations realised through Iliad. As such, Iliad's journey encapsulates a collaborative triumph in technological advancement and environmental consciousness, leading us towards a more informed and sustainable interaction with our planet's oceans.

MGnify: mapping global taxonomic and functional marine biodiversity

Christian Atallah (chrisata@ebi.ac.uk), Ekaterina Sakharova (kates@ebi.ac.uk), Mahfouz Shehu (mahfouz@ebi.ac.uk), Germana Baldi (germanab@ebi.ac.uk), Martin Beracochea (mbc@ebi.ac.uk), Alexander Rogers (sandyr@ebi.ac.uk), Lorna Richardson (lornar@ebi.ac.uk), Robert Finn (rdf@ebi.ac.uk)

EMBL-EBI (UK)

Background

MGnify is EMBL-EBI's resource for microbiome sequence analysis, covering both microbial taxonomic and functional diversity. In collaboration with the European Nucleotide Archive (ENA), MGnify offers publicly accessible services for analysing biome-specific meta'omics sequence data, including sequencing reads, assemblies, and proteins. Users can request the analysis of public or their own private data, which is carried out using standardised and open-source analysis pipelines developed and maintained by MGnify. These pipelines label quality-controlled metagenomic data with rich taxonomic classifications and varied functional annotations, ranging from protein domain and family annotations from InterPro, to predicted biosynthetic gene clusters and their generated secondary metabolites using SanntiS and AntiSMASH.

Of the various biomes users can browse in MGnify, marine-derived studies currently contribute over 37,000 analyses, many of which are from umbrella "superstudies" like Tara Oceans and AtlantECO that aggregate multiple studies. As a key participant in European Commission funded projects like AtlantECO, Blue-Cloud2026 and BlueRemediomics, much of MGnify's recent efforts were within the context of increasing the quantity and the diversity of marine microbiome data on offer, as well as improving the standardised methods of generating these data. Many of these outputs are driving new scientific discoveries that are important to the blue economy, as well as enabling improved modelling of marine ecosystems. In this work, we give an overview of recent and upcoming changes to the MGnify services, and in particular how they pertain to the marine sciences.

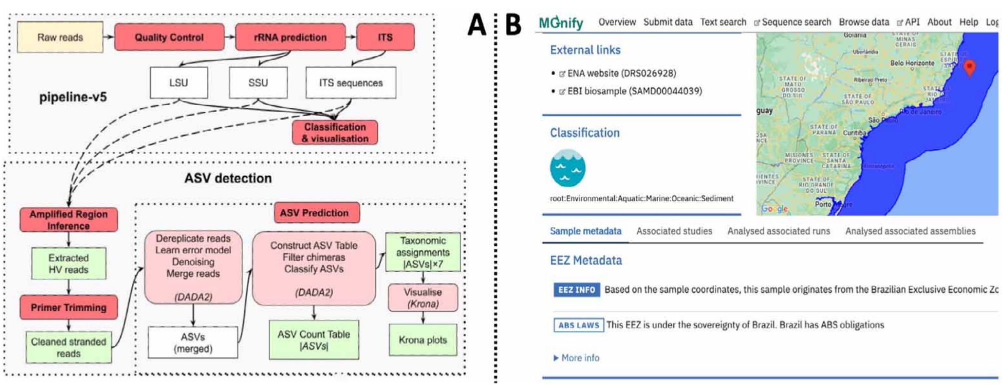


Figure 1 A. Schema of the updated V6 Amplicon pipeline, now containing a module for calling annotated amplicon sequence variants (ASVs). B. Screenshot of the MGnify website displaying the geolocation of a sample and whether it falls in a EEZ (Exclusive Economic Zones).

Taxonomic classification pipelines and data products

An Amplicon Sequence Variant (ASV) is a unique DNA sequence derived from high-throughput analysis of marker genes. Calling ASVs is a more refined method of defining unique sequences compared to traditional operational taxonomic units (OTUs). As ASVs are individual sequences rather than clusters, they also function as consistent keys that allow for easy comparison of taxonomic diversity between different samples and studies. ASV detection was developed within the MGnify service team, as part of the version 6 release of the amplicon analysis pipeline, in addition to its closed-reference method (Figure 1A). This pipeline generates lists of ASV sequences, inferred primer and amplified region metadata, taxonomic assignments from the SILVA ribosomal RNA reference database, and visualisations of taxonomic distributions. Furthermore, as protists are primary producers, assessing their presence and abundance in marine samples is vital for ecosystem modelling. To enable this modelling, the protist-focused PR2 reference database was also added to both the ASV and closed-reference methods, and will be particularly useful for tracking eukaryotic marine diversity, including plankton of varied clades.

For the analysis of raw reads, MGnify has developed a pipeline using mOTUs (molecular Operational Taxonomic Units). This pipeline is written in Nextflow and designed to produce taxonomic profiles by clustering marker genes. It includes several key steps such as quality control, host decontamination, RNA prediction, taxonomic classification, and abundance analysis of mOTUs. Presently, the results are accessible on the MGnify FTP web server. However, the future plan is to deliver these results in the RO-Crate format, corresponding to the respective run, to encapsulate the complete analysis provenance and resulting data products.

MAGs, the marine catalogue, and functional annotations

MGnify currently hosts ten biome-specific, non-redundant Metagenome-Assembled Genome (MAG) catalogues, including a Marine catalogue, containing 1,504 genomes, of which 560 represent novel species that were not present in the r207 release of GTDB. Each catalogue is functionally annotated with EggNOG-mapper and InterProScan and taxonomically annotated with GTDB-Tk. Every MAG is evaluated for completeness and contamination to estimate the quality of the genomic resources, providing access to genomic novelty that is only just starting to be realised via this technique.

In 2021 MGnify launched pipelines to generate and annotate prokaryotic MAGs. All MAG sequence files included in the marine catalogue are submitted to ENA as public data along with taxonomic and coverage metadata. MGnify also provides a public genome_uploader tool to support external researchers in uploading their self-generated MAGs and bins to ENA. MGnify has recently updated and extended the pipeline for generating MAGs to include both prokaryotes and eukaryotes. These MAGs serve as a valuable resource for users, offering an expanded scope for data exploration using our search interfaces, and enabling the extraction of insights about whole communities and the different processes that occur within them.

Exclusive economic zones and Access and Benefit Sharing

Exclusive Economic Zones (EEZ) denote the areas over which coastal nations have exclusive rights regarding the usage and exploration of marine resources. Access and Benefit Sharing laws (ABS) are put in place in EEZs to ensure that monetary and non-monetary benefits that may accrue from the study of samples obtained from marine regions, are shared with the regions from which the samples originate. In an effort to promote awareness and adherence to these policies (i.e., FAIR and CARE data), open access and encourage global participation in data and knowledge generation for marine Biodiversity. MGnify promotes marine data awareness and biobased discoveries by highlighting EEZ and ABS information on our Sample pages (Figure 1B). The darker blue region on the map indicates the sample's EEZ origin, while the EEZ metadata section provides details about the EEZ and its potential ABS status.

EMODnet Ingestion and the operational data exchange examples and hot topics

Antonio Novellino[1] (antonio.novellino@grupposcai.it), Valeria Pizziol[1] (valeria.pizziol@grupposcai.it), Giulia Dapueto[1] (giulia.dapueto@grupposcai.it), Francesco Misurale[1] (francesco.misurale@grupposcai.it), Beatrice Maddalena Scotto[1,2] (beatrice.scotto@grupposcai.it), Rachele Bordoni[1,2] (rachele.bordoni@grupposcai.it), Patrick Gorringe[3] (patrick.gorringe@smhi.se), Dick Schaap[4] (dick@maris.nl), Anastasia Iona[5] (sissy@hcmr.gr)

[1]ETT Solutions S.p.A. (Italy)
[2]Department of Earth, Environment and Life Science, University of Genoa (Italy)
[3]SMHI (Sweden)
[4]MARIS (The Netherlands)
[5]HCMR (Greece)

Introduction

The European Marine Observation and Data Network (EMODnet) represents a long-term marine data initiative, offering a fully functional platform that grants free access to standardized marine data, products, and metadata for both public and private users. In this scenario, EMODnet Ingestion plays a pivotal role in the EMODnet infrastructure, facilitating data holders in the public and private sectors to share their data for safekeeping and open distribution. EMODnet Ingestion services to data holders encompass data submission, which includes an online submission form, user management services, and tracking services; discovery and access, which operate on ingested and completed data submissions together with operational data integration. For real-time data, exchange is facilitated through the implementation of machine-to-machine interfaces like DAB, ERDDAP and others. This exchange is particularly relevant for the EMODnet Physics activities and at the moment, the domains that see the most interesting integrations in terms of the identification of guidelines and the development of automatic protocols for structuring data flows are those described in the following paragraphs.

River flow

EMODnet Physics hosts an operational, open, and FAIR service that provides near real-time outflow data at the nearest river mouth station, by establishing a single-stop-shopping river-runoff data service. This service offers standardized operational river data from national and regional water administrations, gathering information from hydrologic stations to the coastal areas which are unaffected by tides. Currently, it incorporates data from over 700 stations from 35 providers across 17 countries and 3 continents (Europe, North America, and South America). Observations are combined with harmonized metadata based on existing vocabularies (Table 1). The user can access the data operationally by connecting directly to ERDDAP. The user can query timeseries data from the ERDDAP Dataset ID page:
https://data-erddap.emodnet-physics.eu/erddap/tabledap/ERD_EP_TS_RVFL_NRT.html
The Dataset ID page is a web tool to construct and refine queries, such as listing the operational platforms (and their coordinates) in the dataset, by selecting PLATFORMCODE, LAT, LONG. To see the platform page from the mapviewer (dev) the user can generate a script using the following construct:
https://map.emodnet-physics.eu/platformpage/?platformcode=Curranure20002BANDON&source=cp

Metadata field	Vocabulary exists	Link to vocabulary	Vocabulary governance
Platform id	*		EMODnet Physics
Station		https://www.bafg.de/SharedDocs/ExterneLinks/GRDC/grdc_reference_stations_zip.html?nn=201698	GRDC
Owner/provider Institution	Yes	https://edmo.seadatanet.org/	SeaDataNet
Variable names	Yes	http://vocab.nerc.ac.uk/collection/PXX/current/ where XX=02;01;07	BODC:NVS
Unit	Yes	https://vocab.nerc.ac.uk/collection/P06/current/	BODC:NVS
Quality Flag Scheme	Yes	http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf	OceanSites
Time	Yes	ISO8601	ISO
Datum	Yes	WGS84	ISO
Country	Yes	ISO3166	ISO
License	Yes	https://creativecommons.org/	CC
INSPIRE	Yes	ISO 19115	ISO/INSPIRE

Table 1 Applied standards.

**Currently EMODnet Physics is assigning a unique id to river stations, anyhow this inventory will adopt a new convention based on river name and river station, and integrate the GRDC station number.*

Underwater noise

Underwater noise pollution is a growing concern as it can have detrimental effects on marine life, including marine mammals, fish, and invertebrates. By consolidating and standardizing data from various sources - covering both Impulsive and Continuous noise sources - EMODnet aims to contribute to a better understanding of the impact of human activities on the marine environment, including the effects of underwater noise. Looking ahead, the next steps involve forward integrating more data sources, setting up a controlled catalog of underwater sound signatures and implementing tools for automatic soundtracks classification. This last goal can be reached also through the proposal and definition of parameters (metadata) for sharing acoustic files that can be used for training automatic classification systems, as shown in Table 2.

Acoustic signal		Processing-FFT	
Type of data	raw	Min points	512
Data extension	.wav	Max points	4098
Sampling rate	<48 kHz	Window type	Hamming
Dynamic range	16-24 bit	Colour scale	16 greys
Duration	<10 min	Scale	Log
Size	~55 MBytes	Overlap	50%

Table 2 Example of data pre-requisites for automatic soundtracks classification.

Sea ice

In this context, the management of ice-related data is a work in progress. Ongoing efforts are dedicated to expanding the availability of in situ data in polar regions, with initiatives like SOOS (Southern Ocean Observing System) serving as primary stakeholders in this endeavor, driving the need to select themes and types of platforms. While satellite data is well-established as a source for sea ice extensions and quality assessment, collecting in situ data poses more significant challenges. In this context, there is space and potential for using new innovative data types to explore sea ice. For example, the webcams of Antarctic bases may provide continuous information.

Conclusion

With this paper we presented the latest call to action for expanding EMODnet capacities on key parameters while providing new contributors with basic guidelines and indications for joining the challenge.

SHARKtools – Streamline your CTD processing and quality control

Martin Hansson (martin.hansson@smhi.se), Magnus Wenzer (magnus.wenzer@smhi.se), Örjan Bäck (orjan.back@smhi.se)

SMHI (Sweden)

CTD is the most commonly collected data type for oceanographic research and marine environmental monitoring. However, CTD usually suffers from lack of metadata, sensor info and calibration information. Furthermore, the processed end product is most likely in a format that cannot handle data quality information. Addressing these issues consistently and efficiently is crucial to ensuring the long-term accessibility and usability of 'FAIR' (Findable, Accessible, Interoperable, Reusable) CTD data.
The Swedish Meteorological and Hydrological Institute (SMHI), serving as the Swedish National Oceanographic Data Centre (NODC), have developed a solution to streamline this process in the form of SHARKtools. Named after Svenskt HavsARKiv (Swedish ocean archive), SHARKtools is now implemented as the main tool on our national monitoring surveys. The software consists of various functionalities:

- Pre-cruise checks
- Add station, sensor and metadata information before the cast
- Trigger CTD software with added metadata
- Process or reprocess cast
- Convert to NEW CTD DATA FORMAT
- Automatic quality control
- Manual/visual quality control
- Data management

To manage metadata in an efficient way and to avoid manual induced errors the metadata enrichment step includes selecting station, project, orderer, operator, weather codes etc. from predefined lists. Metadata from the onboard event system can be added automatically, including unique IDs for future reference.

New CTD data format
To satisfy our needs of a data format that can hold all necessary data and metadata, and with the possibility to flag data, SMHI have developed a new data format. The format seamlessly integrates with the visualization software Ocean Data View (ODV), allowing for file drag-and-drop functionality without requiring manual import steps. It features quality flag columns for predefined automatic quality checks as well as a "final" quality flag column.
The new format is also compatible with our manual quality control software, providing a rapid visual overview of the data. Scientists can scrutinise the data with expert eyes and apply flags accordingly.

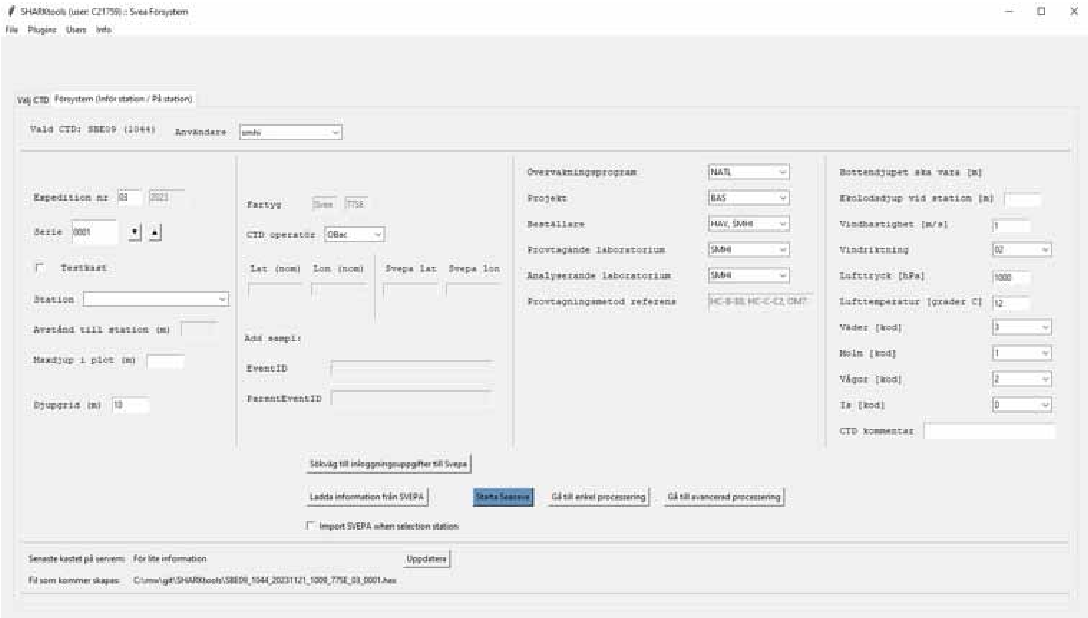


Figure 1 Preparation step in SHARKtools, add station metadata by free text and predefined lists, also load metadata from onboard event system.

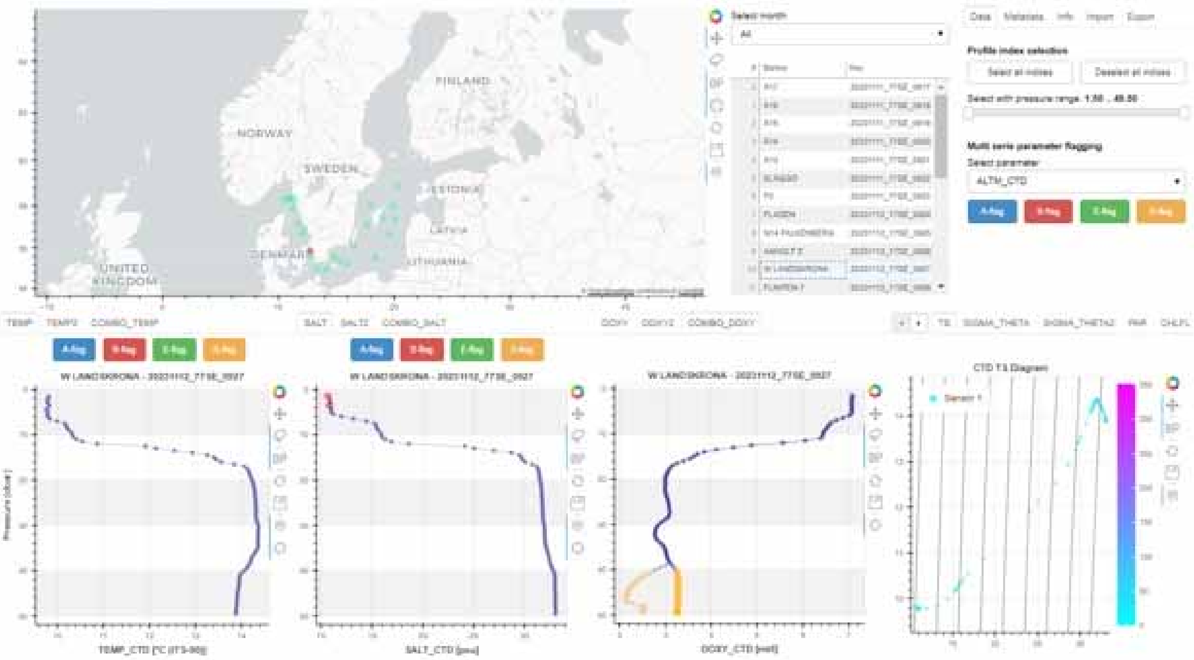


Figure 2 Visualisation of data and automatic quality control results in web browser. Interactive with the possibility to manually flag data.

SHARKtools is written in Python and all source code is available at: https://github.com/sharksmhi

ARCMAP – A database and visualisation tool for Arctic in situ observing systems

Frode Monsen (frode.monsen@nersc.no), Tor Olaussen (tor.olaussen@nersc.no), Hanne Sagen (hanne.sagen@nersc.no), Torill Hamre (torill.hamre@nersc.no)

Nansen Environmental and Remote Sensing Center (Norway)

Abstract

Climate change in the Arctic is expected to have a strong impact on both marine life and sustainable development in the region. In situ observations are therefore essential to estimate the state and changes in the marine environment. As field experiments in Arctic waters are often conducted in short-term research projects, such data becomes time-limited, scattered and lacking a common access point. Based on the INTAROS survey of Arctic in situ observing systems, we have developed the ARCMAP system for gathering updated information about in situ observing systems and their data. This paper presents recent developments and plans for ARCMAP.

Background

The H2020 INTAROS project conducted a survey of Arctic in situ observing systems, in situ and remote sensing data collections [Tjernstrøm et al., 2019]. Based on the questionnaires from this survey we developed a user-friendly web-based system for collecting and maintaining information about in situ observing systems and data collections [Sandven et al., 2022]. This system, ARCMAP, has been used to maintain and extend the INTAROS survey of Arctic observing systems and data collections, as a contribution to GEO and SAON. Furthermore, it can provide a contribution to the upcoming GOOS Regional Alliance (GRA) for the Arctic.

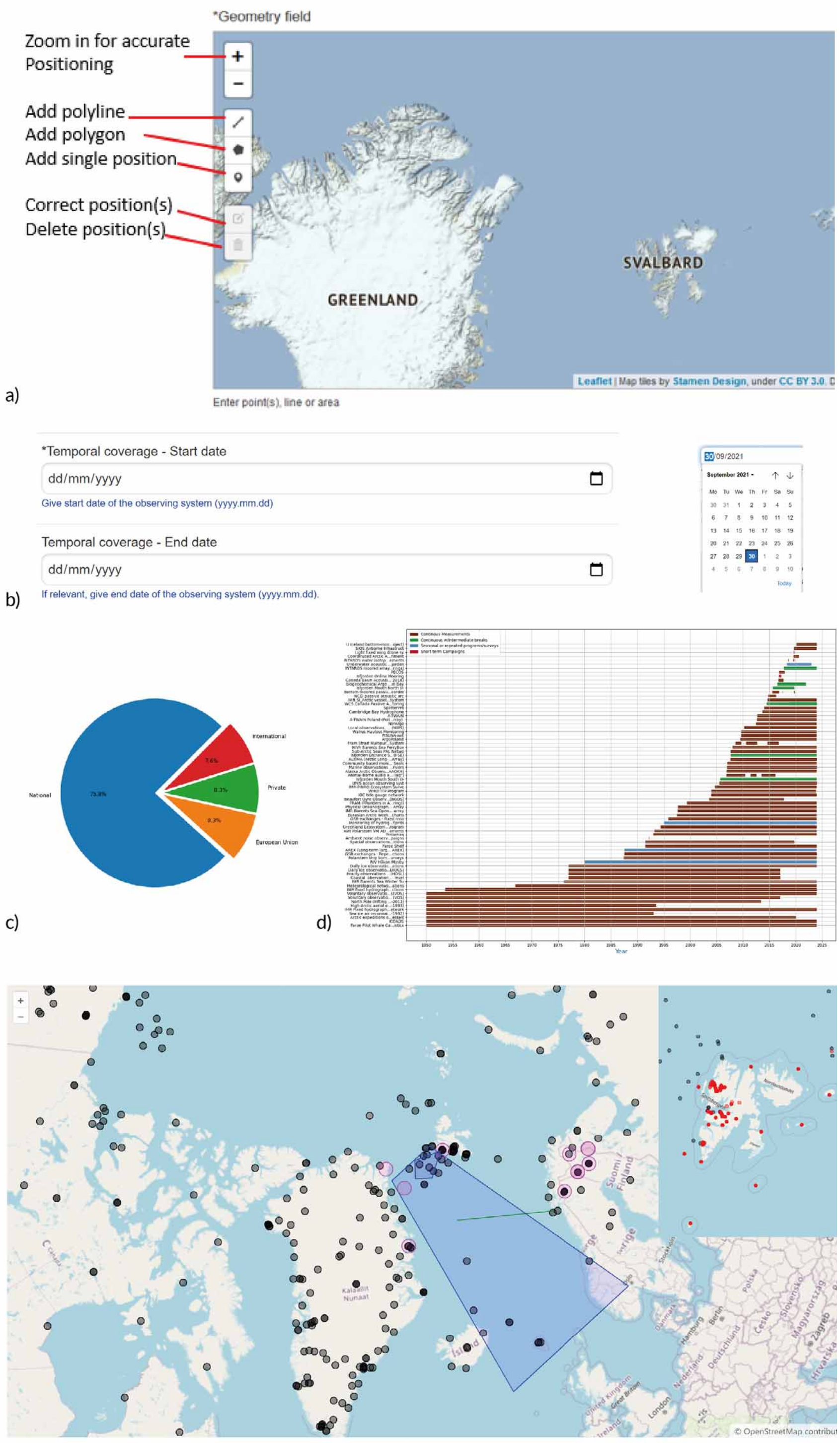
Overview of current functionality

The ARCMAP survey application enables users to register information about in situ observing systems, such as name, owner (organisation), location, duration, and parameters measured. Many of the input fields are free text or values selected from a dropdown list. New capabilities added recently include adding location of a system and its components on a map (Figure1 (a)) or in a text field (longitude, latitude in decimal degrees), and selecting start-stop dates from a calendar (Figure1 (b)). Some new questions have been added, e.g., about the funding sources for the observing system and the periodicity of the measurements. Figure1 (c) and 1 (d) show two examples of new statistics plots generated based on all responses to the survey. These and other statistics plots are updated daily; this can be configured. The map components of ARCMAP have also been updated, allowing users to toggle display of certain observing systems on and off. Figure 1 (e) shows a map of selected observing systems, including some imported from the SIOS Observation Facility Catalogue (OFC).

Future development

Proposed extensions to make ARCMAP more attractive to new user groups: (1) include the INTAROS questionnaire for assessing Community Based Monitoring systems, (2) simplify the data management questions to lessen the burden of scientists registering their observing system, (3) develop a new questionnaire for data managers to assess data systems used for storing and providing access to Arctic in situ data. Other extensions such as exchange of metadata with other observing asset and data systems, further improvements of the reporting subsystem of ARCMAP, and implementation of analytical tools, will also be considered.

Figure 1 Screenshots from ARCMAP (https://arcmap.nersc.no);
a) adding location on a map;
b) defining start-end date for the observing system;
c) distribution of funding sources (a system may have more than one source);
d) observing period of marine observing systems;
e) map of location of selected systems, SIOS systems with red markers.



References

Sandven S. et al., (2022). *Deliverable 1.10 Roadmap for a sustainable Arctic Observing System*. Zenodo. https://doi.org/10.5281/zenodo.7033845

Tjernstrøm M. et al., (2019). *Deliverable 2.10. Synthesis of gap analysis and exploitation of the existing Arctic observing systems*. Zenodo. https://doi.org/10.5281/zenodo.7050807

SatBałtyk System - the Digital Twin of the Baltic Ecosystem

Mirosława Ostrowska (ostra@iopan.pl), Mirosław Darecki (darecki@iopan.pl),
Marcin Wichorowski (wichor@iopan.pl)

Institute of Oceanology PAS (Poland)

The *SatBałtyk System*

An advanced oceanographic monitoring system for the Baltic Sea, incorporates satellite, modeling, and in situ data. It serves as both an innovative research tool in line with the requirements of modern oceanography and a comprehensive platform for delivering environmental data. The system integrates three types of data sources: satellite data for day-to-day monitoring of extensive marine areas, model data utilizing hydrodynamic and ecohydrodynamic models to describe ongoing processes in the marine environment and provide short term forecast, and in-situ measured point data obtained from buoys, shore stations, and other measurement platforms equipped with modern scientific instruments. Additionally, research vessels intermittently contribute data to the system by employing traditional oceanographic measurement techniques.

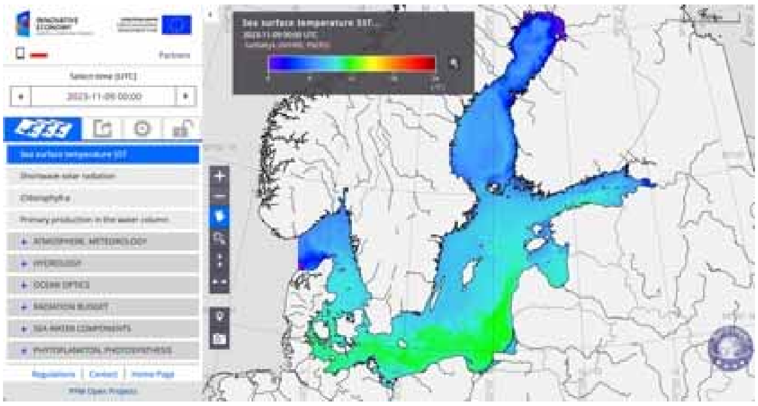


Figure 1 Web portal for data access.

The synergy of diverse data sources is essential for meeting the demands of comprehensive, high-quality, continuous, and near real-time observation of the marine environment. This approach addresses the limitations of ocean color (OC) and sea surface temperature (SST) satellite data, which offer valuable information but are constrained by cloud cover. In instances of cloudy conditions, well-modeled data can serve as a reliable alternative, delivering the necessary information to users in such cases.

On the other hand, satellite radiometers, during favorable weather conditions, provide valuable data that can be assimilated by models to enhance their accuracy and reliability. Through a very complex information flow, the *SatBałtyk System* has been delivering a range of Baltic Sea ecosystem properties via the www.satbaltyk.pl website since 2015, including nowcast and short-term forecasts for certain modeled parameters.

The *SatBałtyk System* establishes the technical infrastructure and implements operational procedures for the effective and regular monitoring of the Baltic environment. This includes assessing its structural and functional characteristics, such as the influx and features of radiant energy (PAR, UV), temperature distributions, the dynamic state of the sea surface (current, wavelength), the concentrations of chlorophyll and other phytoplankton pigments, blooms of algae, the occurrence of upwelling events, and the characteristics of the primary production of organic matter.

The spatial and vertical coverage of the system extends to the entire Baltic. The data systematically gathered without any gaps and with the use of the same methodology throughout the research period, can provide a new perspective on the nature of the processes taking place in various regions of this sea. Open access data are provided as digital maps, and graphs or they can be downloaded as data files for further processing conducted by users on their own. The values of almost one hundred parameters of the Baltic Sea are divided into eight task-oriented groups: Atmosphere, meteorology, Hydrology, Ocean optics, Radiation budget, Sea water components, Phytoplankton, photosynthesis, Coastal zone and Hazards are presented. Apart from the typical oceanographic characteristics of the marine environment, the system determines specialised parameters describing complex natural processes. Many of these characteristics are unique, and their values are not determined, in near real-time, by any other comparable system. The holistic information describes the sea-coast-atmosphere system as a whole, including what is really important and unique, the depth and spectral variability of physical, chemical, and biological parameters. It enables the tracking of long-term changes in the marine environment and comprehensive analyzes of the processes taking place in the Baltic ecosystems (see Figure 2). The collection of data describing the Baltic ecosystem can not only serve for scientific analyses but also aids in decision-making related to the economy, management, and protection of its resources. To enhance access to this data for a broad group of stakeholders, key resources from the *SatBałtyk System* are also integrated into the Oceanographic Data and Information System, eCUDO.pl. *SatBałtyk* streamlines flow of information arriving from accessible remote sensing systems covering the Baltic Sea region, and the Polish economic zone in particular (e.g., the AVHRR instruments on board TIROS N/NOAA, the Seviri instrument on board Meteosat 9, the MODIS instrument on board EOS AQUA, the OLCI instrument on board Sentinel 3a and 3b and many others), and the information available from developed mathematical models of the sea and atmosphere.

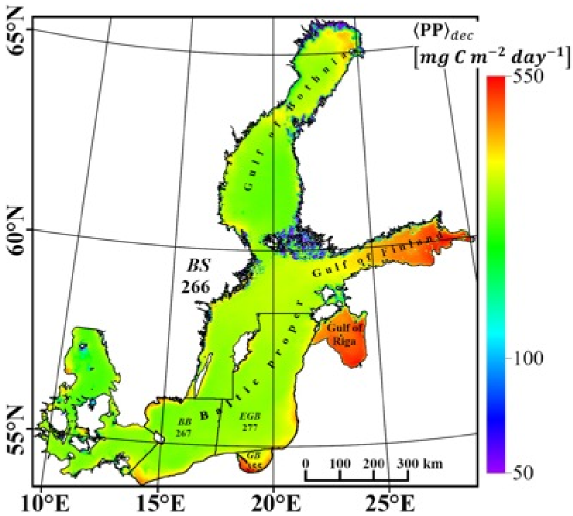


Figure 2 Daily PP of the Baltic Sea, the mean from the period of 2010-2019. The numbers represent mean values determined for the analyzed basins: the Bornholm Basin (BB), the Eastern Gotland Basin (EGB), the Gdańsk Basin (GB) and for the entire Baltic Sea (BS). (Data source: SatBałtyk System, spatial resolution 1x1 km, basin division according to HELCOM).

The *SatBałtyk System* was developed and implemented by the Scientific Consortium SatBałtyk, comprising four scientific institutions: the Institute of Oceanology Polish Academy of Sciences in Sopot (coordinator), the University of Gdańsk, the Pomeranian Academy in Słupsk, and the University of Szczecin.

Automated identification of seafloor deep species

Antoine Lebeaud (antoine.lebeaud@ifremer.fr), Catherine Borremans (catherine.borremans@ifremer.fr), Vanessa Tosello (vanessa.tosello@ifremer.fr)

Ifremer (France)

Faced with the increasing exploitation and artificialisation of the oceans, political players and society recognize the urgent need for a global assessment of the state of marine ecosystems. Our ability to distinguish natural variations in ecosystems from changes resulting from anthropogenic activities requires multidisciplinary monitoring of these environments. This thinking led to the development of monitoring programs at instrumented sites in coastal environments, and to the establishment of observatories that provide continuous access to both coastal and deep-sea ecosystems, in which the Research Infrastructure EMSO plays a leading role at the European level. These advances have been made possible by the technological developments of recent decades. In particular, underwater imaging has rapidly emerged as a non-destructive method for examining biodiversity on unprecedented time and space scales. The exponential increase in resolution and quality of optical sensors means that species can be identified with ever greater precision, and the reduction in their cost facilitates their acquisition. These advancements demonstrate the potential of imaging tools for ecological monitoring of marine ecosystems.

Yet the success of imagery data for scientific purposes leads to new challenges linked to the processing of the exponential amount of data collected, which can be time-consuming and tedious. The advent of Artificial Intelligence (AI) has enabled the development of powerful algorithms that should facilitate the processing of large imaging datasets. However, the capacity of machines to convert large volumes of raw optical signals into usable data for studying marine habitats is conditioned by a learning phase on large standardized reference databases resulting from manual processing based on human decisions. These reference databases are generated by scientists, students, technical staff in laboratories, as well as by citizens through online platforms (e.g. Deep Sea Spy citizen science annotation platform developed by IFREMER). Innovative AI algorithms should be developed to improve the human-intensive work of annotations.

In the framework of the iMagine Horizon Europe project (Grant number 101017567), IFREMER is participating in the “Ecosystem monitoring at EMSO sites by video imagery” use case. This case study aims to create an operational and integrated service, based on AI models, for automatic processing of images collected by cameras at EMSO underwater sites, identifying and analysing fauna and habitats for ecosystem monitoring purposes.

In this context, the primary aim is to develop an automated object detection system for identifying species in the images using supervised machine learning techniques. This supervised learning approach relies on labelled data, and in this case, we used the citizen science annotations from Deep Sea Spy to train the model. Indeed, Deep Sea Spy is a participative science platform launched in 2017, that provides access to images from EMSO-Azores and Ocean Networks Canada observatories for annotation purposes.

Convolutional Neural Networks (CNN) are frequently used for object detection tasks due to their ability to extract relevant features from images. Consequently, a CNN-based object detection algorithm called Yolov8 (iv) was specifically employed to automatically identify species present in the images.

Significant efforts have been made to prepare the training dataset for automatic identification, which has been challenging due to the diversity of species and the variable methods used for labelling the data. Indeed, the annotations were made by multiple annotators (each image is

analyzed by 10 persons), which resulted in varying levels of accuracy and consistency in the labelling of the different species. Furthermore, the data were labelled differently depending on the species (polygons, lines, points). This has required additional effort to standardise and clean the data in order to make it suitable for use with the YOLOv8 model.
The final result of this work is the development of a user-friendly pipeline that comprises several interconnected components and functions, enabling efficient training data preparation, cleaning, and effective model training. Key elements of the pipeline include:
Training data Conversion: The pipeline starts by converting various annotation types (polygons, lines, and points) into regular bounding boxes. Corrections are applied to compensate for potential conversion errors, and adjustments are made based on the original label and species.
Training data Unification: A Python function unifies overlapping bounding boxes, leveraging redundant information from multiple annotators to ensure accurate object labelling. The unification process considers the Intersection over Union (IoU) metric and is species-dependent, reducing incorrect annotations.
Training data Visualization and Validation: A function generates thumbnails from bounding boxes, allowing users to visually inspect and verify data quality. Unreliable or incorrect thumbnails can be deleted, removing their corresponding bounding boxes from the dataset and ensuring precise control over the data used for Yolov8 model training.
Yolov8 learning set formatting: The pipeline prepares the cleaned dataset in the required format for training the Yolov8 model, automating the process and facilitating efficient model development.
As a final step, Yolov8 was trained with a cleaned learning dataset and results are now under revision and validation in order to further optimise the algorithm.

Figure 1 Unification of citizen's bounding boxes.

Once finalized, the pipeline will be integrated into the iMagine platform as a service, providing researchers, marine biologists and other stakeholders with tools for cleaning of datasets (especially citizen science image datasets) and an AI model capable of automatic marine species classification.
The presentation at IMDIS 2024 will give an overview of the final annotation pipeline, from training data preparation and cleaning to model learning and inference results.

References

EMSO (European Multidisciplinary Seafloor and water column Observatory): https://emso.eu/
DeepSeaSpy: https://www.deepseaspy.com/
iMagine Horizon Europe project: https://www.imagine-ai.eu/
Jocher G., Chaurasia A., & Qiu J. (2023). *Ultralytics YOLO (Version 8.0.0)*. Computer software. https://github.com/ultralytics/ultralytics
iMagine platform: https://dashboard.cloud.imagine-ai.eu/marketplace

Marine Environmental Indicators production on the cloud

Giovanni Coppini[1] (giovanni.coppini@cmcc.it), Francesco Palermo[1] (francesco.palermo@cmcc.it), Simona Simoncelli[2] (simona.simoncelli@ingv.it), Alessandra Giorgetti[3] (agiorgetti@ogs.it), Jan Willem Noteboom[4] (janwillem.noteboom@knmi.nl), Matteo Scuro[1] (matteo.scuro@cmcc.it), Rita Lecci[1] (rita.lecci@cmcc.it), Massimiliano Drudi[1] (massimiliano.drudi@cmcc.it), Antonio Mariani[1] (antonio.mariani@cmcc.it)

[1]CMCC (Italy)
[2]Istituto Nazionale di Geofisica e Vulcanologia – INGV (Italy)
[3]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[4]KNMI (The Netherlands)

Ambition & Scope

Started in the pilot phase of Blue-Cloud (2020-2022), *Marine Environmental Indicators* (MEI) VLab (Virtual Lab) allow users to monitor and assess the environmental status of marine areas and support the decision-making process for the ocean management. Multiple data sources are exploited in a unique data analysis service, which will allow the online computation of indicators.
Functionalities developed in the pilot Blue-Cloud are going to be improved, including new data sources (physics, biogeochemistry, biology, chemical data) and new algorithms. The tool will calculate online metocean information and indicators on the environmental quality of the Mediterranean Sea and Global Ocean, using input from BDIs, also improving uncertainty evaluation.
A set of metocean environmental indicators will be defined according to available datasets from BDIs. For each indicator, an analysis and production workflow will be implemented, or further developed and tested, applying big data analysis methods on the multi-disciplinary datasets.

The web application

Marine Environmental Indicators VLab (Virtual Lab) aims to develop a web application that allows users to monitor and assess the environmental status of marine areas, by performing online spatio-temporal analysis with the implemented algorithms, for selected environmental variables.
After a feasibility study to define the new algorithms, they will be implemented inside the VLab for the computation of metocean characteristics and indicators. They can contribute to initiatives such as the Copernicus Marine Ocean State Report or Med-CORDEX.
New proposed indicators for MEI in Blue-Cloud 2026 (lead by CMCC, partners: OGC, INGV, KNMI):

- Ocean Heat Content
- Enhanced Storm Severity Index (SSIv2)
- Trophic index (TRIX)
- Marine Heat Waves

The user can choose many parameters for sending the request (they depend by the algorithm) e.g.,: Method, Output Type, Data Source, Time Range, Area, Depth. The web application interacts with the D4Science CCP Analytics engine to submit the user request. The CCP executes then the selected algorithm code in a customized Docker container. The produced outputs will be then stored in the Cloud Shared Filesystem and made available to the WMS Service. It offers better visualization of results in form of map or offers the data extraction for producing timeseries. The Vlab architecture diagram is visible in Figure 1. The Figure 2 offers the screenshot of the currently available *Marine Environmental Indicators* web application that we developed during Blue-Cloud Pilot and currently operational at the following url on D4Science Blue-

Cloud2026 VRE[43]. Cloud Computing Platform (CCP) is based on containerization, REST APIs and Json is a type of service that offers on-demand and scalable availability of computing resources such as servers, storage, and software over the internet. It eliminates the need for the users to manage their own physical infrastructure, while being able to execute, execute again, demonstrate and share their procedures, algorithms, comuptations and results with their scientific community independently technological and implementation details.

Some definitions:
BDI (Blue Data Infrastructure)
DD/DA (Data Discovery/Data Access)
WMS (Web Mapping Service)

Figure 1 MEI architecture Diagram.

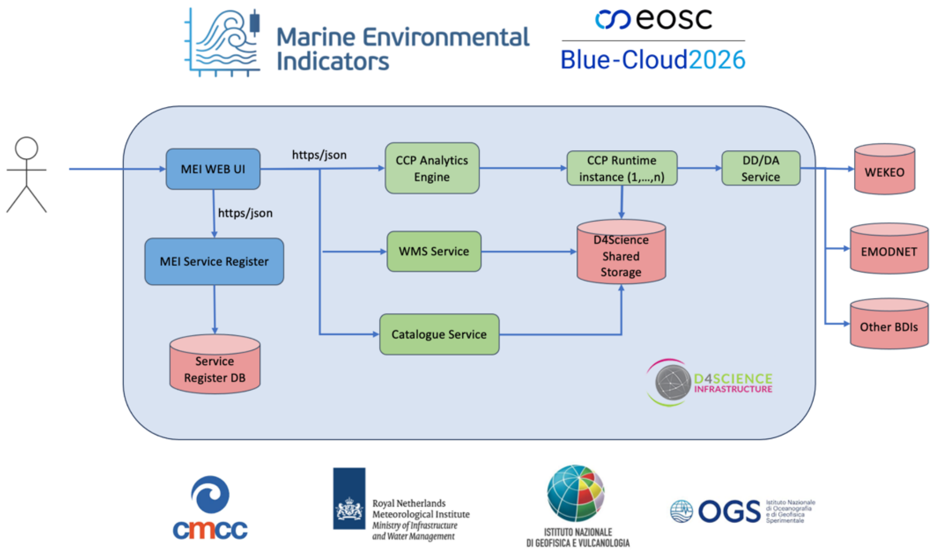
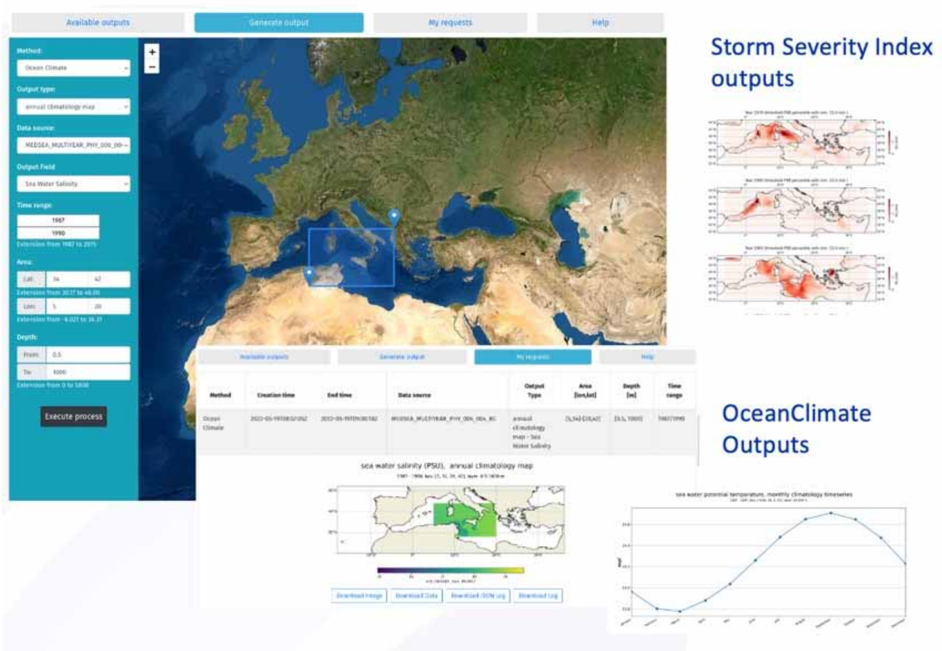


Figure 2 MEI pilot Web App.



Poster session

[43] https://blue-cloud.d4science.org/group/marineenvironmentalindicators/marineenvironmentalindicators

Easier Access to Scientific Data in the Context of the Multi-disciplinary MACMAP Project

Claudia Fratianni (claudia.fratianni@ingv.it), Simona Simoncelli (simona.simoncelli@ingv.it), Antonio Guarnieri Minnucci (antonio.guarnieri@ingv.it)

Istituto Nazionale di Geofisica e Vulcanologia – INGV (Italy)

Introduction

Climate change is a highly cross-disciplinary research domain in which building knowledge as a collaborative endeavor is overwhelmingly important. The issue is not only creating new data but also facilitating knowledge and data sharing across the various working groups to address multifaceted scientific and societal challenges. In this context, proper data management is crucial and following the data throughout the so-called data value chain (https://doi.org/10.1016/B978-0-12-823427-3.00001-3) contributes to make the most efficient and sensible use of them, transforming de facto the data into a valuable, precise and reliable source of scientific information and knowledge.

MACMAP project

A multidisciplinary approach has been adopted in the framework of the project MACMAP (A Multidisciplinary Analysis of Climate change indicators in the Mediterranean And Polar regions), funded by Istituto Nazionale di Geofisica e Vulcanologia (INGV). The project is dedicated to the study of the climate evolution in the Polar and Mediterranean regions by extending and integrating existing data with new observations, model outputs and qualitative historical information. The ultimate objective is the advancement of the comprehension of the present climate evolution considering the Earth System as a whole and producing new and more accurate estimates of climate change indicators.

Since the project covers a heterogeneous pool of data regarding very different disciplines of the geosciences such as oceanography, physics of the atmosphere, seismology, hydrology, geochemistry and more, the integration of such diverse resources requires to be compliant with FAIR principles, adopting standards-compliant formats and controlled vocabularies to facilitate their findability, accessibility, interoperability and reusability.

Methodology

The multidisciplinary nature of the project required a particular effort to enhance the interaction amongst data producers, managers and users. The Data Management strategy adopted within the project aims to: 1) facilitate the data exchange within the project and the users community; 2) adopt the use of common vocabularies and data formats to ensure consistency and interoperability with the existing data portals and services(i.e EMODnet Physics, SeaDataNet); 3) make data easy to discover (Findable), openly available (Accessible) and readily usable (Interoperable, Reusable also machine to machine); 4) promote and encourage the concept of Open Science (OS) by raising the awareness of its value.

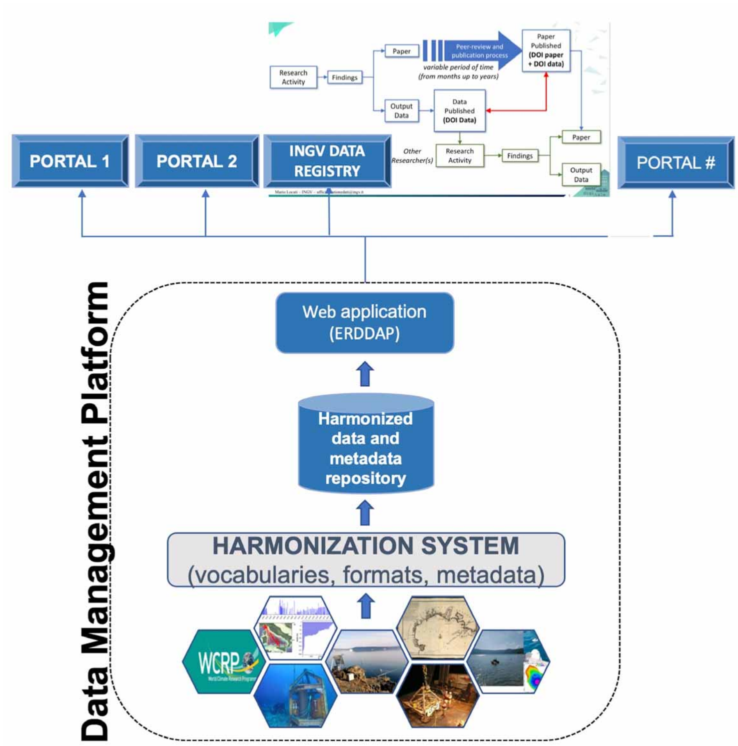
Our innovative strategy was to implement an ERDDAP server to facilitate the availability of subsets of CMIP6 HighResMIP (High Resolution Model Intercomparison Project, https://doi.org/10.5194/gmd-9-4185-2016) climate model data to the partners. This avoided duplication of efforts or resources and it could be considered a first endeavor of a data lake solution. Then the ERDDAP was used to share harmonized research outputs from the different activities (Figure 1) according to international common data and metadata models, allowing their integration into existing data infrastructures. Moreover, assigning Digital Object Identifier

(DOI) through the INGV Data Registry (https://data.ingv.it/) acknowledges the efforts of those who contributed to generating the data and products and assures data availability in further scientific publications.

The data harmonization process encompasses several key elements:

- Semantic: Common Data Models (CDMs), such as netCDF, are advocated since they provide standardised formats for storing and sharing data. Identifying suitable common terms and vocabularies (such as OceanSITES, Climate and Forecast (CF), SeaDataNet, British Oceanographic Data Centre (BODC)) in collaboration with the principal investigators helps avoid ambiguities and ensures consistency in data interpretation.
- Technical: Data publication involves the use of agreed DOIs to enhance data discoverability. Additionally, enabling machine to machine communication is strategic for efficient data access and use;
- Policy: Associating a Creative Commons license with the identified data ensures clarity regarding the terms of use and promotes data sharing while maintaining appropriate attribution. This policy aspect helps establish legal frameworks for data sharing and reuse, fostering a culture of openness and collaboration.

Figure 1 MACMAP Data Management Workflow design.



Conclusions

The MACMAP project's approach intends to implement a comprehensive strategy to ensure data FAIRness. The use of semantic artifacts and the ERDDAP server as a data delivery solution not only enhances the accessibility and interoperability of the data but also commutes the scientific process itself. By promoting the OS principles among the INGV research community, the project fostered a culture focused on transparency, collaboration, and data sharing. This approach facilitates scientific advancement and elevates data to valuable, reusable research contributions, maximizing their impact and usefulness.

Challenges of Tracking Provenance in Marine Data

Tanja Auge[1] (tanja.auge@ur.de), Fajar J. Ekaputra[2,3] (fajar.ekaputra@wu.ac.at), Susanne Feistel[4] (susanne.feistel@io-warnemuende.de), Susanne Jürgensmann[4] (susanne.juergensmann@io-warnemuende.de), Meike Klettke[1] (meike.klettke@ur.de), Laura Waltersdorfer[2,3] (laura.waltersdorfer@tuwien.ac.at)

[1]University of Regensburg (Germany)
[2]TU Vienna (Austria)
[3]WU Vienna (Austria)
[4]Leibniz Institute for Baltic Sea Research Warnemünde (Germany)

Long-term interdisciplinary data studies in the Baltic Sea area

The *Leibniz Institute for Baltic Sea Research, Warnemünde (IOW)* is a non-university research institution dedicated to interdisciplinary marine research in coastal and marginal seas divided into four scientific sections. The IOW focuses on the Baltic Sea ecosystem and holds data from more than 130 years of research, both collected by the institute itself and provided by other research institutes.

Due to their *provenance* in terms of where they were collected, who collected them and when they were collected, the data are in different formats and have different measuring devices, methods and standards for data collection and management. Provenance generally describes the information supporting the reproducibility of research data processes. The Physical Oceanography Section, for example, deals with long-term observations of the marine environment. For decades now, the IOW has deployed a *CTD-Probe* (*Conductivity*, *Temperature* and *Depth*) to explore the water column. The device contains multiple sensors to continuously measure parameters and a varying number of water bottles to take samples from the water column at specific depths.

From Sensor to Publication

Any measurement by sensor is initially a voltage value which gets converted by the operating software *Seasoft* to a numeric value of the measured parameter. To assure quality and accuracy of the measurements several measures are taken. One is the deployment of double sensors for the same parameter. Another measure is the comparison of CTD-measurements with measurements of the same parameter by different methods. A third way to ensure the quality of CTD-measurements is the precise calibration of the sensors before and after a deployment or time period. Calibration factors are determined and have to be applied in the post-processing. The results of a CTD measurement, as well as their analyses, are initially stored locally on the scientists' computers in files of various formats and structures such as *txt*, *csv* or *xlsx*, and then in an evolving relational database (see data science pipeline in Figure 1, highlighted by green arrows). To make the data available as quickly as possible, the raw data will be regularly updated/published in real-time databases such as *Copernicus*[44]. In addition, the data processed in the *IOWDB*[45] will be made available to IOW researchers and the most important part of the data will be made available to the public. In summary, (raw) data is published at different times, in different formats and at a different level of analysis (see orange arrows in Figure 1). To make sure that IOW data can be used by others in the long term, the provision of data at

[44] Coperincus: https://www.d-copernicus.de
[45] IOWDB: https://odin2.io-warnemuende.de

IOW is subject to a *data policy*[46] for handling research data. Therefore, *metadata* (geolocation, time, sensors, calibration, operating persons, processing software, etc.) have to be collected and stored during the entire process to enable comparability in the aftermath. To ensure the collection, storage, and reuse of all data and metadata in accordance with *FAIR (Findable, Accessible, Interoperable, Reusable) principles* a box of easy-to-use, standardized tools that accompany the research process is required.

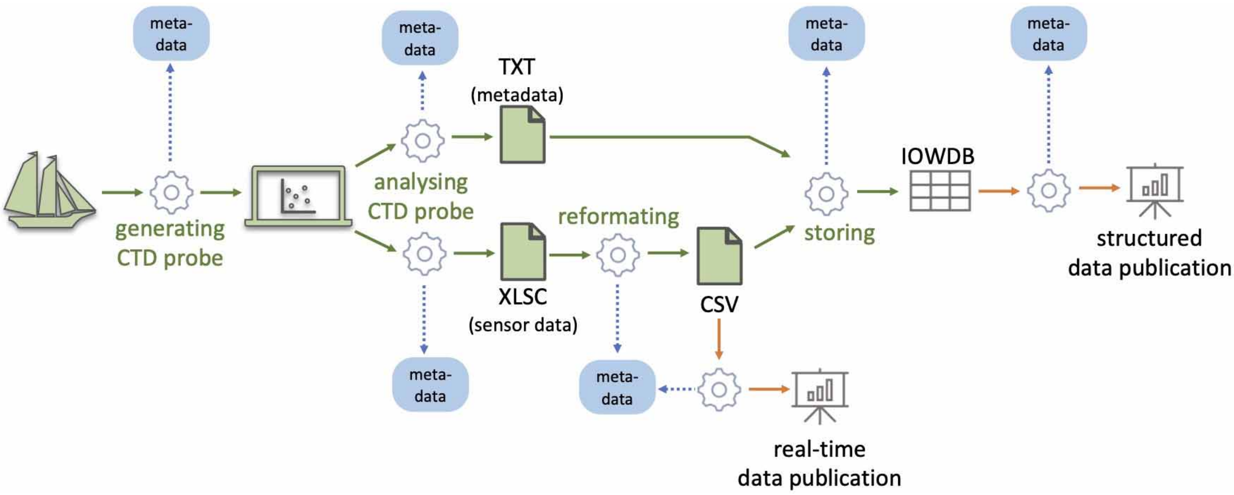


Figure 1 Example for a possible data science pipeline as it might be anchored at the IOW.

Challenges and Use Cases for a Marine Data Provenance Tool

When designing and implementing such a toolbox, several challenges must be addressed:

- As the standards for data collection differ between institutions, countries, and application scenarios, a high degree of heterogeneity within the data is to be expected.
- In long-term studies, schema changes cannot be avoided.
- A centralized collection and management of data is often expensive and time-consuming.
- The storage structures within a project sometimes do not match the desired formats of the funders or the institution-wide storage.
- Data collection is largely automated, depending on the collection method, and manual review of the data collected is simply not possible.
- While metadata is collected throughout the data science process, it is usually collected independently and not collected, managed, and stored in a structured, centered way.

We believe that these challenges are typical for the entire scientific landscape, since in many contexts data are collected in time-limited projects and made available in other formats, in the long term. Our goal is to better support researchers and data managers of institutes like the IOW in their (computational) scientific work to answer specific (provenance) questions such as:

- Q1: Which datasets are affected by an error or bug, and how?
- Q2: Who was involved in generating the data?
- Q3: Which data or scripts are needed to repeat a workflow or (re-)produce a result?

[46] IOW Data Policy:
https://www.io-warnemuende.de/files/forschung/mediathek/iowdb/IOWDataPolicy_20180411.pdf

We therefore aim to capture metadata throughout the whole data science process and preserve it using (workflow- and data-) provenance techniques. For this, we are developing concepts and tools based on existing initiatives (SeaDataNet, EMODnet, EMBRC, ...), standards (W3C PROV, ...) and ontologies in cooperation with the IOW, the University of Regensburg, TU Vienna and WU Vienna.

The AquaInfra VRE: A platform to perform freshwater and marine research online

Merret Buurman[1] (merret.buurman@igb-berlin.de), Dr. Markus Konkol[2] (m.konkol@52north.org), Vanessa Bremerich[3] (vanessa.bremerich@igb-berlin.de), Dr. Sami Domisch[3] (sami.domisch@igb-berlin.de)

[1]Leibniz Institute for Freshwater Ecology and Inland Fisheries (IGB) (Germany)
[2]52°North (Germany)
[3]IGB (Germany)

AquaINFRA: Water scientists, unite!

The AquaINFRA project, which started in 2023, is a pan-European project funded by the Horizon Europe Framework Programme that aims at bringing together freshwater and coastal marine scientists across Europe and provide FAIR and EOSC-compliant data and processing services. A central part of AquaINFRA is a Virtual Research Environment (VRE), which will allow scientists to perform data-intensive computing tasks in a server-based environment.

While in the marine realm large-scale analyses are very common, freshwater research often focuses on small study areas, such as individual lakes, reservoirs or river catchments, requiring relatively small amounts of data. However, more and more high-resolution data is available, and large-scale analyses across the rivers, lakes, coasts and groundwater have become more prevalent. For example, Amatulli et al. [2022] created *Hydrography90m* [https://doi.org/10.5194/essd-14-4525-2022, https://hydrography.org/], a global hydrographic network, containing large rivers as well as small headwaters. It allows for detailed analysis of more than 40 hydrological variables on a high spatial resolution, and for connectivity analyses along a network containing hundreds of millions of segments. Working with such large datasets on local computers requires significant amounts of effort, storage, memory, and download bandwidth. Consequently, scientists need a way to run computations easily and efficiently in a server-based VRE.

The AquaINFRA VRE: Galaxy and OGC processing services

The AquaINFRA VRE allows for web-based processing of freshwater and other datasets. It emphasises usability and reproducibility by combining the Galaxy platform with processing services following the Open Geospatial Consortium (OGC) standards. The Galaxy platform [https://galaxyproject.org/] was initially developed for genomic research. It allows users to choose from a large number of tools to analyse their data. A very convenient feature is the workflow canvas, a graphical surface which allows to chain tools to produce a reproducible workflow which can be run over and over again.

However, Galaxy tools and workflows cannot be easily incorporated into scripts running elsewhere. To allow this, the tools of the AquaINFRA VRE are deployed as OGC processing services [https://ogcapi.ogc.org/processes/], using the Python implementation *Pygeoapi* [https://pygeoapi.io/]. OGC services are programs that run on a server and can be accessed via a standardised REST API. They can thus be incorporated into any routine that allows making HTTP requests and handle their results, increasing the interoperability of the analysis. To support scientists who prefer graphical interfaces over scripting languages, the processing services can be accessed in the Galaxy platform by using Galaxy tools that call these services via HTTP, retrieving the results and making them available in Galaxy for further use.

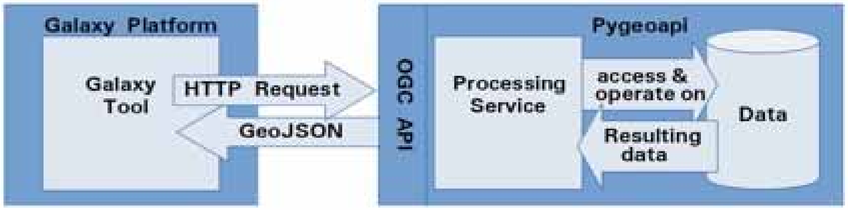


Figure 1 Architecture of the AquaINFRA VRE.

Example workflow: Fish distribution in a Brazilian river

To illustrate this approach, we use a study of species habitat. The necessary steps to compute the fragmentation of a fish habitat in a river (see https://glowabio.github.io/hydrographr/articles/case_study_brazil.html) were completed in R, using the R package hydrographr [Schürz, 2023] [https://github.com/glowabio/hydrographr/], which provides many functions commonly used in the hydrological sciences.

The analysis starts by retrieving species occurrence data from the Global Biodiversity Information Facility (GBIF). These coordinates are then spatially filtered using the outline of the study area, the drainage basin of the river São Francisco, available from the global hydrological dataset *Hydrography90m*. Next, these filtered locations are snapped to the river network: When coordinates measured in the field do not coincide precisely with the segments of the digitized river network, shifting them to the nearest stream segment facilitates the subsequent analysis, which will not be described here in more detail.

Each of these steps can be carried out via HTTP, returning the result as GeoJSON. They can also be carried out visually in Galaxy, providing the input parameters via entry fields and dropdown menus. And finally, the tools can be chained using the graphical canvas, so that all steps can be run with one click (see Figure 2).

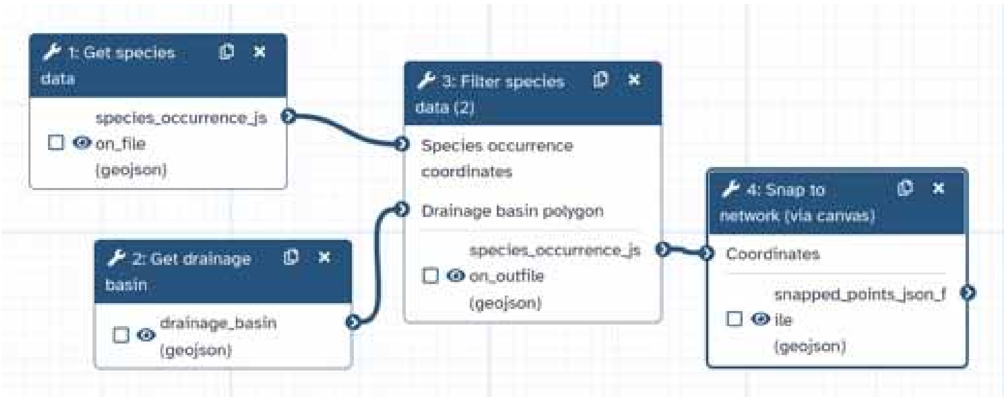


Figure 2 Screenshot of the workflow canvas in Galaxy, showing the first steps of a larger fish habitat analysis.

References

Amatulli G., Garcia Marquez J.R., Sethi T., Kiesel J., Grigoropoulou A., Üblacker M., Shen L., Domisch S. (2022). *Hydrography90m: A new high-resolution global hydrographic dataset*. *Earth System Science Data*, 14, 4525–4550. https://doi.org/10.5194/essd-14-4525-20225-2022

Schürz M., Grigoropoulou A., Garcia Marquez J.R., Tomiczek T., Floury M., Schürz C., Amatulli G., Grossart H.-P., Domisch S. (2023). *hydrographr: an R package for scalable hydrographic data processing*. Methods in Ecology and Evolution, https://doi.org/10.1111/2041-210x.14226

The DTO-Bioflow Project: Bringing Biodiversity Data into the European Digital Twins of the Ocean

Klaas Deneudt[1] (klaas.deneudt@vliz.be), Laura Slaughter[2] (laura.slaughter@sintef.no), Arne Berre[2] (Arne.J.Berre@sintef.no), Syrine Ben Aziza[3] (syrine.benaziza@tno.nl), Carlota Muniz[1] (carlota.muniz@vliz.be), Tomas Martinovic[4] (tomas.martinovic@vsb.cz), Ute Brönner[2] (Ute.Broenner@sintef.no), Alain Arnaud[5] (aarnaud@mercator-ocean.fr), Jesse Harrison[6] (jesse.harrison@csc.fi), Francisco Hernandez[1] (francisco.hernandez@vliz.be)

[1]VLIZ, Flanders Marine Institute (Belgium)
[2]SINTEF (Norway)
[3]TNO (The Netherlands)
[4]IT4Innovations (Czech Republic)
[5]Mercator Ocean (France)
[6]CSC (Finland)

The Digital Twin of the Ocean (DTO)-BioFlow Project will transform access to ocean biodiversity data by enabling the sustainable integration of biodiversity monitoring data flows into the Digital Twin of the Ocean (EU DTO)[47]. The project unlocks the societal value of these data, thereby enriching digital tools and services in cooperation with the European Marine Observation and Data Network (EMOD)

EMODnet Project[48] and various DTO projects, including the European Digital Twin of the Ocean (EDITO)[49], EDITO Model[50], and the Iliad consortium[51]. To fully exploit the potential of these new and existing data flows, DTO-BioFlow will develop the biodiversity component of the DTO. The DTO fulfils the objectives of the EU Biodiversity Mission and Strategy by bringing together data, models, and new algorithms to support the development of policy-relevant tools and services for effective monitoring, restoration, and protection of marine biodiversity.

Figure 1, read left-to-right, shows how DTO-BioFlow provides support for the various steps in a digital twin pipeline. These steps move from data acquisition to data curation and representation, to models, and finally analysis. The twins support stakeholder requirements allowing the user to examine prediction-type queries to the twin, conduct what-if analyses, and interact with various types of visualisations. This will be developed through demonstrator use cases for the following seven areas: (1) invasive species management; (2) adaptive offshore construction and energy harvesting; (3) assessing pelagic biodiversity in relation to human impact; (4) spatial planning of sustainable mariculture; (5) ecosystem-based marine spatial planning and Marine Protected Areas (MPA) management; (6) low impact fisheries; and (7) biodiversity ecosystem services including carbon sequestration.

The key challenge is bringing new biodiversity data streams into the European DTO through a suitable interoperability architecture. New technology and data types (imaging, acoustics, DNA based, satellite) produce huge amounts of readings that require processing, aggregation, and translation into usable data products such as species occurrences, densities, biovolumes, species migration routes, etc. For many of these data types this hampers the flow of data from the biodiversity sensor networks to the biodiversity data integrators used to deal with more

[47] https://digitaltwinocean.mercator-ocean.eu
[48] https://emodnet.ec.europa.eu/en
[49] https://edito-infra.eu
[50] https://edito-modellab.eu
[51] https://ocean-twin.eu

traditional types of observation data. Sustainable ingestion procedures for these data towards the DTO will be provided, with the use of EMODnet Biology[52] as a primary focus. EMODnet is the portal that provides open and free access to interoperable data and data products on temporal and spatial distribution of marine species (angiosperms, benthos, birds, fish, macroalgae, mammals, reptiles, phyto- and zooplankton) from European regional seas. The infrastructure and data flow used within EMODnet Biology is based on the European Ocean Biogeographic Information System[53] (EurOBIS/OBIS), the Global Biodiversity Information Facility[54] (GBIF), and their underlying biodiversity ontologies. Other aggregators will also be considered when appropriate. The collected data will further be transformed to appropriate digital representations suitable for the data lakes of relevant digital twins with a transformation to analysis-ready data. Considered data formats for the digital twin models are Parquet/GeoParquet[55], SpatioTemporal Asset Catalogs[56] (STAC), and Zarr[57] drive, in particular for the analysis-ready data.

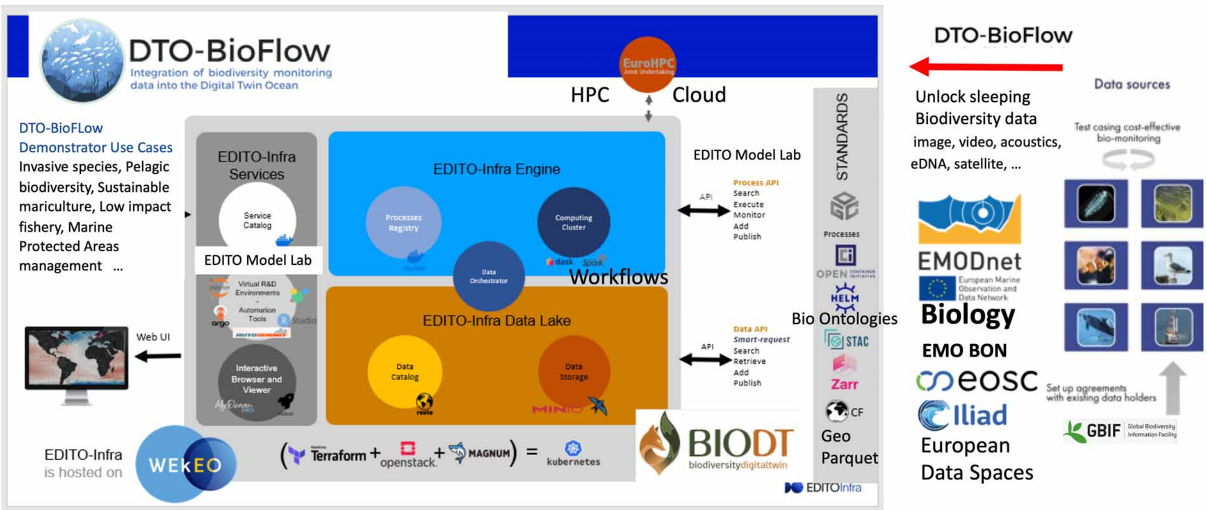


Figure 1 DTO-BioFlow Biodiversity data into Digital Twins of the Ocean.

The digital twin processing will be supported through the EDITOInfra data lake and the EDITO digital twin engine. Further support for High Performance Computing (HPC) model processing for digital twins is carried over from the Biodiversity Digital Twin (BioDT) Project[58], a land-based digital twin. Partners who are jointly involved in BioDT as well as DTO-BioFlow are responsible for transferring HPC-processing services into the marine biodiversity digital twins. This reuses existing HPC infrastructure services such as the LEXIS Project[59] combined with AI/Machine learning Graphics Processing Unit (GPU) clusters, workflow engines, and languages such as the common workflow language (CWL). Both projects take advantage of processing and service

[52] https://emodnet.ec.europa.eu/en/biology
[53] https://www.eurobis.org
[54] https://www.gbif.org
[55] https://geoparquet.org
[56] https://stacspec.org/en
[57] https://zarr.readthedocs.io/
[58] https://biodt.eu
[59] http://lexis-project.eu

support through the European Open Science Cloud (EOSC)[60]. Biodiversity digital twins will also be able to exploit the EDITO ModelLab services for various ocean models and relevant Virtual Research Environments like Jupyter Notebook Hubs[61] and the Galaxy platform[62].
The approach will further be harmonised with the interoperability architectures of supporting initiatives like the Iliad consortium, the United Nations Ocean Decade Initiative[63], DITTO (Digital Twins of the Ocean)[64] with the Turtle Interoperability Framework (TIF)[65]. The basis of the levels of legal, organisational, semantic, and technical interoperability of the European Interoperability Framework (EIF) will be followed.

Acknowledgements
We acknowledge the efforts of the entire DTO-Bioflow "Integration in DTO Infrastructure" team: Clément Villeviere, Elena Lazovik, and Jeroen Broekhuijsen from TNO (The Netherlands); Frederic Leclercq, and Bart Vanhoorne from VLIZ (Belgium); as well as Tuomas Rossi and Jarmo Makela from CSC-IT Center for Science LTD (Finland).

Poster session

[60] https://eosc-portal.eu
[61] https://jupyter.org/hub
[62] https://usegalaxy.org
[63] https://oceandecade.org
[64] https://ditto-oceandecade.org
[65] https://doi.org/10.5194/egusphere-egu23-15620

IPMA's Marine Data Portal

Tanya Silveira (tanya.silveira@ipma.pt), Mafalda Carapuço (mafalda.carapuco@ipma.pt), Elisabete Dias (elisabete.dias@ipma.pt), Conceição Santos (onceicao.santos@ipma.pt)

Portuguese Institute for the Sea and Atmosphere (Portugal)

Introduction

The Portuguese Institute for the Sea and Atmosphere (IPMA) detains a rich and valuable collection of marine data for its Atlantic coast, and maintains important marine monitoring programmes that have been nourishing long-time data series, important for understanding long-term processes and trends, especially for climatic studies.

IPMA is developing a marine spatial data infrastructure for the acquisition, aggregation and provision of oceanographic data, information and services, in order to enhance research, monitoring and data dissemination of the Atlantic Ocean, the "We are Atlantic" data platform. The main objective is to create a quality and reliable data repository that can effectively facilitate the storage, access and interoperability of geographic data and spatial information on the marine environment.

The infrastruture is being developed in the scope of the Atlantic Observatory – Data and Monitoring Infrastructure project, in partnership with Norway and Icelandic partners, funded by the EEA Grants.

Data acquisition platforms

The Ocean observation relies mostly on data acquisition conducted onboard research vessels during dedicated multidisciplinary surveys, and on long term monitoring at fixed stations. It includes seawater properties, geophysical and fisheries data, as well as other metocean parameters.

IPMA is part of major European and international ocean observation initiatives, and has been enlarging its observation network, for example with sea bottom observatories in the scope of EMSO-ERIC, and with Argo floats, in the scope of the EuroGOOS, contributing to the global Argo network.

Products and services

IPMA has national responsability in the areas of meteorology and climate monitoring, and provides a number os services that must be disseminated adequately, ensuring, for example that warnings are issued and widespread whenever adverse meteorological events are predicted or observed.

Likewise, IPMA is responsible for the National Programme for Biological Sampling, that assures data collection, management and analysis related to studies on fish biology, supporting the scientific advice for fish stock and fisheries management under the Common Fisheries Policy. Most information contribute to the Marine Strategy Framework Directive which aims to achieve a good marine environmental status.

Digital Infrastructure

The digital infrastructure is currently being developed and will include a repository, metadata catalogue and web interface (Figure 1). It will rely on standard metadata and data services, based on a common data policy and following the Findable, Accessible, Interoperable and Reusable principles. In line with the current European directives, the "We are Atlantic" data portal will be a user-driven initiative, facilitating the access of economic players and the citizens to information

related with the Atlantic Basin. It will provide data products and services for multiple stakeholders and end users within the Blue Economy and Blue Growth sectors, whilst responding to the major EU frameworks and strategies. It will be interoperable with the major ocean data management initiatives, such as GOOS, EMODnet and SeadataNet.
The present work aims to present IPMA's metocean data and services chain value, concentrated on the "We are Atlantic" digital infrastructure that will act as a single storage and access datapoint, for both data collectors and end-users seeking information and services associated to the ocean.

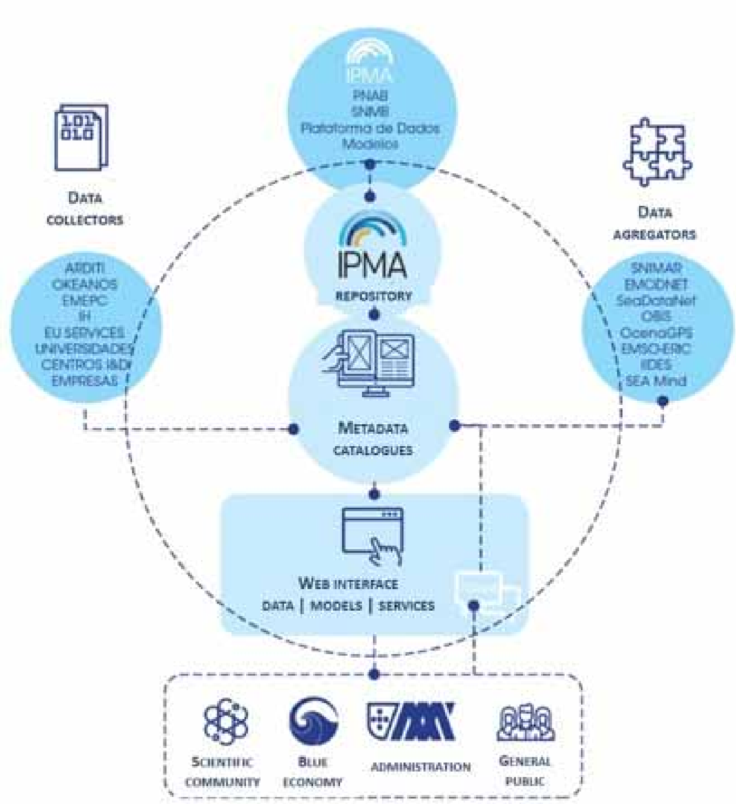


Figure 1 The "We are Atlantic" Digital Infrastructure general workflow.

Capacity development in digital twins of the ocean: Interoperability, standards and best practices

Bente Lilja Bye[1] (bente@blb.as), Sigmund Kluckner[2] (sigmund@kluckner.eu), Jay Pearlman[3] (jay.pearlman@ieee.org), Pauline Simpson[4] (p.simpson@unesco.org), Piotr Zaborowski[5] (pzaborowski@ogc.org)

[1]BLB (Norway)
[2]IEEE France Section (Austria)
[3]IEEE France Section (USA)
[4]UNESCO-IOC (Belgium)
[5]OGC (Poland)

Innovative techniques and new technologies are required to achieve interoperability in digital twins of the ocean, particularly at regional and global scales. New technologies, underpinned by best practices and standards, call for new skills to be addressed by capacity sharing and capacity development throughout the ocean value chain - every step along the pipeline from sensors, data provisions, data preparation, to analytics, ML, AI, and visualization and end user presentation.

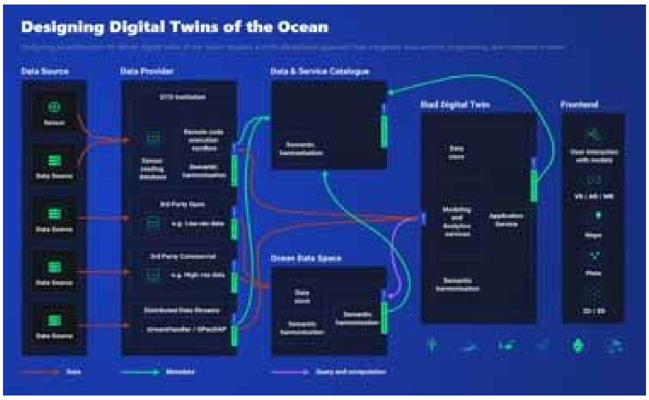


Figure 1 The Iliad logical architecture shows the development pipeline/value chain from data collection on the left to user interface on the right and indicate some of the technologies that are required to build digital twins of the ocean.

Iliad federated system of systems approach.

Iliad has a systems of systems approach to digital twins. The architecture is shown in Figure 1. This architecture promotes interoperability while accepting constructs needed for thematic and local digital twins. With the local twins responding to the needs of a specific environment and application, the details of their structure cannot be dictated from above, but these local twins can adopt a simple foundation for interfacing with a system of system framework. Figure 2 illustrates a cross-cutting foundation system architecture as well as a cloud/HPC environment with the digital twin module interfaces glued together through common standards and best practices.

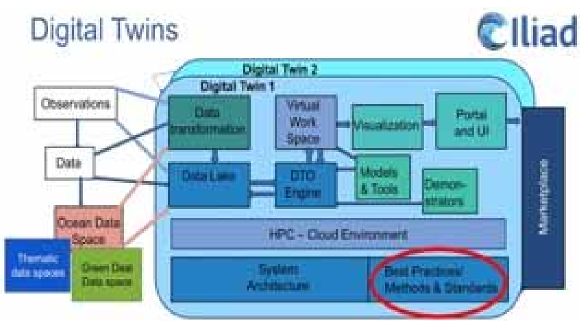


Figure 2 Each individual thematic/local twin has an architecture that is interoperable with others and the overall Iliad logical architecture. In this figure, methods, standards, and best practices are shown as part of the overall twin architecture.

Overlaying the entire digital twin framework of Figure 2 is an approach for capacity development where support is provided to individual twins to further their ability to integrate into the systems of systems. This general approach is in the designs of Iliad, EDITO and, on a larger scale, the global designs of the DITTO Ocean Decade Programmes. Each of these draws from the OGC, ISO, IEEE or CEN for standards and the Ocean Best Practices System for common methods.

Capacity Building in Digital Twin Development

The common architecture of the digital twins, combined with the application of standards and best practices and the creation of new methods provide a coherent basis for capacity development and training for a wide range of professionals in the ocean AND in the digital twin field.

Oceanographers: Use DT to learn about ocean systems, interaction, etc (example oil spill); Can be used in universities for what-if scenarios, different approaches to data, and to train people on improved decision making (example aquaculture and disease).

Software Engineers/DT Developers: Reuse components as libraries and service: need to understand, but don't need to produce all code from scratch; Understand data exchange frameworks, bidirectional in relation to the tools and services; Put 'generic' technology standards in place within/for the simulation; Test out different models, see their values and challenges.

Others: Understand best practices around ethics, visualization, etc. for a specific domain and how they can be different from 'textbooks' Understand the approaches to decision making and societal impacts.

Iliad is aiming at developing new standards and best practices for the ocean and digital twin communities, in order to support work at the intersection of these fields. The strength of the standards lies in the usability leveraged by their wide adoption and tools support. One example is the Application Package concept combined with refurbished OGC Processes extended with the specification of the service description and products compatible with CF conventions and based on STAC draft standard, already widely adopted. The other one is the definition of the Ocean Information Model, gluing observation models with coverages' representations popular within the marine community and adding new, web-friendly format support and formal API definition. This includes an understanding of both standards and best practices. Then, the major work is to test and prove the fit-for-purpose of the proposed solutions in particular cases, supporting their implementations and building persistent demos with tooling support. The implementation is done following capacity development roadmaps tailored for each pilot along internal/external perspectives, see Figure 3.

Figure 3 The different components and processes of ILIAD forming the basis for the method of capacity development needs assessment and creation of roadmaps for pilot digital twins and digital twin of the ocean framework.

Finally, we build a curriculum of the learning materials that could be used in Iliad Academy and exploited beyond the project in persistent, maintained, and open access manuals, guidelines and working examples.

Blue-Cloud Workbenches for the creation of highly qualified physical and chemical data collections

Alessandra Giorgetti[1] (agiorgetti@ogs.it), Simona Simoncelli[2] (simona.simoncelli@ingv.it), Nydia Catalina Reyes Suarez[1] (nreyessuarez@ogs.it), Enrico Baglione[2] (enrico.baglione@ingv.it) and the Workbenches 1 and 2 teams

[1]Istituto Nazionale di Oceanografia e Geofisica Sperimentale - OGS (Italy)
[2]Instituto Nazionale di Geofisica e Vulcanologia - INGV (Italy)

Introduction

Blue-Cloud 2026 will develop and test a series of thematic Workbenches (WBs) for selected physical, chemical and biological Essential Ocean Variables (EOVs). Although the WBs for Blue-Cloud 2026 are new, their establishment will benefit from the experience gained in the Blue-Cloud (BC) pilot project, which generated a pilot cyber platform enabling researchers to discover and access a wide range of multidisciplinary datasets and products managed in Blue Data Infrastructures (BDIs), analytical services and computational facilities essential for blue science. With the Blue-Cloud Open Science Platform, a collaborative environment has been developed in which various services are available. The VRE orchestrates the computational and analytical services in specific integrated and managed applications that mobilise and make available important additional data resources on top of the data managed by Copernicus Marine Service (CMEMS) and EMODnet in collaboration with federated Research Infrastructures (Ris). The innovation potential of BC is explored and unlocked through the dedicated WBs as toolboxes developed in collaboration with senior marine data managers to create efficient workflows that demonstrate the power of the BC Open Science platform for collaborative data analysis, integration and validation. Federated BC services made available to researchers for the discovery of and access to data resources (DD&AS) and for the creation, execution and sharing of analytical processes are expanded and optimised for the orchestration of WBs focused on a set of selected Essential Ocean Variables (EOVs).

Essential Ocean Variables (EOVs) Workbenches

The aim is to develop, validate and document new analytical BC big data WBs that can be adopted by EMODnet, CMEMS and selected RIs to produce a series of harmonised and validated data collections at regular intervals for a selection of EOVs in the fields of physics, chemistry and biology. The WBs cover a wide range of topics related to marine ecosystems and data types (physical, biological, chemical, in situ and remote data, etc.) that are also relevant for multi-disciplinary research. The existing gap in EOVs in the fields of physics, chemistry and biology will be filled with three WBs that will generate new data collections that will be available to EU operational services. Here we will focus on the physics and eutrophication WBs, whose data sources are multidisciplinary and fit-for-integration from selected research and data infrastructures that provide validated and structured data collections (e.g. EMODnet, CMEMS, SeaDataNet).

- **EOV Workbench for physics: temperature and salinity**. A cloud-based workflow is currently being developed to create harmonised and validated data collections for temperature and salinity, integrating multiple datasets from different EU and non-EU data infrastructures such as CMEMS, SeaDataNet and NOAA World Ocean Database. The integration of the datasets relies on the interoperability of the sources BDIs and Ris and the metadata associated with the temperature and salinity observations, which need to be categorised into a common schema to identify and handle potential duplicate observations and finally be linked to the new derived EOV product. The merged dataset is used to calculate gridded climatologies and demonstrate their added value compared to climatologies obtained from the individual input datasets. The WB will be developed and tested for the Mediterranean Sea with the aim of extending it to the global ocean. The WB development will rely on the synergy with the international IQuOD (https://www.iquod.org/) initiative, organized by the oceanographic community that involves experts in data quality and management. In addition, the WB pipeline will be available and extensible for adoption in other big data infrastructures.

- **EOV Workbench for eutrophication: chlorophyll, nutrients and oxygen.** The aim of this workbench is to create harmonised and validated EOV data collections for chlorophyll, nutrients and oxygen by integrating multiple datasets from different EU and non-EU data infrastructures such as Copernicus Marine Service, EMODnet Chemistry and the World Ocean Database as well as the EU RIs such as Argo, ICOS and SOCAT. The approach will be the same as for the physical workbench. Interoperability services of data infrastructures, common vocabularies and brokering services are used to enable the aggregation and harmonisation of datasets, with great attention paid to semantic aspects. EMODnet Chemistry and Copernicus Marine Service experts are working together to agree on a set of quality control procedures to be applied on top of the existing ones. A specific protocol will be jointly developed to identify and handle possible duplicate observations. The Workbench will be developed and tested for the Northeast Atlantic with the aim of extending it to the global ocean.

Figure 1 presents a generalized workflow for the physical and eutrophication WBs. The different datasets (SeaDataNet, CMEMS, EMODnet and WOD) will be harvested by the DD&AS, then subsets are compiled and metadata harmonised so that an integrated data set is created. It is expected that the BEACON tool developed by Maris B.V. will be used extensively for rapid access to large quantities of climate data and their subdivision. The output will be pushed into the Virtual Lab for analytics and additional checking (detection of duplicates, additional QC and validation), using different methods depending on the requirements of each WB. Finally, BC data collections are created and published via the webODV (https://webodv.awi.de/) for scientific research, plot creation and extraction.

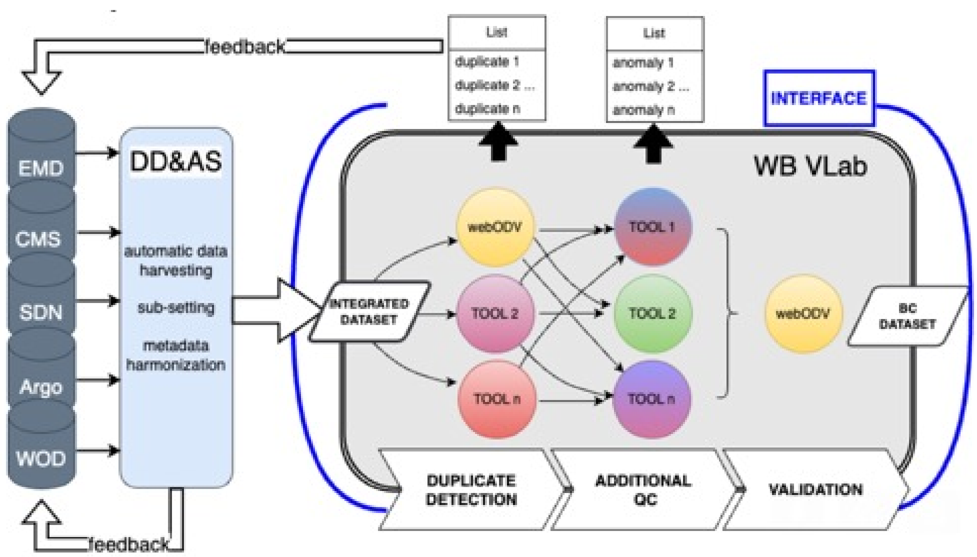


Figure 1 General physics and eutrophication workbench workflow.

Marine observatories match Internet of Everything

Loïc Lagadec[1] (loic.lagadec@ensta-bretagne.fr), Charbel Aoun[2] (charbel.aoun@icam.fr), Ciprian Teodorov[1] (ciprian.teodorov@ensta-bretagne.fr)

[1]ENSTA Bretagne (France)
[2]ICAM (France)

Marine observatories

Marine data are massively heterogeneous. Consequently, collecting data requires specific sensors, formats, and tools for consistency checking, exchange, aggregation, and combination. Furthermore, there are many users and uses, and their numbers are expected to increase even further. This will lead to major changes, one should be ready for, what means being able to capture what-why-how-then so that to play what-if scenarios with impact prediction, for well-aware decision making.

While traditional approaches have long focused on end-point solutions, ever-increasing complexity makes this approach not further practicable. Global requirements (e.g., interoperability) are now central when designing a system. Such a reversal of a process is a real mind shift. Though, there are many ways to express the requirements. Natural language-based specifications (NLB) have been in common use for years, but they suffer from intrinsic limitations: a text is likely to become out of sync with the existing system over time, or to be ambiguous, especially if the underlying context is omitted, keeping persistent no longuer valid constraints.

Instead, in model-based engineering, models capture a domain and focus on features of key interest. The model acts as the sole source of truth, avoiding misunderstandings and promoting collaborative work between experts (scientific, technical, commercial, decision-makers, etc.). Models support transformations: refinement (from abstract to practical), purpose-oriented transformation (e.g., testing, deployment, etc.). Data are massively heterogeneous. Consequently, collecting data requires specific sensors, formats and tools for consistency checking, exchange, aggregation and combination. What's more, users and usages yet are, but will be even more numerous.

This will lead to major changes, one shoud be ready for, what means being able to capture what-why-how-then so that to play what-if scenarios ahead of schedule, with impact measurement, for decision making to be relevant.While, traditional approaches have long focused on end-point solutions, ever-increasing complexity makes this approach increasingly inaccessible. Constraints must be taken as starting point, then converted into requirements (e.g. interoperability).

Ways of expressing requirements are many. Natural language-based specifications (NLB) have been in common use for years (still present in existing systems), but they suffer from intrinsic limitations: a text in likely to become out of sync with the existing system over time, and can be ambiguous (e.g., the underlying context, likely to change over time, may be omitted). Instead, in model-based engineering, models capture a domain and focus on features of key interest. Models act as sole source of truth, avoiding misunderstandings during collaborative work among experts (technical, commercial, decision-makers, etc.); they support transformations: abstraction (from practical to general), refinement (from general to pratical), purpose-oriented model-to-model transformation (e.g., testing).

IoT and IoE

In marine observatories, the services, provided to users by the infrastructure, often appear as compositions of data acquisition through devices, routines, and visualization. The Internet of Everything (IoE) (https://www.cisco.com/c/dam/en_us/about/business-insights/docs/ioe-value-index-faq.pdf) is based on the same four pillars when modeling complex systems: objects, data, processes and people. As so, our claim is that marine observatories are a typical IoE use case.

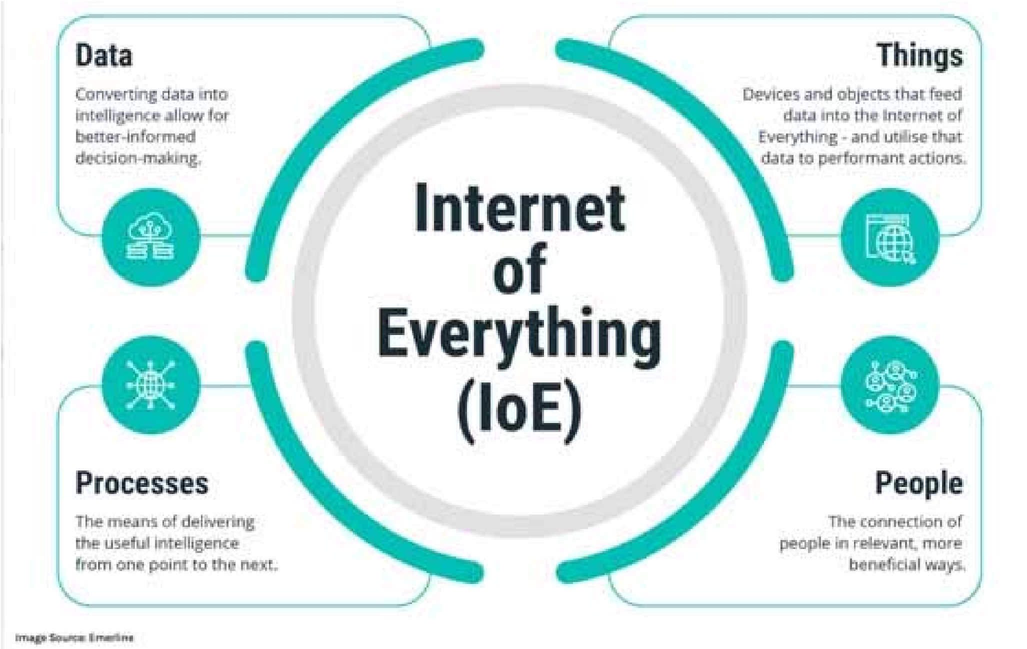


Figure 1 IoE overview.

Acquisition systems (Things: various sensors, with constraints such as hardness) are now moving towards low-end embedded devices (low cost, while retaining computing power). These devices are mission-oriented (People), capable of handling complex data structure (with properties such as ACID[66]), and communicating through various protocols with a balance between long-range / low-power / high bandwidh objectives, and are gradually becoming SCADA[67] compliant (Process: from things to users and back, with data acquisition and analysis, then configuration). This goes far beyond simple IoT (Internet of Things) perspective but fits to IoE.

Contributions

We developed an IoE framework [Aoun et al., 2022] (Figure 2) to model marine observatories. The logical and physical components used in these observatories provide data format and exchange procedures between the various environmental devices (e.g., smart sensors, data fusion servers) or user roles. Figure 3 illustrates a compliant example (underwater sensors network with environmental constraints [Aoun et al., 2017]). Deployed sensors (hydrophones) record underwater activity then transfer it to more advanced components. The model provides captures the technical infrastructure that captures data, converts them into information, and triggers the actuators (e.g., motors). This allows one to properly dimension the infrastructure throught stressing the model against assets (Data), usage (People) and how-to (Process).
Based on this model, we focus on the generation of the corresponding simulation code to help reduce the complexity and time of the design activity, by providing agile refactoring capabilities.

[66] Atomicity, Consistency, Isolation, Durability

[67] Supervisory Control and Data Acquisition

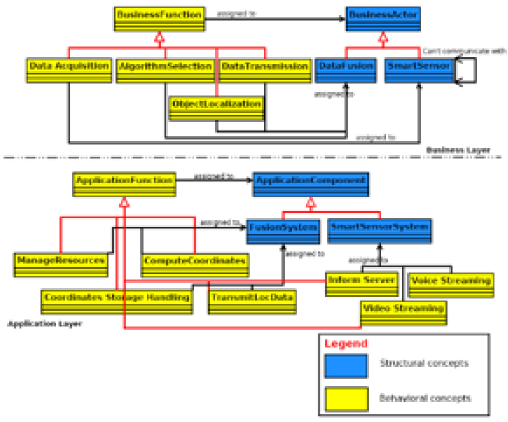


Figure 2 IoE framework.

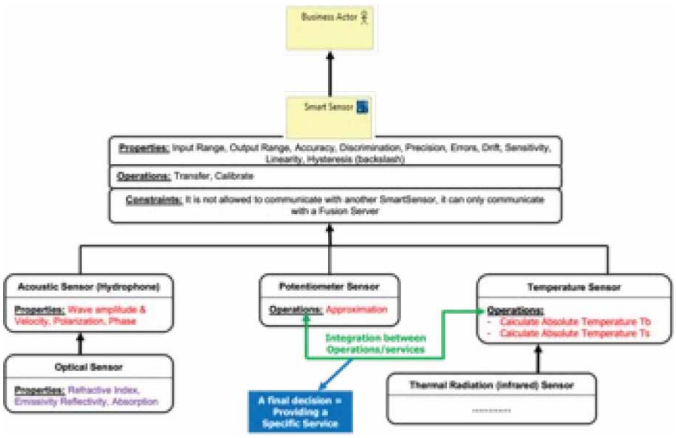


Figure 3 Modelling sensors network.

References

Aoun C.G., Lagadec L., Habes M., (2022). *An Extended Modeling Approach for Marine/Deep-Sea Observatory*. In: Hassanien, A.E., Rizk, R.Y., Snášel, V., Abdel-Kader, R.F. (eds) The 8th International Conference on Advanced Machine Learning and Technologies and Applications (AMLTA2022). AMLTA 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 113. Springer, Cham. https://doi.org/10.1007/978-3-031-03918-8_42

Aoun C.G., Lagadec L., Champeau J., Moussa J. and Hanna E., (2017). *A High Abstraction Level Constraint for Object Localization in Marine Observatories*, International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2017, pp. 605-611, https://doi.org/10.1109/CSCI.2017.105

QUADERNI di GEOFISICA

ISSN 1590-2595 http://istituto.ingv.it/it/le-collane-editoriali-ingv/quaderni-di-geofisica.html/

I QUADERNI DI GEOFISICA (QUAD. GEOFIS.) accolgono lavori, sia in italiano che in inglese, che diano particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari che necessitano di rapida diffusione nella comunità scientifica nazionale ed internazionale. Per questo scopo la pubblicazione on-line è particolarmente utile e fornisce accesso immediato a tutti i possibili utenti. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi. I QUADERNI DI GEOFISICA sono presenti in "Emerging Sources Citation Index" di Clarivate Analytics, e in "Open Access Journals" di Scopus.

QUADERNI DI GEOFISICA (QUAD. GEOFIS.) welcome contributions, in Italian and/or in English, with special emphasis on preliminary elaborations of data, measures, and observations that need rapid and widespread diffusion in the scientific community. The on-line publication is particularly useful for this purpose, and a multidisciplinary Editorial Board with an accurate peer-review process provides the quality standard for the publication of the manuscripts. QUADERNI DI GEOFISICA are present in "Emerging Sources Citation Index" of Clarivate Analytics, and in "Open Access Journals" of Scopus.

RAPPORTI TECNICI INGV

ISSN 2039-7941 http://istituto.ingv.it/it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html/

I RAPPORTI TECNICI INGV (RAPP. TEC. INGV) pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico come manuali, software, applicazioni ed innovazioni di strumentazioni, tecniche di raccolta dati di rilevante interesse tecnico-scientifico. I RAPPORTI TECNICI INGV sono pubblicati esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi.

RAPPORTI TECNICI INGV (RAPP. TEC. INGV) publish technological contributions (in Italian and/or in English) such as manuals, software, applications and implementations of instruments, and techniques of data collection. RAPPORTI TECNICI INGV are published online to guarantee celerity of diffusion and a prompt access to published data. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

MISCELLANEA INGV

ISSN 2039-6651 http://istituto.ingv.it/it/le-collane-editoriali-ingv/miscellanea-ingv.html

MISCELLANEA INGV (MISC. INGV) favorisce la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV. In particolare, MISCELLANEA INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc. La pubblicazione è esclusivamente on-line, completamente gratuita e garantisce tempi rapidi e grande diffusione sul web. L'Editorial Board INGV, grazie al suo carattere multidisciplinare, assicura i requisiti di qualità per la pubblicazione dei contributi sottomessi.

MISCELLANEA INGV (MISC. INGV) favours the publication of scientific contributions regarding the main activities carried out at INGV. In particular, MISCELLANEA INGV gathers reports of scientific projects, proceedings of meetings, manuals, relevant monographs, collections of articles etc. The journal is published online to guarantee celerity of diffusion on the internet. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

Coordinamento editoriale
Francesca DI STEFANO
Istituto Nazionale di Geofisica e Vulcanologia

Progetto grafico
Barbara ANGIONI
Istituto Nazionale di Geofisica e Vulcanologia

Impaginazione
Barbara ANGIONI
Patrizia PANTANI
Massimiliano CASCONE
Istituto Nazionale di Geofisica e Vulcanologia

Istituto Nazionale di Geofisica e Vulcanologia
Via di Vigna Murata, 605
00143 Roma
tel. +39 06518601

www.ingv.it

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA