Vol. 62

Bollettino di Geofisica teorica ed applicata

An International Journal of Earth Sciences

IMDIS 2021
International Conference on Marine Data and Information Systems

12-14 April, 2021
Online

Book of Abstracts

Guest Editors: Michèle Fichaut, Vanessa Tosello, Alessandra Giorgetti

SUPPLEMENT 1

imdis

OGS National Institute of Oceanography and Applied Geophysics

free colour printing

Bollettino di Geofisica
teorica ed applicata

An International Journal of Earth Sciences

Scope of the journal

The "Bollettino di Geofisica Teorica ed Applicata" is an international journal dedicated to the publication of original papers dealing with Deep Earth Geophysics, Near Surface Geophysics, Exploration Geophysics, Borehole Geophysics, Geodynamics and Seismotectonics, Seismology, Engineering Seismology, Geophysical Modelling, Geodesy, Remote Sensing, Seismic and Geodetic Networks, Oceanography, and their application in the fields of Energy, Natural Resources, Environment and Climate, Policies and Regulations, Risk and Security, Technological Development.
Issues of about 80 pages are published quarterly. Papers dealing with Solid Earth and Oceanography are grouped into two parts, A and B, respectively. The journal is published by OGS (Istituto Nazionale di Oceanografia e di Geofisica Sperimentale), Trieste, Italy.

The "Bollettino di Geofisica Teorica ed Applicata" is indexed in: Science Citation Index Expanded (also known as SciSearch®), SCOPUS, Current Contents® /Physical Chemical and Earth Sciences, Journal Citation Reports/Science Edition, Cabell's International.

The "Bollettino di Geofisica Teorica ed Applicata" is included in the product of EBSCO Publishing Inc.

SUBSCRIPTIONS 2021

	Italy	Other countries
Annual subscription rate	€ 65	€ 130
Back volumes	€ 75	€ 150
Separate issue	€ 22	€ 45
Back issue	€ 25	€ 50

All prices include post and packaging. For private persons (excluding institutes, laboratories, libraries, etc.) subscribing directly, the subscription rates are reduced by 20%.

Claims for missing numbers will be honoured only if received within 100 days of normal delivery date.

Subscription orders, changes in address and other business correspondence should be sent to the following address:

Bollettino di Geofisica Teorica ed Applicata
c/o Istituto Nazionale di Oceanografia
e di Geofisica Sperimentale
Borgo Grotta Gigante, 42/c
34010 Sgonico, Trieste, Italy
e-mail: bgta@ogs.trieste.it
http-server: bgta.eu

Publication B.A.M. exempt
Article 4 - 1° comma n. 6 - D.P.R. n. 627 of 6/10/87

Vol.62 – SUPPLEMENT 1

Bollettino di Geofisica
teorica ed applicata

An International Journal of Earth Sciences

IMDIS 2021

International Conference on Marine Data and Information Systems

12-14 April, 2021 - Online

Organised by
NIOZ
jointly with IFREMER, OGS, IOC/IODE, in the frame of SeaDataCloud project
(EU H2020 - Grant Agreement 730960).

OGS
National Institute of Oceanography and Applied Geophysics

TABLE OF CONTENTS

Posters

Posters

Posters

Posters

SESSION PRODUCTS:

Data products, information and knowledge

- Quality Check procedures and data consistency assessment for historical data collections for climatic studies
- Machine learning, Artificial Intelligence and application to data collection and data quality check
- Regional and global data collections and marine data aggregations
- Validation of data products and uncertainty estimation
- Distribution and publication of data products
- Data products for multiple stakeholders/end users including the Blue Economy/Blue Growth/Maritime Spatial Planning
- Data integration and evaluation of data portal solutions
- Transfer of technologies and knowledge
- Systems and products for outreach and dissemination, including Ocean Literacy
- Metadata analysis
- Data (and metadata) rescue and integration into data centre systems

ORAL PRESENTATIONS

World Ocean Database in 3D: Development, Dissemination, Deliverables

Alexey Mishonov, NCEI/NOAA (USA), alexey.mishonov@noaa.gov
Tim Boyer, NCEI/NOAA (USA), tim.boyer@noaa.gov
Ricardo Locarnini, NCEI/NOAA (USA), ricardo.locarnini@noaa.gov
Hernan Garcia, NCEI/NOAA (USA), hernan.garcia@noaa.gov
Dan Seidov, NCEI/NOAA (USA), dan.seidov@noaa.gov
Melissa Zweng, NCEI/NOAA (USA), melissa.zweng@noaa.gov
Christopher Paver, NCEI/NOAA (USA), christopher.paver@noaa.gov
Olga Baranova, NCEI/NOAA (USA), olga.baranova@noaa.gov
James Reagan, NCEI/NOAA (USA), james.reagan@noaa.gov
Igor Smolyar, NCEI/NOAA (USA), igor.smolyar@noaa.gov
Alexandra Grodsky, NCEI/NOAA (USA), Alexandra.grodsky@noaa.gov

The World Ocean Database (WOD) is a collection of scientifically quality-controlled ocean profile and plankton data that includes measurements of temperature, salinity, oxygen, phosphate, nitrate, silicate, chlorophyll, alkalinity, pH, pCO_2, TCO_2, Tritium, Δ^{13}Carbon, Δ^{14}Carbon, Δ^{18}Oxygen, Freon, Helium, Δ^{3}Helium, Neon, and plankton. WOD - the world's most extensive collection of ocean profile data, which is updated four times per year and available without restriction.

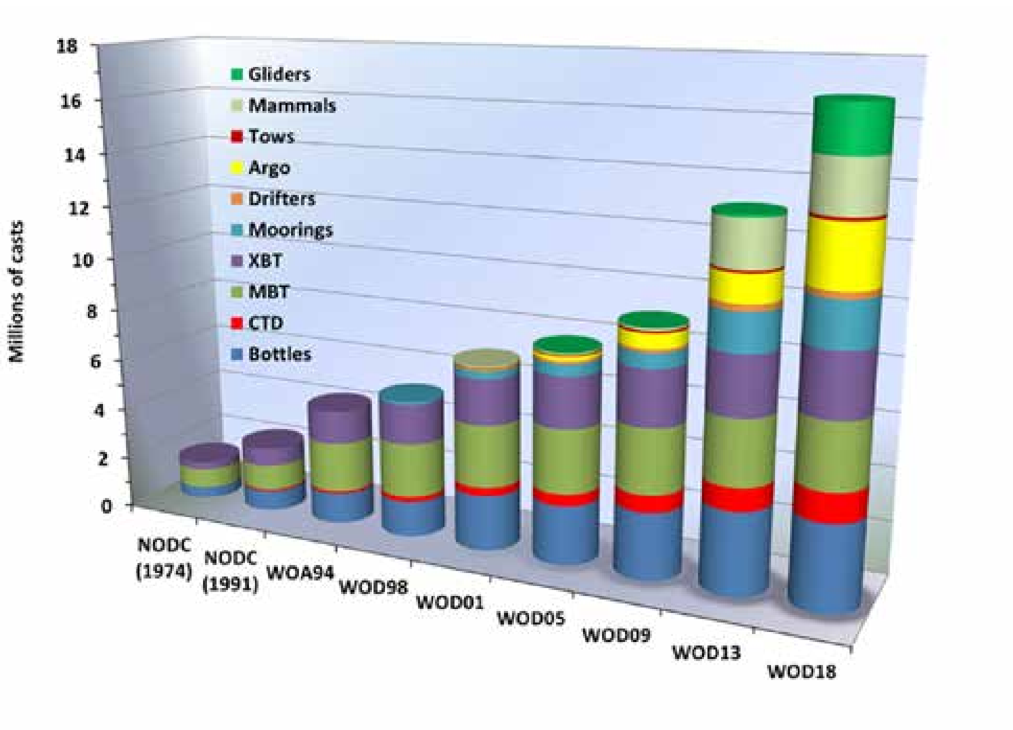


Figure 1: WOD development: in situ seawater Temperature data holding growth

Development. The work on the WOD began in 1992 by Sydney Levitus. Six major versions of WOD has been released since then in 1998, 2001, 2005, 2009, 2013, and 2018. The WOD team at NOAA has been ingesting data from multiple countries, many institution, and various platforms and instruments. The WOD can be considered as the final step in gathering oceanographic profile data and preparing them for public dissemination. The inventors, oceanographers, and engineers who conceived, designed, and tested the oceanographic instrumentation and measurement techniques are responsible for the plethora and variety of oceanographic data. The primary investigators, marine technicians, ship's crew, and volunteers who made and continue to make many of the oceanographic measurements, often under harsh conditions, are responsible for the quality and quantity of the oceanographic data. The institutions, which maintain the platforms and the projects, which plan, fund, and execute the field campaigns and operational ocean monitoring are responsible for the spatial and temporal coverage of the oceanographic profile data. Finally, the data managers are responsible for the preservation and reusability of the data. This is a vast network, maintained and updated over time, which should receive the credit for the aggregated WOD. Every cast, which in essence is a central granule of WOD, contains (when supplied) information on the instrumentation, platform, project, institution, and data management entity. The archive at NCEI and those who populate and maintain it also deserve credit for the continual availability of historical oceanographic data. Finally, international organizations such as the Intergovernmental Oceanographic Commission's (IOC) International Oceanographic Data and Information Exchange (IODE) and the World Data System (WDS) for Oceanography should be credited for creating and facilitating a global culture of data exchange and preservation. Over 40+ years of development more than 17 million casts of oceanographic parameters has been collected, quality controlled and uniformly formatted. The WOD makes these data available for all to work with confidence and convenience. Figure 1 illustrates the WOD seawater temperature data holding grows over the time.

Dissemination. The WOD and the products based on it go through different stages of preparation and dissemination techniques, which reflects the technological evolution in oceanographic observations and processing. It started from 8-tracks mainframe tapes to HD-floppy disks to CDs, to DVDs, and, now completely moved to the web and preparing to be finally transferred to the cloud for being accessible in real time. Currently, the entire collection of data in the WOD is accessible via WODSelect web-portal (https://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html) where data selection can be made based in different user-defined criteria. The data selected by a user request are prepared automatically and can be downloaded from a NOAA server.

Deliverables. Aside from quality controlled and uniformly formatted oceanographic data, WOD is the foundation for several stand-alone products. The major product is the World Ocean Atlas – a set of global climatological fields of major oceanographic variables – temperature, salinity, etc., calculated on two spatial grids at the 33 (before 2013) and 102 (after 2013) standard depth levels with one- (before and after 2013) and quarter-degree (after 2013) spatial resolutions. These climatologies calculated based on entire data collection as well as on decadal sub-sets. The top product based on WOD and WOA is Global Ocean Heat and Salt Content anomalies accompanied by Seal level changes data. These assessments are updated quarterly. For several regions of the world ocean where data density is sufficient for data analyses on one-tenth-degree grid, the high regional climatologies created. As for now, there are seven regional climatologies has been prepared for Southwest North Atlantic, Greenland-Iceland-Norwegian Seas; Northeast Pacific; Northern North Pacific; Northwest Atlantic; Arctic; East Asian Seas; and Gulf of Mexico.

Time-Series Data Products from the Australian Integrated Marine Observing System

Marton Hidas, Australian Ocean Data Network (Australia), marty.hidas@utas.edu.au
Eduardo Klein Salas, Australian Ocean Data Network (Australia), eduardo.kleinsalas@utas.edu.au
Guillaume Galibert, Australian Ocean Data Network (Australia), guillaume.galibert@utas.edu.au
Sebastien Mancini, Australian Ocean Data Network (Australia), sebastien.mancini@utas.edu.au
Craig Steinberg, Australian Institute of Marine Science (Australia), C.Steinberg@aims.gov.au
Bernadette Sloyan, Commonwealth Scientific and Industrial Research Organisation (Australia), bernadette.sloyan@csiro.au
Michael Hemming, University of New South Wales (Australia), m.hemming@unsw.edu.au
Ana Redondo Rodriguez, South Australian Research and Development Institute (Australia), ana.redondorodriguez@sa.gov.au
Miaoju Chen, Commonwealth Scientific and Industrial Research Organisation (Australia), Miaoju.Chen@csiro.au
Simon Spagnol, Australian Institute of Marine Science (Australia), S.Spagnol@aims.gov.au
Peter Jansen, Commonwealth Scientific and Industrial Research Organisation (Australia), Peter.Jansen@csiro.au

Australia's National Mooring Network, a facility of the Integrated Marine Observing System (IMOS), has been operating moored instruments at over 60 coastal sites for up to 13 years. At the seven National Reference Stations, water-column profiles and water samples are also obtained with up to a monthly frequency. From these, a wide range of biogeochemical analyses are made.

All the IMOS observations are available to the public via the Australian Ocean Data Network (AODN) Portal (https://portal.aodn.org.au). However, as the data are served in separate files per instrument and deployment, visualising and analysing the complete time series from a single mooring site requires the downloading hundreds of files, and significant technical expertise to combine them. This is limiting the use and impact of the data in the broader user community.

To address this problem, a Working Group has been set up, comprising technical experts from the IMOS moorings facilities, members of the user community, and AODN data scientists and managers. The group has identified a range of data products required by distinct user groups and use cases. While there are numerous possibilities for high-level products (e.g. mixed-layer depth), the group has been focusing on developing foundational aggregations of oceanographic parameters (e.g. temperature, salinity, currents) to serve broader use cases. The higher-level products will then be easier to build from these.

An initial year-long project was run in 2019-20 to design and implement these foundational products. The scope of the project was defined in an overall business case document, reviewed and accepted by the key stakeholders. For each individual product, the development process began with a more detailed, product-specific business case document, followed by an implementation plan. After the business case was accepted, the implementation plan, along with initial code and prototype files, were reviewed by a small team of technical experts from each regional moorings group. Where feasible, feedback from the review was implemented immediately. However, in order to keep the overall project on track, more significant changes were instead noted for future development. Finally, the first version of the product was created, published, and advertised to the user community.

This project delivered five products, addressing the initial challenges of combining many data sets from individual moored instruments, deployed at varying depths and using different sampling regimes. The products were built up in three levels:

1. Aggregating all the original single-instrument files from a mooring site, without modification;
2. Binning the aggregated time series into a common hourly time interval; and
3. Interpolating onto a site-specific set of depths (where there are enough measurements through the water column), resulting in a time-depth gridded product.

In order to simplify the initial code design, we split the input data into two categories: 1) "non-velocity" parameters, which are always single-point time series; and 2) water velocity observations, measured on most IMOS moorings using Acoustic Doppler Current Profilers (ADCPs), and thus inherently two-dimensional. In combination with the above levels, this results in a total of six possible products. However, the last of these (velocity gridded) did not fit into the scope of the initial project.

The five products have been published on the AODN THREDDS Catalogue (http://thredds.aodn.org.au/thredds/catalog/catalog.html). The hourly binned and gridded products are also available via the AODN Portal. The Python code to generate the products is open source (https://github.com/aodn/python-aodntools). All the products will be updated regularly to incorporate new data.

We present an overview of these products, demonstrate how they can be accessed and used, and describe plans for future work.

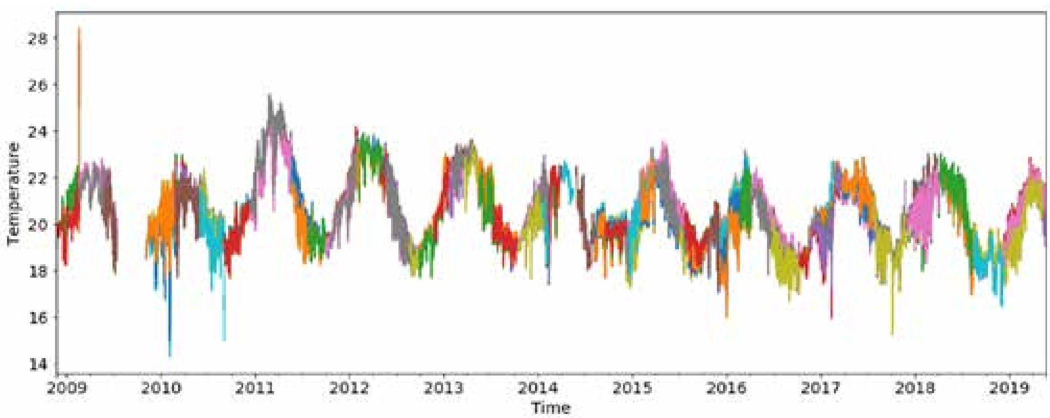


Figure 1: Aggregated time series of water temperature observations at Rottnest Island (Western Australia). Distinct colours indicate each ~4-month long deployment of an instrument, originally stored in a separate file. Up to 7 instruments are deployed simultaneously at various depths.

The Ocean InfoHub Project

Lucy Scott, UNESCO/IOC Project Office for IODE (Belgium), L.Scott@unesco.org
Tobias Spears, Fisheries and Oceans Canada – Bedford Institute of Oceanography (Canada), Tobias.Spears@dfo-mpo.gc.ca
Peter Pissierssens, UNESCO/IOC Project Office for IODE (Belgium), p.pissierssens@unesco.org
Arno Lambert, UNESCO/IOC Project Office for IODE (Belgium), a.lambert@unesco.org
Pier Luigi Buttigieg, Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung (Germany), pbuttigi@mpi-bremen.de

The IOC Ocean InfoHub Project (OIH) is a new initiative to help realise a global digital commons to empower ocean science for sustainable development across scales. Funded by the Government of Flanders (Kingdom of Belgium), the project started in April 2020 and will run for three years. The OIH is coordinated by the IOC Project Office for IODE (Oostende, Belgium), building on its 59-year history of supporting member states in the exchange and management of marine data and information. The OIH will establish and anchor a network of regional and thematic nodes that will improve online access to and synthesis of existing global, regional and national data, information and knowledge resources. The OIH will center on an openly accessible web platform designed to support interlinkages and interoperability between distributed resources including existing clearinghouse mechanisms.

The OIH will be initialised by consolidating IOC-associated online resources - including OceanExpert, OceanDocs, the Ocean Best Practices System, the Ocean Biodiversity Information System (OBIS), the World Ocean Database (WOD) and Ocean Data Portal (ODP) – extended by partnerships with EurOcean, Marinetraining.eu, EMODNET, and other sources in the IOC ODIS Catalogue of Sources (ODIScat). The initial focus of OIH will thus be on (i) experts, (ii) institutions/organizations, (iii) research data and information infrastructures and their capabilities and services offered, (iv) projects, (v) research vessels, (vi) education and training opportunities, (vii) funding programs and other opportunities, (viii) documents and publications, (ix) manuals, guidelines, standards and best practices, (xi) metadata catalogue for specific variables and (xii) access to data sets and/or data products relevant to particular program priorities (e.g. the SDGs).

Subsequently, the OIH will expand its capabilities to other global and regional partners, including international and regional organisations, private sector entities and NGOs. Through these actions, the OIH will help usher in a digital ecosystem where users, from any entry point, can discover and avail of the content and services that they require, while having opportunities to become content creators themselves. Further, the project will support collaboration between knowledge brokers at regional scales, across themes and across disciplines.

Other foci will align to the IOC's Capacity Development (CD) strategy and include the transfer of local knowledge and technologies, support of early-career scientists and remedying gender disparity. With these priorities in mind, the OIH will include automated and human-brokered capacities to match needs to capacities, allowing - for example - study and training or vessel survey opportunities to be identified by young ocean professionals from nations with nascent capacities. A peer-to-peer service will support scientific collaboration, and an automated/self-serve service will allow the search for specific human or institutional expertise.

To support regional representation in global systems, the OIH will also federate data and information held by stakeholder institutions within the three focus regions: Latin America, Africa and Pacific SIDS. Communities of practice for these pilot regions will be formed to co-develop the OIH (e.g. the ongoing CMA2 project run by INVEMAR in Colombia), alongside formal partnerships with other UN agencies and key international partners. By responding to capacity development requests from these regions, the OIH will be tailored to regional needs while meeting global objectives.

As it is co-developed, the OIH will increasingly contribute towards meeting national and regional requirements for coordinating marine data and information, while supporting global reporting requirements for the UN Sustainable Development Goals (particularly goals 4, 9, 14 and 17). In doing so, it will contribute to key Orientations and Objectives of the UN Decade of Ocean Science for Sustainable Development, as well as aspects of the Paris Agreement, the Sendai Framework for disaster Risk Reduction and agreements on Marine Biological Diversity of Areas Beyond National Jurisdiction. The OIH will also assist IOC member states to report on ocean science capacities through the Global Ocean Science Report (GOSR).

A scalable interoperability framework is needed to support the OIH's objectives and is being developed in the form of a reference architecture for the IOC Ocean Data and Information System (ODIS) and the resources it federates. During the project's initial phases, the "ODIS-Arch" will allow the three regional nodes to project and receive digital products to and from the OIH infrastructure, while maintaining their native technical implementations that are meeting the needs of their unique user communities (e.g. themes and language). Capacity for thematic products (e.g. data on the GOOS Essential Ocean Variables; EOVs) will be included at a later stage, and they will be managed by their parent organisation, linked through the ODIS-Arch. Together, this system will nucleate a web-based, global knowledge base driven by linked open data technologies.

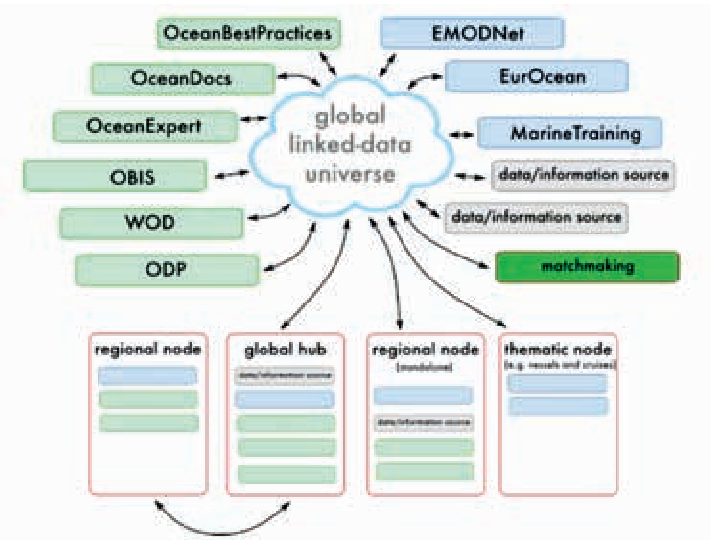


Figure 1: The e-environment underlying the Ocean InfoHub

The development of the OIH and ODIS will embrace an open approach, with the work being undertaken through co-design with the stakeholder community (contributors, users, developers), and will foster activity within the IOC and the community at large. The project will employ many conventions and standards already present in the open source community to facilitate the development, testing, and adoption of ODIS-Arch within established and developing systems.

In conclusion, the Ocean InfoHub represents a key enabling project, both in the context of advancing UN priorities related to capacity development, but also as an enabler to modernizing data sharing within the global marine data and services communities.

EMODnet Chemistry: progresses and challenges along the path to the European marine litter data management

Matteo Vinci, OGS (Italy), mvinci@inogs.it,
Maria Eugenia Molina Jack, OGS (Italy), mmolina@inogs.it,
Elena Partescano, OGS (Italy), epartescano@inogs.it,
Alessandra Giorgetti, OGS (Italy), agiorgetti@inogs.it,
Alessandro Altenburger, OGS (Italy), aaltenburger@inogs.it,
Alexia Cociancich, OGS (Italy), acociancich@inogs.it,
Erik Geletti, OGS (Italy), egeletti@inogs.it
Keywords: marine litter data management, Marine Strategy Framework Directive, data quality

EMODnet Chemistry project is facing the challenge of European marine litter data management to support the reporting of the descriptor 10 of the Marine Strategy Framework Directive (MSFD), since its third phase (2016-2018). The data management activities are focused on the following topics: beach, seafloor (trawlings), and floating microlitter. These data are managed in a PostgreSQL relational database that provides PostGIS spatial capabilities for beach and seafloor litter. For floating microlitter the SeaDataNet infrastructure has been adapted.

The development of data submission tools has facilitated the tasks related to data formatting, validation and database population.

Data product (Fig.1) and aggregated and validated datasets are provided with free access.

Figure 1: An example of free available products through the online mapviewer service showing the distribution of beach and seafloor litter surveys

Data quality and reliability has always been a key point of the data management challenge. Presently, the quality control (QC) for marine litter data is performed before and after database ingestion. The first phase of the QC process checks that the input files are provided in the correct format and that contents are consistent (syntactic and semantic) with the guidelines and common vocabularies used to describe data. The online validation tools handle these kind of controls and additionally, the database population web service monitors the existence of potential duplicates. In case of suspects, further manual checks are accomplished with the support of GIS tools.

After database ingestion further quality checks take place during metadata, data publication and products generation. At this stage of the QC the use of GIS tools is very relevant to support database queries. Data analysis during products generation are a further step of data handling that can highlight errors or anomalies previously missed.

Differently from what happens with eutrophication and contaminants data types, in the marine litter topic there is not structured QC loop that aggregates, homogenizes and controls the received data at regional level.

The correction of errors recognized after database ingestion or even after data, metadata, and products publication is very expensive in terms of time and resources. In view of the previous considerations, to optimize the use of resources in the workflow that goes from data collation to data publication, there is a focus on continuous update and upgrade of warnings during the online validation phase, before data ingestion. This can be done taking advantage of the previous experience in litter data handling, the support of the National Oceanographic Data Centres and the positive interactions with other relevant stakeholders like: JRC and TG-ML (EU Marine Beach Litter Baselines, Hanke et al., 2019), ICES (Datras database), EEA (Marine Litter Watch) and Regional Sea Conventions acting at regional and EU level.

New litter data types are under evaluation for its inclusion in EMODnet Chemistry data management. Seafloor litter images are already available from several initiatives and floating macrolitter data is accessible thanks to a Mediterranean area project. The addition of these new litter types would be a good complement for the available information at European scale. Thanks to the participation to OceanObs'19 conference, EMODnet Chemistry had the chance to be in contact with the most relevant initiatives on marine litter data at global level. This is bringing already some important results like the active interaction with the Japanese environmental agency JAMSTEC that is providing knowledge and lessons learned from the JAMSTEC deep sea debris online database experience.

In conclusion the EMODnet Chemistry experience in European marine litter data management led to improve the data quality procedures. Great efforts have been made to improve the data management tools provided for users. These efforts are finalized to streamline the data flow from submission to publication and to save resources to be invested in the new marine litter datatypes management.

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

Thierry Schmitt, Shom (France), thierry.schmitt@shom.fr
Dick M.A. Schaap, MARIS (Netherlands), dick@maris.nl
George Spoelstra, GGSgc (Netherlands), george@ggsgc.eu
Benoit Loubrieu, IFREMER (France), benoit.loubrieu@ifremer.fr
Cecile Pertuisot, IFREMER (France), cecile.pertuisot@ifremer.fr
Martin Verlaan, Deltares (Netherlands), martin.verlaan@deltares.nl
Sandra Gaytan Aguilar, Deltares (Netherlands), sandra.gaytanaguilar@deltares.nl
Ricard Campos, Coronis Computing S.L. (Spain), ricard.campos@coronis.es
Knut Hartmann, EOMAP (Germany), hartmann@eomap.de

Access to marine data is a key issue for the **EU Marine Strategy Framework Directive** and the **EU Marine Knowledge 2020 agenda** and includes 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. The DTMs are produced from survey and aggregated data sets (Composite DTMs) that are referenced with metadata adopting the SeaDataNet Catalogue services. SeaDataNet is a pan-European infrastructure for marine and ocean data management, developed and operated by NODCs in Europe, and connecting already more than 110 major oceanographic data centres around the European seas. Bathymetry survey data sets are gathered and populated by national hydrographic services, marine research institutes, and companies in the SeaDataNet CDI Data Discovery & Access service. This amounts to more than 27.500 datasets, while more than 190 composite DTMs and Satellite Derived Bathymetry 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. The number of data providers amounts to 51 from 24 countries. 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 2020.

The current EMODnet Digital Terrain Model (DTM) was released in September 2018 and has a grid resolution of 1/16 * 1/16 arc minutes (circa 115 * 115 m2), covering all European seas including part of the Arctic Ocean and Barents Sea. This DTM is based upon circa 9,400 in situ datasets and composite DTMs. Overall, this DTM contains approximately 12.3 billion grid nodes. The DTM and other relevant layers can be viewed in the Bathymetry Viewing and Download service, which also includes 3D viewing capability, based upon Cesium. However, current open source implementations of Cesium all make use of a height map (raster based) tile structure. Although workable, the performance on an average computer is not optimal and in

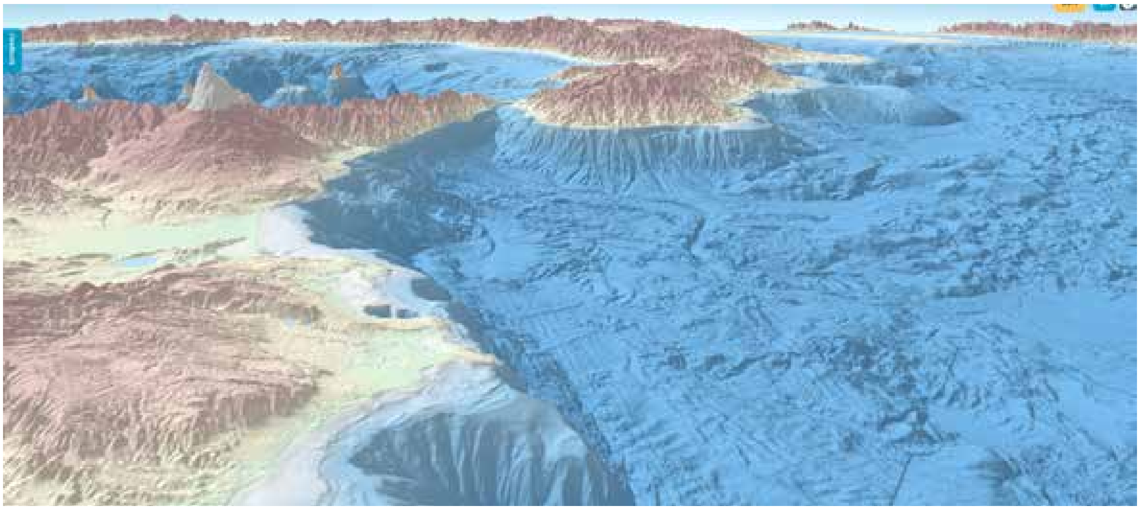


Figure 1: 3D View from the South along Sicily and the Southern part of mainland Italy

areas below sea level, artefacts (tile joints) may be visible. Therefore, to overcome these issues EMODnet Bathymetry developed a data structure based on a triangulated irregular network (TIN). This enables faster representation of the complexity of the map. As there was no open-source tool available for creating tiles in quantized mesh format out of a raster format such as used in EMODnet, this software has been developed by the project and is now available by GitHub.

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 200 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 21 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 portal and its DTM viewing and download service are very popular and in 2019 more than 50.000 portal visitors were counted, of whom circa 4.000 users from 2.100 organisations from 115 countries have downloaded altogether 38.000 EMODnet DTM tiles in 11.000 transactions. The user community roughly consists of 20% industry, 20% research, 40% universities, and 20% others. In 2019 a major milestone was reached of more than 200.000 DTM tile downloads since the start of EMODnet Bathymetry in 2008. These statistics together with many direct contacts and received mail feedbacks indicate that EMODnet Bathymetry serves a large number of users and use cases and is very much appreciated by all sectors. This is amplified by the intense use of the OGC web services (machine-to-machine) as many users also apply the EMODnet DTM and its various information layers as base layers in their own portal and applications.

The presentation will highlight key details of the EMODnet Bathymetry process, results and the ongoing activities for further improving the digital bathymetry and services.

Blue-Cloud Demonstrator: A machine learning approach to derive plankton biomass and diversity products from the Global Ocean.

Patricia Cabrera, Flanders Marine Institute (Belgium), patricia.cabrera@vliz.be
Viviana Otero, Flanders Marine Institute (Belgium) viviana.otero@vliz.be
Gert Everaert, Flanders Marine Institute (Belgium) gert.everaert@vliz.be
Raphaëlle Sauzède, Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), raphaelle.sauzede@imev-mer.fr
Renosh Pannimpullath Remanan, Laboratoire d'Océanographie de Villefranche – Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), renosh.pr@obs-vlfr.fr
Hervé Claustre, Laboratoire d'Océanographie de Villefranche – Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), claustre@obs-vlfr.fr
Julia Uitz, Laboratoire d'Océanographie de Villefranche – Institut de la Mer de Villefranche, CNRS – Sorbonne Université (France), julia.uitz@imev-mer.fr
Alexander Barth, GHER/ULiège (Belgium), a.barth@uliege.be
Charles Troupin, GHER/ULiège (Belgium), ctroupin@uliege.be
Lennert Schepers, Flanders Marine Institute (Belgium), lennert.schepers@vliz.be

The European marine data landscape and the Blue-Cloud project

Oceanographic and marine data management in the European landscape has progressed considerably during the last three decades, developing standards, services, and establishing dedicated infrastructures. As such, providing discovery and access to multidisciplinary data sets is becoming essential to provide analytical frameworks and ultimately acquire a better understanding of the ocean system. Large amounts of multidisciplinary marine data are available in an interoperable or harmonized manner in established leading European marine data infrastructures (e.g. SeaDataNet, EurOBIS, Euro-Argo, EMODnet, ELIXIR-ENA, EuroBioImaging, Copernicus Marine Environment Monitoring Service). As new technologies become available, there is a challenge to adopt new methodologies and optimise the technologies to expand the data management services. In this context, the operators of these European marine data infrastructures have joined forces to explore and demonstrate the power of a Blue Cloud virtual research environment.

Blue Cloud is a thematic European Open Science Cloud (EOSC) project that aims to implement a practical approach to explore and demonstrate the potential of cloud-based open science. The purpose is to exploit the already existing resources and to develop and deploy, through a pragmatic workplan, the Pilot Blue Cloud as a cyber platform bringing together and providing access to (1) multidisciplinary data from observations and models, (2) analytical tools, and (3) computing facilities essential for key blue science use cases. The Pilot Blue Cloud will support research to better understand and manage aspects of ocean sustainability, through a set of five pilot Blue-Cloud demonstrators. The demonstrators build upon a range of oceanographic data from multiple streams, made interoperable and integrated through Blue Cloud services. The Blue Cloud Services developed can contribute to unlocking the innovation potential of the Blue Cloud, showcasing its potential in promoting the blue economy shortening the time span between research and innovation in frontier fields, such as micro-organisms and genomics-enabled innovations.

Plankton EOV demonstrator

The demonstrator on plankton Essential Ocean Variables (EOVs) aims at rationalizing data collection, facilitate data dissemination and maximize data utilization. As biological EOVs on phytoplankton and zooplankton are still in conceptual or pilot phase, this demonstrator can contribute directly to their further development and application on large scales. Plankton plays an important role in the marine ecosystem and comprises an integral and vital component at the base of the pelagic food web. Plankton communities provide an indication of the health of marine ecosystems and their response to anthropic stressors. Moreover, they are used within several descriptors of the Marine Strategy Framework Directive (MSFD); D1: Biodiversity; D4: Food webs; D6: Sea Floor Integrity; D5: Eutrophication. In particular, phytoplankton EOVs will contribute to the understanding of the environmental conditions and top-down factors that shape the global distribution of phytoplankton community biomass and diversity. Zooplankton EOVs showing spatial trends and long-term anomalies in abundances will provide the basis for studying the dynamics of food availability for commercially exploited fish species.

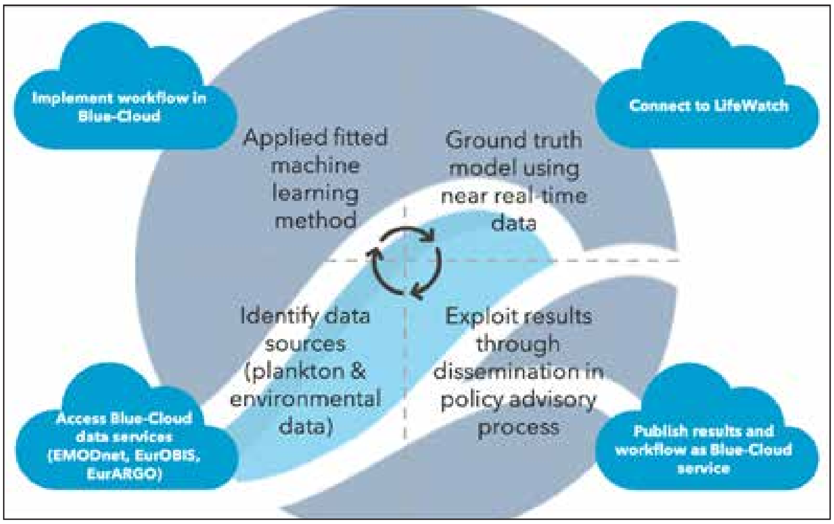


Figure 1: Plankton EOVs demonstrator workflow

The workflow of the plankton EOV demonstrator (Figure 1) consists of (1) compiling and processing several large-scale plankton and environmental data that are available from multiple European marine networks, (2) applying big data analysis and machine learning (e.g. neural networks) methods to create novel, synergistic ocean products, (3) including models using near real-time observations from LifeWatch data, (4) and exploiting the results that will be made available to a variety of users through a Blue Cloud Virtual Lab. The virtual EOVs produced will contribute to improve knowledge and quantitatively reduce uncertainty regarding the present state of the marine plankton ecosystems and their response to climate change. More specifically, the phytoplankton products will consist of global 4D fields of key phytoplankton products, namely EOVs of phytoplankton chlorophyll biomass and diversity, the latter being expressed as chlorophyll contribution of several phytoplankton groups (PFTs). These EOVs will be obtained from multidisciplinary (physics, biogeochemistry, biology) data acquired from oceanographic platforms with complementary space-time resolution and coverage (e.g. vessels, satellites, Argo/BGC-Argo floats). Whereas, the zooplankton products will show spatial trends and long-term anomalies in zooplankton distributions, focusing on the most abundant Copepod species from the North East Atlantic. A neural network model will be built with the observed zooplankton abundances complemented with environmental parameters (oxygen, temperature, salinity, chlorophyll, depth, nutrients) to create spatio-temporal products.

We verify our products using a nutrient–phytoplankton–zooplankton model and near-real time observations from LifeWatch data. This will enable us to detect and explain anomalies in long-term trends and quantify the relative contribution of the environmental drivers for these EOVs. The results from the demonstrator will be implemented in a Blue-Cloud Virtual Lab, providing its users access to blue multidisciplinary data for exploring the methodology and data used in the products. Ultimately, the products will help marine advisory process and serve the wider Blue community.

SeaDataCloud temperature and salinity climatologies for the European marginal seas and the Global Ocean

Simona Simoncelli, Instituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna (Italy), simona.simoncelli@ingv.it
Christine Coatanoan, Ifremer (France), christine.coatanoan@ifremer.fr
Volodymyr Myroshnychenko, METU, (Turkey), volodymyr@ims.metu.edu.tr
Örjan Bäck, SMHI (Sweden), orjan.back@smhi.se
Helge Sagen, IMR (Norway), Helge.Sagen@imr.no
Serge Scory, RBINS (Belgium), sscory@naturalsciences.be
Nadia Pinardi, University of Bologna (Italy), nadia.pinardi@unibo.it
Alexander Barth, Ulg (Belgium), a.barth@uliege.be
Charles Troupin, Ulg (Belgium), ctroupin@uliege.be
Kanwal Shahzadi, University of Bologna (Italy), kanwal.shahzadi@studio.unibo.it
Paolo Oliveri, INGV, Bologna (Italy), paolo.oliveri@ingv.it
Reiner Schlitzer, AWI (Germany), reiner.schlitzer@awi.de
Michèle Fichaut, Ifremer (France), michele.fichaut@ifremer.fr
Dick Schaap, MARIS (Netherlands), dick@maris.nl

SeaDataCloud Temperature and Salinity historical data collections covering the time period 1900-2018 were released in 2020 for each European marginal sea (Arctic Sea, Baltic Sea, Black Sea, North Sea, North Atlantic Ocean, and Mediterranean Sea). A Quality Assurance Strategy (QAS) was developed and continuously refined in order to improve the quality of the SeaDataNet database content and derive the best data products through an iterative approach, which allows the versioning of data products.

Regional Temperature and Salinity climatologies (see Figure 1) have been produced using DIVAnd software (Barth et al. 2014) and integrating for the first time SeaDataNet data with external data sources, such as CMEMS in situ TAC (Coriolis Ocean Dataset for Reanalysis) that highly increased the temporal and spatial data coverage. Regional climatologies were designed with a harmonized initial approach and all cover the time period after 1955, when marine data start to be sufficient for mapping. All regional products are characterized by monthly fields over the whole time span 1955-2018 and seasonal decadal fields on the same vertical standard levels of the World Ocean Atlas (WOA18, Garcia et al., 2019).

A global SDC climatology has been created for the first time, which contains two different monthly climatology for temperature and salinity, one covering the time period 1900-2017 and the other with a different time coverage 2003-2017, computed from World Ocean Database (WOD2018, Boyer et al., 2019). This choice has been made because spatial coverage of SeaDataNet data at global scale is still too sparse.

A consistency analysis of all SDC climatologies versus the WOA has been performed to demonstrate the differences and the value added of SDC products.

SDC team worked to optimize the data integration process with external sources, to better tune the DIVAnd parameters, the background field estimation and to improve the final consistency analysis with the available multi-year products from WOA and CMEMS.

An overview of the methodology applied to compute the SDC climatologies and their main characteristics will be presented together with the main results achieved by the SDC products team.

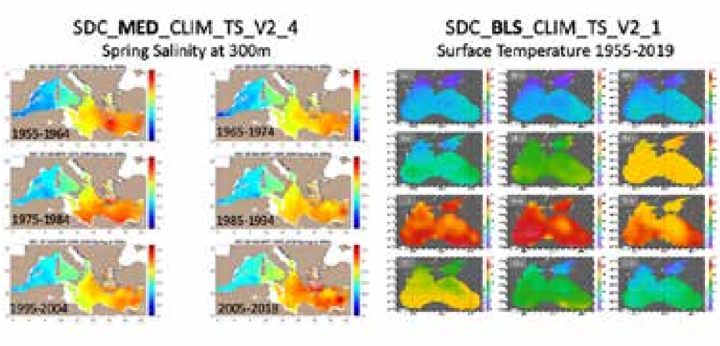


Figure 1: Example of climatological maps from Mediterranean and Black Seas.

Products availability

SDC products, data collections and climatologies, are available through a dedicated web catalogue (https://www.seadatanet.org/Products/) together with their Digital Object Identifier (DOI) and the relative Product Information Document (PIDoc), containing all specifications about product's generation, quality assessment, technical details and usability to facilitate users' uptake.

References

Boyer, T. P., O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, R. A. Locarnini, A. V. Mishonov, C. R. Paver, J. R. Reagan, D. Seidov, I. V. Smolyar, K. Weathers, M. M. Zweng,(2019): World Ocean Database 2018. A. V. Mishonov, Technical Editor, NOAA Atlas NESDIS 87.

Barth, A., Beckers, J.-M., Troupin, C., Alvera-Azcárate, A., and Vandenbulcke, L.: divand-1.0: n-dimensional variational data analysis for ocean observations, Geosci. Model Dev., 7, 225-241, doi:10.5194/gmd-7-225-2014, 2014.

Garcia H.E., T.P. Boyer, O.K. Baranova, R.A. Locarnini, A.V. Mishonov, A. Grodsky, C.R. Paver, K.W. Weathers, I.V. Smolyar, J.R. Reagan, D. Seidov, M.M. Zweng (2019). World Ocean Atlas 2018: Product Documentation. A. Mishonov, Technical Editor.

To patch, to rebuild, or to build a new? The genesis of the ICES portal

Carlos Pinto, ICES (Denmark), carlos@ices.dk
Mehdi Abbasi, ICES (Denmark), mehdi.abbasi @ices.dk
Anna Osypchuk, ICES (Denmark), anna.osypchuk@ices.dk
Joana Ribeiro, ICES (Denmark), joana.ribeiro@ices.dk
Neil Holdsworth, ICES (Denmark), Neilh@ices.dk

Introduction

The International Council for the Exploration of the Sea (ICES) operated EcoSystemData portal[1] for 12 years, allowing users to search, view, and download data available at ICES. The EcoSystemData portal became a focal point for the cross-disciplinary research based on ICES resources, but with time, it also revealed several flaws and limitations. Therefore, a new project was started to create a new ICES Data Portal[2] to replace the EcoSystemData portal. The new data portal uses recent technologies, such as microservices architecture, and is made in AngularJS for a good and responsive user experience. ICES Data Portal standardises the services that each dataset makes available to the community and minimises the impact on the resources that each portal/dataset has available. In addition to the features previously provided by the EcoSystemData portal, the new portal aims for a more intuitive interface by visualizing data, has an extended data portfolio, and provides both the base data and the ICES community data products for users.

Figure 1: Search data by theme. In the figure we can see a screenshot of the top topics where the user can drill down into

Visualization of data in ICES Data Portal

The main component of the new data portal is the visualization of data, either in a tree map or in a geographic map. We provide the tools to allow a good search, query and visualization of the dataset. Data are organized in topics, which comprise one or more theme(s) (Figure 1). Themes, in turn, include one or more dataset(s) and/or data product(s). When the user selects a theme, the corresponding data will be added to the map. The tool also allows an advanced search, where the user can filter the data by defining their preferred data criteria. Data criteria consists of two parts:

[1] mandatory fields that are: location and date range and [2] fields defined by the data provider, some examples are species, country, survey, station name.

After selecting one or more themes, the user can visualize the data in the map (**Figure 2**). The user can return to the search (**Figure 1**) by clicking the search button.

A user has the possibility to save a custom map (as an image) or to download the selected data. When a download is requested, each data provider microservice sends the data to the data portal from the thematic databases, and only when all the data are available, does the service create a zip file and sends the download link by email to the user. RabbitMQ (an open-source message-broker software) was implemented to manage the messaging between all microservices and ensure the user receives the requested data. This allows the fail-free processing of complex queries for big data layers without overloading any of the data providing databases or the ICES Data Portal itself.

Microservices architecture

ICES Data Portal is built with AngularJS and with microservices architecture. Microservices are an architecture, where all the components of the system are put into individual components that can be built, deployed, and scaled individually. This brings more scalability, fault to tolerance and allows the option of moving to the cloud in the future. The microservices also allow services to be fine-grained and the protocols are lightweight. For each of the data streams (base data or data products) that feed the new data portal, the data provider has to provide the address of the microservices. In the microservice, the data managers have a possibility to customize options, such as (1) map style, (2) data criteria, (3) summary of data, and (4) download service.

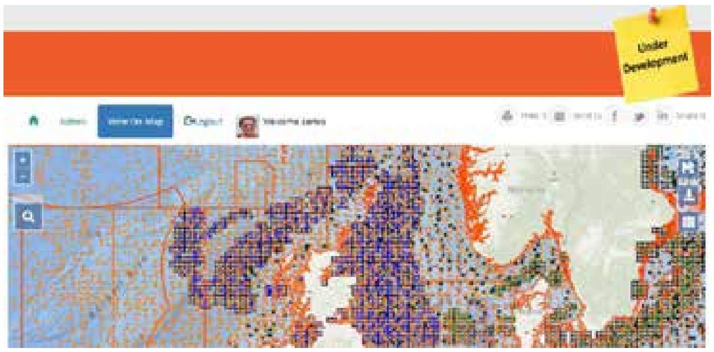


Figure 2: Mapped data, where several layers (polygon, point, and base data and data products) are displayed together.

Summary

The ICES Data Portal facilitates and fortifies the exchange of data and the organizational data communication of ICES and its community. In this version we open the option of the ICES community to share their products.

The development of the new ICES Data Portal (https://data.ices.dk) was done in an agile way and with the option of moving the data portal to the cloud.

Future developments would include user authorisation for data types with separate data licenses restricting data access, not covered by the ICES data policy (https://www.ices.dk/marine-data/Documents/ICES-Data-policy.pdf), which allows full access and re-use rights.

References

1 - C. PINTO, & H. Jensen & N. Holdsworth. ICES EcoSystemData – Visualising data for the ecosystem approach. In Nishida, T., and Caton, A.E. (Editors) 2010. GIS/Spatial Analyses in Fishery and Aquatic Sciences (Vol. 4). Fishery-Aquatic GIS Research Group, Saitama, Japan, 401-416.

2 - https://data.ices.dk

The concept of featured data services in the PORTO interface

Aldo Drago, University of Malta, Physical Oceanography Research Group, Dept. of Geosciences (Malta), aldo.drago@um.edu.mt,
Audrey Zammit, University of Malta, Physical Oceanography Research Group, Dept. of Geosciences (Malta),
Adam Gauci, University of Malta, Physical Oceanography Research Group, Dept. of Geosciences (Malta),
Anthony Galea, University of Malta, Physical Oceanography Research Group, Dept. of Geosciences (Malta),
Joel Azzopardi, University of Malta, Department of Intelligent Computer Systems (Malta)
Raisa Tarasova, University of Malta, Physical Oceanography Research Group, Dept. of Geosciences (Malta),
Gianfranco Calise, University of Malta, Physical Oceanography Research Group, Dept. of Geosciences (Malta),
Giuseppe Ciraolo, University of Palermo, Department of Engineering, (Italy)
Salvatore Aronica, Istituto per lo studio degli impatti Antropici e Sostenibilità in ambiente marino, CNR-IAS (Italy)
Rosario Sinatra, University of Catania, Department of Civil Engineering and Architecture, (Italy),
Elisabetta Paradiso, University of Catania, Department of Civil Engineering and Architecture, (Italy)
and Fulvio Capodici, University of Palermo, Department of Engineering, (Italy)

In the last decade, the valuation of ocean information and the drive to make marine data more freely accessible to a wider range of users have been on the agenda of many initiatives with some tangible results. In Europe, the COPERNICUS and EMODnet web platforms have been consolidated with an increasing set of data products and user support services appealing to an audience that goes beyond the research community. SeaDataNet and subsequently SeaDataCloud have contributed to synergise national ocean data centres on a common operationally robust data infrastructure providing access to ocean and marine metadata, data and data products, based on common data management standards, promoting data archaeology, and providing secure data archival.

There has been a proliferation of web data services focusing on the delivery of essential datasets targeted to specialised users, and often linked to international and regional initiatives such as with observational programmes or thematic networks. The IOC Ocean Data and Information System (https://catalogue.odis.org/) provides an online browsable and searchable broad **worldwide catalogue of existing ocean related web-based data and information sources, including products and services.** The evolution of these web data services has matured to provide user-friendly systems, adopting increasing functionalities to aid the users in data discovery, online visualisation, data subsetting and download criteria including variable data formats.

The digital era has opened new realms for ocean data delivery. As more users become dependent on reliable information deriving from multiple sources data, the non-professional users are increasing in numbers whose demands are different from those of professional users. Technology is leading to a step shift in the value addition chain of data to information, knowledge and intelligence, providing sophisticated user experiences online, with faster delivery and service elaborations on a wider range of more affordable smart mass media like iPhone, iPad and other wireless devices. While providing more elaborated deliveries to expert users, the digital environments add a new dimension to data services by matching user experience, expectations and demand.

This has led to the concept of featured data services. In contrast to the traditional generic services, the featured services are usually dedicated to a category of users with specific needs, providing routine data-derived products to support their operational day-to-day activities or production lines. User categories can be from industry, such as in relation to the tourism sector to provide a higher-level information service according to user location, demanded facility, and specific real-time request. Compared to the static delivery of pre-prepared products determined at source, the featured web services use a dynamic added value production composed directly by the online user who can customise the service to fit a specific need or query. Four essential ingredients of featured services are: (i) GIS-like web tools for geographic rendering and mapping; (ii) user specified dynamic content delivery; (iii) online functionalities to elaborate data online; and (iv) real-time updating of data and information.

The CALYPSO South project (www,calypsosouth.eu) is pioneering in featured data services through the PORTO interface which follows on the experience gained in the KAPTAN online and smartphone application. PORTO is an integrated online service of met-ocean information delivered to especially aid harbour masters, port authorities and operators in the shipping and maritime services in the proximity of the Maltese Islands and south of Sicily. The initiative follows the trail of efforts to deliver services deriving from operational oceanography and meteorology to dedicated users. The PORTO service package entails the use of web-based tools that give the user a direct handle to view, analyse and compare different datasets online such as through user-selected geographic sub-domains, transects, inter-comparison of sites and overlaying of data layers. The novelty of PORTO is that it allows the user to combine data on user-dedicated request employing augmented web features to customise the use of the online data. A drawing board allows the user to adapt the viewing the content. The interface also provides early warning alerts on extreme events such as gale winds, high waves or strong currents, serving as important indicators to operators who can avail of such information to mitigate adverse conditions.

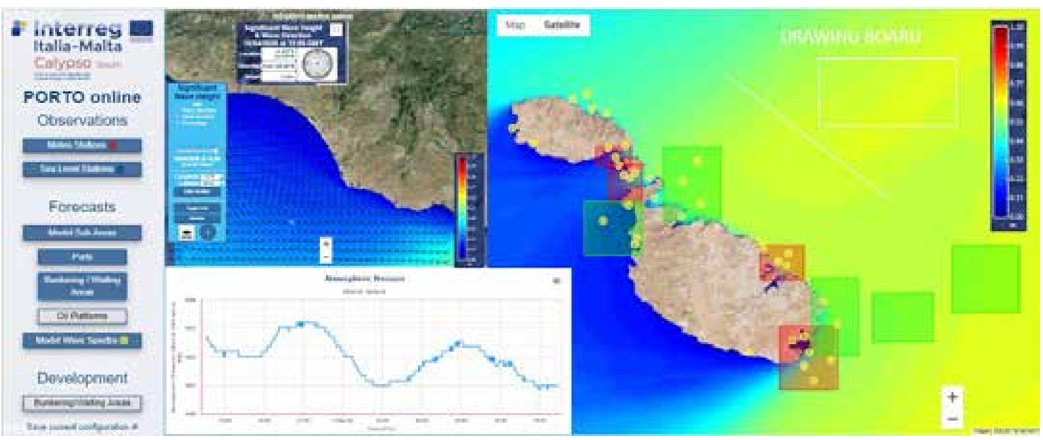


Figure. 1: Snapshots of the PORTO online service delivery

The PORTO interface is based on Google Maps JavaScript API. Using this API it is possible to display geo-referenced data as an image over various types of maps, such as satellite images of the area, terrain or street maps. The control panel on the interface allows users to select from various layers of data that are available, while the data panel displays details for the selection and elaboration of selected data points. Clicking on the various map features such as sub-area categories opens new windows displaying more detailed information and further data pertaining to the specific feature.

The interface was developed using Bootstrap, and is therefore responsive and mobile friendly. It is deployed as a Progressive Web App allowing users to eventually download it onto their mobile device for easy access. It utilises HTML5's local storage facility to be able to save the user's preferred configuration from one session to the next. Interaction between the interface and the server is through one of two Simple Object Access Protocol (SOAP) web-based services. These web services expose a number of functions, which can be called using AJAX (Asynchronous JavaScript And XML). When a user selects a feature to view, the interface calls the appropriate function on the web service, with elaborations on the area of interest and the time frames required. On the server side, the function retrieves the relevant data, either by sending an SQL request to a database, or by running a script to read the data from a text file. The data is then packaged as an XML (eXtensible Mark-up Language) object and transmitted back to the interface.

Data for the interface comes from various sources, including both observations and numerical model data. Key data in PORTO are real time observations from ten meteorological stations, four sea level gauges and seven HF radars belong to the CALYPSO network; numerical model data for atmospheric parameters, waves, currents sea surface temperature and salinity; satellite data from different platforms. The raw model data outputs are further elaborated to extract secondary data, which is also available to the user. It is intended to package the PORTO interface as a Progressive Web App to deliver the service on mobiles without the need of developing a dedicated smartphone application.

Acknowledgements. The PORTO interface is developed within the CALYPSO South project which is partially funded by the Interreg V-A Italia-Malta programme, Cohesion Policy 2014-202020. Besides the Physical Oceanography Research Group (project lead) from the University of Malta, the partnership comprises four Sicilian partners: University of Palermo (Polo Universitario della Provincia di Trapani) as Sicilian Focal Point, Istituto per lo studio degli impatti Antropici e Sostenibilità in ambiente marino (IAS-CNR), University of Catania (CUTGANA) and ARPA SICILIA (Agenzia Regionale per la Protezione dell'Ambiente), and two other Maltese partners: Transport Malta and Civil Protection Department.

References

A. Drago (2018), A smartphone application delivering meteo-marine data to the public, in The Ever Growing Use of COPERNICUS across Europe's Regions - A selection of 99 user stories by local and regional authorities, pp 150-1, NEREUS, EC and ESA, 2018.

Drago A., A. Zammit, R. Tarasova, A. Gauci, A. Galea, J. Azzopardi, G. Ciraolo, F. Capodici (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., G. Ciraolo, F. Capodici, S. Cosoli, M. Gacic, P.-M. Poulain, R. Tarasova, J. Azzopardi, A. Gauci, A. Maltese, C. Nasello and G. La Loggia (2015), 'CALYPSO – An operational network of HF radars for the Malta-Sicily Channel', *Proceedings of the Seventh International Conference on EuroGOOS, 28-30th October 2014, Lisbon, Portugal.* Edited by H. Dahlin, N.C. Fleming and S. E. Petersson. First published 2015. EuroGOOS Publication no. 30. ISBN 978-91-974828-9-9.

The CYCOFOS new Web GIS

Elena Zhuk, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), alenixx@gmail.com
George Zodiatis, ORION Research (Cyprus), oceanosgeos@gmail.com

CYCOFOS forecasting models

The Cyprus coastal ocean forecasting system, known as CYCOFOS provides operational hydrodynamical and sea state forecasts in the Mediterranean and the Black Sea since early 2002 (Zodiatis et al. 2003; 2008). Recently, it has been improved with the implementation of new hydrodynamic and wave models with the objective of targeting larger and higher resolution domains, at sub-regional and regional scales (Zodiatis et al. 2018a; 2018b). For the new CYCOFOS hydrodynamic modeling system a novel parallel version of the POM-Princeton Ocean Model has been implemented. The new CYCOFOS hydrodynamical model covers the entire Eastern Mediterranean with a resolution of 2 km and the Levantine Basin with 600 m, nested in the Copernicus Marine Environmental Monitoring Service of the Mediterranean Forecasting Center-CMEMS Med MFC, see Table 1 for the CYCOFOS hydrodymanical models configuration.

The new ECMWF wave model WAM CY46R1 (Ardhuin et al. 2010) implemented recently for CYCOFOS, covers the Mediterranean and the Black Sea using the SKIRON winds and provides hourly wave forecast on a daily base, see Table 1 for the CYCOFOS WAM configuration.

Both, hydrodynamical and wave models were validated against remote sensing SST, ARGO floats T/S profiles, in-situ wave time series and with parent models (Zodiatis et al. 2018).

CYCOFOS hydrodynamical models	Eastern Mediterranean	Levantine Basin	CYCOFOS wave model	Mediterranean & Black Sea
Model's domain	29N – 47N, 6W – 42E	31N – 37 N, 30E-36E	Model's domain	29N – 47N, 6W – 42E
Horizontal Resolution	2 km	600m	Horizontal Resolution	0.05 x 0.05 degrees
Vertical layers	30 sigma layers	30 sigma layers	Frequencies	25, range 0.0417 - 0.54764 Hz logarithmically spaced)
Initial/lateral boundaries	CMEMS Med MFC or NOAA	CYCOFOS East Med.	Directions and time steps	24 (equally spaced) 120 sec
Surface fluxes	SKIRON or ECMWF	SKIRON or ECMWF	Surface fluxes	SKIRON wind
Forecasting details	Daily: 6 hourly for 5 days	Daily: 6 hourly for 5 days	Forecasting details	Daily: hourly for 5 days

Table 1. The new CYCOFOS models configuration.

CYCOFOS WEB GIS

The CYCOFOS Web GIS provides access and visualization of the forecasting sea currents, temperature, salinity, waves and winds data from the CYCOFOS modelling systems. The

CYCOFOS WEB GIS http://cycofos.orioncyprus.org/ was developed at the base of client-server architecure. CYCOFOS WEB GIS server provides the archiving of the CYCOFOS forecasting data from a dedicted ftp. Daily run python scripts extracting and preparing the forecasting parameters for visualization (Figure 1).

The CYCOFOS Web GIS User Interface was developed using jQuery and openLayers tools and provides the selection of the parameter and its visualisation. To access the CYCOFOS forecasting data it is necessary to select the region: Levantine Basin, East Mediterranean, Mediterranean Sea (includes the Black Sea) and the forecasting parameters, such as waves, currents, temperature, salinity, hour, date and water depth.

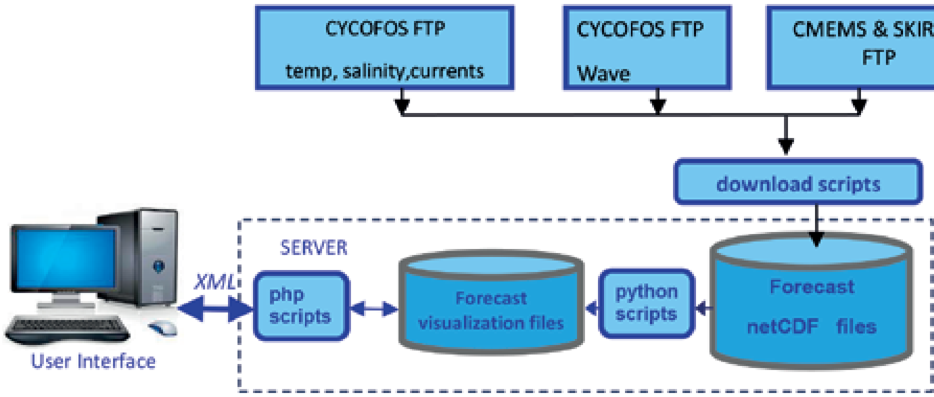


Figure 1: CYCOFOS WEB GIS structure

The selected CYCOFOS forecasting parameter is presented as a map (Figure 2). The CYCOFOS forecasting data are used also for the HERMES WEB GIS http://hermes.orioncyprus.org/ proviiding data in the coastal areas of the North Aegeam, SE Cyprus, Albania and Bulgaria and in addition will be used for the Black Sea WEB GIS, which is developed at the Marine Hydrophysical Institute, Russian Academy of Sciences.

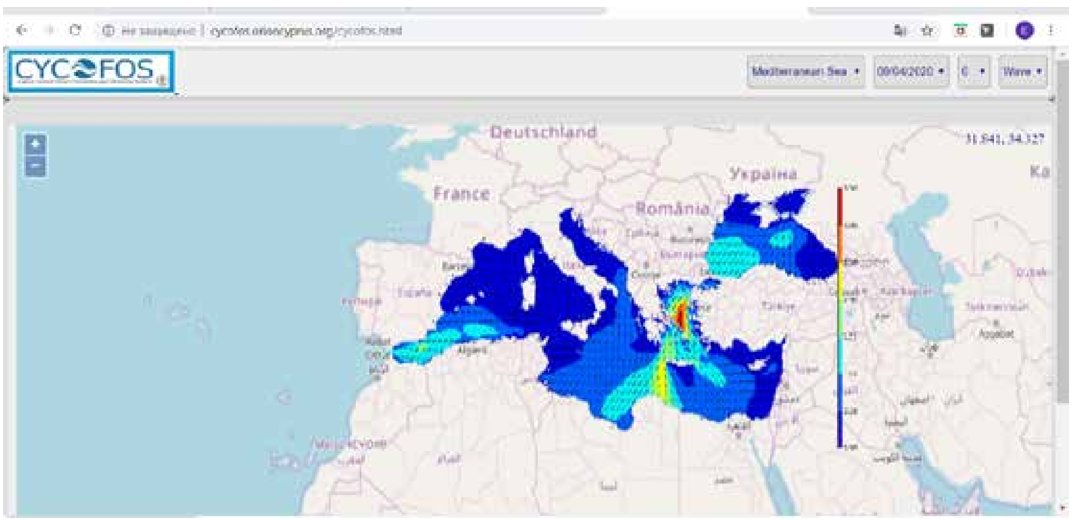


Figure 2: CYCOFOS WEB GIS User Interface**: example of the waves visualisation.**

Some modules were developed in the framework of the state task on theme No. 0827-2019-0002 “Operative Oceanology”.

References

ARDHUIN F. et al. (2010), Semiempirical Dissipation Source Functions for Ocean Waves. Part I: Definition, Calibration, and Validation, Journal of Physical Oceanography, Vol. 40, pp. 1917-1941.

ZODIATIS et al. (2018b). Downscaling the Copernicus marine service in the Eastern Mediterranean. OM14A: Advances in Coastal Ocean Modeling, Prediction, and Ocean Observing System Evaluation. AGU, Ocean Science meeting, 11-16 February 2018, Portland, Oregon.

ZODIATIS et al. (2018a). Downscaling the Copernicus CMEMS Med-MFC in the Eastern Mediterranean: The new CYCOFOS forecasting systems at regional and sub-regional scales. *Proceedings of the 8th EuroGOOS International Conference* "Operational Oceanography serving Sustainable Marine Development, 3-5 October 2017, Bergen, Norway", 305-310pp.

ZODIATIS et al. (2008), Operational ocean forecasting in the Eastern Mediterranean: implementation and evaluation, Ocean Science, 4, 31-47.

ZODIATIS et al. (2003). High resolution nested model for the Cyprus, NE Levantine Basin, eastern Mediterranean Sea : implementation and climatological runs, Annales Geophysicae, 21, 221-236.

Are the pan-European seas a promising source for critical metals supply? GeoERA-MINDeSEA Marine Data and Information Management Best Practices

Trevor Alcorn, Geological Survey Ireland (Ireland), trevor.alcorn@gsi.ie
Xavier Monteys, Geological Survey Ireland (Ireland), xavier.monteys@gsi.ie
Iker Blasco, Geological Survey of Spain (Spain), ikerblasco12@gmail.com
Ana Lobato, Geological Survey of Spain (Spain), ab.lobato@igme.es
Javier González, Geological Survey of Spain (Spain), fj.gonzalez@igme.es

Abstract

Covering 15 million km^2, the pan-European seas represent a promising new frontier for the exploration for mineral resources while sustainably managing the marine resource. An enormous challenge in terms of research, technological innovation, environmental protection, spatial planning and social acceptance is facing European and international research and sustainable development plans related to this potential marine resource. The GeoERA-MINDeSEA project is an ERA-NET action, Horizon2020 project, involving collaboration between eight GeoERA Partners and four Non-funded Organisations at various points of common interest for exploration and investigation on seafloor mineral deposits. MINDeSEA aims to develop harmonised data models, datasets and information products to assess seabed minerals potential. MINDeSEA data assets are based on detailed studies and compiled data on geology, geochemistry, mineralogy, environmental and regulatory issues of hydrothermal mineralisations, polymetallic nodules, ferromanganese crusts, phosphorites, marine placer deposits and exploration activities. Cobalt, lithium, tellurium, nickel, rare earth elements, copper, and other strategic and critical metals are being investigated in several seabed mineral deposits under the jurisdiction of European coastal States, looking for alternative sources to land-based mined deposits. MINDeSEA will identify areas for sustainable resourcing and information to support decision-making on management and Maritime Spatial Planning regarding extraction in pan-European seas as part of its core actions.

Figure 1: GeoERA Raw Materials MINDeSEA

Key to understanding the seabed minerals resource potential in pan-European seas is the acquisition, processing and management of reliable data and information on this unique marine asset. It is the position of this oral presentation to introduce 10 critical data management best practices guided by initiatives such as the World Wide Web Consortium (W3C), the INSPIRE Directive on spatial data infrastructures and SeaDataNet standards. These cross-thematic data management best practices are transferable across various marine domains where data and information exchange plays a critical role in understanding the marine resource.

MINDeSEA results are closely linked to the Seabed mineral deposits (WP7) of EMDOnet Geology,

one of several EMODnet portals with the purpose to strengthen blue growth in Europe. The European Marine Observation and Data Network (EMODnet) is a network of organisations supported by the EU's integrated maritime policy and aims to standardise data, services and products across the marine domain.

This position paper delivered by MINDeSEA partners under Work Package (WP) 8 Link to Information Platform has been harnessed from the W3C recommendations for publishing data on the Web known as Data on the Web Best Practices (DWBP) which support the development and encouragement of the continued expansion of the Web as the medium for the exchange of data. These Best Practices (BP) provide an ideal framework in which to build a guide for data management for MINDeSEA as a seabed mineral marine data and information platform as the ultimate goal of seabed minerals data management is delivery of this data on the GeoERA Web platform.

The 10 Best Practices for MINDeSEA marine and information data management to be presented are as follows:

1. **Metadata**
   - Provide metadata
   - Provide descriptive metadata
   - Provide structural metadata
1. **Licence** : Provide information about the license standard used, its caracteristics, permissions and restrictions.
2. **Provenance** : Information about the origins of the data and any changes made in the metadata lineage statement.
3. **Quality** : Information on the established quality standards and the applied quality plan.
   - Provide metadata quality information, following ISO 19115
   - Provide data quality information, following ISO 19157
4. **Versioning** : Information about the versioning policy to know the timing of the data and the changes made
   - Provide a version indicator
   - Provide version history
5. **Identifiers** : Use persistent identifiers, following INSPIRE data publication guidelines.
6. **Formats** : Use machine-readable and human-redable open and standardised formats.
7. **Vocabularies** : Reuse vocabularies, preferably standardised ones and normalize those that have not yet been registered.
8. **Access** : Provide bulk download.
9. **Enrichment** : Enrich data by generating new data.

These 10 Best Practices in MINDeSEA WP8 are being applied to data sourced and supported by MINDeSEA partners for seabed mineral occurrences for Seafloor Massive Sulphide Deposits (WP3); Ferro-manganese crusts, phosphorites and critical raw materials (WP4) ; Marine Placer Deposits (WP5) ; Polymetallic Nodules (WP6) and Exploration in the Atlantic, Mediterranean, Baltic and Black Sea (WP7). Since the INSPIRE codes recorded so far have always responded to the terrestrial domain, they are not complete enough to respond to the diversity of deposits that exist, for example, on the ocean floor and therefore should be expanded.

The authors would like to acknowledge the standardised and harmonised data contribution of MINDeSEA work package leaders from Geological Survey of Norway (NGU), Instituto Geológico y Minero de España (IGME), Hellenic Survey of Geology & Mineral Exploration (HSGME) and Laboratorio Nacional de Energia e Geologia I.P. to the project.

Supporting the essential - Recommendations for the development of accessible and interoperable marine biological data products

Dan Lear, Marine Biological Association (United Kingdom), dble@mba.ac.uk
Peter Herman, Deltares (Netherlands), peter.herman@deltares.nl
Gert Van Hoey, Research Institute for Agriculture, Fisheries and Food (Belgium), gert.vanhoey@ilvo.vlaanderen.be
Lennert Schepers, Flanders Marine Institute (Belgium), lennert.schepers@vliz.be
Nathalie Tonné, EMODnet Secretariat (Belgium), nathalie.tonne@emodnet.eu
Marina Lipizer, Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), mlipizer@inogs.it
Frank Edgar Muller-Karger, University of South Florida (United States of America), carib@usf.edu
Ward Appeltans, Intergovernmental Oceanographic Commission of UNESCO (Belgium), w.appeltans@unesco.org
W. Daniel Kissling, University of Amsterdam (Netherlands), wdkissling@gmail.com
Neil Holdsworth, International Council for the Exploration of the Seas (Denmark), neilh@ices.dk
Martin Edwards, Marine Biological Association (United Kingdom), maed@mba.ac.uk
Ellen Pecceu, Research Institute for Agriculture, Fisheries and Food (Belgium), ellen.pecceu@ilvo.vlaanderen.be
Henrik Nygård, Finnish Environment Institute (Finland), henrik.nygard@ymparisto.fi
Gabrielle Canonico, National Oceanic and Atmospheric Administration (United States of America), gabrielle.canonico@noaa.gov
Silvana Birchenough, Centre for Environment, Fisheries and Aquaculture Science (United Kingdom), silvana.birchenough@cefas.co.uk
George Graham, Marine Biological Association (United Kingdom), geogra@mba.ac.uk
Klaas Deneudt, Flanders Marine Institute (Belgium), klaas.deneudt@vliz.be
Joana Beja, Flanders Marine Institute (Belgium), joana.beja@vliz.be

We outline stakeholder-led approaches in the development of data products based on marine biological observations to support effective conservation, management and policy development.

The European Marine Data and Observation Network (EMODnet) Biology project has undertaken unparalleled activity in the collation and standardisation of marine biological data from all European seas and the wider North East Atlantic region since 2009. EMODnet Biology has collated and standardised in excess of 25 million species observations as of April 2020. Significant efforts have been invested in the adoption and further development of standards, guidelines and best practices for the initial acquisition of biological and biodiversity observations, their curation, and publication. A major focus are strategies to meet future, as yet unknown, challenges.

We showcase and analyse EMODnet Biology efforts in the development of stakeholder-led data products to support conservation, management, and advisory decisions by regulators, industry and policy makers. Following a definition of the term “Data Product”, we explain the steps taken to ensure the resultant products are meeting specific use-cases. These are derived from a range of stakeholder engagement exercises and through a review of the availability and applicability of marine biological data. We outline the process to develop these products and illustrate possible improvements.

The EMODnet Biology Engagement Lifecycle (Fig. 1) summarises the ongoing approach and methodology.

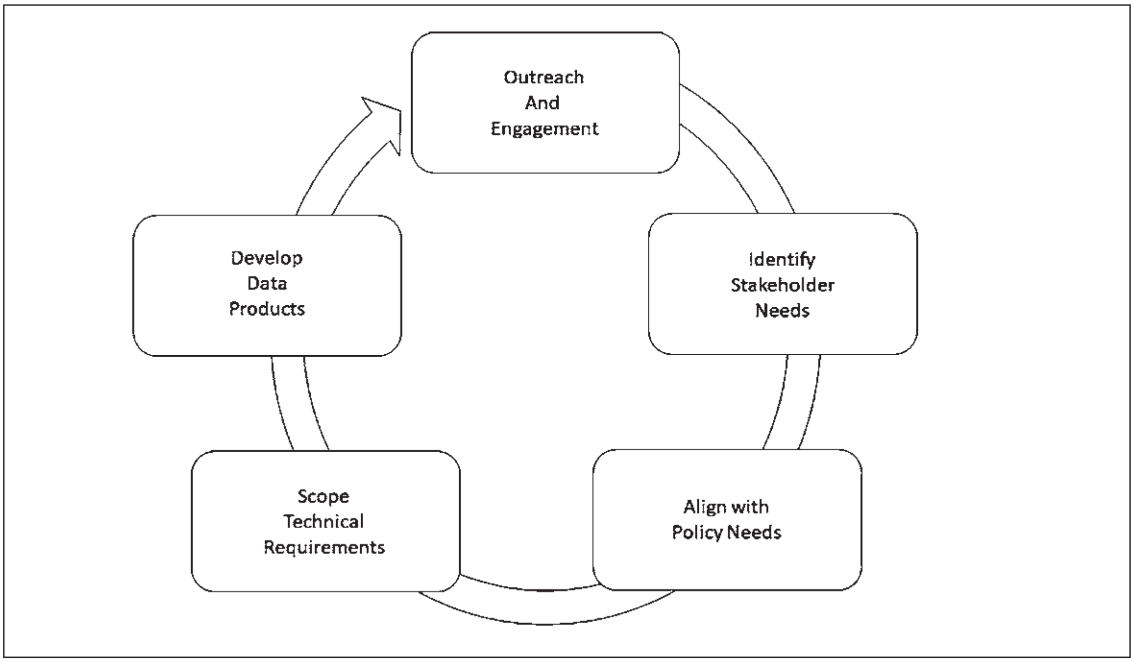


Fig.1 EMODnet biology data product engagement lifecycle

The requirements of a broad range of stakeholders and iterative, structured processes framed the development of tools, models and maps that support the FAIR (Findable, Accessible, Interoperable, Reusable) data principles. The products are also based on open data and are openly distributed. By structuring the resultant data products around the emerging biological Essential Ocean Variables (EOVs) of the Global Ocean Observing System (GOOS), and through the engagement with a broad range of end-users, the EMODnet Biology project has delivered a suite of demonstration data products. These products are presented in the European Atlas of Marine Life, an online resource demonstrating the value of open marine biodiversity data and help to answer fundamental and policy-driven questions related to managing the natural and anthropogenic impacts in European waters. Examples of products available in the Atlas include: Phytoplankton community analysis in the Northern Adriatic, Thermal affinities for European marine Species groups, Neural network modelling of Baltic zooplankton abundances, and two products relating to invasive species in European harbours and specifically within the Baltic Sea.

Predicting the spread of invasive marine species with open data and machine learning: Process and Challenges

Adrian Bumann, Chalmers University of Technology (Sweden), adrian.bumann@chalmers.se
Robin Teigland, Chalmers University of Technology (Sweden), robin.teigland@chalmers.se
Jannes Germishuys, Combine AB (Sweden), jurie.germishuys@combine.se
Benedikt Ziegler, Combine AB (Sweden), benedikt.ziegler@combine.se
Martin Mattson, Medins Havs och Vattenkonsulter AB (Sweden), martin.mattsson@medinsab.se
Eddie Olsson, RISE - Research Institutes of Sweden (Sweden), eddie.olsson@ri.se
Robert Rylander, RISE - Research Institutes of Sweden (Sweden), robert.rylander@ri.se
Markus Lindh, Swedish Meteorological and Hydrological Institute (Sweden), markus.lindh@smhi.se
Yixin Zhang, University of Gothenburg (Sweden), yixin.zhang@ait.gu.se
Torsten Linders, University of Gothenburg (Sweden), torsten.linders@gu.se

Background.

One of the world's most complex marine challenges is the spread of invasive species. Invasive species cause severe harm to marine ecosystems and the people who depend on them, with economic impact alone amounting to several billion dollars annually. Recent advances in data science and artificial intelligence (AI) along with the increasing availability of free marine and other data online are improving the possibility to tackle these challenges. This paper presents the efforts by Ocean Data Factory Sweden (ODF Sweden), a data-driven innovation consortium in Gothenburg, to apply machine learning (ML) to one use case – the prediction of the spread of the Killer Shrimp, or Dikerogammarus Villosus, into the Baltic Sea (Figure 1). We discuss our process to address this use case as well as some reflections on the process and its challenges, in particular when taking into consideration the FAIR (findable, accessible, interoperable and reusable) principles in data science.

Figure 1: Visualizing the prediction of the spread of the Killer Shrimp into the Baltic Sea

Developing the Killer Shrimp Use Case.

Defining the questions for investigation. Killer Shrimp have already invaded Europe (e.g., Bollache et al. 2004) presumably through the ballast water of cargo ships as ocean expanses are too vast for the shrimp to traverse. The shrimp have been recorded in rivers in Western Europe, perhaps by travelling through inland waterways from the Black Sea, and more recently, it has been detected in the Baltic Sea. As this shrimp devastates the local ecosystems it invades, ODF Sweden and its partner, the Swedish Agency for Marine and Water Management (SwAM), decided to investigate whether ML could help predict the areas of the Baltic Sea suitable for the Killer Shrimp. In particular, we decided to explore these questions: 1) what are the factors that could lead to the spread of the shrimp into the Baltic Sea region?, 2) how might various scenarios, such as changes in climate or shipping routes affect these factors?, and 3) how might this species' spread Baltic affect ecosystems and local industry?

Preparation of training and test datasets. We collected the following open datasets: 1) port locations in Europe (EMODNET), 2) ocean surface temperatures and salinity for Baltic Sea (SMHI) and North Sea regions (SeaDataNet), 3) presence data of D. Villosus from observations from 1928-2019 (GBIF), 4) marine data layers (Bio-Oracle), and 5) ocean temperature and salinity (CMEMS). The features of temperature, salinity, depth, substrate and wave activity were selected based on input of marine experts. Missing data were removed, and features were visualized. We noticed that the data were very skewed towards the absence class, i.e. there was extreme high-class imbalance. To address this, we used oversampling to increase the instances of the "presence" class by creating synthetic cases based on the original presence cases. We split the data 80/20 into a training set and a test set for evaluation.

Building and refining the model. We used primarily tree-based models, a single decision tree and a Random Forest but also included Deep Neural Networks for more complex feature extraction. All models were trained with their standard configurations in scikit-learn and fast.ai Python libraries for easy replication. Models were scored on their ability to correctly predict the locations where the killer shrimp would be present. We were able to get a probability that a particular point belongs to our presence class. Throughout the entire process, we adapted our methods as new data became available and we learned more about the nature of the problem.

Visualizing our results. Since our features come in the form of rasters, using our trained models we were able to make predictions for each cell in the raster grid. The model output is then the probability of "presence" in that cell. We built a web application that helps visualize the probabilities from some of these models as well as the impact of future climate changes on these probabilities in the Baltic Sea. Specifically, we noticed the increased suitability of Åland and the Eastern coast of Sweden under future climate condition forecasts from SMHI (figure 1).

Discussion.

Following the FAIR principles, ODF Sweden worked with a principle of openness. Only open datasets were used; the project was documented on Jupyter notebooks on Kaggle; and Github was used to host our code repository. This resulted in several challenges. First, while the data are plentiful on open data platforms, they lie in multiple siloed systems without central access point or methodology. As a result, extracting and converting data took the bulk of the time. Second, when shifting to geospatial ocean data, we found that it was easy to be overconfident in our model predictions. When we simplified sample data points from a large area in the ocean and then split our datasets into training and test sets, the distribution of the training and test data were so similar that the test set effectively "leaked" into the training set. Third, each data provider has its own preferred **coordinate reference systems (CRS)**. Since the Earth is spherical, each CRS represents a projection onto a flat 2D surface for visualization. Re-projecting between these systems is often necessary when performing comparisons and calculations. Python packages, such as GDAL and Rasterio, helped simplify this. Finally, another major challenge was **interpreting inland data**. Since we had no information available about inland water sources, we had to match these to the closest body of ocean water. This proved to be difficult and inaccurate, and we had to make assumptions such as "inland water is just as salty as sea water". This led to large biases in our initial results and led us to revisit this assumption and ultimately abandon this when we obtained additional presence data in the Baltic Sea. The model output also revealed the importance of appropriate data input in answering our questions. In our case, our results reveal that we need additional data to answer more insightful questions, such as future migration patterns to predict species abundance and data on other species in the area to understand the interaction effects on overall biodiversity in areas where the Killer Shrimp has been detected. Looking forward from a user perspective, we recommend that data providers improve and align documentation standards. We also hope that datasets will become more searchable and that new datasets will be promoted to boost research efforts to answer these important questions.

References

Bollache, L., Devin, S., Wattier, R., Chovet, M., Beisel, J. N., Moreteau, J. C., & Rigaud, T. (2004). Rapid range extension of the Ponto-Caspian amphipod Dikerogammarus villosus in France: potential consequences. *Archiv für Hydrobiologie, 160* (1), 57-66.

A picture is worth a thousand data points: Making videos and images from marine environmental monitoring available to all

Markus Lindh, Swedish Meteorological and Hydrological Institute (Sweden), markus.lindh@smhi.se
Patrick Gorringe, Swedish Meteorological and Hydrological Institute (Sweden), patrick.gorringe@smhi.se
Johan Stål, Swedish Meteorological and Hydrological Institute (Sweden), johan.stål@smhi.se

"*A picture is worth a thousand words*" is a common proverb to describe the advantage of using a picture over words in advertisement. In China there is also the famous proverb: "*it is better to see something once, than to hear about it a thousand times*". In marine environmental monitoring, video and images are often the foundation of data, where one video or image can be worth thousands of data points. Yet, in many cases only analysed data is publicly available and the video and image data are kept locally inaccessible for potential users and sometimes even discarded after analysis. There are now also increasing international and national demands to make video and image data publicly available. For example, the INSPIRE directive, FAIR principles and the public sector are putting high demands on marine data collected. Taken together, there is

MYEAR	SDATE	LATIT	LONGI	LATNM	SFLAG	COVER%	
2019	2020-04-07	56.94000	12.21170	**Zostera marina**	SPP	85	
2019	2020-04-07	56.94000	12.21170	**Polysiphonia**	SP	10	
2019	2020-04-07	56.94000	12.21170	**Mytilus edulis**	SP	5	

Figure 1: A picture of an eelgrass meadow (*Zostera*; top panel) and a simplified mock example of analysed data from the same picture in a table format (bottom panel). In this example, a diver performs regional or national monitoring and aims to observe and identify different species by collecting videos and images. Analysed data (species observations) is sent to the national archive for oceanographic data for quality control and eventually becomes publicly available. However, the original raw data, i.e. images and video that forms the basis of analysed data have up until now not been archived or published.

a fundamental gap between data sampling and data access where video and image data from the marine environmental monitoring community needs to be made publicly accessible with good metadata descriptions.

The Swedish Meteorological and Hydrological Institute (SMHI) is the National Oceanographic Data Centre (NODC; UNESCO/IOC/IODE) and hold the Swedish archive for oceanographic data (commissioned by the Swedish agency for marine and water management; SwAM). We collect, archive, perform quality controls and publish data from national and regional marine environmental monitoring. SMHI also provide data for international data aggregators and regional sea conventions.

SMHI have now also developed a state-of-the-art system for archiving and publishing image and video data from marine environmental monitoring. Here, we will present our solution for a system that is capable of receiving, archiving, quality controlling and packaging data.

By making the video and image data available for the public we aim to:

- Increase data quality -> Data quickly gets outdated where analysis and re-analysis of data will keep data updated so the user has the highest available quality.
- Meet future expected and unexpected needs -> Changes in analysis methods and interpretation of data, including taxonomic re-annotation is problematic where data quickly becomes outdated.
- Utilize emergent technologies (machine learning and other AI methods) -> State-of-the-art video and image analyses can automate workflows and free resources while decreasing human bias.
- Easy data interpretation aimed toward decision-makers -> Stakeholders and decision makers involved in marine monitoring are continuously met by data reports often including complicated tables and graphs of analysed data; easy access to collections of video and image data may be a way to simplify interpretation of environmental change.
- Increase data availability and transparency -> One of SMHIs most important missions is to increase the availability and use of data including its transparency, i.e. making the underlying original raw data that is the basis for the table of analysed data publicly available.

The novel system and data management will be documented and used to create an entry for Ocean Best Practices within the JERICO-RI project and hence contribute to the Ocean Best Practices System Repository (OBPS-R). Video and image data collected at SMHI include monitoring of epibenthos, phytoplankton and megafauna (seals and harbour porpoise). We are using a metadata description applied from the INSPIRE directive and include the use of digital object identifiers (DOI) for published data packages so that data is findable. All data, including metadata, goes through a high-throughput quality control and administration system using software developed in-house at SMHI with open source code (MIT license) and is available both for Microsoft, Linux and Apple users. The data will also be made available to ongoing international projects such as JERICO-RI and for data aggregators such as EMODnet, SeaDataCloud and ICES. We anticipate that the data will be used by the regional sea conventions HELCOM and OSPAR and potentially also the European Environmental Agency.

POSTERS

EMODnet Physics from data to use cases

Antonio Novellino, ETT SpA (Italy), antonio.novellino@ettsolutions.com
Marco Alba, ETT SpA (Italy), marco.alba@ettsolutions.com
Patrick Gorringe, Swedish Meteorological and Hydrological Institute, SMHI (Sweden), Patrick.Gorringe@smhi.se
Luca Bonofiglio, ETT SpA (Italy), luca.bonofiglio@ettsolutions.com

Introduction

EMODnet-Physics map portal (www.emodnet-physics.eu/map) has been developed for the past 10 years and nowadays provides a single point of access to in situ datasets, products and their metadata. Time series and datasets are made available, as recorded by fixed platforms (moorings, tide gauges, HF radars, etc.), moving platforms (ARGO, Lagrangian buoys, ferryboxes, etc.) and repeated observations (CTDs, etc.). Apart from these data, EMODnet Physics is developing interoperability services to facilitate machine-to-machine interaction and to provide further systems and services.. These features range from widgets to WxS OCG compliant services.

Products

Datasets are organized in products and are discoverable by many interfaces: the catalogue (catalog.emodnet-physics.eu), map-viewer (emodnet-physics.eu/map), thredds (thredds. emodenet-physics.eu), erddap (erddap.emodnet-physics.eu) and geoserver (geoserver.emodnet-physics.eu). The following paragraph lists the key available products:

a. *Temperature and Salinity in the water column*

Temperature, in the water column, is a vital component of the climate system and its variability. Salinity observations contribute to monitoring the global water cycle, ocean density and mass, etc. In situ observations available in EMODnet Physics are taken from a variety of catalogues (e.g. CMEMS INSTAC, SDN, IOOS, IMOS, etc.) linking platforms with a large range of spatial and temporal scales.

b. *Sea Surface Currents*

Ocean surface general circulation is responsible for significant surface transport of heat, salt, passive tracers and ocean pollutants. The existing surface current observing systems (moorings, Lagrangian drifters) capture much of this range. EMODnet Physics is combining these observations together with land-based HF radars observation that offer a high-resolution tool (with limited spatial coverage) for improved understanding of surface currents, eddies and air-sea fluxes, and exchange between coastal waters and the open ocean.

c. *River Runoff Data*

River runoff exert a strong influence in their neighbouring coastal area in several ways, modifying water stratification, introducing significant fluctuations in circulation patterns and modulating the impact of upwelling events. EMODnet Physics developed a dedicated data infrastructure to manage and give access to in situ river runoff operational data (~170 stations) and historical trends (~550 stations - based on the Global Runoff Data Base collection).

d. Total Suspended Matter

Total Suspended Matter (unit: % of suspended particles, not dissolved) is a gridded product based on the CoastColour L2W Concentrations Data, obtained from the OC4 algorithm for clear and moderate turbid waters, and from the CoastColour v1 neural network. The L2W product is then remapped on a regular grid, maintaining 300m full resolution, in order to obtain products over the European sea basins and monthly averaged. The product covers the period 2012 – 2013.

e. Sound Maps and acoustic pollution

EU Technical Group on Underwater Noise (TG-NOISE) has made progress on this concept and has lately started discussing how to implement the MSFD indicator. It is also likely to combine and correlate sound/noise maps vs biodiversity maps. The noise map can be based on ship density maps combined to a model of noise propagation at sea. This product should be calibrated (or the map uncertainty should be assessed) by in situ data. EMODnet Physics is implementing such an approach by running the pilot for the areas in which in situ data is available, and it is developing a common method to manage and federate the in situ operational SPL data.

f. Wave and winds – Sea State

Sea State is the characterisation of wave and swell, typically in terms of height, wavelength, period, and directional wave energy flux. These data are accessible in EMODnet Physics, integrating several data sources (Data Buoy Cooperation Panel, regional observations in Europe – CMEMS INSTAC, US - IOOS, Australia – IMOS, etc.) into one single catalogue. Operational data are aggregated into a synoptic dynamic view.

Use cases

Users are at the centre of EMODnet's services and resources. The data and the data products provided by EMODnet physics are important for many applications. To understand the full benefits of EMODnet, users have been asked to describe how EMODnet supports them in their daily work and activities; if they have developed an application using EMODnet products or if they use EMODnet data for other purposes. Thanks to the developed features in EMODnet Physics, many different users have started to link and use the data. . Results and significant use cases come from SINDBAD Operational service to support navigation, CMEMS in Situ TAC where EMODnet Physics has developed a user-friendly interface to improve the viewing service developed for outreach and promotion activities. Other examples are DLR's German Remote Sensing Data Center (DFD) and Remote Sensing Technology Institute (IMF), DHI and HyMOLab-University of Trieste and others that will be presented and can be viewed at https://www.emodnet.eu/en/use-cases.

Declassification of naval data: the steps!

Sara Almeida, Instituto Hidrográfico (Portugal), sara.almeida@hidrografico.pt
Paulo Nunes, Instituto Hidrográfico (Portugal), antunes.nunes@hidrografico.pt

The Instituto Hidrográfico, as navy unit, has centralized over the years numerous data sets related to military missions. Some of them were naval fleet exercises that brought together several ships from different countries, with the common goal of carrying out oceanographic studies - examples of these were the MILOC campaigns realized between 1965 and 1969.

Those datasets collected by military ships were usually, classified and remain archived in secure stores.

Following the guidance of the Intergovernmental Oceanographic Commission (IOC) of UNESCO, and the principles of data sharing as a path for a better ocean knowledge and awareness, the Instituto Hidrográfico requested naval staff to start the declassifying process of historical BT and XBT profiles.

More than twenty-five thousand bathy messages (temperature versus depth), since 1957, soon will be made available for marine data users and scientific community.

During the military campaigns, it was not rare to use expendable bathythermographs (XBT). Each time these were launched, and after the conclusion of the station, a bathy message was sent to the national reference bodies. The information was encoded and, in addition to the temperature profile, the position and the meteorological conditions for the station itself were added, usually wind direction and speed (figure 1). After receiving the message, it was decoded and archived per year. When analysing these 50 year's data, the geographical coverage is the Portuguese EEZ: 45°N to 30°N and 40°W till 7°W, with higher concentration of coastal stations up to 10°W and some transects between the continental zone and the Islands: Azores and Madeira.

After the completion of a catalogue with all the metadata regarding the campaigns that produced bathy messages, the next step is the processing and quality control of those files. This is a huge phase!

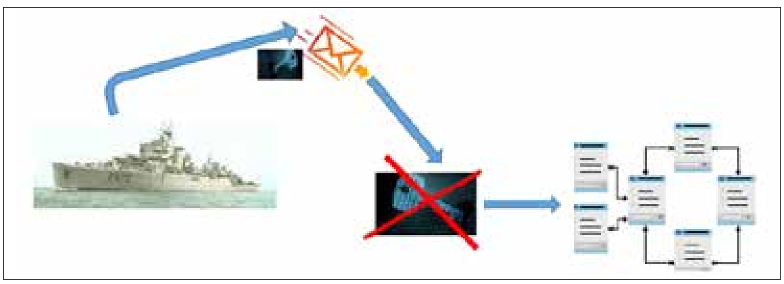


Figure 1: The BATHY messages flux

Our main goal, after authorization to declassify, is the fully availability of this historical datasets.

Data Management

Although BATHY messages were constructed by selecting significant data points in a temperature-depth profile, corresponding therefore to what we may call low-resolution data, these datasets can contribute to the knowledge of the ocean.

To ensure significance of the results one must insure the quality of the data stage, so it is important to submit it to the procedures and protocols that have proved reliable.

Looking at the existing documentation, there are several relevant publications on this topic. Therefore, we will produce a working method leading to the creation of formatted profiles compatible with the usual tools, for instance, the ones from SeaDataNet and applied on EMODnet.

A special case

More than the campaigns with a wide geographical distribution, there are some interesting observations in this set. This is the case of several years of observations at the same location. During seven years, between 1962 and 1968, a fixed station was held at Lisbon with regular launches per day and along the years (figure 2). This is a very special case that will be challenging to deal with, but we hope will bring forward new insights!

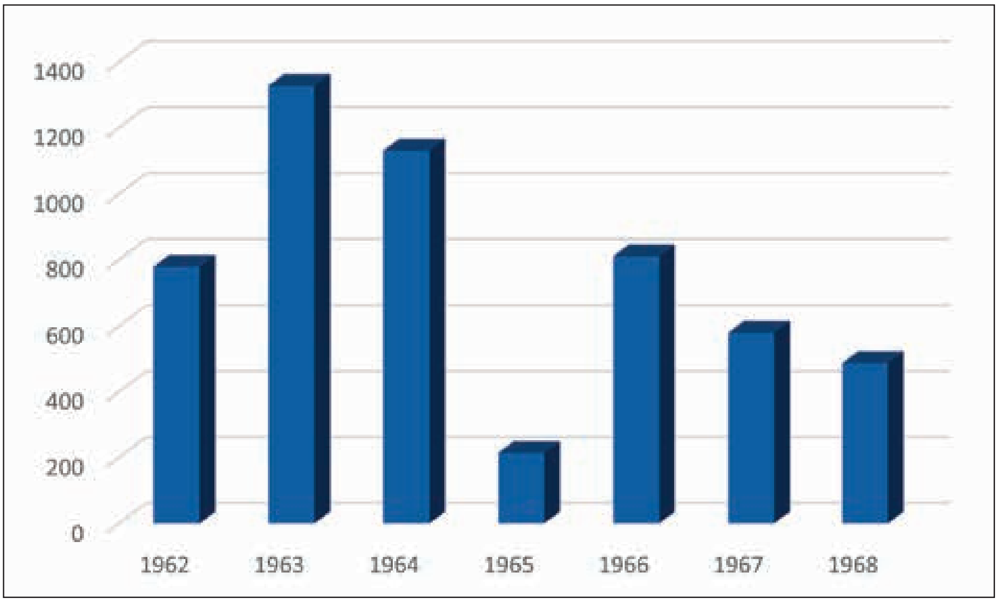


Figure 2: Observations at Lisbon for 7 years

References:

https://www.oceanbestpractices.net/, 24march2020

Bailey, R.; Gronell, A.; Phillips, H.; Tanner, E. and Meyers, G. - Quality Control Cookbook for XBT Data. CSIRO Marine Laboratories, Report 221, Version 1.1

ICES Data and Information Group - Data_Guidelines_XBT_v7_revised_2006.pdf

IOC Circular Letter nº 1439 – Declassification of Oceanographic Data

JCOMM_Technical Report-1-Status-Reports, July 1999

Jerico-Next - Report on data management best practice and Generic Data and Metadata models. Version 2.1

Automated extraction and fusion of the intertidal and subtidal bathymetry from the Landsat and Sentinel satellite data

Gennadii Donchyts, Deltares (NL), Gennadii.Donchyts@deltares.nl
Giorgio Santinelli, Deltares (NL), Giorgio.Santinelli@deltares.nl
Arjen P. Luijendijk, Deltares (NL), Arjen.Luijendijk@deltares.nl
Antonio Moreno Rodenas, Deltares (NL), Antonio.MorenoRodenas@deltares.nl
Martin Verlaan, Deltares (NL), Martin.Verlaan@deltares.nl

Monitoring of the intertidal and subtidal bathymetry at large spatiotemporal scales using traditional surveying methods is a challenging and costly task. With the abundance of freely available satellite data in the last years and the availability of parallel processing platforms like Google Earth Engine, an automated derivation of bathymetry from satellite data sounds very attractive. A number of methods exist to extract bathymetry from satellite data, including methods to derive bathymetry for the **intertidal** zone [1], by combining water/land boundary dynamics with the water level measurements. At the same time, the light attenuation in a water column, observed by optical satellite sensors, can be used to infer water depth from spectral reflectance, providing a way to estimate **subtidal** bathymetry [2].

In this research, we will discuss both of these methods and will explore how they can be combined to generate consistent intertidal and subtidal bathymetry data. Our algorithm is based on the use of multitemporal optical satellite data to estimate water occurrence for the intertidal zone and the use of inverse-depth methods to derive bathymetry for the subtidal zone. Furthermore, we combine both bathymetric products to derive a single consistent bathymetry dataset (Figure 1).

The algorithm has been applied to derive bathymetry for the Dutch coast but it is transferable to other coastal areas. A great advantage in the estimation of satellite-derived bathymetries is the high temporal frequency (3-5 days) of sampling provided by the satellite sensors, addressing the need for data on short-duration morphological changes.

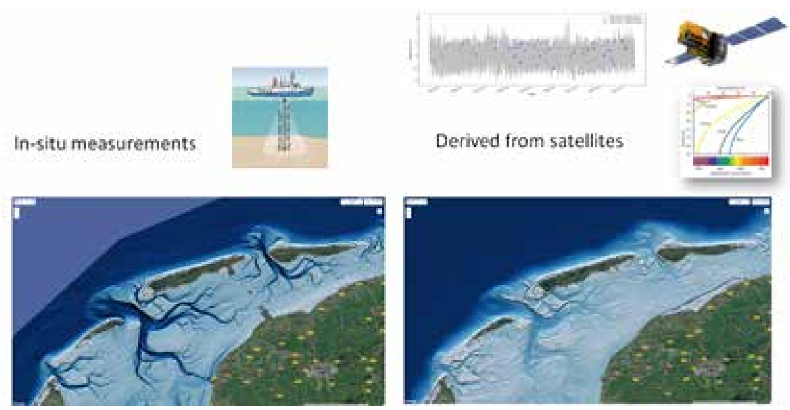


Figure 1: Vaklodingen Bathymetry (in-situ data source for Dutch waters) on the left and Satellite-Derived Bathymetry on the right. Source: https://earthengine.google.com.

Our method is implemented within the Google Earth Engine parallel processing platform and makes use of NASA/USGS Landsat 8 and ESA/Copernicus Sentinel-2 satellite imagery. The outline of the method is shown in Figure 2. We apply a two-step approach to remove cloudy pixels from the input data. The first step includes filtering of cloudy images by means of computing cloudiness of observed TOA reflectance over small image patches. This metric is combined with the cloud frequency estimated from MODIS data [3]. The second step includes per-pixel cloud masking followed by intertidal and subtidal bathymetry estimation algorithms. After calibration of the subtidal bathymetry using in-situ measurements, the end product is a fusion of high-resolution, high temporal intertidal and subtidal bathymetry.

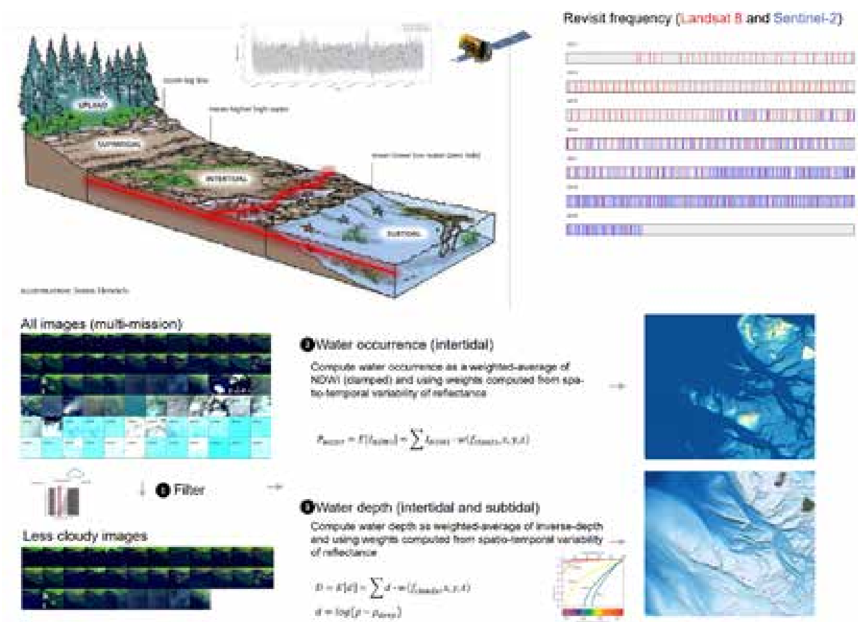


Figure 2: Overview of the method used for the processing of Landsat 8 and Sentinel-2 satellite imagery, combining intertidal (2) and subtidal (3) bathymetry generation.

The results of this research were used to improve coastline detection in the European Marine Observation and Data Network (EMODnet) bathymetry. Future developments of the algorithm will include fusion of intertidal and subtidal bathymetry for large extents, or even globally, as well as more extensive calibration and validation with in-situ data.

Potential applications include automated monitoring and mapping of bathymetry, operating as a service allowing to turn Earth Observation (EO) data into bathymetric products, assuming the presence of inter-tidal zone for intertidal and sufficient water clarity for sub-tidal bathymetries.

The research was supported by the European Commission and the Dutch Ministry of Infrastructure and Water Management.

References

Sagar, Stephen, et al. "Extracting the intertidal extent and topography of the Australian coastline from a 28 year time series of Landsat observations." Remote sensing of environment 195 (2017): 153-169.

Lyzenga, David R. "Passive remote sensing techniques for mapping water depth and bottom features." Applied optics 17.3 (1978): 379-383.

Wilson AM, Jetz W (2016) Remotely Sensed High-Resolution Global Cloud Dynamics for Predicting Ecosystem and Biodiversity Distributions. PLoS Biol 14(3): e1002415. doi:10.1371/journal. Pbio.1002415

Marine Databases produced by the UK Met Office

Rachel Killick, Met Office Hadley Centre (UK), rachel.killick@metoffice.gov.u k
Nick Rayner, Met Office Hadley Centre (UK), nick.rayner@metoffice.gov.uk
Chris Atkinson, Met Office Hadley Centre (UK), chris.atkinson@metoffice.gov.uk

The Met Office Hadley Centre produces numerous observational datasets. The EN4 and HadIOD (Hadley Centre Integrated Ocean Database) datasets are two of those focusing on ocean data. Both are freely available for research and private use from https://www.metoffice.gov.uk/hadobs. Present uses of these datasets include, but are not limited to, the creation of ocean reanalyses, model and forecast verification and ocean heat content monitoring.

The EN4 dataset provides subsurface temperature and salinity profiles from 1900 to the present day alongside globally complete, infilled analyses which are accompanied by uncertainty estimates down to 1000m. The current version of EN is EN.4.2.1 (Good et al., 2013), with EN.4.3.0 due for release in 2021.

The EN4 dataset takes input data from four sources, the World Ocean Database, Argo, the Arctic Synoptic Basin wide Oceanography database and the Global Temperature and Salinity Profile Programme (GTSPP). When monthly updates are performed, with approximately half a month lag, data are taken from Argo and GTSPP only. Monthly netCDF files are available, with data contained at a higher resolution within these files for profiles.

The profiles in EN4 are fully quality controlled using an automatic QC procedure, with flags indicating which checks have been failed, allowing users to interrogate the data in greater detail. If a profile has failed any QC checks it will not be passed into the analyses. For eXpendable BathyThermographs (XBTs) and Mechanical Bathythermographs (MBTs) bias corrections are also provided.

The EN4 analyses are on a one by one degree horizontal grid with 42 depth levels, of increasing size, down to 5500m. These analyses are produced by combining background fields with the quality controlled observations and will relax back to climatology in the persistent absence of observations. The 'observations weights' variables provide information on the weighting of observations to background. These observation weights fields can be seen for three months of the temperature analyses in Figure 1, illustrating, alongside the profile coverage plots on the left, the increase in spatial coverage over time. The profile plots also indicate the increase in depth coverage.

EN.4.3.0 will include a number of changes relative to EN.4.2.1, focusing on improving uncertainty representation. A recent machine learning project providing a probabilistic estimate of XBT types in the case of missing metadata will be used to expand the bathythermograph correction ensemble. The method of calculating uncertainties in the analyses will be revisited, making use of the additional decade of data since EN4 was first produced. Measurement uncertainty estimates from HadIOD will be incorporated into the EN4 profiles and then feed into the analyses and the representivity errors encountered when going from point based observations to infilled analyses will be investigated.

HadIOD combines surface temperature data (from 1850 onwards) with subsurface temperature and salinity data (from 1900), providing an integrated database ideal for use where both surface and subsurface data are required, for example, in coupled reanalyses. HadIOD.1.2.0.0 (the current version) takes subsurface data from EN4 and surface data from ICOADS release 2.5.1 (Woodruff et al., 2011), the Copernicus Marine Environmental Monitoring Service (CMEMS) and a few additional smaller sources. For each observation HadIOD provides quality flags and measurement uncertainty estimates, as well as bias corrections where these are available (for MBTs, XBTs and ships). HadIOD data are released as daily files, updated on a monthly basis with a lag of a few months.

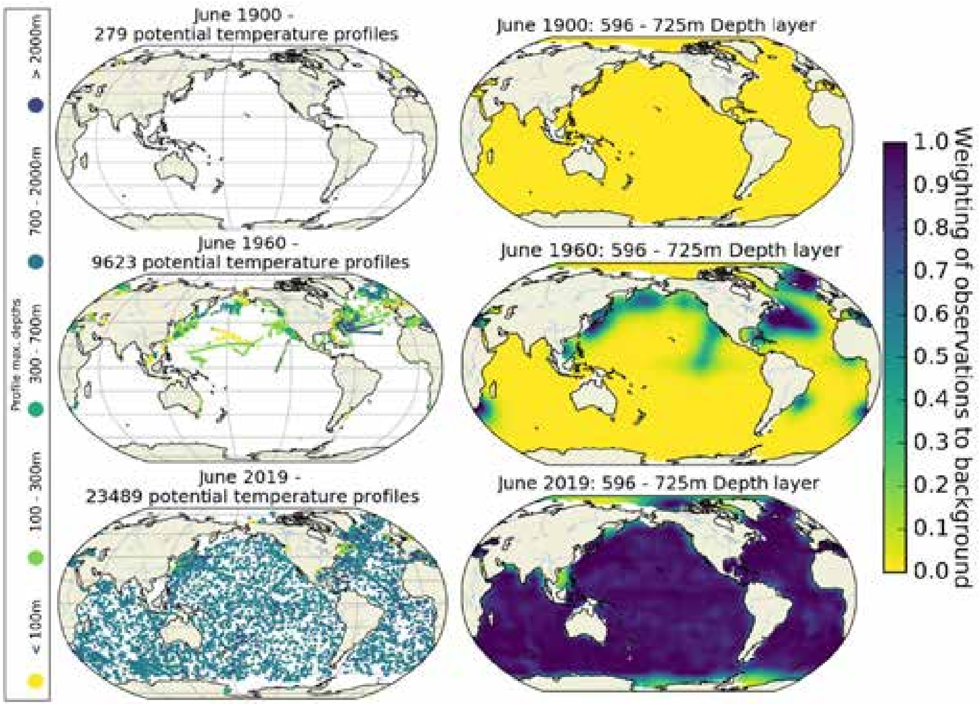


Figure 1: EN4 over time. Plots on the left indicate profile coverage, colour coded by profile maximum depth, whilst plots on the right indicate the relative weighting of observations to background fields in the analyses at an intermediate depth layer.

References:

Argo: http://wo.jcommops.org/cgi-bin/WebObjects/Argo

ASBO: See Good et al., 2013 for which input sources of this dataset are taken

Atkinson et al., 2014: C. P. Atkinson, N. A. Rayner, J. J. Kennedy and S. A. Good (2014): An Integrated Database of Ocean Temperature and Salinity Observations, Journal of Geophysical Research: Oceans, 119, 10, pp 7139 – 7163, doi: 10.1002/2014JC010053

CMEMS: http://marine.copernicus.eu/. Product INSITU_GLO_NRT_OBSERVATIONS_013_030 (Global Ocean In Situ Near Real Time Observations). Drifting Buoys.

Good et al., 2013: S. A. Good, M. J. Martin and N. A. Rayner (2013): EN4: Quality Controlled Ocean Temperature and Salinity Profiles and Monthly Objective Analyses with Uncertainty Estimates, Journal of Geophysical Research: Oceans, 118, 12, pp 6704-6716

GTSPP: Sun, C. & Co-authors (2010): "The Data Management System for the Global Temperature and Salinity Profile Programme" in Proceedings of OceanObs.09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21 – 25 September 2009, Hall, J., Harrison, D. E. & Stammer, D., Eds., ESA Publication WPP-306, doi:10.5270/OceanObs09.cwp.86

WOD: Boyer, T. P., J. I. Antonov, O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, D. R. Johnson. R. A. Locarnini, A. V. Mishonov, T. D. O'Brien, C. R. Paver, J. R. Reagan, D. Seidov, I. V. Smolyar, and M. M. Zweng, 2013: World Ocean Database 2013, NOAA Atlas NESDIS 72, S. Levitus, Ed., A. Mishonov, Technical Ed.; Silver Spring, MD, 209 pp., http://doi.org/10.7289/V5Nz85MT

ICOADS: S. D. Woodruff, S. J. Worley, S. J. Lubker, Z. Ji, J. E. Freeman, D. I. Berry, P. Brohan, E. C. Kent, R. W. Reynolds, S. R. Smith and C. Wilkinson (2011): ICOADS Release 2.5: Extensions and enhancements to the surface meteorological archive, Int. J. Climatology, 31, 951-967

Increasing quality in ICES commercial fisheries data using the Regional Database and Estimation System

David Currie, Marine Institute (Ireland), david.currie@marine.ie
Liz Clarke, Marine Scotland Science (UK), liz.clarke@gov.scot
Edvin Fuglebakk, Institute of Marine Research (Norway), edvin.fuglebakk@hi.no
Kirsten Birch Håkansson, National Institute of Aquatic Resources (Denmark), kih@aqua.dtu.dk
Henrik Kjems-Nielsen, ICES, Henrikkn@ices.dk
Nuno Prista, Swedish University of Agricultural Sciences (Sweden), nuno.prista@slu.se

Assuring quality 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. One of its key roles is to provide scientific advice for a wide range of recipients including its 20 member countries, international organisations including the European Commission, and other end-users.

ICES expert groups require a user-friendly framework to archive data, methods, and results used in fish stock assessments. For recurrent assessments and advice, it is important to ensure that the outputs can be replicated at a future date and also be re-run the following year. ICES assessments are carried out using various disparate data sources so it is important to ensure transparent input data processing and reproducible assessment methods, while also automating and documenting the outputs which form the evidence base. The aim should be that data are available at the highest possible resolution whilst taking into account confidentiality constraints. Assuring quality is a key element of the ICES advice plan (ICES, 2019) and the commercial fisheries Regional Database & Estimation System (RDBES) will be an essential tool in the overall quality assurance framework.

Regional Database and Estimation System (RDBES)

The Regional Database and Estimation System (RDBES) will replace the current ICES Regional Database (RDB) and InterCatch systems (Currie et al. 2018, ICES 2020) – its aims are: 1) to ensure that commercial fisheries data can be made available for the coordination of regional fisheries data sampling plans, including for the EU DCF Regional Coordination Groups (RCGs); 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 RDBES will store aggregated landings and fishing effort data, and detailed biological sample data. The significant novelty in the RDBES data model is that it provides a common structure to describe both the disaggregated sampling data and, most importantly, the sampling design underlying how those data were obtained. For the purposes of fish stock assessment it is necessary to combine detailed biological data with census data of fishing fleet activity to produce an estimate of the removals from the fish stock due to fishing mortality. The RDBES data model allows a variety of different estimation techniques to be used including unbiased design based estimation methods. The RDBES should be seen as part of the movements towards Statistically Sound Sampling Schemes (4S), greater regional coordination, and improved estimates to ICES stock assessments and advice.

Transparent Assessment Framework (TAF)

The RDBES web application will provide certain functionality such as data uploading, and managing permissions but stock estimation and imputation will be performed within the ICES Transparent Assessment Framework (TAF).

The TAF is an open framework for organising stock assessments built on Git repositories. All data inputs and outputs are fully traceable and versioned. The open framework enables stock assessment scientists to easily find, reference, download, and run the assessment from any stage in the process leading to the published ICES advice for a given stock. Anyone will also be able to find, reference, and download the estimation method behind the assessment.

Basing the stock estimation functions of RDBES on the TAF has a number of advantages: the TAF exists and users are already gaining expertise in it, there is technical and content support available, version control of data and scripts is established, and it provides strong linkages to stock assessment groups.

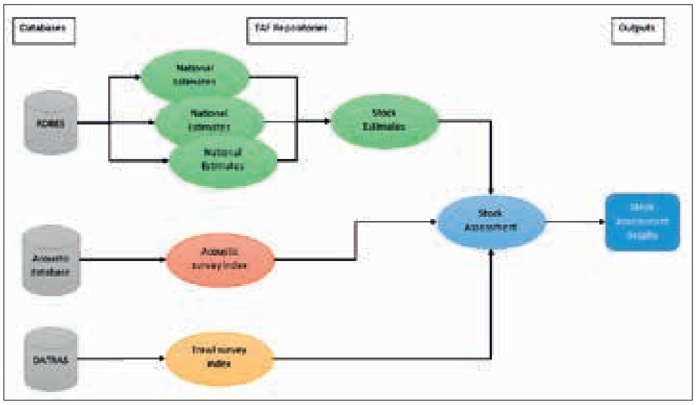


Figure 1 Example of a data processing pipeline using different TAF repositories as building blocks

Next steps

The RDBES will continue to be developed during 2020 and 2021 and it is scheduled to move into production in 2022. To this end a number of workshops have been planned which will both help data submitters with the transition to the new system and encourage more people to be involved in the process. The upcoming availability of an ICES database that explicitly incorporates important details of sampling design will allow the development of improved fisheries estimates and result in improvements in sampling at both national and regional levels. The success of the RDBES will rely on the effort and contributions from a large number of people in the wider ICES/EU commercial fisheries data collection community.

References

ICES Advisory Plan, 2019 https://issuu.com/icesdk/docs/ices_advisory_plan

D. Currie, L. Dubroca, E. Fuglebakk, K. Birch Håkansson, H. Kjems-Nielsen, T. Leijzer and N. Prista, Towards a Regional Database and Estimation System for Fisheries, International Conference on Marine Data and Information Systems 5-7 November, 2018

ICES 2020. Steering Committee of the Regional Fisheries Database (SCRDB; outputs from 2019 meeting). ICES Scientific Reports. 2:24. 57 pp. http://doi.org/10.17895/ices.pub.5992

ICES Transparent Assessment Framework https://www.ices.dk/marine-data/assessment-tools/Pages/transparent-assessment-framework.aspx

Development of the ICES Continuous Underwater Noise reporting format and database

Joana Ribeiro, International Council for Exploration of the Sea – ICES (Denmark), joana.ribeiro@ices.dk
Carlos Pinto, ICES (Denmark), carlos@ices.dk
Mehdi Abbasi, ICES (Denmark), mehdi.abbasi@ices.dk
Neil Holdsworth, ICES (Denmark), NeilH@ices.dk

Introduction

Underwater noise is an anthropogenic pressure that can have a negative effect on the organisms that are exposed to it[1], and it is recognised as a pressure on the marine environment by the EU Marine Strategy Directive (MSFD). Underwater noise can be of impulsive or continuous nature. The latter, also referred to as ambient underwater noise, is a low frequency noise that occurs over an extended period of time, as opposed to being associated to a particular event.

Ambient underwater noise has been monitored in the North Atlantic as part of time limited projects such as the Baltic Sea Information on the Acoustic Soundscape (BIAS)[2] and the Joint Monitoring Programme for Ambient Noise North Sea (JOMOPANS)[3]. As part of the BIAS project, a Soundscape tool for the Baltic Sea was delivered. The Soundscape tool was used in the Helsinki Commission (HELCOM) Second Holistic Assessment of the Ecosystem Health of the Baltic Sea (HOLAS II). In 2019, the HELCOM expert group on underwater noise (EN-Noise) proposed to establish a continuous noise database, and transfer the hosting of the Soundscape tool to an international data platform. After a tender specification was published, the International Council for Exploration of the Sea (ICES)[4] was selected as the preferred contractor. The database has the purpose of hosting continuous underwater noise monitoring data collected by HELCOM countries. These data are to be used in the upcoming HOLAS III assessment.

Reporting format and data submission

The HELCOM countries (bordering the Baltic Sea) collect continuous or duty cycled sound recordings from stationary monitoring stations, which is referred to as raw data. Prior to submission to ICES, data are processed and reported in mean sound pressure levels (dB re. 1 uPa) at regular intervals and in regularly spaced 1/3-octave bands. The data are organised around deployments, and countries are to report data on how the deployment was made along the mean sound pressure levels. The ICES Continuous Underwater Noise format[5] is based on the data structures used by BIAS and JOMOPANS, and it was developed by the ICES Data Centre in cooperation with EN-

1 https://ec.europa.eu/environment/marine/pdf/MSFD_reportTSG_Noise.pdf
2 https://biasproject.wordpress.com/
3 https://northsearegion.eu/jomopans/
4 https://ices.dk/Pages/default.aspx
5 https://ices.dk/data/Documents/ContinuousNoise/ICES-Continuous-Underwater-Noise-format.zip

Noise. Data are submitted to the Continuous Underwater Noise database through the website[6] (figure 1) in HDF5 files, which are organised in three groups: a) File Information, b) Metadata, c) Data. Each of these groups consists of several datasets. During submission, data undergo quality checks aimed at controlling for file integrity, format compliance and quality control. The Continuous Underwater Noise database was opened for testing in March 2020, and it is planned to open for HELCOM submissions as soon as the community finish the tests.

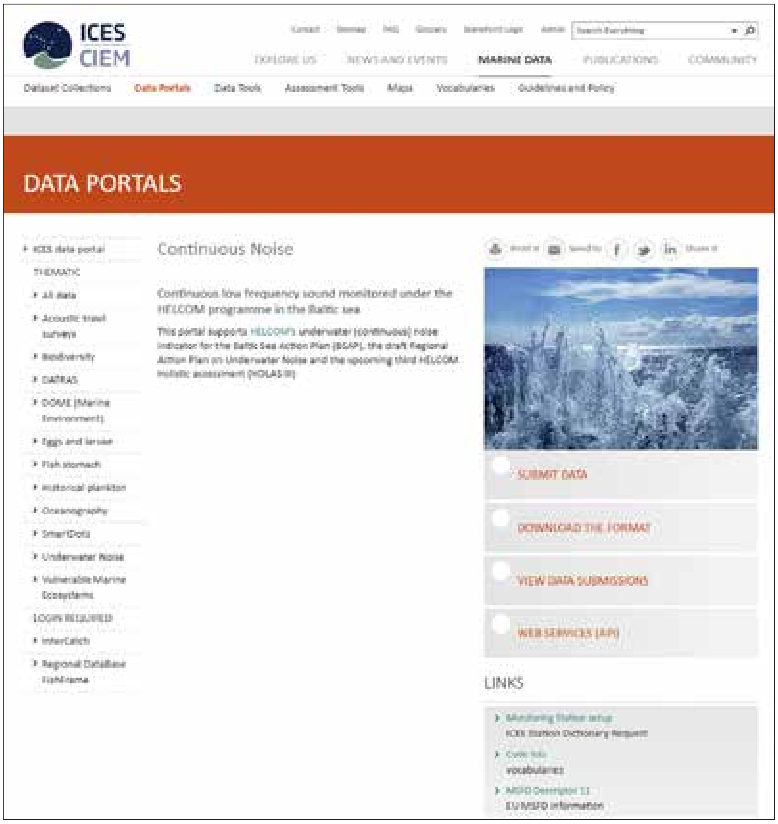


Figure 1: Continuous underwater noise website

The database consists of two components: an SQL server database, and data files. Currently, the files are kept outside of the database and this segregation is in place due to the size and non-flat nature of the HDF5 files. To submit files up to 2 GB, the data can be uploaded to ICES using an upload form. To submit files larger than 2 GB, the user has to provide a link to submit the file to the database.

Data extraction and visualisation

Users will be able to download data from the database. The data will be delivered in HDF5 format and the user will able to filter data based on a time period, an area and range of channels. If the period of time covers more than one deployment, the user will receive as many files as deployments. If the period of that covers only part of a deployment, the user will receive a file that has been trimmed to include only the data corresponding to the requested time period. Users will be able to download data under the conditions detailed in the ICES data policy[7] as soon as the database is populated. After the database is populated, maps of summary data per station will be produced as a visualization tool for the data hosted by ICES.

6 https://underwaternoise.ices.dk/continuous

7 http://ices.dk/marine-data/Documents/ICES-Data-policy.pdf

Using citizen science to rescue tide gauge data

Andrew Matthews, National Oceanography Centre (United Kingdom), antt@noc.ac.uk
Elizabeth Bradshaw, National Oceanography Centre (United Kingdom), elizb@bodc.ac.uk
Kathy Gordon, National Oceanography Centre (United Kingdom), kmg@noc.ac.uk

The need for tide gauge data rescue

Tide gauges are the main source of data that underpin our knowledge of coastal sea level. They have been essential to our study of tidal science, but also how sea level means and extremes will change as sea level rises. They make a vital contribution to the study of ocean currents and local land movement, and also navigation around harbours (Marcos et al., 2019). Unfortunately, the distribution of stations in the global tide gauge network is heavily weighted towards the Europe and North America, particularly when considering long term installations (Figure 1).

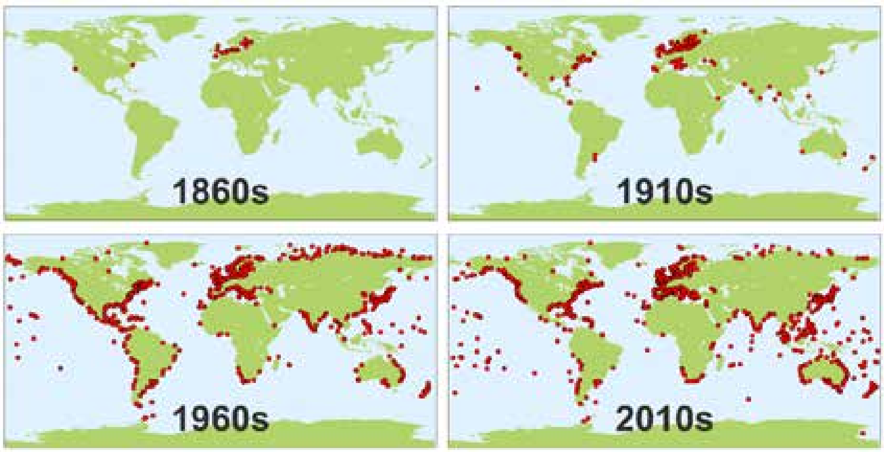


Figure 1: The development of the global tide gauge network over time, as held by PSMSL (https://psmsl.org)

There is a clear need for extra data to improve our knowledge in data sparse areas, but also to improve the quality of our predictions by extending the length of the time series available. Such data already exists in the form of charts and ledgers that have never been converted into a digital format (Figure 2). The International Oceanographic Commission's Global Sea Level Observing System (GLOSS) programme has been encouraging organistations to search for such material, and investigating efficient methods of digitisation (Bradshaw et al., 2015).

Digitisation Difficulties

There have recently been several successful studies extending long records via the digitisation of old data, for example, Talke et. al (2018) digitised 50 years of historical data to extend the record at Boston, USA to cover the period 1825-2018. However, the amount of effort can be significant: that example required the recovery of data and metadata from several national, state, and university libraries and archives, and a team of student digitisers to process them. The use of automated scanning using optical character recognition enchanced by machine learning seems an obvious solution, however current attempts have been thwarted by regular changes in formats of historical material, and by the variations of handwritten characters and numerals, particularly in older documents.

Citizen Science: a possible solution ?

One possible avenue to explore is the use of large groups of volunteers to digitise data, an approach which has been successful in related areas. For example, Hawkins et al. (2019) describes an effort to digitise 1.5 millions of observations taken from the summit of Ben Nevis (the point of maximum elevation in the UK) between 1883-1904: the project was completed in under 3 months by over 1500 volunteers. These projects require careful thought in how to make the project understandable, attractive, and interesting to potential volunteers: in this case the tale of a team of three observers living at the top of Scottish mountain throughout the year undoubtedly helped.

Large online platforms such as https://www.zooniverse.org/ provide the infrastructure required to create these projects and accumulate the volunteers' efforts. The team behind the Ben Nevis project are currently operating a project to digitised monthly rainfall totals from the UK between the 1820s and the 1960s (https://www.zooniverse.org/projects/edh/rainfall-rescue), which at the time of writing has digitised about 4 millions observations in a week with the help of about 15,000 volunteers.

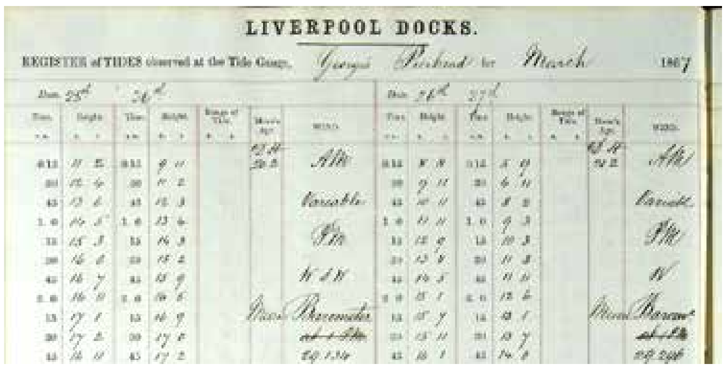


Figure 2 : An example of a ledger containing tide gauge data

We are in the process of creating a citizen science project to digitise sea level data. The project is in currently in active development, and will focus on the recovery of data from approximately 15,000 ledgers recording either 15 minutes observations or the height and time of high and low waters from two sites in the Merseyside area: the historic George's Dock in Liverpool, and Hilbre Island, a tidal island off the Wirral peninsula. We will need to define the process volunteers will undertake, tackle challenges in converting outputs to usable data, and create a plan to archive it in a long term data bank. Possible future extensions of the project include digitising other areas, and including tidal charts in future citizen science projects.

References

Bradshaw, E., Rickards, L. & Aarup, T. (2015). Sea level data archaeology and the Global Sea Level Observing System (GLOSS). GeoResJ, 6, 9-16. https://doi.org/10.1016/j.grj.2015.02.005

Hawkins, E., Burt, S., Brohan, P. et al. (2019). Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. Geoscience Data Journal, 6 (2), 160-173. https://doi.org/10.1002/gdj3.79

Marcos, M., Wöppelmann, G., Matthews, A. et al. (2019) Coastal Sea Level and Related Fields from Existing Observing Systems. Surv Geophys 40, 1293–1317. https://doi.org/10.1007/s10712-019-09513-3

Talke, S. A., Kemp, A. C., & Woodruff, J. (2018). Relative sea level, tides, and extreme water levels in Boston harbor from 1825 to 2018. Journal of Geophysical Research: Oceans, 123, 3895– 3914. https://doi.org/10.1029/2017JC013645

HarmoNIA project: Searching for data and tools to support pollution assessment and response

Maria Eugenia Molina Jack, OGS Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), mmolinajack@inogs.it
Marina Lipizer, OGS Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), mlipizer@inogs.it
Giordano Giorgi, Istituto Superiore per la Protezione e la Ricerca Ambientale (Italy), giordano.giorgi@isprambiente.it
Stefanos Papazisimou, Region of Western Greece (Greece), s.papazisimou@pde.gov.gr
Boris Petelin, National Institute of Biology (Slovenia), boris.petelin@nib.si
Luka Žilić, Ruđer Bošković Institute (Croatia), Luka.Zilic@irb.hr
Ivan Vučić, IOF Institute of Oceanography and Fisheries (Croatia), vucic@izor.hr
Christina Zeri, HCMR Hellenic Centre for Marine Research (Greece), chris@hcmr.gr
Ana Castelli, Public institution University of Montenegro - Institute of Marine Biology (Montenegro), ana.bulatovic24@gmail.com
Rigers Bakiu, Agriculture University of Tirana, Department of Plant Protection (Albania), rigers.bakiu@ubt.edu.al
Mateja Poje, Slovenian Environment Agency (Slovenia), mateja.poje@gov.si

HarmoNIA project

The Adriatic and Ionian Seas are crucial for the blue growth of both EU and non-EU coastal states. However, increased human use of the marine and coastal space might compromise marine ecosystems through several kinds of physical, chemical and biological disturbances.

In order to promote the sustainable use of the seas, the Marine Strategy Framework Directive (MSFD) and the Protocols for the Protection of the Mediterranean Sea against Pollution of the Barcelona Convention require EU Member States and all UNEP/MAP Contracting Parties to take measures to maintain or achieve Good Environmental and Ecological Status in the European seas by 2020. However, the level of coherence in the Adriatic – Ionian (ADRION) marine subregions in relation to the implementation of the environmental policies is considered low, particularly in the case of pollution from hazardous substances.

In this framework, the Interreg project HarmoNIA has focused on improving available information related to pollution by hazardous substances in the ADRION area and on implementing tools to support pollution assessment and response.

Pollution assessment

Due to the legal regulations in place for the Mediterranean, there is already a wide coastal and marine monitoring undertaken in the Adriatic and Ionian Seas. However, there is a high heterogeneity in the elements related to contaminants sampling and measurements.

Environmental assessment needs information about monitoring protocols applied in the station network of the different countries, in order to evaluate the comparability of the data that are actually being collected.

In order to facilitate understanding of the specific sampling procedures in a marine region (e.g. variables measured, monitoring purpose, frequency of observations, type of sampling stations…), the analysis of metadata from monitored stations is needed.

Regional data collections contain a substantial amount of information from an area, provided in a harmonised way. Within HarmoNIA project, a QA/QC dataset focused on contaminants, composed of all unrestricted data available for the Adriatic – Ionian Seas and data specifically gathered during the project, was produced. The metadata of this collection were analysed and processed to obtain information about the sampling efforts in the ADRION area.

GIS tools to gather information about sampling stations

Visualisation is a useful tool to organize and synthesize monitoring data to produce practical and understandable information for several kinds of stakeholders and decision makers. The production of informative and visually engaging outputs is as important as monitoring and data management.

The use of geoprocessing operations to manipulate the metadata of the sampling locations can derive information addressed to cover needs from different users or stakeholders: monitoring/ research purpose, sampled matrix, group of substances, sampling frequency.

The structure and the information considered useful was defined according to the needs and suggestions of HarmoNIA partnership and taking into account best practices adopted in the framework of Regional Sea Conventions (e.g. Barcelona Convention, OSPAR, HELCOM).

The format of the initial QA/QC controlled dataset was ODV transposed, decomposed spreadsheet file. One file for each matrix (water, sediment and biota) was composed using the profile and timeseries files, in order to have only one file for each matrix. The datasets were processed to harmonise the information about time and depth. A join of the datasets with information about water body types using EEA reference spatial datasets was performed.

Once data processing was finished, information was aggregated for each station. The identification of stations based on position is challenging because SeaDataNet formats have a different definition for the concept of station. However, the labelling of the long -term sampled stations is relevant for stakeholders. For this reason, an aggregation of the data to obtain single information for each sampling position was performed.

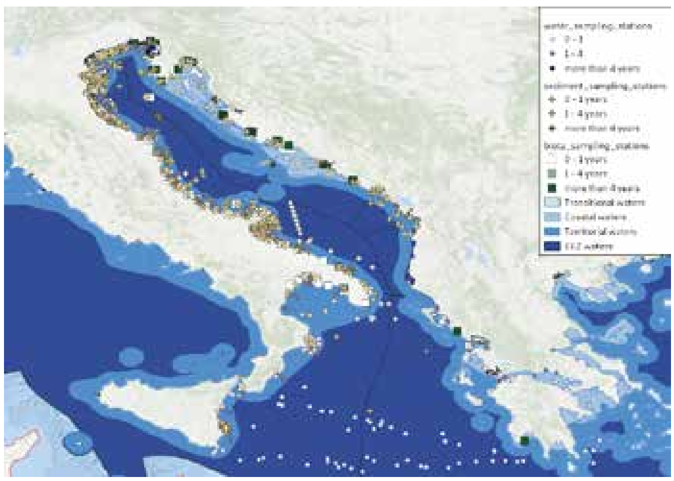


Figure 1: Sampling stations of the Adriatic-Ionian sub-region

The final layers can be displayed in different visualisation systems and overlapped with other layers of information such as human activities, vulnerable areas… This process will yield new information that will contribute to increase the knowledge of the area as well as to evaluate the appropriateness of the current monitoring network.

Data time-series formation and studies of long-term trends for marine water quality indicators in the coastal zone regions of the north-western Black Sea

Yuriy Ilyin, Ukrainian Hydrometeorological Institute (Ukraine), ypilyin@gmail.com

Introduction.

Variability of hydrological and chemical indicators in the marine coastal and estuarine waters is the product of both natural and anthropogenic influences. Multi-annual (decadal) trends may be connected with climate change effects (global warming, fresh water inflow, sea level rise, acidification, etc.) as well as systematic reducing/increasing of pollutant emissions or improvement/worsening of waste water treatment. The aim of this work was to identify linear trends in water quality indicators by means of non-parametric statistical testing with account of serial correlation of time series. Results are demonstrated for the important sites of the Black Sea coastal zone: harbor of city Odessa; harbor and seashore of city Chernomorsk (former Ilyichevsk); Kiliya delta of the Danube river.

Data and methods.

In the framework of the international project EMODnet Chemistry [1], the array of hydrological and chemical data obtained in 1992-2016 was created. Sampling and primary data processing were performed by marine units of the Ukrainian hydrometeorological network in the marine coastal regions (listed above and others). For each of the monitoring areas spreadsheets were created containing the meta-information and data on all physical and chemical parameters that were measured. Data quality control was performed and the metadata bases for the specified areas on the websites of SeaDataNet and EMODnet Chemistry projects were replenished. Data of these observations in the format of Ocean Data View [4] were uploaded on the server of Ukrainian Hydrometeorological Institute.

Further analysis considers three groups of quality indicators of sea water that are regularly researched during 1992 – 2016, i.e. for them a homogeneous time series can be built:

- general water quality (temperature, °C, salinity, PSU, dissolved oxygen concentration, mmol/kg, and saturation, %, pH and alkalinity, mmol/m^3);
- content of dissolved nutrients (silicates, phosphates, total phosphorus, nitrites, nitrates, ammonium and total nitrogen, mmol/m^3);
- content in water of technogenic pollutants (total petroleum hydrocarbons (TPHs), anionic detergents and total phenols, mg/m^3).

For the studies of long-term (decadal) variability, every observational site may be considered as the single point, while shallow coastal waters can be described by the averaging of measurements on surface, bottom and intermediate depths. So, the following algorithm of the time-series construction and analysis was adopted for each region.

- To identify all oceanographical stations where observations were executed during the entire 1992-2016 years period.
- To determine the mean annual values of all indicators for the water column together with other statistics (standard deviation, errors) by means the data averaging for all the selected stations and all sampling depths.

- To estimate the linear trends and their statistical significance for each water quality indicator time series.

Reliable estimates of the significant long-term tendencies of variability can only be obtained through statistical methods applied for time series, the main feature of which is usually serial correlation, i.e. each successive value of a series depends somewhat on previous values (this is especially true for the series of averaged values). This means that the residuals after extracting the difference between the initial and estimated values are not statistically independent and standard methods for trend assessing should not be applied without additional tests or correction procedures [5]. Moreover, indicators of chemical status and pollution only theoretically correspond to the normal (Gaussian) probability distribution, and in practice almost never. To avoid the serial correlation effect, Prais-Winsten regression method was applied to the trend parameters estimation while non-parametric Mann-Kendall test was used for the detection of significant trends [2]. Both procedures are implemented in a freely accessible statistical analysis software PAST [3].

Results.

Results obtained by means the described technologies are illustrated on the Fig. 1.

	Water temperat.	Salinity	Dissolved oxygen	Oxygen saturation	pH	Alkalinity	Silicate	Phosphate	Total phosphor.	Nitrite	Nitrate	Ammon.	Total nitrogen	TPHs	Detergen.	Total Phenols
Danube, Kiliya delta							+	−	−			−	+	−	−	−
Chernomorsk harbor and seashore	+	−	−	−			−			−	−	−	−	−	−	−
Odessa harbor		−	+		+		−	−	−	−	−	−	−	−	−	−

Figure 1: Trend analysis results for the NW Black Sea coastal zone regions.
+/- means the increasing/decreasing long-term trend significant on the 95 % level

It should be noted that only Chernomorsk's harbor and open sea has the clear climate change manifestation: water temperature increasing and decreasing of salinity and dissolved oxygen's content and saturation. No one coastal site does not characterized by acidification – there are not pH decreasing and the rise of alkalinity. For nutrient contents, significant long-term negative trends were revealed in Odessa and Chernomorsk regions (except phosphorus in Chernomorsk). In the Danube delta the long-term rise of silicate and total nitrogen contents were discovered while phosphorus species and ammonium concentrations decreased. Technogenic pollutants were steady decreased everywhere probably because of better waste water treatment and other marine environment protection measures implemented on both national and regional levels.

References.

EMODnet Chemistry website. *https://www.emodnet-chemistry.eu/about*

Gilbert R.O. Statistical methods for environmental pollution monitoring. *Van Nostrand Reinhold*, New York. 1987. 320 p.

Hammer Ø., Harper D.A.T., Ryan P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica.* 2001. 4(1). 9 pp. *http://palaeo-electronica.org/2001_1/past/issue1_01.htm*

Schlitzer R. Ocean Data View. 2018. *https//odw.awi.de*

Wooldridge J.M. Introductory Econometrics – a Modern Approach (5th ed.). *South-Western Cengage Learning.* 2012. 882 p.

The Killer Shrimp Invasion Challenge on Kaggle: An online competition tackling the spread of invasive marine species through machine learning

Adrian Bumann, Chalmers University of Technology (Sweden), adrian.bumann@chalmers.se
Robin Teigland, Chalmers University of Technology (Sweden), robin.teigland@chalmers.se
Jannes Germishuys, Combine AB (Sweden), jurie.germishuys@combine.se
Benedikt Ziegler, Combine AB (Sweden), benedikt.ziegler@combine.se
Martin Mattson, Medins Havs och Vattenkonsulter AB (Sweden), martin.mattsson@medinsab.se
Eddie Olsson, RISE - Research Institutes of Sweden (Sweden), eddie.olsson@ri.se
Robert Rylander, RISE - Research Institutes of Sweden (Sweden), robert.rylander@ri.se
Yixin Zhang, University of Gothenburg (Sweden), yixin.zhang@ait.gu.se
Torsten Linders, University of Gothenburg (Sweden), torsten.linders@gu.se

Background.

The world faces numerous complex marine challenges, such as overfishing, the spread of invasive species, and rising sea levels. These challenges are interconnected with the 17 UN Sustainable Development Goals, and in particular Goal #14 - Life below water. A paradox for any action related to the ocean is that while there is an enormous lack of ocean data, there is also an abundance of online marine and geo data that could be used to develop solutions through the application of artificial intelligence (AI). Open innovation and crowdsourcing could be a solution to such complex problems; however, applying data science to solve marine challenges through these open strategies is limited.

Responding to the above, Ocean Data Factory Sweden (ODF Sweden), a data-driven innovation consortium in Gothenburg, developed an online competition, The Killer Shrimp Invasion Challenge (Figure 1). The competition was launched on the data science competition platform, *Kaggle,* in the spring of 2020 to address the spread of the invasive Killer Shrimp through applying machine learning (ML) to online data. This paper will describe the Killer Shrimp use case, the launch of the Kaggle competition, competition results and reflections on this form of tackling marine challenges.

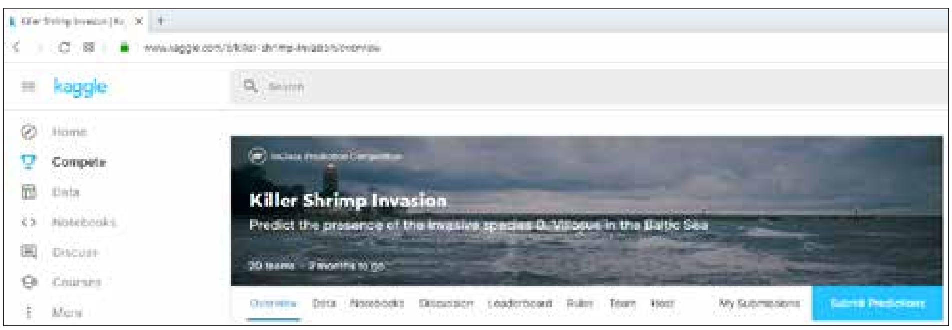


Figure 1: Landing page for the Killer Shrimp Invasion Challenge on Kaggle

Developing the Killer Shrimp Use Case.

In the fall of 2019, ODF Sweden started its first use case: predicting the spread of the invasive species, *Dikerogammarus Villosus* (aka the Killer Shrimp) in the Baltic Sea region. As this shrimp devastates any local ecosystem it invades, ODF Sweden and one of its partners, the Swedish Agency for Marine and Water Management (SwAM), decided to investigate whether ML methods could help predict the areas of the Baltic Sea suitable for the Killer Shrimp.

Throughout the use case, ODF Sweden applied a principle of openness. Primary data were national operational monitoring data from the Pan-European Copernicus Marine Services (marine.copernicus.eu). Datasets included presence data from the North Sea and Baltic Sea (roughly 3000 data points), pseudo-absence data from the Baltic Sea (2.8 million data points), and environmental rasters for key environmental drivers informed by subject experts. The project was documented on Kaggle Jupyter notebooks and Github was used as a code repository.

Launching the Kaggle Challenge.

Once the team was satisfied with the first use case results, we explored how to launch the Kaggle challenge. Our motivation for a Kaggle challenge was threefold: 1) to create awareness of our efforts to monitor and predict marine invasive species, 2) to encourage others to improve our ML solution, and 3) to build an engaged global community from diverse backgrounds to contribute to our future challenges. We chose Kaggle as it is one of the largest, most diverse data science communities. Kaggle is specifically designed as a platform for data scientists from anywhere "to solve predictive modeling problems through data competitions". Founded in 2010, Kaggle was acquired by Google LLC in 2017. A featured Kaggle competition costs between $80,000 and $200,000 including both rewards and Kaggle's hosting fees. An example challenge is the Deepfake Detection Challenge by AWS, Facebook, Microsoft, Partnership on AI's Media Integrity Steering Committee, and academics with 33,007 entries by 2281 teams competing for $1 million in prizes. In the marine area, NOAA Fisheries is offering $12,000 in a three-year contest to develop algorithms to count sea lions in aerial photos.

Kaggle also supports free in-class competitions for educational purposes. Due to our interest in exploring the Kaggle platform for innovation, we chose to run an in-class prediction competition. We conducted the following steps to launch our challenge: 1) defined our machine learning problem as "to predict the presence of the Killer Shrimp in areas of the Baltic Sea", 2) prepared and uploaded a training dataset with locations, physical parameters and shrimp presence to enable competitors to develop their ML models, 2) prepared and uploaded an example of a submission file to enable competitors to understand what they should submit (a simplified set of predictions assuming all locations with temperatures above 5 degrees contain the Killer Shrimp), 4) chose the evaluation metric of "Area Under Receiver Operating Curve" (AUROC) between the predicted probability and the observed target, 5) uploaded a test dataset to evaluate the competitors' models on unseen data, i.e. predicting the presence of the Killer Shrimp in various areas, with 30% of the test set used to calculate the public leaderboard during the competition and the remaining 70% used to calculate the private leaderboard, i.e., the winners, at the end of the competition, 6) decided on a first prize of €150 and the opportunity for the winner to present their solution at an ODF Sweden grand meeting, and 7) determined the timeline to be just under three months. We then launched the challenge through our networks and on social media channels, such as Twitter, LinkedIn, Facebook, and Instagram.

Kaggle Competition Results and Reflections.

After one month of the three-month competition, our Challenge leaderboard showed that we had received 114 entries from 20 competitors in 20 teams from 12 different countries. Shortly after the launch, two competitors achieved a perfect score of 1.0000, with one achieving this after only one entry. As a perfect score is very unusual, this suggested that competitors found ways to overfit their models. We therefore added the requirement for competitors to upload their code as well. This decision was met with positive feedback on the discussion board. One top scorer explained which method he used and that he had submitted the perfect score to point out issues with the data setup. The competitor had also submitted legitimate models before and encouraged other users to do the same. This revealed some of the difficulties in hosting such a competition.

In summary, we are positively surprised and encouraged by the results to date despite the relatively small prize money. Our next steps will be to increase our understanding of Kaggle and other open innovation platforms, such as Zooniverse, as well as how we can better incorporate open innovation in the ML design process to tackle marine challenges.

References

Dick, J. T., Platvoet, D., & Kelly, D. W. (2011). Predatory impact of the freshwater invader Dikerogammarus villosus (Crustacea: Amphipoda). *Canadian Journal of Fisheries and Aquatic Sciences, 59*(6), 1078-1084.

Majchrzak, A., & Malhotra, A. (2019). *Unleashing the Crowd: Collaborative Solutions to Wicked Business and Societal Problems.* Springer Nature.

Analysis of in-situ temperature and salinity data in the Levantine Sea

Cem Serimozu, Middle East Technical University Institute of Marine Sciences (Turkey), cem.serimozu@metu.edu.tr

The Levantine Sea (LS), part of the Eastern Mediterranean exposed to rising anthropogenic pressure, and four of its sub-regions: Cilician (CB) and Levantine Basins (LB), Coastal Nile Delta (CND) and Rhodes Gyre (RG) is the focus of this study. On account of the relative scarcity of scientific studies of these regions, the aim is to contribute to the investigation of Sea Water Temperature and Salinity, two critical oceanographic parameters in the context of climate change and to explore the statistical distribution of data for modelling purposes. This work uses the "Mediterranean Sea - Temperature and Salinity Historical Data Collection SeaDataCloud V1" (https://sextant.ifremer.fr/record/2698a37e-c78b-4f78-be0b-ec536c4cb4b3/) aggregated dataset for a descriptive analysis at multiple depth layers and showcases that missing data can be replaced using approximations obtained through chained equation method and gives an overview of the water column in the LS.

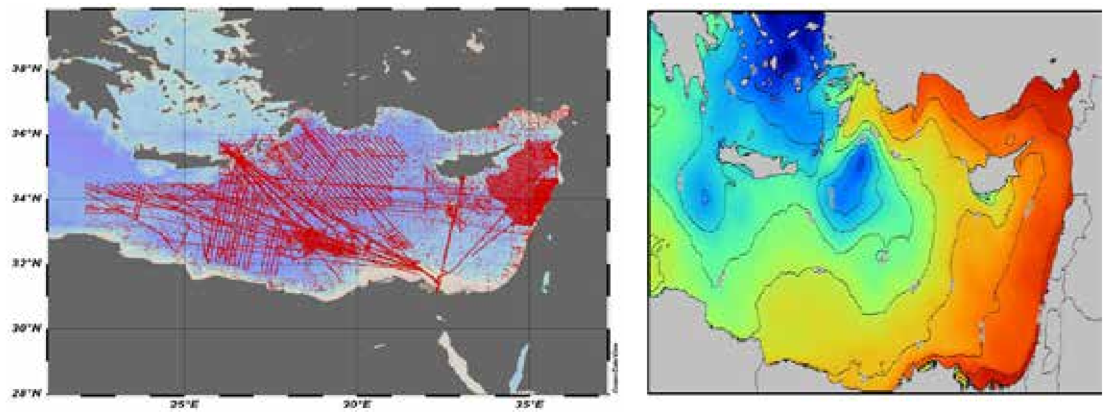


Figure 1: Map of the region showing: Spatial distribution of the data (Left); Isothermal sea surface temperature map in September using WOA18 (Right).

The investigation is on the in-situ data available in LS between 1960 and 2017, for a total number of 81,317 stations and 10,590,891 individual entries. Boundaries of the sub-regions are delimited with climatological maps for SWT and SWS obtained in Ocean Data View (ODV, https://odv.awi.de/) with Data-Interpolating Variational Analysis (DIVA) and supported by isothermal sea surface temperature maps employing World Ocean Atlas 2018 (WOA18, https://www.nodc.noaa.gov/OC5/woa18/) data (Fig. 1). Cluster analysis is applied to identify the existence of distinct regional properties. Missing observations are replaced with approximations computed from chained equations to increase the robustness of the analysis. The descriptive study is conducted for monthly, seasonal and yearly values for both SWT and SWS at the following depth ranges: Surface (0-10m), 25m, 50m, 100m, 200m, 500m, 1000m and 2000m.

The distribution of data shows a significant change in data collection capacity and ratio of absent measurements (Fig. 2). Testing reveals that replacing missing values does not make a significant change in the resulting plots, even in the case of SWS with 27% missing values beside less yearly gaps in sub-regions, presenting an avenue to improve modelling in the area. Comparing the monthly mean SWT and SWS values prior and post 1980 it is discernible that average values are higher in the later period, sometimes as much as 5°C difference in summer maxima SWT in August/September at the surface layer.

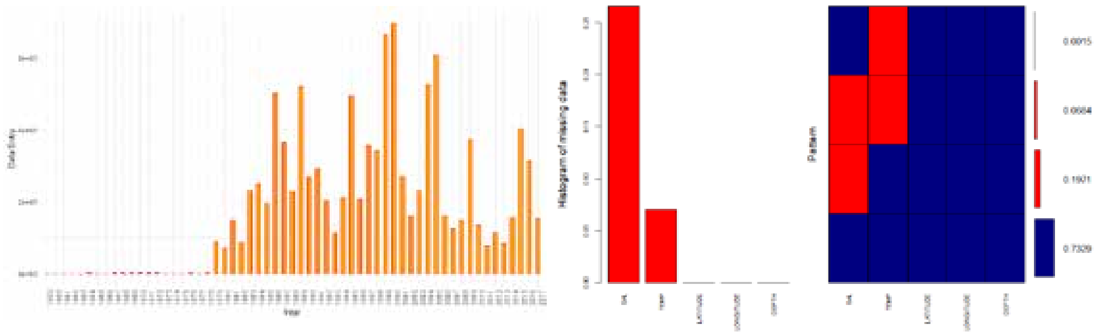


Figure 2: Histogram of data and missing entries per year showing: Increase in data collection capacity in the region after 1980 (Left); 27% SWT and 7% SWS of all entries are absent (Center); Percentage of both SWT and SWS missing entries is 6.8% (Right).

For example, Fig. 3 demonstrates a slow but steady increase in SWS and SWT in LB for the summer months in the surface waters, a process that is mirrored in the area with different degrees of significance per sub-regions. Anomalous events are more prevalent with higher SWT values than the seasonal average, especially in surface to 100m range interacting with the atmosphere.

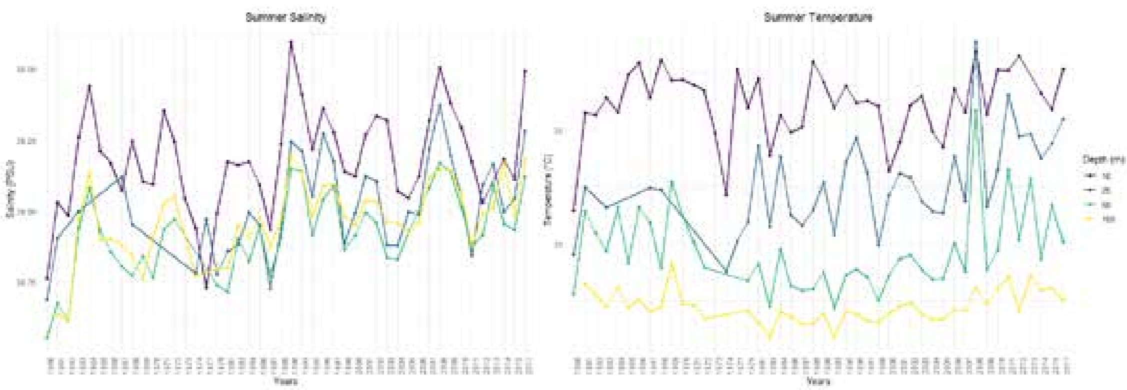


Figure 3: Annual summer mean values in the LB for: Salinity (Left) and Temperature (Right)

The most studied region in the LS is LB, followed by RG. Meanwhile, the least explored areas are the CB and CND. The CB and LB show consistently higher SWT, especially at the surface layers with substantial seasonal variability and overall higher SWS compared to LS. On the other hand, the RG in almost all seasons is considerably colder than LS as a result of the upwelling on the Eastern Mediterranean Deep Waters (EMDW). The study of the data also demonstrates an apparent lack of information in the broader LS contrasting with the rest of the Mediterranean Sea, especially the Western Mediterranean.

SWT normalises between the 200m and 500m layers for all sub-regions and no longer shows any seasonality corresponding with the Levantine Intermediate Water. The distribution pattern of the data for SWT and SWS obtained through density plots is bimodal corresponding to atmospheric temperature differences at the surface before taking a mounded appearance around 200m to 500m indicating the homogeneity of the Levantine Intermediate Water. An interesting find is that SWS becomes bimodal at the 2000m layer which might potentially indicate two slightly different water masses with different SWS properties instead of the traditionally accepted more homogenous distribution of the EMDW in the literature.

Mobilizing historical marine data within EMODnet Biology

Vasilis Gerovasileiou, Hellenic Centre for Marine Research (Greece), vgerovas@hcmr.gr
Stamatina Nikolopoulou, Hellenic Centre for Marine Research (Greece), snikolo@hcmr.gr
Dimitra Mavraki, Hellenic Centre for Marine Research (Greece), dmavraki@hcmr.gr
Christos Arvanitidis, Hellenic Centre for Marine Research (Greece), arvanitidis@hcmr.gr
Leen Vandepitte, Flanders Marine Institute (Belgium), leen.vandepitte@vliz.be
Gizem Poffyn, Flanders Marine Institute (Belgium), gizem.poffyn@vliz.be
Joana Beja, Flanders Marine Institute (Belgium), joana.beja@vliz.be
Ruben Perez Perez, Flanders Marine Institute (Belgium), ruben.perez@vliz.be
Laura Boicenco, National Institute for Marine Research and Development "Grigore Antipa" (Romania), lboicenco@alpha.rmri.ro
Marina Lipizer, National Institute of Oceanography and Applied Geophysics (Italy), mlipizer@inogs.it
Menashè Eliezer, National Institute of Oceanography and Applied Geophysics (Italy), meliezer@inogs.it

Historical marine data are often scattered in old publications, reports and expedition logbooks, either in the form of hard copies or in simple and unorganized spreadsheets in electronic storage media. Such sources provide an invaluable resource of biological and environmental information that could be used for reconstructing and modelling past environmental conditions or predicting future trends and shifts in distribution range, biological invasions and regional species extinctions. This is extremely important for regional European Seas and adjacent marine regions, which are vulnerable to an ever-increasing number of human activities and pressures.

Recognizing the importance for mobilizing historical marine data, EMODnet Biology has developed a long-term data archaeology and rescue strategy, under a dedicated work package (WP3). The overall objective of this initiative is to fill spatial and temporal gaps in aquatic species occurrences and make the rescued historical data available freely through the EMODnet portal and global biogeographic and biodiversity information systems (e.g. OBIS and its regional nodes, WoRMS and its sub-registers, LifeWatch Species Information Backbone). The standard WoRMS taxonomic mapping makes these data usable for long term time series. Special focus is given to understudied, "data poor" regions which are particularly susceptible to environmental alterations and biological invasions such as the South-Eastern Mediterranean Sea and the Black Sea.

A non-exhaustive pool of approximately 240 historical marine archaeological (1890s to 1950s) and rescue datasets (1960s to 2000s) was assembled by HCMR and is continuously being updated and annotated with metadata. A similar process was carried out by VLIZ, resulting in 95 datasets being identified for rescue; the datasets range from 1930s to mid-1990s, with the bulk of covering the 1970s decade and describing data collected mainly in the Belgian North Sea area. In addition, paper archives, primary protocols and grey literature from the Romanian Black Sea waters were scanned and rescued by NIMRD, containing data on phytoplankton, zooplankton and macrozoobenthos dating back to the 1950s, which remained unavailable to the wider research community.

A set of purpose-build criteria was adopted for the prioritization and selection of datasets to be digitized and mobilized, including thematic and taxonomic cover, temporal and geographic scope, language and readiness in their availability (e.g. online digital files versus hard copies in libraries). So far 90 historical datasets, published during the period 1868-1999, which included 110,230 occurrence records have been rescued and mobilised (Figure 1). The experience gained has revealed several challenges at all stages of the data "life-cycle", from dataset identification and metadata extraction to the digitization, standardization and quality control of historical datasets, such as: lack of standardization, georeferencing accuracy, taxonomic inconsistencies and updates, misspellings of taxa and locations and poorly documented or missing sampling protocols (Figure 2). Nevertheless, facing these challenges is of paramount importance since loss of such data equals to loss of unique resources required to understand global changes and ultimately to the loss of our natural wealth.

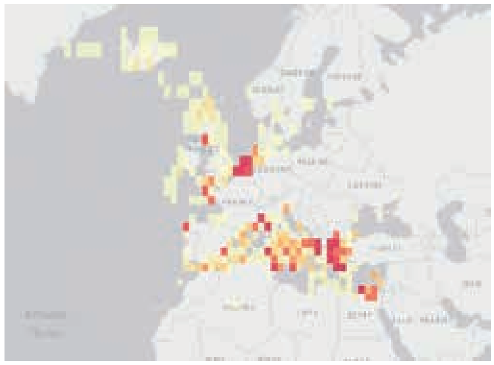
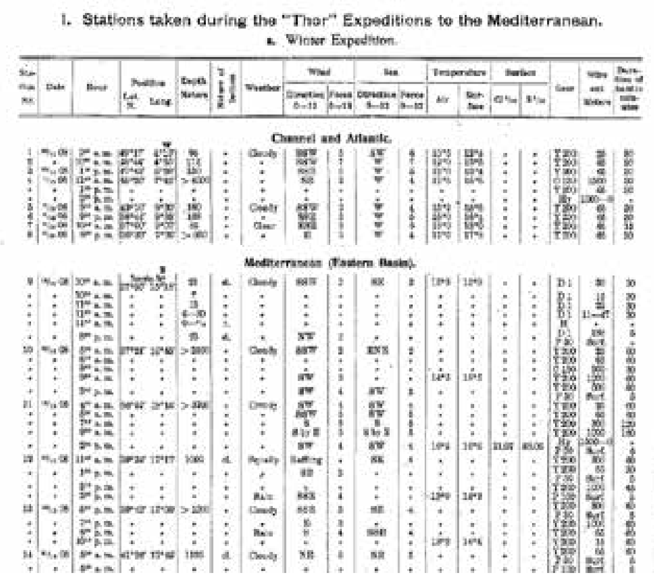


Figure 1: Distribution of historical occurrence data mobilised in OBIS (in total 110,230 occurrences from 90 datasets). Colour shades represent the number of occurrence records (higher values in red)



I. Stations taken during the "Thor" Expeditions to the Mediterranean.

a. Winter Expedition.

Channel and Atlantic.

Mediterranean (Eastern Basin).

Figure 2: Original presentation of sampling metadata of the historical "Thor" Expedition to the Mediterranean Sea and adjacent waters (from: Schmidt 1912 Høst & Son, Copenhagen, 49 pp.)

Acknowledgments

This work has been financially supported by the EC DG-MARE (EMODnet Observation and Data network - Lot5 – Biology: EASME/EMFF/2016/1.3.1.2/Lot5/SI2.750022) and by the Greek Government under the General Secretariat of Research and Technology and the National Strategic Reference Framework (LifeWatchGreece Research Infrastructure (ESFRI) project - MIS 384676). The Belgian archaeology work was carried out in the framework of LifeWatch, of which the Flemish contribution is funded through FWO (Research Foundation Flanders). Sarah Faulwetter, Nicolas Bailly and collaborating data managers who were involved in the first stages of this activity are highly acknowledged.

A supporting marine information system for marine oil spill emergencies in the Italian seas

Roberto Sorgente, IAS Consiglio Nazionale delle Ricerche (Italy), roberto.sorgente@cnr.it
Mario Allegra, ITD Consiglio Nazionale delle Ricerche (Italy), mario.allegra@itd.cnr.i t
Marco Arrigo, ITD Consiglio Nazionale delle Ricerche (Italy), marco.arrigo@itd.cnr.it
Andrea Cucco, IAS Consiglio Nazionale delle Ricerche (Italy), andrea.cucco@cnr.it
Antonia Di Maio, IAS Consiglio Nazionale delle Ricerche (Italy), antonia.dimaio@cnr.i t
Manuel Gentile, ITD Consiglio Nazionale delle Ricerche (Italy), manuel.gentile@itd.cnr.it
Dario La Guardia, ITD Consiglio Nazionale delle Ricerche (Italy), dario.laguardia@itd.cnr.it
Gennaro Oliva, ICAR Consiglio Nazionale delle Ricerche (Italy), gennaro.oliva@cnr.it
Federica Pessini, IAS Consiglio Nazionale delle Ricerche (Italy), federica.pessini@ias.cnr.it
Angelo Perilli, IAS Consiglio Nazionale delle Ricerche (Italy), angelo.perilli@cnr.it
Giovanni Quattrocchi, IAS Consiglio Nazionale delle Ricerche (Italy), giovanni.quattrocchi@ias.cnr.it
Alessandro Signa, ITD Consiglio Nazionale delle Ricerche (Italy), alessandro.signa@itd.cnr.it
Alberto Ribotti, IAS Consiglio Nazionale delle Ricerche (Italy), alberto.ribotti@cnr.it

Introduction

Promptness and quality with which decisions are taken to respond to marine pollution events are fundamental in the positive or negative final outcome of the adopted measures. In this context, decision support tools, like numerical oil-spill prediction models and risk maps, may significantly improve and make more effective the decision-making process of the competent national or regional authorities. Over the years, commercial organizations or scientific institutions have developed a wide variety of supporting tools, such as WITOIL (Liubasrteva et al., 2016) or MEDESS-4MS (Zodiatis et al., 2016). Since the Intervention Convention of 1969, the United Nations Convention on the Law of the Sea (UNCLOS) imposes a general obligation on Countries to protect and preserve the marine environment, taking individually or jointly all the measures to prevent, reduce and control pollution from any source. UNCLOS also asks to take measures beyond the territorial sea proportionate to the actual or threatened damage to protect their coastline or related interests. The increasing sensibility of the Italian public opinion on marine oil pollution, led the Italian Ministry of the Environment and Protection of Land and Sea to entrust the Italian National Research Council (CNR) the development of an oil-spill forecasting system for the Italian seas. In 2016 the project SOS Piattaforme e Impatti Offshore started with the aim to implement a numerical prediction service for hydrocarbons dispersion within the Italian seas. The objective of this project is to provide to the Italian authorities a Decision Support System (DSS), giving the near-real-time spatial and temporal evolution of oil spills in case of emergency situations or planned exercises, allowing the optimization of the resources and emergency management.

DSS: the numerical system

The main components of a DSS are a numerical system and a Graphical User Interface (GUI). The prediction of the oil slick movement due to the currents and wind is carried out by a computational processing of the sea forecast conditions. Our numerical system includes the forecast of the marine circulation at different space scales (Sorgente et al., 2011) and the transport and transformation of the spill of hydrocarbon through the MEDSLIK-II model, a Lagrangian community numerical model (De Dominicis et al., 2013_a). It allows the estimation of the spatial and temporal evolution (within the next 96 hours) of the oil spill fate at sea, as forced by weather forecast (wind), sea conditions (temperature and currents) and chemical and physical processes affecting the spilled oil. The system also permits to estimate hazards of oil slicks from the seven active oil platforms in the Sicily Channel and in the Adriatic Sea under different weather and sea conditions.

DSS: the Graphical User Interface

The DSS is managed through a GUI, where the users can set the parameters for the simulation (oil type, physics of seawater, oil weathering processes), the computation (time steps per hour and dimension of the array), the control (rate of evaporation, increase of water fraction, spreading of the slick, its dispersion), and other general parameters like date and length of simulation, spill coordinates, type, rate, and duration, all necessary to the numerical component. The choice of the domain (Western or Central Mediterranean Sea and Adriatic Sea) is made by a rule-based system, which takes into account the coordinates of the spill, the availability of the forecasting for the selected date and the length of the simulation. Once the simulation starts, the user is redirected to a grid view representing the evolution. Finally, the DSS produces a gallery of images showing the trend of the characteristic parameters of the oil during the simulated period and the oil spill's evolution frame-by-frame. The access to the DSS is from http://www.seaforecast.cnr.it/sos-piattaforme/ but actually limited to competent national or regional authorities.

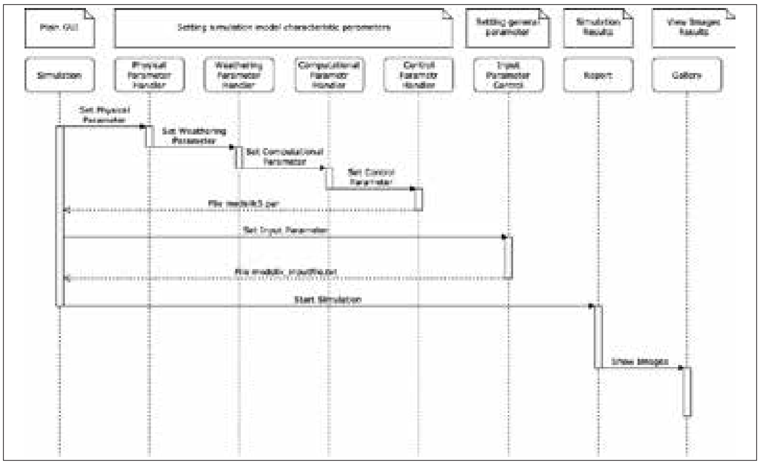


Figure 1: Simulation Flow Diagram

Bibliography

De Dominicis, M.; Pinardi, N.; Zodiatis, G.; Lardner, R. MEDSLIK-II, a Lagrangian marine surface oil spill model for short-term forecasting. Part 1: Theory. Geosci. Model Dev. 2013, 6, 1851-1869, https://doi.org/10.5194/gmd-6-1851-2013.

Sorgente, R.; Olita, A.; Oddo, P.; Fazioli, L.; Ribotti, A. Numerical simulation and decomposition of kinetic energies in the central Mediterranean: insight on mesoscale circulation and energy conversion. In: Special Issue, ECOOP (European Coastal-shelf sea Operational Observing and forecasting system Project), Ocean Sci. 2011, 7, 4, 503-519, https://doi.org/10.5194/os-7-503-2011.

Liubasrtseva, S., Coppini, G., Pinardi, N., De Dominici, M., Lecci, R., Turrisi, G., Creti', S., Martinelli, S., Agostini, P., Marra, P. and F. Palermo. Dicision support system for emergency management of oil spill accidents in the Mediterranean Sea. Nat. Hazards Earth. Sci., 16, 2009-2020, 2016. doi:10.5194/nhess-16-2009-2016.

Zodiatis, G., De Dominicis, M., Perivoliotis, L., Radhakrishnan, H., Georgoudis, E., Sotillo, M., Lardner, R.W., Krokos, G., Bruciaferri, D., Clementi, E., Guarnieri, A., Ribotti, A., Drago, A., Bourma, P., Padorno, E., Daniel, P., Gonzalez, G., Chazot, C., Gouriou, V., Mancini, M.. (2016). The Mediterranean Decision Support System for Marine Safety dedicated to oil slicks predictions. Deep Sea Res. Pt II, 133. 4-20. 133. 10.1016/j.dsr2.2016.07.014.

SeaDataCloud quality control of data collections

Christine Coatanoan, Ifremer (France), christine.coatanoan@ifremer.fr
Simona Simoncelli, INGV (Italy), simona.simoncelli@ingv.it
Volodymyr Myroshnychenko, METU, (Turkey), volodymyr@ims.metu.edu.tr
Örjan Bäck, SMHI (Sweden), orjan.back@smhi.se
Helge Sagen, IMR (Norway), Helge.Sagen@imr.no
Serge Scory, RBINS (Belgium), sscory@naturalsciences.be
Reiner Schlitzer, AWI (Germany), reiner.schlitzer@awi.de
Michèle Fichaut, Ifremer (France), michele.fichaut@ifremer.fr
Dick Schaap, MARIS (Netherlands), dick@maris.nl

During the SeaDataNet II (SDN) EU-project, a Quality Assurance Strategy (QAS) was implemented, aiming at improving the data quality and creating the best products. The QAS was applied to produce the first aggregated datasets released in SeaDataCloud (SDC) in 2018. It has been continuously reviewed and was upgraded regarding the step for harvesting data for the second version of the aggregated datasets. These last regional temperature and salinity data collections for 6 European sea regions (Arctic Ocean, Baltic Sea, Black Sea, Mediterranean Sea, North Atlantic, and North Sea) covering the time period 1900-2019 (SDN infrastruture content at July 2019) have been released within the SeaDataCloud (SDC) project in 2020. A general description of these datasets, their data quality control and assessment procedures and results are presented.

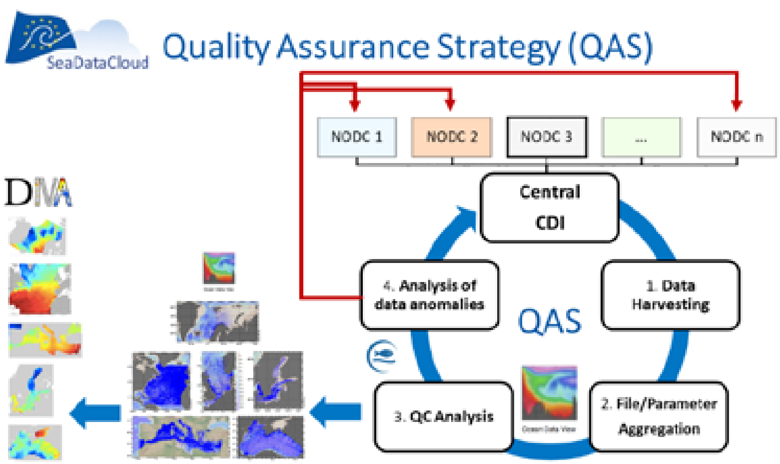


Fig. 1: Quality Assurance Strategy implemented during SDN.SDC project

The specific procedure implemented during SDN II allows assuring and certifying high quality of the datasets. Comparing to the first version, the new procedure has been put in place for the step corresponding to data harvesting from the central CDI catalogue (Fig. 1). Only new and updated data have been harvested to focus on this subset of data for the quality control (QC). This adjustment allowed to skip the quality check of data already validated in the previous QA (Quality

Assurance) loop and contained in the preceding data product version. The QC is performed at regional levels in a coordinated way, using the ODV software (5.2.1) as common and basic QC analysis tool. In SDC the additional checks have been performed per basin to consider the specific water masses characteristics, per instrument type to investigate data completeness and consistency, per data provider to better identify data anomalies. After performing the quality control on this specific dataset, the second version of the regional data collections consist of a merge between the previous version (delivered in 2018) and the new subset.

The QAS describes how to highlight doubtful data and to organize the data anomalies in lists that are sent to each concerned SDC partners together with guidelines explaining the expected corrections. The SeaDataNet partners check their data based on these lists, and eventually correct the original data and resubmit them in the SDC dataflow. The iterative procedure has been designed to facilitate the update and improvement of SDC database content.

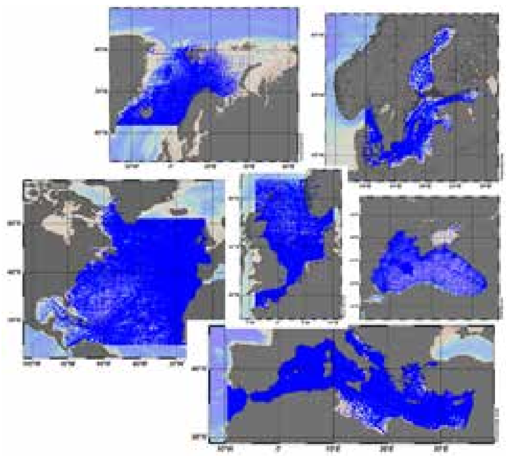


Fig. 2: Regional data distribution maps of SDC temperature and salinity data collections

A detailed description of each regional dataset (Fig. 2) is published in a Product Information Document (PIDoc). It includes general product's characteristics (space-time coverage, resolution, format), its quality (description of QC applied, validation methodology results), experts' recommendations for its usability and main changes from previous version. ODV qualified dataset collections and PIDocs are available at:

https://www.seadatanet.org/Products.

Development and Application of Teaching Materials for Ocean using JOISS

Mirinae Kim, KESTI (South Korea), malji527@kesti.co.kr
Taeydoon Song, KESTI (South Korea), tysong@kesti.co.kr
Minyoung Kim, KESTI (South Korea), myk0907@kesti.co.kr

High school students in Korea are learning about the ocean phenomenon in subjects such as ‘Integrated Science’, ‘Science Inquiry Experiment’, ‘Earth Science I’, ‘Earth Science II’. The maritime curriculum is a small part of each subject and although there are some textbooks with practices using serial oceanographic observation from NIFS(National Institute of Fisheries Science) or Argos around Korea, it is difficult to find the use of various other marine observation data. Recently, it has been a trend to provide teaching•learning methods for students to collect and analyze data. Collecting and processing scientific data are also a necessary capability for students in the fourth industrial revolution. Therefore, the teaching materials for Ocean, which allow students to use their own observed data, will make it easier for students to increase their understanding and interest. JOISS(http://joiss.kr) is a portal that collects and provides marine research data in Korea, and students will be able to easily access the data that fit the subject of the class at JOISS.

This study aims are to introduce the contents of teaching materials for the curriculum using the marine data provided by JOISS and analyze the results of applying them. 16 teachers in the Earth Science Education & Learning Community reviewed each subject, extracted the contents related to the ocean and divided into 5 parts: marine physics, marine chemistry, marine biology, marine geology/geophysics and marine weather. The teaching materials were developed by selecting themes suitable for each part and reviewing the JOISS data for the themes. Finally, we developed teaching materials for a total of 31 themes: 11 marine physics, 7 marine chemistry, 5 marine biology, 3 marine geology/geophysics and 5 marine weather. The contents of the materials consist of 1) outline, 2) open thinking, 3) using JOISS, 4) inquiry activities, 5) result report and 6) reference. In the case of climate change among the

Figure 1: Teaching materials for Ocean (coastal upwelling)

various themes, students can download the water temperature and salinity data around Korea for 30 years from JOISS and draw a time series graph with trend line to interpret data. And JOISS supports it by serving the education application.

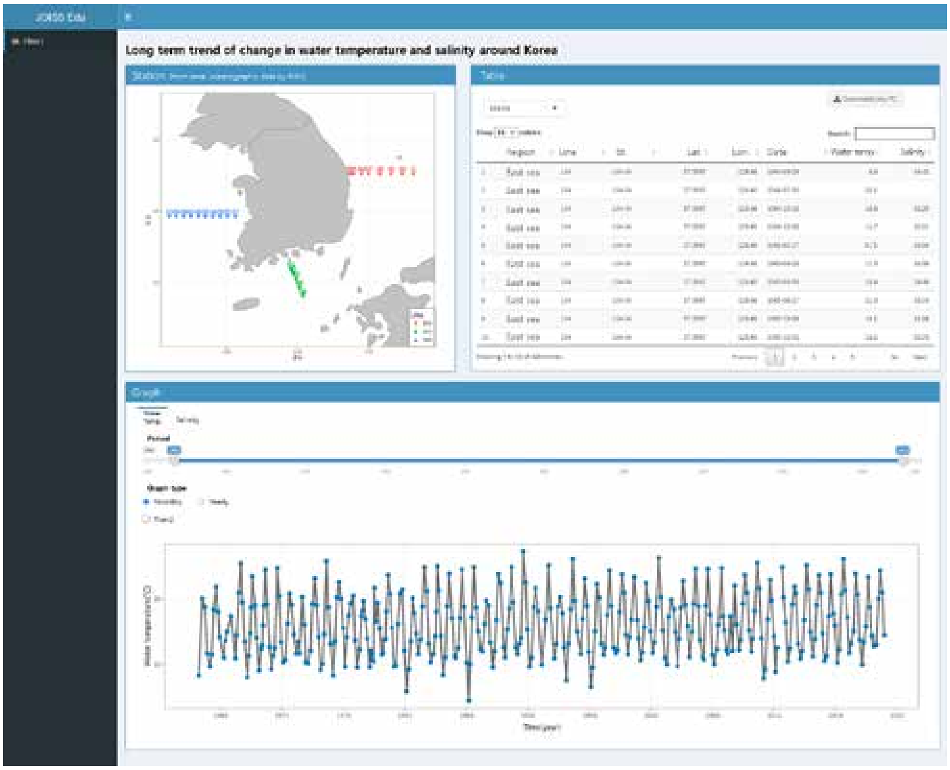


Figure 2: JOISS Education App. (Long term trend of change in water temperature and salinity)

The developed materials have been distributed to every middle and high schools located in Incheon, and has been provided at JOISS for students and teachers. As a result, 36 schools and 568 students joined JOISS and are using these materials. The developed teaching materials for ocean are applied in several schools and seminars, and some of the developed contents are published in various reference book for teachers. As schools can switch to online classes because of the COVID-19 pandemic, the number of classes that use the teaching materials and the education application in JOISS is increasing. Therefore, it is expected that teaching materials for ocean using JOISS will be used in various ways in schools.

Figure 2: A case of class using the developed teaching materials

Comparison of long-term SST around the Korea using the datasets from JOISS, Korean ocean data repository

Minyoung Kim, KESTI (south Korea), myk0907@kesti.co.kr
Taeyoon Song, KESTI (south Korea), tysong@kesti.co.kr
Wooram kim, KESTI (south Korea), kwr@kesti.co.kr
Jiyoon Lee, KESTI (south Korea), jiyoon4550@kesti.co.kr

The IPCC (Intergovernmental Panel on Climate Change) announced in the 4th report (2007) that climate change represented by global warming is in progress, and suggested temperature, water temperature, and sea level rise as evidence of the progress of climate change.

In establishing a response policy to ocean change, it is fundamental to forecast the long-term trend of change.

The period of observation data for long-term changes in sea surface temperature on the Korean Peninsula provided by Papers published in Korea to date and the Korea Meteorological Agency is from 1968 to 2017.

This study identified the recent trend of surface water temperature changes in the waters of the Korean Peninsula and compared it with the upward tendency of surface water temperature data in the waters of the Korean Peninsula observed overseas.

The domestic data used for the study was the serial oceanographic observation data of KODC (Korea Oceanographic Data Center) of JOISS Portal, which collects and provides research data observed in the Korean Peninsula and public data from domestic and foreign institutions.

In the case of Serial Oceanographic observation data, it is marine scientific data produced regularly and systematically in the sea area around the Korean Peninsula. From 1961 to the present, it is surveyed six times a year, and 17 items such as water temperature and salinity in the standard water layer are investigated.

For foreign data, NOAA's ERSST v5 (Extended Reconstructed Sea Surface Temperature) data were used.

The ERSST v5 (NOAA) data provided ssta values, whereas the Serial Oceanographic observation (KODC) data provided only water temperature, so [annual average - average Korean peninsula (16.9°C)] was calculated and anomaly was calculated and used for comparison.

As a result of the study, it was found that the annual average surface water temperature from 1968 to 2017 in the Serial Oceanographic observation(KODC) data increased at an increase rate of 0.025°C/year.

Comparing ERSST v5 data with Serial Oceanographic observation data, the average annual growth rate of surface seawater water temperature in ERSST v5 data between 1968 and 2017 was 0.02°C/year, showing a tendency similar to Jeongseon ocean observation data.

As a result of the analysis of the anomaly regime shift in serial oceanographic observation data, it was confirmed that the anomaly increased significantly from 1988.

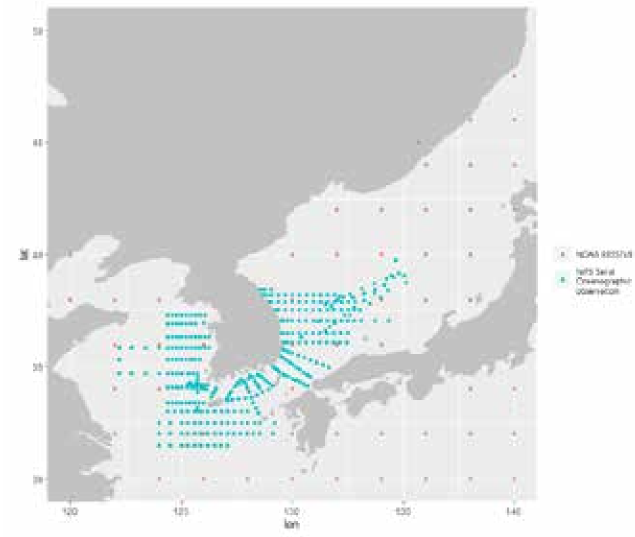


Figure 1: NOAA ERSSTv5 & NIFS Serial Oceanographic observation station

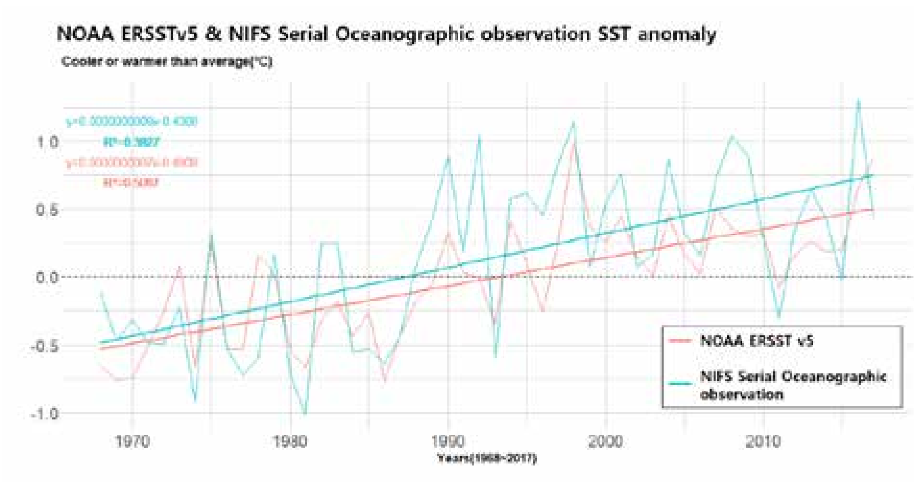


Figure 2: NOAA ERSSTv5 & NIFS Serial Oceanographic observation SST anomaly

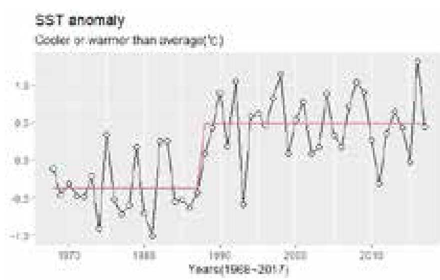


Figure 3: Serial Oceanographic observation SST anomaly regime shift analysis results

Black Sea Temperature and Salinity climatologies computed with DIVAnd tool

Volodymyr Myroshnychenko, Middle East Technical University, Institute of Marine Sciences (Turkey), volodymyr@ims.metu.edu.tr
Alexander Barth, GHER, University of Liege (Belgium), a.barth@uliege.be
Charles Troupin, GHER, University of Liege (Belgium), c.troupin@uliege.be
Simona Simoncelli, Istituto Nazionale di Geofisica e Vulcanologia (Italy), simona.simoncelli@ingv.it

General description

The Black Sea (including Sea of Azov) Temperature (T) and Salinity (S) gridded climatologies for the period 1955 – 2017 were created in the framework of the SeaDataCloud (SDC) project from in situ observations extracted from 3 major sources: 1) SeaDataNet infrastructure, 2) World Ocean Database (Boyer et al., 2018), and 3) COriolis Ocean Dataset for Reanalysis (Szekely et al., 2016). Computing was performed with DIVAnd (Data-Interpolating Variational Analysis - n-dimensional) (Barth et al., 2014) on the following grid: geographical extent 27.5 - 41.875°E, 40.875 - 47.25°N; horizontal resolution: 1/8°; vertical resolution: 67 depth levels from 0 to 2000 m - same as in World Ocean Atlas (WOA, 2018).

The data product SDC_BLS_CLIM_TS_V1 (https://doi.org/10.12770/ad2d0efc-7191-4949-8092-796397106290) is published at the SeaDataNet web site along with the Product Information Document (https://doi.org/10.13155/61812). The product contains:

- monthly T and S fields for the periods 1955 – 2017, 1955 – 1994, and 1995 – 2017;
- seasonal fields for 6 decades starting from 1955 and for the same 3 periods as monthly.

Observational data

The input dataset for the computation of the Black Sea T and S climatological fields was integrated from two internal SDC datasets and two external datasets:

1. SDC Temperature and Salinity Historical Data Collection for the Black Sea (Version 1) – product SDC_BLS_DATA_TS_V1 (Myroshnychenko et al., 2018).
2. SDC Restricted Temperature and Salinity Historical Data Collection for the Black Sea.
3. Data extracted from the World Ocean Database 2018 - WOD18.
4. Data extracted from the COriolis Ocean Dataset for Reanalysis - CORA 5.1.

The quality controlled (QC-ed) SDC datasets were taken as primary. Data integration was performed through the following steps: excluding internal duplicates; identifying and excluding overlapping data; applying additional QC to non-overlapping data from external datasets; merging non-overlapping data; excluding climatically non-relevant data, i.e. those acquired in river estuaries or in adjacent lakes (called “limans” in Black Sea region).

Table 1: Content of the input dataset

	SDC_BLS_DATA_TS_V1	SDC restricted	WOD18	CORA 5.1	Total
# of stations	130466	10285	48227	57847	246825
	53%	4%	20%	23%	

CORA dataset mainly supplied the underway data. They are distributed irregularly in time and space and further were heavily subsampled in order to eliminate "trajectories effect" from climatic fields.

Methodology

Computation was done with DIVAnd tool implemented in the Julia programming language with a Jupyter notebook interface. DIVAnd allows the interpolation of observations onto orthogonal grids in an arbitrary high dimensional space by minimizing a cost function. The fields produced with DIVAnd are the results of 3D (or 4D if time is considered) interpolation and this is the innovation and also the main difference of SDC T-S climatologies from similar products (e.g. WOA), where 3D fields are combined from 2D slices.

Table 2: DIVAnd settings

DIVA parameter	Background field		Climatology	
	T	S	T	S
Horizontal correlation length (km)	200	200	150	150
Vertical correlation length [m]	depth dependent for 67 levels			
epsilon2	0.5	0.1	0.3	0.1
epsilon2 adjustment	With weights of the observations			

In the first iteration, the residuals obtained with DIVAnd for global (1955 – 2017) monthly fields were analysed against Three-Sigma criteria, the outliers were discarded, the second iteration was run to obtain the final fields.

Results

The Black Sea T and S climatological fields well represent Black Sea oceanographic and circulation features (Oguz et al., 1993) as well as their seasonal variability. For example, the inflow of rivers fresh water and its propagation is well visible in surface salinity field all over the year but mostly in spring months, while the Black Sea Rim Current, Batumi eddy, Bosporus plume are more pronounced in 70-130m layer (Figure **1**).

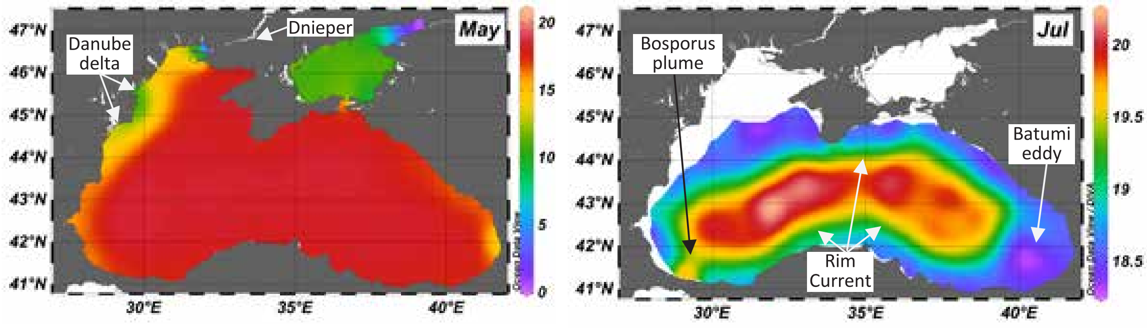


Figure 1: (left panel) Surface Salinity and (right panel) Salinity at 70 m (DIVAnd analysis for 1955-2017)

Despite overall good quality of the climatological fields there are known issues such as: anomalies in some decadal fields – mainly due to scarcity of data; anomalies in upper layer (DIVAnd issue that later was resolved in new version of the software); violation of vertical stability in some T-S profiles that are combined from separately calculated fields.

A consistency analysis was performed against WOA18 climatological fields (objectively analysed means) available at a 1/4° resolution. Most of the differences between the two products are observed in the upper 300 m layer. The WOA18 maps are smoother and do not capture some important Black Sea features (e.g. Batumi eddy, Bosporus plume), while SDC maps, even noisy, are more realistic.

Conclusions

The SDC Black Sea Temperature and Salinity climatologies are computed from the most complete dataset combining in-situ data from the tree major data sources – SeaDataNet, WOD, and CORA. Performing real 3-d interpolation with DIVAnd allowed us to obtain realistic climatological fields however with some known issues, which will be addressed in next release of the product.

References

Boyer, T. P., O. K. Baranova, C. Coleman, H. E. Garcia, A. Grodsky, R. A. Locarnini, A. V. Mishonov, C. R. Paver, J. R. Reagan, D. Seidov, I. V. Smolyar, K. Weathers, M. M. Zweng, (2019): *World Ocean Database 2018.* A. V. Mishonov, Technical Editor, NOAA Atlas NESDIS 87. Resource web address: https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html.

Szekely Tanguy, Gourrion Jerome, Pouliquen Sylvie, Reverdin Gilles (2016). *CORA, Coriolis Ocean Dataset for Reanalysis. SEANOE.* https://doi.org/10.17882/46219.

Barth, A., Beckers, J.-M., Troupin, C., Alvera-Azcárate, A., and Vandenbulcke, L.: *DIVAnd-1.0: n-dimensional variational data analysis for ocean observations,* Geosci. Model Dev., 7, 225-241, doi:10.5194/gmd-7-225-2014, 2014.

WORLD OCEAN ATLAS 2018 (WOA18) – https://www.nodc.noaa.gov/OC5/woa18/.

Volodymyr Myroshnychenko, Dick Schaap, Reiner Schlitzer (2018). *Black Sea - Temperature and salinity Historical Data Collection SeaDataCloud V1.* https://doi.org/10.12770/2287615d-1977-479f-8d5b-439960bcb21a.

Oguz, T., V. S. Latun, M. A. Latif, V. V. Vladimirov, H. I. Sur, A. A. Makarov, E. Ozsoy, B. B. Kotovshchikov, V. Eremeev, and U. Unluata, *Circulation in the surface and intermediate layers of the Black Sea,* Deep Sea Res., Part I, 40, 1597–1612, 1993.

Development of Digital Atlas "Physical Oceanography of the Black Sea – 2020"

Vladimir Belokopytov, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), belokopytov.vn@mhi-ras.ru
Eugeny Godin, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), godin_ea@mhi-ras.ru
Andrey Ingerov, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), ingerov@mhi-ras.ru
Elena Zhuk, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), alenixx@gmail.com

In the framework of international and national projects, considerable experience in development and creation of various maps and atlases reflecting the state of the Black Sea environment has been accumulated in Marine Hydrophysical Institute, Russian Academy of Sciences. Selected maps and entire sections of such atlases as digital atlas "Physical Oceanography of the Black Sea" (2003), "Atlas of the Black Sea and Sea of Azov Nature Protection" (2006), "National Atlas of Ukraine" (2007), "Oceanographic Atlas of the Black Sea and the Sea of Azov" (2009), etc. can be given as examples. Data held in the Oceanographic data bank of Marine Hydrophysical Institute, Russian Academy of Sciences (MHI BOD) served as the informational basis while building maps for relevant sections of the atlases.

At present, digital atlas "Physical Oceanography of the Black Sea" is available online on the MHI BOD website (http://bod-mhi.ru/climaticAtlas.php) and contains about 350 maps in its three sections. The first version of the atlas was designed in 2003 in the framework of project "Rescue of Black Sea Hydrological Data, Creation Digital Atlas, and Studies of Seasonal and Interannual Variability of Black Sea". The atlas was highly appreciated by users from different marine research institutions in Russia, Ukraine, and other states and remains to be actively used.

However, in more than fifteen years since the atlas release, the array of oceanographic data that can be used for building maps has increased from 105,000 to more than 165,000 stations; methods of mapping have been considerably improved; a new interface has been developed to provide users with more opportunities when working with the atlases. All of that promoted the development of a new digital atlas "Physical Oceanography of the Black Sea – 2020" with its possible location on the MHI BOD website.

Structurally, the new atlas will consist of four parts (temperature, salinity, density, geostrophic circulation) and include above 500 maps and sections.

The maps for the atlas "Physical Oceanography of the Black Sea – 2020" will be built based on the array of reanalyzed temperature and salinity data for the entire period of observations in the Black Sea calculated on the data kept in the MHI BOD. Spatial resolution of maps in the atlas will be improved in comparison with the previous one. Furthermore, special attention will be paid to the contemporary period (1991–2020) marked by a clear trend to warming and desalination of the upper layers in the Black Sea. For the 30-year period mentioned above, climatological averaging will be made, and seasonal empirical orthogonal functions (EOF) to compute average monthly thermohaline fields and oceanographic derivatives will be determined. It should be noticed that the time range under consideration corresponds to the next reference period set by the World Meteorological Organization to calculate climatological standard normals.

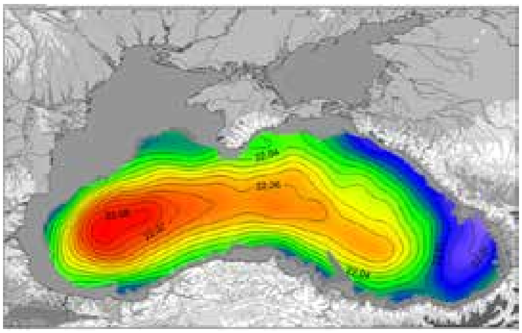


Figure 1: Salinity at the depth 500 m.

The atlas “Physical Oceanography of the Black Sea – 2020” will allow clarifying the understanding of oceanographic peculiarities of the Black Sea and further improve the information management of academic studies and marine economy in the Black Sea basin.

The work on creating the atlas “Physical Oceanography of the Black Sea – 2020” and developing MHI BOD databases is carried out within the state task on theme No. 0827-2018-0002 “Operative Oceanology”.

10 years of EMODnet Biology: past, present & future

Joana Beja, Flanders Marine Institute (Belgium), joana.beja@vliz.be
Leen Vandepitte, Flanders Marine Institute (Belgium), leen.vandepitte@vliz.be
Ruben Perez Perez, Flanders Marine Institute (Belgium), ruben.perez@vliz.be
Gizem Poffyn, Flanders Marine Institute (Belgium), gizem.poffyn@vliz.be
Dan Lear, Marine Biological Association (UK), dble@mba.ac.uk
Vasilis Gerovasileiou, Hellenic Centre for Marine Research (Greece), vgerovas@hcmr.gr
Peter Herman, Deltares (Netherlands), peter.herman@deltares.nl
Bart Vanhoorne, Flanders Marine Institute (Belgium), bart.vanhoorne@vliz.be
Lennert Tyberghein, Flanders Marine Institute (Belgium), lennert.tyberghein@vliz.be
EMODnet Biology consortia

EMODnet Biology is part of a network of more than 150 organisations working towards making European Marine data easily accessible and available across seven thematic disciplines. It focuses on species, distributions, species attributes, sampling methods and biological indicators data from 9 functional groups: algae, angiosperms, benthos, birds, fish, mammals, phytoplankton, reptiles and zooplankton.

The infrastructure and data flow used within EMODnet Biology is based upon that of EurOBIS (the European OBIS node), hosted at the Flanders Marine Institute (VLIZ); the fact that the two initiatives are interconnected means that any technical changes and/or tools developed in one initiative can be used by the other. Due to this symbiotic approach, EMODnet Biology data can also be seamlessly integrated with wider networks such as OBIS (Ocean Biogeographic Information System) and GBIF (Global Biodiversity Information Facility).

A first stage of preparatory actions took place from 2009-2012 and focused on building a marine biological portal, completing an inventory of existing holdings of marine data, performing a gap analysis to determine the shortcomings in data quality and geographical and taxonomic coverage and developing a sustainable strategy for the portal.

From 2013-2016 the network expanded from 10 to 23 partners and saw not only the enhancement of the portal but also the development of tools to improve data access and download.

The performance phase that ensued (2017-2019) saw the transition to the OBIS-ENV data exchange format, an expansion of Darwin Core, allowing for even greater degrees of interoperability. Eight new associated data partners also joined the consortia via a grant call. The current phase is building on the knowledge and experience acquired over the past ten years and working towards the changes due in the near future, a single point of access to data, data products and web services via the EMODnet Central Portal. The number of associated data partners in this phase has also doubled and will see the addition of 85 new datasets to the initiative. Figure 1 exhibits the number of datasets containing functional group's records before EMODnet Biology started and throughout the different phases.

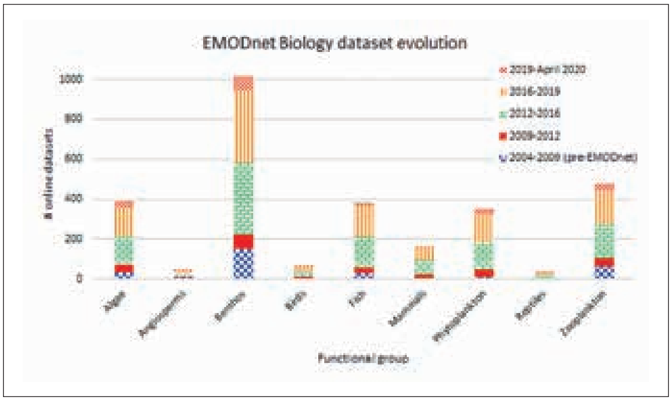


Figure 1: Number of online datasets containing functional group records (per phase from pre-EMODnet Biology implementation to April 2020)

The development of a suite of data products was framed in the context of the biological Essential Ocean Variables, illustrating the reuse value of open marine biological data to support policy development, conservation and management.

The implementation of FAIR principles has been at the core of all the work developed since 2009. By following the FAIR paradigm in making data, tools, products and services Findable, Accessible, Interoperable and Reusable, EMODnet Biology is ensuring that these unprecedented volumes of marine biological data have the highest possible utility and value to the widest range of stakeholders. Since mid-2019 EMODnet Biology is registered with FAIRsharing.org allowing not only a wider dissemination of the work done within EMODnet Biology, but also working to enable the FAIR Principles and to make Standards, Knowledge Bases, Repositories and Data Policies FAIR.

Acknowledgments

This work has been financially supported by the EC DG-MARE (EMODnet Observation and Data network – Lot° V – Biology: EASME/EMFF/2016/1.3.1.2/Lot5/SI2.750022)

Open access datasets from the Antarctic Circumnavigation Expedition

Jen Thomas, Swiss Polar Institute (Switzerland), jenny.thomas@epfl.ch
Carles Pina Estany, Swiss Polar Institute (Switzerland), carles.pinaestany@epfl.ch

Antarctic Circumnavigation Expedition (ACE)[1]

During the austral summer of 2016/2017, the R/V Akademik Tryoshnikov embarked on an ambitious voyage around Antarctica. In December 2016 it set off from Cape Town, in South Africa making stops at Marion Island, Crozet Islands and Kerguelen Islands, then passing by Heard Island before completing the first leg of its journey in Hobart, Australia. The second leg of the journey saw ACE head directly to the Mertz Glacier, followed by the Balleny Islands, Scott Island, Mount Siple region, the rarely visited Peter I Island and the Chilean islands of Diego Ramírez before a short stop in Punta Arenas, Chile. Final stops were made at South Georgia, the South Sandwich Islands and the remote island of Bouvetoya before returning to Cape Town; a journey of 33,463 km over a period of 90 days.

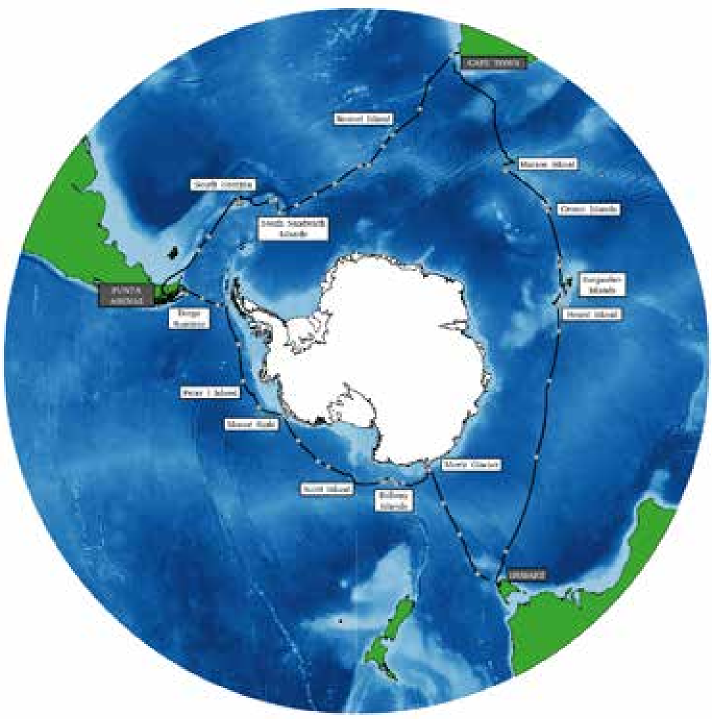


Figure 1: Track and sampling of the Antarctic Circumnavigation Expedition

With more than 150 scientists representing 21 countries participating in the expedition, ACE was an international expedition that brought together twenty-two distinct science projects which collected almost 30,000 air, water and biological samples and over 20 TB of data.

Here we present an overview of the data and samples that were collected during this expedition and where they can be accessed, as well as the tool that was used to record the expedition metadata on board.

Figure 1 displays the oceanographic sampling stations where CTD deployments for profile data and water sampling, neuston and bongo net deployments for micro-plastic sampling, profiles of optical properties and benthic organism sampling took place. To complement these observations continuous physical parameters of the seawater were recorded from the on-board Ferrybox and water samples taken from the underway line every three hours. Specific sampling sites on the upwind and downwind sides of island groups provided an interesting insight into island effects.

Simultaneous atmospheric and meteorological measurements were made throughout the cruise, providing a unique opportunity to complement the circumpolar oceanographic measurements.

Open source, on-board data management tool

During ACE a Python and Django tool, science-cruise-data-management[2], was built to record all sample and data collection events, metadata and provide simple data visualisation, reports and tools for researchers on board. This tool still today provides a very useful interface for checking the metadata behind the datasets collected on board, forming the basis of all data publications arising from ACE.

Open, discoverable and re-usable data

Over 80 datasets have been published as of mid 2020, 70 % of which are open access. As further sample processing and data analysis take place over the coming months and years, the number of openly and freely accessible datasets will increase.

A Digital Object Identifier (DOI) can be used to identify and cite published datasets, the majority of which can be found in the Zenodo data repository[3]. Datasets are also discoverable using Google dataset search and repositories that harvest DataCite metadata[4] (such as DataCite themselves and Mendeley Data).

In some cases, data are being worked on and published in RENKU[5], an open-source tool that captures dataset provenance, thereby improving reproducibility. In addition, these datasets will be accessible through the RENKU platform, providing them to a wider user base.

Data will be integrated over the coming months into the EMODnet Physics web portal[6], consequently also making it discoverable through the Southern Ocean Observing System (SOOS) map[7].

References

[1] Walton, D.W.H and J. Thomas. (2018). Cruise Report - Antarctic Circumnavigation Expedition (ACE) 20th December 2016 - 19th March 2017 (Version 1.0). Zenodo. http://doi.org/10.5281/zenodo.1443511

[2] Zenodo. (2009-2017). CERN. https://zenodo.org

[3] DataCite Metadata Working Group. (2019). DataCite Metadata Schema for the Publication and Citation of Research Data. Version 4.3. DataCite e.V. https://doi.org/10.14454/f2wp-s162

[4] EMODnet Physics projet. European Marine Observation and Data Network. www.emodnet-physics.eu/map

An interactive atlas to display information on Atlantic marine economic activities

Erwann Quimbert, Ifremer (France), Erwann.quimbert@ifremer.fr
Régis Kalaydjian, Ifremer (France), Regis.Kalaydjian@ifremer.fr
Mickael Treguer, Ifremer (France), mickael.treguer@ifremer.fr

Maritime, Ocean Sector and Ecosystem Sustainability: Fostering Blue Growth in Atlantic Industries

MOSES is an EU multinational research project involving eight partners representing the five member states of the Atlantic Area. The partners of MOSES (universities, marine research institutions, and regional bodies) include representatives from France, Ireland, Portugal, Spain, and the United Kingdom.

MOSES seeks to understand and quantify the sustainability dimension of Blue Growth with respect to the potential environmental impacts of key strategic marine sectors, and will result in the development of integrated marine sustainability assessment toolkits and sustainable transition plans using case studies.

As part of work package 4 activities, MOSES has developed an economic database of marine activities in the five countries of the Atlantic Area. Each economic activity is characterized by a set of indicators, with the objective of having comparable data for the five countries. The purpose of this poster is to present the different steps that led to the creation of this database and atlas.

The methodology and framework used for building the MOSES economic database are based on the work performed for MARNET, which was an INTERREG project (2009-2012) to analyse marine activities in the Atlantic Area. The three dimensions of the MOSES database are similar to that of MARNET: a selection of marine activities and of yearly indicators characterizing each activity; and a common time frame for the indicators as time series.

- Marine activities are identified by NACE code, i.e. by business category. This hierarchical classification is exhaustive (all activities are classified by NACE class, with one code per class), avoids double counting (each activity has one code) and allows readily accessing economic documentation by activity.
- Marine activities are equally identified by territorial unit, based on the NUTS European statistical classification of territorial units. The units used in the database include countries (level 0 of the NUTS) and Atlantic regions (levels 2 and 3), i.e. units with an Atlantic shoreline. Level 1 is less necessary as it includes countries or groups of regions.

- Collecting reliable data requires using official European sources whenever possible: the project gives priority to the databases from Eurostat and the National Statistical Institutes of the five Atlantic countries. The EC statistical administration, Eurostat, collects data from the National Statistical Institutes under EU regulation.
- Economic indicators are selected among those currently used by these sources for developing the Structural Business Statistics and National Accounts. Such indicators are available at NUTS 0 only. At higher NUTS levels (regions and sub regions), only establishments can be documented with much fewer indicators, e.g. number of establishments and employment.
- Specific "proxies" are collected to supplement economic indicators, and are mostly available at local level: proxies are non-monetary indicators characterizing important features of certain marine activities.
- The time frame of the MOSES database (2013-2015) follows that of MARNET (2005- 2012).

The MOSES database

The partners collected the required data from member states' national statistical institutes and from Eurostat, as said above. In several cases, data sources included state or private agencies with an expertise on specific maritime activities such as fisheries.

The poster will present the different steps from integration of these data within a PostgreSQL database until the mapping of these data in an interactive tool using OGC standards for displaying and downloading results and products.

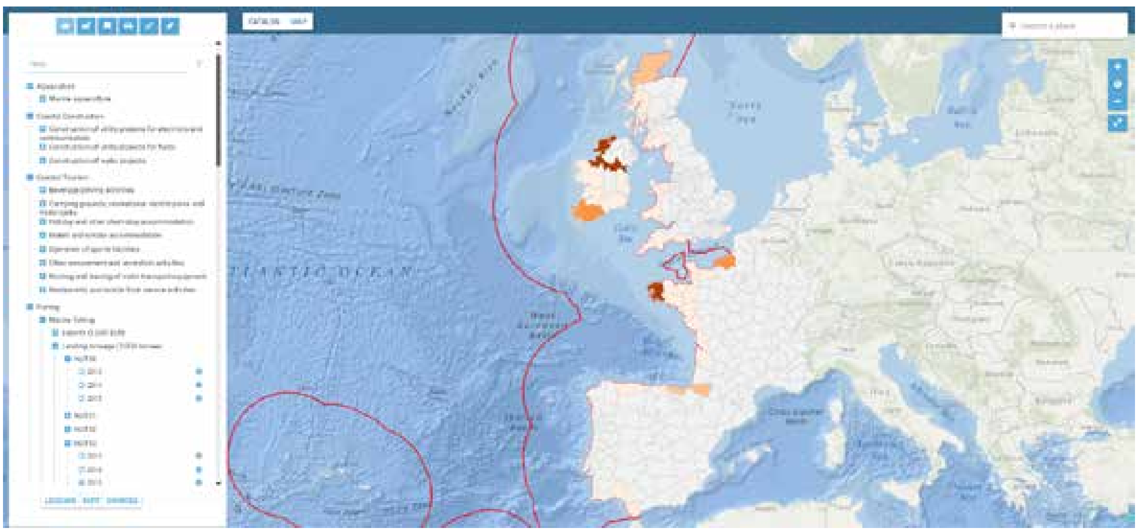


Figure 1: Map representing Marine fishing /Landing tonnage (1000 tonnes) at Nuts 3 level in 2015

Contribution of EMODNET Bathymetry Project to Marine Strategy Framework Directive (MSFD). Case of use: The IEO experience.

Olvido Tello, Spanish Institute of Oceanography IEO (Spain), olvido.tello@ieo.es
Dulce Mata, Spanish Institute of Oceanography IEO (Spain), dulce.mata@ieo.es
Alberto Cabezuelo, Spanish Institute of Oceanography IEO (Spain), alberto.cabezuelo@ieo.es
Gerardo Bruque, Spanish Institute of Oceanography IEO (Spain), gerardo.bruque@ieo.es
Luis M. Agudo, Spanish Institute of Oceanography IEO (Spain), luismi.agudo@ieo.es

The Spanish Institute of Oceanography is involved in the EMODNET Bathymetry project from its beginning. The results of this project are very useful in several projects carried out by researchers of this institution, emphasizing the usefulness in developing the work in the framework of MSFD.

EMODnet: The European Marine Observation and Data Network

The European Union, in line with its policy and commitment to ensure sustainable growth and the conservation of biodiversity and marine resources, has carried out several initiatives in this respect, such as the Integrated Maritime Policy in 2007, the Directive for Marine Strategies (2008/56/EC), the "Blue Growth" to support sustainable growth in the marine environment, the "Marine Knowledge 2020" to increase knowledge of the marine environment and stresses the importance of knowing what data exists and providing access to it, the INSPIRE Directive 2007/2 / CE, to provide interoperability to information and promotes the sharing of information through Spatial Data Infrastructures, the Marine Spatial Planning Directive /2014/89/EU), etc.

With the purpose of encouraging these policies, Marine Knowledge 2020 promotes the open access to marine data through the EMODnet. https://emodnet.eu/

The general objectives of EMODnet are to find, collect, harmonize and standardize the existing data on the marine environment, scattered throughout different European institutions, as well as to provide interoperability into the existing data and metadata. Another objective includes the creation of a Spatial Infrastructure that allows access to the data and metadata. EMODnet provides access to European marine data set in seven fields: Bathymetry, Geology, Seabed Habitats, Chemistry, Biology, Physics and Human Activities.

In order to achieve these aims, the DGMARE launches a "call" under its guidelines searching a consortia of institutions. Since the beginning of this initiative, several calls have been launched, in which the IEO has been participating and working in a consortium.

In the case of bathymetry call, the main objective is to create a muti-resolution DTM with extended coverage which would include European zones. In the current development phase (2017-2020) the IEO in consortium are developing a DTM of ¼ of an arc minute, about 100 meters of resolution. Additionally, the satellite-derived bathymetry data contributes to cover the gaps in the coastal area. Moreover, a quality index is provided at the grid node level in order to inform the user about the quality of the DTMs.

MSFD: Marine Strategy Framework Directive (2008/56/EC)

The MSFD aims to achieve Good Environmental Status (GES) of the EU's marine waters and to protect the resource base upon with marine-related economic and social activities depend.

The Directive enshrines in a legislative framework the ecosystem approach to the management of human activities having an impact on the marine environment, integrating the concepts of environmental protection and sustainable use. https://ec.europa.eu/

MSFD establishes a monitoring program of different descriptors for continuous evaluation and periodic updating of the objectives. In Spain, the Ministry of Ecological Transition (MITERD) is responsible and coordinator of carrying out the MSFD, but the Spanish Institute of Oceanography (IEO) performs the research and study of the different indicators and therefore the tasks of collecting oceanographic data.

The process is cyclical, a review of the different elements of the strategy is produced every six years. The first cycle started in 2012 and the second cycle started in 2018. IEO is working in all monitoring programmes of this cycle.

EMODnet and MSFD. Case of Use: The IEO experience.

With this work we want to highlight that all data collected in EMODnet is essential information in order to have a database with georeference information about the marine environment to achieve the objectives of MSFD.

Focusing on EMODnet bathymetry, the researchers of IEO are participating as partner on the development of the EMODnet bathymetry data (continous DTMs for all european seas). IEO is contributing with new bathymetric surveys, following the common standards of the project, to integrate the regional GRID into the global result. The spanish contribution to EMODnet Bathymetry is also increasing the quality of the DTMs provided. The high resolution of the datasets represent a remarkable improvement to see in more detail the seafloor characteristics and to increase the efficiency of seabed mapping.

On the other hand, the IEO researchers are involved in the development of research in the framwork of MSFD in Spain and they are using the EMODnet bathymetry as important element of reference. The bathymetric data is a fundamental contribution to the study of marine habitats, it is also one of the physical factors on which most of the ecosystems are structured. For the MSFD works is essential to have a continuos DTM bathymetry in order to the stablishment of indicators for each descriptor. For instance, in the MSFD "Biodiversity descriptor", it is essential to know the seabed relief in order to know the habitat situation from the coastal zone to the deep sea. The distribution of marine habitats normally responds to bathymetric criteria, thus determines the existence of certain species and associated communities at different ranges of depth. It is very important to define bathymetrical thresholds to identify the benthic habitats domain as infralitoral, circalitoral, etc.

Moreover, the bathymetry it is especially valuable in the indicators as spatial extent of loss of seabed, spatial extent of physical disturbance to seabed, distribution of habitat adversely altered by physical disturbance. Other utilities of bathymetry are to know the marine currents and therefore in order to know species distribution. In line with the EU MSFD, the marine spatial databases are an essential tool to analysis and to understand the ocean dynamics.

The Bathymetic Model from EMODnet has also presence on other descriptors such as seafloor integrity, where the seafloor plays an essential role to evaluate and to monitor the Good Enviromental Status of the european waters. Hence, It represents an special key in the assesment and management of the marine resources.

Searching for common ground at sea Diving deeper into the process of delivering bathymetry data to different user groups

Ellen Vos, Hydrographic Service of the Royal Netherlands Navy (the Netherlands), em.vos@mindef.nl
Eliza Etman, Hydrographic Service of the Royal Netherlands Navy (the Netherlands), ea.etman@mindef.nl

The need for bathymetry.

The Netherlands Hydrographic Service (NLHS) collects bathymetry data at sea to produce navigational charts for safe navigation. However, other than only for charting this data can serve many other purposes. This re-use of the data for other purposes is an efficient use of resources. Both the EMODnet program and INSPIRE Directive are European initiatives deal with the re-use of bathymetry. The NLHS is committed to both of them and started serving it as open data.

From 'ex usu nautae' to 'ex usu communi'.

The traditional role of a hydrographic agency has changed. Still, the main added value of bathymetric data is its contribution to safe navigation. 'Ex usu nautae', means 'serving the sailor'. However, congruent with the open data trends and needs, bathymetric data is not exclusive for the sailor anymore. It is a common good, and thus 'ex usu communi'. This means that hydrographic organisations need to adapt their mindsets by developing additional skills and adopt new ways of working to meet the challenges posed by these emerging needs from different worlds.

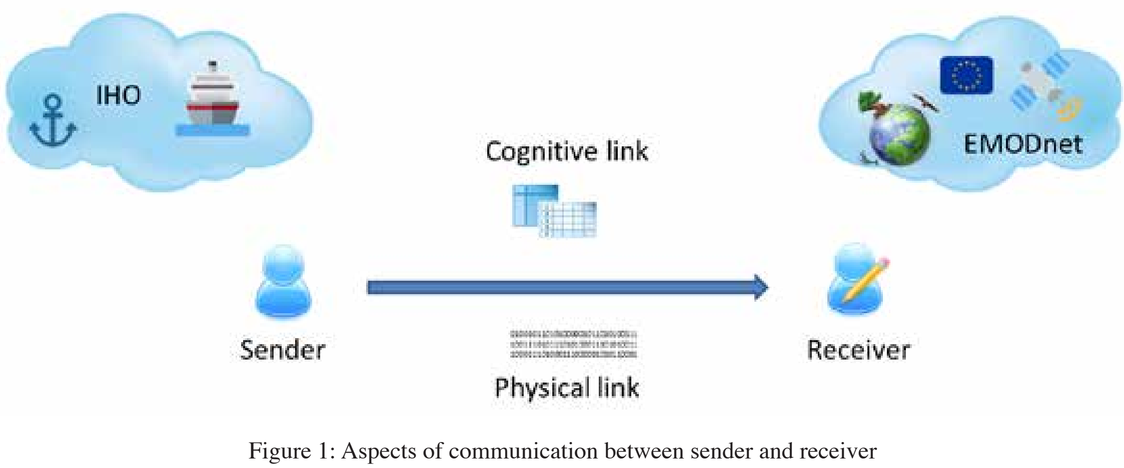


Figure 1: Aspects of communication between sender and receiver

The efforts it takes to produce different data packages for different types of users.

Figure 1 shows that the new types of users (EMODnet and INSPIRE users) speak different languages then the traditional users (navigators on the bridge). In addition, they are used to different standards for packaging data and use (different) portals to find the data. This means that the Hydrographic Service needed to adjust in many different ways to connect to the perspective of this new audience.

For example, metadata is evident content in many portals, as it enables the data to be found. No metadata, no data! In the current situation, for civil purposes only, the bathymetric grid must be described using up to three metadata-profiles, to be found on three different portals: the National Geo-Portal NGR, the European INSPIRE Portal and the EMODnet Portal. Although these profiles are all based on the ISO 19115 metadata specifications, there are both semantic differences and procedural differences in acquiring the metadata files.

Recently, Flemish Hydrography has initiated the metadata profiles of EMODnet to become INSPIRE compliant. This European example helps us to investigate to what degree metadata-profiles could be translatable to other profiles.

Breakdown of the new composite process in structural building blocks.

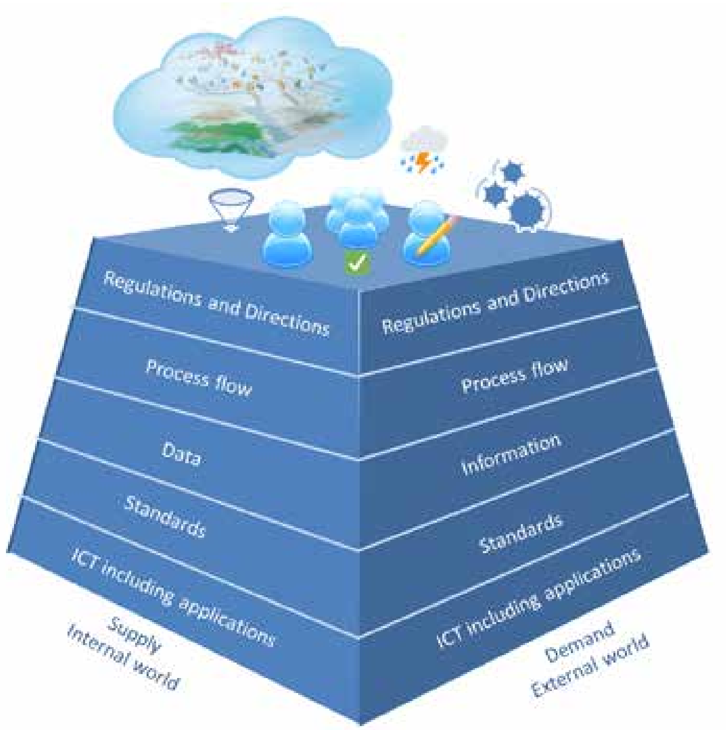


Figure 2: Framework to enable common understanding

In our presentation, we will dive deeper into the process. We will use the framework presented in Figure 2 to explain how we changed our internal structures to pass the bathymetry over to EMODnet and INSPIRE users in a meaningful way.

Sharing solutions looking for Common Ground.

Our experience at the Netherlands Hydrographic Service is that there does not exist a ‘one size fits all’ solution for Hydrographic Offices or other organizations to deliver good quality bathymetry to the navigator and both the EMODnet and INSPIRE user. Nevertheless, we intend to share our story of our ‘obstacles and solutions’ looking for Common Ground with other organizations to share stories about pain and glory and to share solutions together. Besides, applying the framework together as a community helps to strengthen our shared learning capacity, which helps to continue to meet future needs: serving Blue Growth.

A Non-linear Quality Control Procedure for Representativeness errors in Ocean Historical Datasets

Kanwal Shahzadi, University of Bologna (Italy), kanwal.shahzadi@unibo.it
Nadia Pinardi, University of Bologna (Italy), nadia.pinardi@unibo.it
Vladyslav Lyubartsev, CMCC (Italy), vladyslav.lyubartsev@cmcc.it

In order to compute world ocean climatologies representing the low frequency variability of ocean essential variables, such as temperature and salinity, there is a need to consider the concept of uncertainty in the input profiles. Uncertainty in the input data arises from different kinds of errors, mainly categorised into gross and representativeness errors (Janjic et., 2017). Systematic errors, parts of the gross error are attributed to incorrectly calibrated instruments, incorrect archiving or technical problems in data transmission, instrument sensitivity and accuracy. Representativeness errors instead represent the inadequate resolution of the space and time scales of variability of the ocean fields. In particular for ocean climatologies, the representativeness error is given by the inability of the input data to represent the low frequency signals.

A non-linear Quality Control (NQC) procedure is developed (Shahzadi et al., 2020) similar to Jia et al. (JAOT, 2016) in order to eliminate non-representative or high frequency signals of the World Ocean Database 18 (WOD18) ARGO profiling floats Temperature and Salinity profiles. The procedure requires the subdivision of the domain of interest into regime oriented regions in order to make the statistical estimates of the mean and standard deviation as reliable as possible, i.e. representative of the climatology of that sub region. In addition, mean and standard deviations will be computed in the sub regions by gridding the intermediate results of the NQC procedure with an objective analysis (OA) algorithm developed by Carter and Robinson (1987). At each iteration of the NQC procedure, data are compared to standard deviations calculated from the gridded field in the sub regions and rejected if their value exceeds 2-3 standard deviations. The procedure flags the rejected data until a convergence is obtained, i.e. no more data is rejected by comparison with the standard deviations in the sub regions (Fig.1).

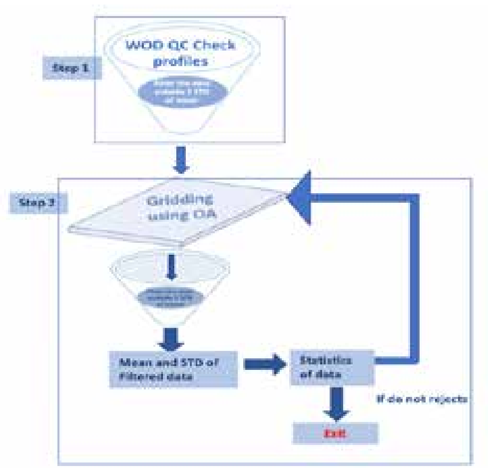
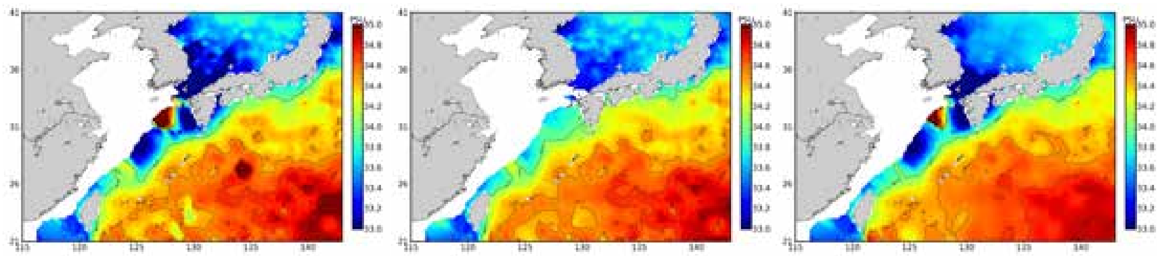


Figure. 1: Nonlinear Quality Control Schematics

The procedure is applied in the North West Pacific (NWP), North Atlantic (NA) and South Atlantic (SA) as different test areas using ARGO profiling floats for the period 2003-2017. Here, we show the results for NWP and for August at 10m, along with the sensitivity of the NQC to the regional subdivisions, i.e., a dynamical and a regular, 5x5 degrees subdivision. In Fig. 2, we show the results of the NQC as compared with the simple standard deviation quality control as defined in WOD18. It is clear that the NQC subtracts the high frequency signals from the climatological estimate more efficiently. Furthermore, it is found that dividing the area in dynamical sub regions makes the NQC procedure converge, rejecting a relatively low number of profiles as compared to regular subdivisions.



(i) (ii) (iii)

Fig. 2: Salinity Field at 10m, (i) only WOD statistical check and (ii) NQC check (regime oriented division) and (iii) NQC check (regular division) after 3rd iteration

References

1- W. Jia, D. Wang, N. Pinardi and S. Simoncelli, 2016. *A Quality Control Procedure for Climatological Studies Using Argo Data in the North Pacific Western Boundary Current Region,* 33, 2717-2733.

2- T. Janjic, N. Bormann, M. Bocquet, 2017, *On the representation error in data assimilation,* 144:1257–1278.

3- K. Shahzadi, N. Pinardi, V. Lyubartsev, 2020, A *Nonlinear Quality Control for large-scale Ocean Climatologies* (in preparation).

4- E. F. Carter, and A. R. Robinson, 1987, *Analysis model for the estimation of oceanic fields,* 4, 49–74.

Assurance offshore monitoring, a cross-disciplinary approach.

Guttorm Alendal, University of Bergen (Norway), Guttorm.alendal@uib.no,
Jerry Blackford, Plymouth Marine Laboratory (UK), JCB@pml.ac.uk,
Elisa Calignano, TNO (The Netherlands), elisa.calignano@tno.nl,
Stefan Carpentier, TNO (The Netherlands), stefan.carpentier@tno.nl,
Dorothy J. Dankel, University of Bergen (Norway), dorothy.dankel@uib.no,
Marius Dewar, Plymouth Marine Laboratory (UK), made@pml.ac.uk,
Bjarte Fagerås, OCTIO Environmental (Norway), bjarte.fageraas@octio.com,
Sarah E. Gasda, NORCE (Norway), sgas@norceresearch.no,
Kristian Gundersen, University of Bergen (Norway), kristian.gundersen@uib.no,
Raphael Heffron, University of Dundee (UK), r.heffron@dundee.ac.uk,
Martha Lien, OCTIO Environmental (Norway), martha.lien@octio.com,
Anna Olyenik, University of Bergen (Norway), Anna.Oleynik@uib.no,
Abdirahman Omar, NORCE (Norway), abom@norceresearch.no,
Rajesh Pawar, Los Alamos National Laboratory (USA), rajesh@lanl.gov,
Katherine Romanak, University of Texas at Austin (USA), katherine.romanak@beg.utexas.edu,
Darren Snee, Plymouth Marine Laboratory (UK), dsn@pml.ac.uk,
Sigrid E. Schütz, University of Bergen (Norway), Sigrid.Schytz@uib.no,
Parisa Torabi, University of Bergen (Norway), Parisa.Torabi@uib.no.

We outline the approach chosen in an ongoing project, ACTOM[1], aiming to develop procedures for design and execution of appropriate, rigorous and cost-effective monitoring of offshore geological carbon storage (OGCS), aligning industrial, societal and regulative expectations with technological capabilities and limitations.

An interdisciplinary consortium applies methods to critically assess secure offshore CO_2 storage as this technology becomes implemented and upscaled internationally as a key greenhouse gas emissions reduction strategy. Reserving marine areas for subsea CO_2 storage raise new and difficult questions on policy and legal frameworks. International law, regional (like EU) and national regulations come into play in the CO_2 storage permit procedure, including preliminary Risk and Project Impact assessments, procedural requirements regarding stakeholder and public participation, and public access to information and integrated assessments. In the last 20 years, more strategic and holistic instruments for management of (sea) areas and resources have been developed, compared to the traditional license system.[2] Examples might be Marine Spatial Planning (MSP) requirements for opening marine areas for storage based on Strategic Environmental Assessments (SEA).[3]

Formulating appropriate monitoring programs, from either a regulatory or operator viewpoint, is very difficult to achieve without a properly quantified cost-benefit analysis of what that monitoring could achieve. In addition, communicating risks and uncertainties is a challenge for offshore storage projects, and tools assisting in dialogue with stakeholders, governments and public at large will be of value.

Currently, regulations require assurance of zero or minimal loss of storage integrity, but generally do not specify what constitutes acceptable assurance. Providing assurance of storage conformance with an acceptable standard is philosophically challenging. Whilst we seek evidence of absence (of leakage), that cannot be delivered by an absence of evidence. Any given monitoring observation has a finite relevance in both space and time. The question is therefore – what is a sufficient number and type of observations for assurance?

Transparency of decision-making, open and credible science technology, social, legal and ethical considerations are prerequisites for high public acceptance of new technologies.[4,5,6] The monitoring programs will have a role in communicating risks and benefits for storage projects and assure against unjustified accusations for having adverse environmental effects, but cannot be seen in isolation from the multi-leveled Carbon Capture and Storage (CCS) management systems.

Evaluations of CO_2 storage monitoring techniques usually aim to determine the suitability to user-defined project scenario (e.g. IEAGHG monitoring selection tool[7]) or to assess the availability of sensors that can measure variables that are likely to fluctuate under a seepage scenario, or processes that are sensitive to CO_2-related stress. Less focus has been on how they perform relative to regulatory requirements, cost efficiency, and user friendliness.

We can use observations and models to characterise the natural variability of the marine system, or, the noise from which an anomalous signal must be detected. Here we focus on detecting changes in pH caused by CO2 seeps through the seafloor from unknown locations. The strength of the signal will be reduced very quickly away from the source and will be hidden within the natural pH variability.

We can use models to simulate hypothetical leak events thereby defining the monitoring target(s). We have algorithms that assess the cost-benefit of a range of anomaly criteria – i.e. a signal that would provoke a more concerted monitoring campaign and finally algorithms that can derive the optimal deployment strategy – i.e. where to monitor and when. The challenge is to collate these abilities into a coherent whole, which then allows the presentation of an evaluated monitoring system that can be judged against regulatory and societal expectations.

In the project, we establish a toolbox capable of simulating "what if" seep scenarios, as well as monitoring deployments, that can be used to deliver environmental impact assessments as required under the CCS and EIA directives[8]. As a result, recommended monitoring strategies could be delivered autonomously and be largely dependent on established generic operational marine simulation models, both factors reducing costs.

The overall framework that we use to connect the models and scenarios to society is *Responsible Research and Innovation (RRI)*[9]. In RRI societal actors work together during the whole research and innovation process (legitimacy) in order to better align both the process and its outcomes (control), with the values, needs and expectations of society (orientation). In accordance with the RRI approach researchers, citizens, policy makers, businesses and organizations will be engaged in discussions of how the CCUS monitoring innovation process and the outcomes best can be aligned with society's climate change mitigation objective in an iterative dialogue.

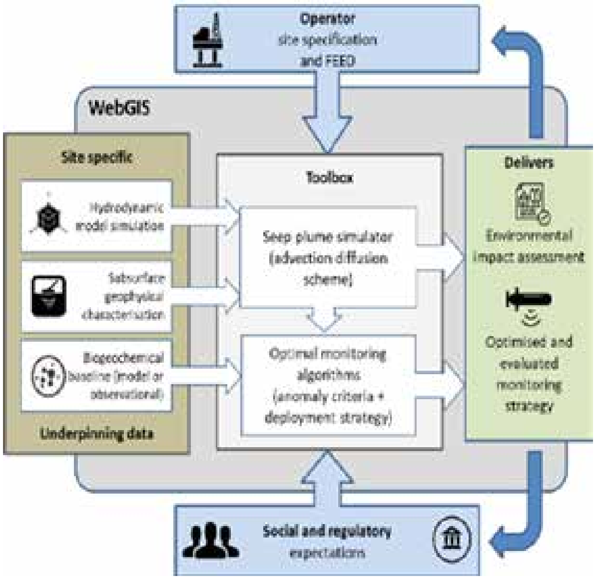


Figure 1: Conceptual sketch of the monitoring tool kit.

References

1 - https://actom.w.uib.no

2 - Schütz in Raphael J., Heffron; Little, Gavin F.M.(ed), "Renewable energy production in marine areas and coastal zone – the Norwegian model" in Delivering Energy Law and Policy in the EU and the US, Edinburgh University Press 2016.

3 - Directive 2001/42 EU on the assessment of the effects of certain plans and programmes on the environment.

4 - Funtowicz, S.O., and Ravetz, J.R. 1993. Science for the post-normal age. Futures, 25(7): 739-755.

5 - Funtowicz, S.O., and Ravetz, J.R. 1990. Uncertainty and quality in science for policy. Springer Science & Business Media. 231 p.

6 - Kloprogge, P., and van der Sluijs, J. P. 2006. The inclusion of stakeholder knowledge and perspectives in integrated assessment of climate change. Climatic Change, 75: 359–389.

7 - https://ieaghg.org/ccs-resources/monitoring-selection-tool

8 - Following inclusion of capture installations in EIA Directive (85/337/EEC), now repealed by Directive 2011/92/EU, environmental impact assessment has to be carried out in the capture permit process, cf. preamble CCS Directive sec. 27.

9 - https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

Interoperability of institutional data management systems

Hela Mehrtens, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany), hmehrtens@geomar.de
Pina Springer, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany), pspringer@geomar.de
Carsten Schirnick, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany), cschirnick@geomar.de
Claas Faber, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany), cfaber@geomar.de
Doris Maicher, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany), dmaicher@geomar.de
Lisa Paglialonga, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany), lpaglialonga@geomar.de
Sören Lorenz, GEOMAR Helmholtz Centre for Ocean Research, Kiel (Germany), slorenz@geomar.de

Motivation

The knowledge about FAIR data requirements is rising. As an institutional data management team, we implement, develop and maintain infrastructures, tools and services at GEOMAR for more than a decade now. We have systems to make data findable, accessible and reusable but interoperability is the remaining challenge. Here we describe the infrastructure, challenges and next steps to come closer to this aim.

The GEOMAR data management systems

In 2009 GEOMAR started to establish an institutional data management platform enabling collaborative data management in research projects by integration of several components (Fig. 1): (1) OSIS, (2) OceanRep, (3) Liferay portal, (4) THREDDS / OPeNDAP, (5) Gitlab. The challenge was to come up with data management solutions that were fast to implement, easy to use and highly accepted by users. (1) The Ocean Science Information System **OSIS** (osis.geomar.de) is our metadata and data exchange system for research data, which was developed in-house. It serves both as the instrument to collect the necessary metadata and also as an information system, allowing the search and dissemination of metadata and data. For several projects it is used to collect data of expeditions, experiments and model experiments and to display and remind for deliverables such as to upload or publish data in a domain specific data centre according to the timeline of an agreed data policy. It serves as our central data information hub. (2) **OceanRep** is the institutional repository at GEOMAR, based on the EPrints software. It contains quality controlled entries of the research output at GEOMAR, links to data and software repositories and project information, shows collaboration with research partners and can be harvested via OAI-PMH. (3) We use the **portal** software Liferay as our central data management platform. It comes out of the box with a lot of useful collaboration features. It also regulates the individual or project based access for researchers to a series of data management applications. More than 100 separate sites serve researchers as internal workspaces for their project collaboration. It also offers public websites and our services can be integrated into these sites (project specific views of OSIS, publication lists from OceanRep). (4) An **OPeNDAP** server provides access to model data referenced in peer-reviewed papers or papers in discussion. OPeNDAP is an established protocol for sharing access to large gridded datasets and allows subsetting on the server side. We are currently using a THREDDS server to provide our OPeNDAP service. (5) Software is increasingly regarded as integral part of sustainable research and as a research output in its own right. To enable collaborative software development in research projects, we operate **Gitlab** as a service within GEOMAR and also for external project partners. Our Gitlab instance on premise not only offers revision control for software code but also collaboration tools (e.g. issue tracker), a container registry as well as development tools (e.g. Continuous Integration pipelines).

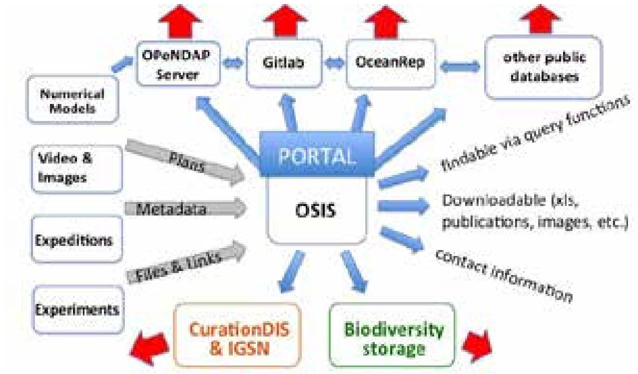


Figure 1: Relation of data and sample management structures

Current challenges

To open our collaborative data management system to the community, we face organisational and technical challenges. The integration and interoperability of the components on the institutional level is established, as shown in Fig. 1. However, to reach interoperability for the marine community, locally established workflows and vocabularies have to be opened to discussion within the community and adapt to external requirements. Local operational constraints need a technical redesign. The institutional data management service portfolio has, at least partly, to be transformed into a community service portfolio. To achieve these efforts, we plan: (1) Machine readable provision of metadata has to be implemented on an early stage of data acquisition. (2) Design tools for researchers to allow them to provide accurate and relevant attributes for their data from the beginning.(3) Enhancing the use of persistent identifiers along the data and sample collection and managing cycle.

Next steps raising interoperability

The given institutional implementations of data management infrastructures and services are diverse. They typically have discipline specific and technical specialisation that fit perfectly for the institution. This is true for example at GEOMAR for its established workflows and interoperable components (1. to 5. described above) which foster collaboration and exchange of the most diverse expertise. This strength is counteracted by a weak interoperability with research data management (RDM) environments of other institutional or community-specific systems. On the long run, and to preserve productive diversity, only a federated community infrastructure will meet the requirement of interoperability. This requires trusting and trusted services by the providing institution and a negotiation among the community on a federated RDM service portfolio. GEOMAR aims at providing a collaborative RDM platform for marine projects as a community service. How to achieve this on the operational level is currently developed jointly with other marine research institutions in community projects, e.g. within the DAM (German Alliance of Marine Research) and MareHub, the marine-oriented initiative within the Helmholtz research area Earth and Environment which is reaching out to other disciplines, for example atmosphere and earth and even beyond. We will present our ideas on how to meet the challenges and a roadmap towards them.

Operational results of Temperature and Salinity Quality Control at Coriolis for CMEMS based on improved MinMax approach

Jérôme Gourrion, OceanScope (France), jerome.gourrion@ocean-scope.com
Christine Coatanoan, Ifremer-SISMER (France), christine.coatanoan@ifremer.fr
Delphine Dobler, delphine.emilie.dobler@protonmail.com

Since 2019, a novel quality control (QC) procedure for temperature and salinity observations is implemented and active at the Coriolis data facility in the near-real-time (NRT) qualification framework. The MinMax approach compares an observation to a local validity interval with bounds estimated from the historical variability through minimum/maximum values (Gourrion et al. 2020). A significant improvement of both the efficiency (total number of good detections) and the robustness (ratio of bad to good detections) of the procedure is obtained through an adhoc extrapolation of the sample minimum and maximum values ; this can be understood as a simulation of the missing variability in the historical observations used as reference.

Figure 1 shows some examples of ARGO salinity profiles as screenshots of the inspection tool used operationally to visualize the automatically raised detections. The left panel likely illustrates a case of biofouling in the conductivity cell that perturbates the measurement in some ascending profile at all pressure levels above 1100 dbar. The right panel displays a case of platform for which the conductivity cell is experiencing a positive and increasing temporal drift.

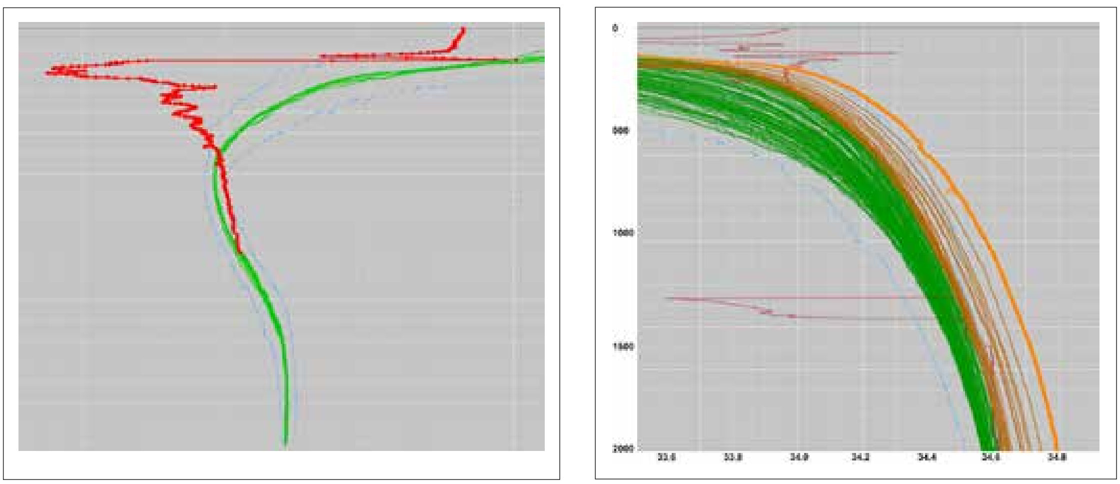


Figure 1: Left panel : Salinity profiles from ARGO profiler (green), one cycle is flagged as 'bad' with QC = 4 (red). Right panel : Salinity profiles from ARGO profiler (green) with cycles flagged as 'bad potentially correctable' with QC = 3 (orange).

Here, the case of conductivity drift is of particular interest as the Min/Max method has allowed since early 2019 earlier detections of the ARGO platforms affected by such a sensor failure with increased occurrence in the last years (see http://www.argodatamgt.org/Data-Mgt-Team/News/Message-to-Argo-users-about-an-increased-occurrence-of-salinity-erros-in-the-real-time-Argo-data-stream).

For CMEMS, the ARGO profiles with automatic NRT detections are not distributed ; in a second step, an operator checks the detections and may confirm them, especially in the cases illustrated in Figure 1. Beside these useful error detections, the method may also provide erroneous detections. As an illustration, Figure 2 shows a time series over about one year of the number of good and bad detections. For ARGO, it appears that 85 % of detections are confirmed, while 15 % are erroneous detections. When the operator faces an erroneous detection, the distribution is unlocked, and the data are finally made available to CMEMS with a few days delay after observation. If the bad detection is found to be caused by an erroneous Min or Max value in the reference fields, he/she may modify manually the operational reference fields in the concerned geograpical/depth grid cell.

Figure 2: time series of number of total/good/bad Min/Max alerts for the CMEMS-INSTAC ARGO dataset.

In the presentation, the latest method developments made at OceanScope in order to further reduce the ratio of bad/good detections will be presented ; updated operational statistics over 2020 will follow, demonstrating the benefits on the quality of the main datasets concerned, the CMEMS-INSTAC and ARGO NRT T/S products.

Acknowledgement

This work has received financial support from CMEMS-INSTAC and from the European Maritime and Fisheries Fund (EMFF) under the MOCCA grant with agreement N° EASME/EMFF/2015/1.2.1.1/SI2.709624 Monitoring the Oceans.

References

J.Gourrion, T.Szekely, R.Killick, B.Owens, G.Reverdin, B.Chapron: Improved Statistical Method for Quality Control of Hydrographic Observations, J. Atmos. Ocean. Tech, 2020. DOI: 10.1175/JTECH-D-18-0244.1

SESSION SERVICES:

Data services and tools in ocean science

- Services and tools for discovery, visualisation, dissemination (considering the potentially diverse range of stakeholders/end users) and education
- Services and tools for data processing and data analysis
- Services and tools for data ingestion
- Services and tools for data management (data reformatting, quality checks, data access…)
- Virtual Research Environment (VRE), Cloud environment
- European Open Science Cloud (EOSC) and Blue Cloud pilot
- Open data for the marine domain
- Federated search

ORAL PRESENTATIONS

Virtual Research Environments supporting sustainability of global fisheries

Marc Taconet, FAO (Italy), Marc.Taconet@fao.org
Aymen Charef, FAO (Italy), Aymen.Charef@fao.org
Anne-Elise Nieblas, FAO (France), Anne.Nieblas@fao.org
Aureliano Gentile, FAO (Italy), Aureliano.Gentile@fao.org
Emmanuel Blondel, FAO (France), Emmanuel.Blondel@fao.org
Julien Barde, IRD (France), Julien.Barde@ird.org
Anton Ellenbroek, FAO (Italy), Anton.Ellenbroek@fao.org

Introduction

Under the wider umbrella of the **European Open Science Cloud (EOSC),** 'The Future of Seas and Oceans Flagship Initiative', aims for the development of a thematic Marine EOSC serving the Blue Economy, Marine Environment and Marine Knowledge agendas. The Blue-Cloud project is a broad collaboration that builds this cloud, and the Food and Agriculture Organization of the UN (FAO) is using Virtual Research Environments (VREs) in two Blue-Cloud demonstrators; one for fisheries, one for aquaculture.

In the 2030 Agenda for Sustainable Development, 193 States pledged to ensure that "no one will be left behind" to achieve the Sustainable Development Goals (SDGs). For SDG Goal 14 "Life below water", FAO aims to support data sharing, data analysis and data dissemination with Blue-Cloud VREs, and 3 example VREs supporting these steps for the SDG indicator 14.4.1 'Proportion of fish stocks within biologically sustainable levels' will be introduced.

Tuna Atlas VRE; a cloud enabled database and OGC compliant Spatial Data Infrastructure

The Global Tuna Atlas VRE is a tool to integrate and analyse public domain fishery statistical data from various tuna Regional Fisheries Management Organizations. Data include catches and fishing efforts and size frequencies of the species managed by the five tuna RFMOs. It was developed with co-funding from H2020, FAO and IRD, and its strength is an ISO/OGC compliant spatial data infrastructure in the D4Science / Blue-Cloud. This ensures high reliability, re-usability and sustainability, and similar cases requiring the management of spatial data can quickly establish core services. Enabling these services is made possible by open source tools such as *geoflow*[1], a R package for managing and automating (meta)data publication workflows, and *OpenFairViewer*[2]; a re-usable map visualization web component, based on international standards (ISO, OGC) supporting FAIR data implementation, and offering standard data access and sharing fully driven by metadata. For instance, these tools will allow other Blue-Cloud VREs such as the GRSF to publish and disseminate information about stocks and fisheries.

1 https://doi.org/10.5281/zenodo.3138920

2 https://doi.org/10.5281/zenodo.2249305

In H2020 Blue-cloud, the Tuna Atlas will grow to include information from other fisheries, and add new datasets to obtain a better view over global fisheries. The cloud environment will facilitate to have for-purpose maps, for instance for a subset of the ocean, and add maps of fishing effort and Essential Ocean Variables.

SDG 14.4.1 VRE; an integrated R Shiny Application

The SDG 14.4.1 VRE plays a key role in FAO support to the global community by offering web-based computational facilities with considerable processing resources to execute a specially-developed Stock Monitoring Tool (SMT) with R-shiny. The SMT can be used to run some methods of fish stock determination and thus to evaluate the sustainability of fish stocks. The VRE is used in FAO capacity development workshops as an operational e-training environment coupled with hands-on interactive session to test the SMT. The training allows users to develop their skills using sample datasets and use their own datasets with the tool. The VRE is related to a wider overall FAO e-Learning module on SDG 14.4.1 that introduces basic fisheries concepts and definitions, illustrates some technical aspects of stock assessment and provides detailed guidance on the process and tools for the analysis and reporting of the Indicator.

The SDG 14.4.1 VRE is available to authorized users after the training, and is available to the global community upon request. It is a community-oriented, flexible, and secure working environment that was developed through a collaboration between FAO and CNR-ISTI. It is integrated in the D4Science e-Infrastructure (www.d4science.org) that is also used for Blue-Cloud. The infrastructure fosters collaboration between users and regulate users’ access to data and services. Downloadable reports are automatically saved to the private workspace, enabling users to trace their tests, and create reproducible results.

Access to the SDG 14.4.1 VRE is now controlled; it is available to authorized users within the window of a few days prior to and after the training. However, is available to the general public upon request.

Global Record of Stock and Fisheries VRE; an on-line registry combining CKAN and semantic technologies

In an ever more data-dependent world national, regional and global institutions need the best available scientific evidence to ensure sustainable fisheries. The Global Record of Stock and Fisheries (GRSF) is the digital innovative public-private partnership offering data services in support of: 1) keeping a record of the state of stocks and fisheries for reporting on SDG indicators and 2) harmonizing public and private data on stocks and fisheries for seafood traceability, ecolabelling and food safety. Each stock and fishery reported through the SDG 14.4.1 process gets a unique identifier and all data that describe the area, management authority, and production that are needed for the monitoring of fish stocks and fisheries status are harmonized. GRSF identifiers enable a seamless data flow from data collection, through harmonization, to dissemination. The GRSF is both a collaboration and a global repository of stocks and fisheries resulting from collation and merging of records across (currently) three data sources; FAO Fisheries and Resources Monitoring System (FIRMS), Sustainable Fisheries Partnership, and University of Washington.

The GRSF is a FAO initiative and developed with H2020 funding. It is part of the iMarine Data Catalogue and is public data are accessible through a registry https://i-marine.d4science.org/web/grsf/data-catalogue, and a map-viewer https://i-marine.d4science.org/web/grsf/map-viewer.

H2020 Blue-Cloud will add more data-services to GRSF to add data from other domains, such as time series download (e.g. csv files of catch and effort data), Essential Ocean Variables from capture areas, or effort maps. The data will also be made available as Blue-Cloud Web services (APIs in JSON format).

The infrastructure; D4Science

The e-Infrastructure (www.d4science.org) for these VREs is operated by CNR-ISTI and also used for Blue-Cloud. It uses Liferay portals as the portlet-hosting platform, as a complete platform for building web apps, mobile apps, and web services. It is JSR 168 and JSR 286 compliant. These VREs can take various forms including web interactive user interface, web applications, and pluggable standalone user interface. There are different levels of user authorization: VRE Manager, VRE Designer, and VRE User.

General features that are shared by all VREs include a social networking collaboration platform to use the common facilities typical of social networks e.g., posting news, commenting on posted news of common interest. The community space allows users to interact both publicly and privately, and provides a sense of community between fisheries scientists worldwide. The VRE provides private and public workspaces where data and documents can be stored and shared. Every user can have up to 100 GB storage space. An important feature of the VRE is the interaction between the VRE and the private workspace.

Conclusion

FAO is committed to mobilizing technologies and services to support restoring overfished stocks to sustainable levels. VREs provide an online environment to the global fisheries science community with access to dedicated tools but with robust data policies.

The VRES can deliver innovation for communities that want to collaborate across domains to achieve the SDGs, and. FAO strives to expand the range of these VRE tools and services for the global SDG community.

SeaDataCloud Virtual Research Environment: Implementation and Technical Aspects

Merret Buurman, German Climate Computing Centre (Germany), buurman@dkrz.de
Charles Troupin, GHER/ULiège (Belgium), ctroupin@uliege.be
Alexander Barth, GHER/ULiège (Belgium), a.barth@uliege.be
Leo Bruvry-Lagadec, IFREMER (France), Leo.Bruvry.Lagadec@ifremer.fr
Sebastian Mieruch, Alfred-Wegener-Institute (Germany), sebastian.mieruch@awi.de
Naranyanan Krishnan, UKRI STFC (United Kingdom), naranyanan.krishnan@stfc.ac.uk
Giorgio Santinelli, Deltares (The Netherlands), Giorgio.Santinelli@deltares.nl
Fedor Baart, Deltares (The Netherlands), fedor.baart@deltares.nl
Peter Thijsse, MARIS (The Netherlands), peter@maris.nl
Filip Waumans, Vlaams Instituut voor de Zee (Belgium), filip.waumans@vliz.be
Themis Zamani, Greek Research and Technology Network (Greece), themis@grnet.gr

What is a Virtual Research Environment?

A Virtual Research Environment (VRE) is a collaborative environment to perform data-driven research. Gone are the days when you had to download all input files to process them on a slow, memory-lacking desktop computer! Instead, perform your analyses with ample data and the latest versions of the software tools on performant hardware in the cloud! The results can be shared and the processing workflow can be reproduced by other users.

The SeaDataCloud VRE (availble to authorized users at https://vre.seadatanet.org/) is a pilot application and supports research with marine data and SeaDataNet tools in the cloud. Its user-facing functionalities are described in the abstract "*Working with the SeaDataCloud VRE: what can we do for you?*". This abstract focuses on the architecture and technical details.

How is the Virtual Research Environment deployed?

The VRE's components are deployed on servers across various data centres to distribute the load. The shared central components as well as the various processing services are deployed as Docker containers. Containers provide a standardized operating environment across all data centres. This significantly facilitates the deployment and the integration of heterogenous services. The services are loosely coupled to allow for easy extension: The VRE provides central services such as Single Sign-On and central web storage for users, but the services can also run as standalone services.

Shared components: Dashboard, private workspace and file selector

The dashboard component is the heart of the VRE, responsible for managing diverse aspects of the VRE. It provides the main user interface with entrypoints to the different services, and access to the user's private workspace, an online storage space for users to store their own datasets. Behind the scenes, it triggers data synchronisation mechanisms and ensures user authorization for the various services by a token-based authorization system. The dashboard is a web application based on the PHP framework Laravel, which includes a user management system and state-of-the-art security measures. Marine-ID [https://www.marine-id.org/] is used to provide Single-Sign-On based on EUDAT's *B2ACCESS* service used by several European projects [https://eudat.eu/services/b2access].

The private workspace is a customized instance of the well-known open-source software NextCloud [https://nextcloud.com/], running on SeaDataCloud's servers. **The file selector** is a bridge between most VRE services and the NextCloud, allowing users to select data from their private workspace to be handed over to the processing service instances by the backend.

Two types of processing services

The actual processing services are what researchers use to perform their analyses. Two types of services are deployed in the VRE. The first type of service operates like a normal web server, where one instance serves many users. The *BioQC* tool, the *VIZ-ualisation* tool, and *webODV* are examples for this type. For these services, one Docker container is running continuously and waiting for users.

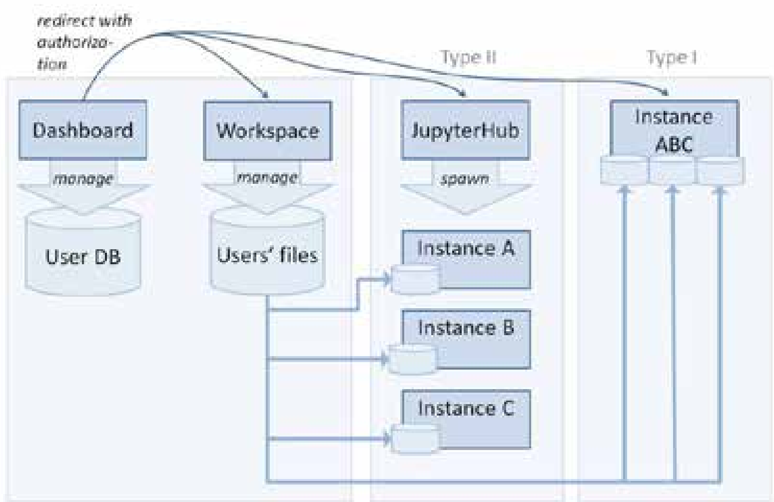


Figure 1: Loosely-coupled structure of VRE services

The second type of service is for the more computationally intensive tools, or for those that are not optimized for handling multiple users' sessions (this frequently occurs when desktop software is ported to the cloud). For this type of service, every user works with an individual instance. As services are packaged as Docker containers, this means that we need to deploy one container per user. We use JupyterHub [https://jupyter.org/hub] to spawn the container when a user logs in. JupyterHub was developed to serve JupyterNotebooks, but it can be used for any tool that is "dockerized" and interacts with the user via HTTP. JupyterHub's benefits include instance management, authentication/authorization, and web security measures such as reverse proxying and SSL termination. This solution is used for the services DIVAnd and ERDDAP subsetting.

DIVAnd is a tool designed to perform interpolation of oceanographic observations onto a n-dimensional grid. The code is written in the Julia programming language and is complemented by a set of Jupyter notebooks. The Docker image is publicly available on DockerHub includes the installation of the Julia language along with packages for plotting and manipulation of netCDF files, the master versions of the DIVAnd Notebooks, which users can adapt and use to prepare data products, and of DIVAnd itself [https://github.com/gher-ulg/DIVAnd.jl].

The VRE's **subsetting service** embeds an instance of ERDDAP [https://coastwatch.pfeg.noaa.gov/ erddap/index.html], configured dynamically and started on the fly according to a dataset selected by the user via the file selector. Its docker image includes the the ERDDAP web application running in an Apache Tomcat web server and a visualization frontend using VueJS framework.

This was a short introduction to the technical side of our VRE. The container-based deployment has proved stable and convenient to maintain. In future, we hope to be able to extend and improve the VRE. Besides incorporating user feedback, possible areas of improvement concern the container orchestration (e.g. by introducing technologies such as *kubernetes* or *Apache Mesos & Marathon*) and the synchronization of data between the data centres.

Working with the SeaDataCloud Virtual Research Environment: what can we do for you?

Charles Troupin, GHER/ULiège (Belgium), ctroupin@uliege.be
Alexander Barth, GHER/ULiège (Belgium), a.barth@uliege.be
Leo Bruvry-Lagadec, IFREMER, Plouzané (France), Leo.Bruvry.Lagadec@ifremer.fr
Merret Buurman, DKRZ, Hamburg (Germany), buurman@dkrz.de
Seppo Kaitala, SYKE (Finland), Seppo.Kaitala@ymparisto.fi
Narayanan Krishnan, UKRI STFC (United Kingdom), narayanan.krishnan@stfc.ac.uk
Sebastian Mieruch, AWI, Bremerhaven (Germany), sebastian.mieruch@awi.de
Giorgio Santinelli, Deltares, Delft (The Netherlands), Giorgio.Santinelli@deltares.nl
Peter Thijsse, MARIS, Nootdorp (The Netherlands), peter@maris.nl
Filip Waumans, VLIZ (Belgium), filip.waumans@vliz.be
Themis Zamani, GRNET, Athens (Greece), themis@admin.grnet.gr

Motivations to build a Virtual Research Environment

For researchers, the conventional work on their desktop computer to analyse data and generate products is getting more and more complex due to the increasing amount of data to process. Three main hurdles often prevent them from performing the required operations:

1. Insufficient CPU power, leading to too long computation times
2. Lack of available memory to process large datasets
3. Insufficient disk storage, making the use of the full datasets impossible.

To overcome these, cloud services are becoming common practice, such as the European Research Data Infrastructure (EUDAT, https://eudat.eu/), the European Open Science Cloud (EOC, https://www.eosc-portal.eu/). With such services, no more downloading of data is needed: all the processing is performed *close* to the data. In addition, the cloud allows cooperative research on a common platform, with a guarantee that the latest versions of the software tools are installed, and enhances the reproducibility of the scientific work. In order to provide the oceanographic community with seamless access to the SeaDataNet data, standard and software tools, the SeaDataCloud (SDC) Virtual Research Environment (VRE) has been developed.

This paper describes the tools made available to users. The implementation and deployment are detailed in "*SeaDataCloud Virtual Research Environment: Implementation and technical aspects*".

SeaDataCloud VRE tools

Users login with their MarineID (https://www.marine-id.org/) and have access to a dashboard that provides access to a private workspace as well as the different services described hereinafter.

webODV consists of a suite of online services based on Ocean Data View (ODV, https://odv.awi.de/), designed to interactively perform analysis, exploration and visualization of ocean data. webODV allows users to aggregate large numbers of SeaDataNet data files and perform quality control. More details about webODV are presented in "webODV – operational and ready for the community" and "Ocean Data View goes Online".

DIVAnd (Data Interpolating Variational Analysis in *n* dimensions; https://github.com/gher-ulg/DIVAnd.jl) is a cutting-edge software tool designed to efficiently interpolate in-situ observations onto a regular grid, in an arbitrary number of dimensions (for instance longitude, depth and time). A set of *Jupyter notebooks* (https://github.com/gher-ulg/Diva-Workshops/) provides a guideline to the user on how to prepare the data, optimise the analysis parameters and perform the interpolation.

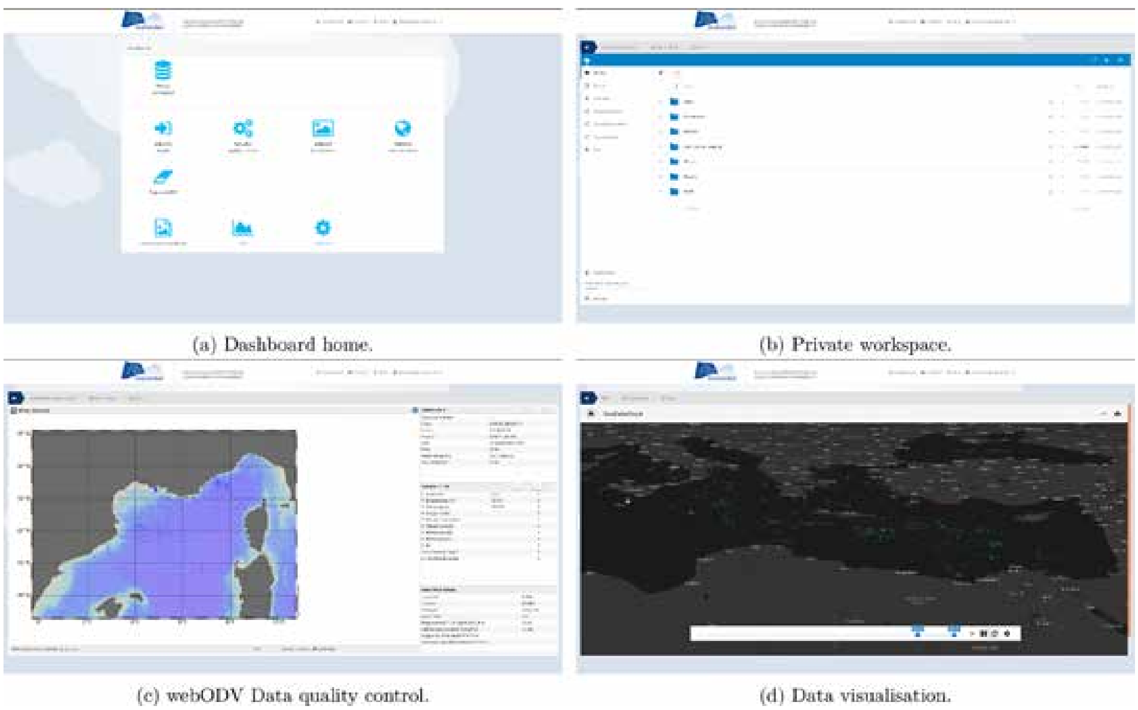
BioQC is a tool to process and to run quality control on biological datasets. BioQC helps researchers to evaluate whether a particular biological occurrence record within the input file is useful for their analysis. It also helps the data providers to identify possible gaps and errors in their datasets.

The tool returns the input file with quality information attached for each occurrence record and a detailed report. This result file will enable the users to filter for suitable records.

VIZ (https://github.com/openearth/sdc-visualization) is a modern and dynamic visualisation service to explore datasets on a map. By clicking on data points, the users see a plot of the full profile prepared with WebODV, and metadata of the input ODV files. A time selector permits to limit the data for the period of interest. Additionally, the visualisation service provides the possibility to explore 4D gridded products prepared with DIVAnd.

The **Subsetting service** strives to make data access easier, by providing services to subset, download and plot data. It returns datasets in various data formats such as CSV, MATLAB, netCDF, ODV and more. An interactive visualization tool allows users to discover and browse through the subset results with modern web technologies.

All together, these services make up the SeaDataCloud VRE, which we hope users will find useful and enjoy using.



(a) Dashboard home. (b) Private workspace.

(c) webODV Data quality control. (d) Data visualisation.

Figure 1: Different components of the VRE.

webODV – operational and ready for the community

Sebastian Mieruch, Alfred-Wegener-Institute (Germany), sebastian.mieruch@awi.de
Reiner Schlitzer, Alfred-Wegener-Institute (Germany), reiner.schlitzer@awi.de

A tool for the online analysis of marine and environmental data

webODV is based on the Ocean Data View (ODV; https://odv.awi.de) software, which is widely used for the analysis, exploration and visualization of oceanographic and other environmental data. It has been developed in the SeaDataCloud (SDC) project and has reached now a state of maturity for the operational application for marine and environmental online services, portals, data providers and many more.

The concept

webODV comprises a suite of different partly customizable compartments. On the server backend we have the fully functional ODV software additionally equipped with a WebSocket server. On the frontend we have state-of-the-art responsible, interactive browser interfaces. Via the WebSocket technology, which provides a fast bi-directional connection between server (ODV) and client, the browser interface communicates with ODV. Thus we can send WebSocket messages from the browser to ODV requesting for data processing, creating maps, plots, extracting and exporting data and much more.

Operational applications

webODV is deployed operationally until now in three different applications / projects. It is used in the SDC Virtual Research Environment (VRE; https://vre.seadatanet.org) for importing marine data, extracting data, performing quality control and visualization. For the EMODnet Chemistry project it is used as a fast, interactive and easy tool for data extraction shown exemplarily in Fig. 1. Similar extraction services are operational in the GEOTRACES project (https://webodv.awi.de), where we have up to now more than 800 registered users and daily downloads.

Services out of the box

The webODV suite provides three different services out of the box. We have the data extraction, which offers a simple, intuitive, responsive and fast browser interface for the data selection and extraction without requiring any knowledge about the underlying ODV software. Similar we provide the easy to use import service without the need of prior ODV knowledge. An additional application, which we call ODV-online is provided. Here we mimic the original ODV interface in the browser with approximately covering 99 % of the original features, including the interactive generation of publication ready maps and plots. ODV-online is perfectly suited for the large global ODV community, which is already familiar with the desktop standalone version. Detailed documentation is available at https://odv.awi.de.

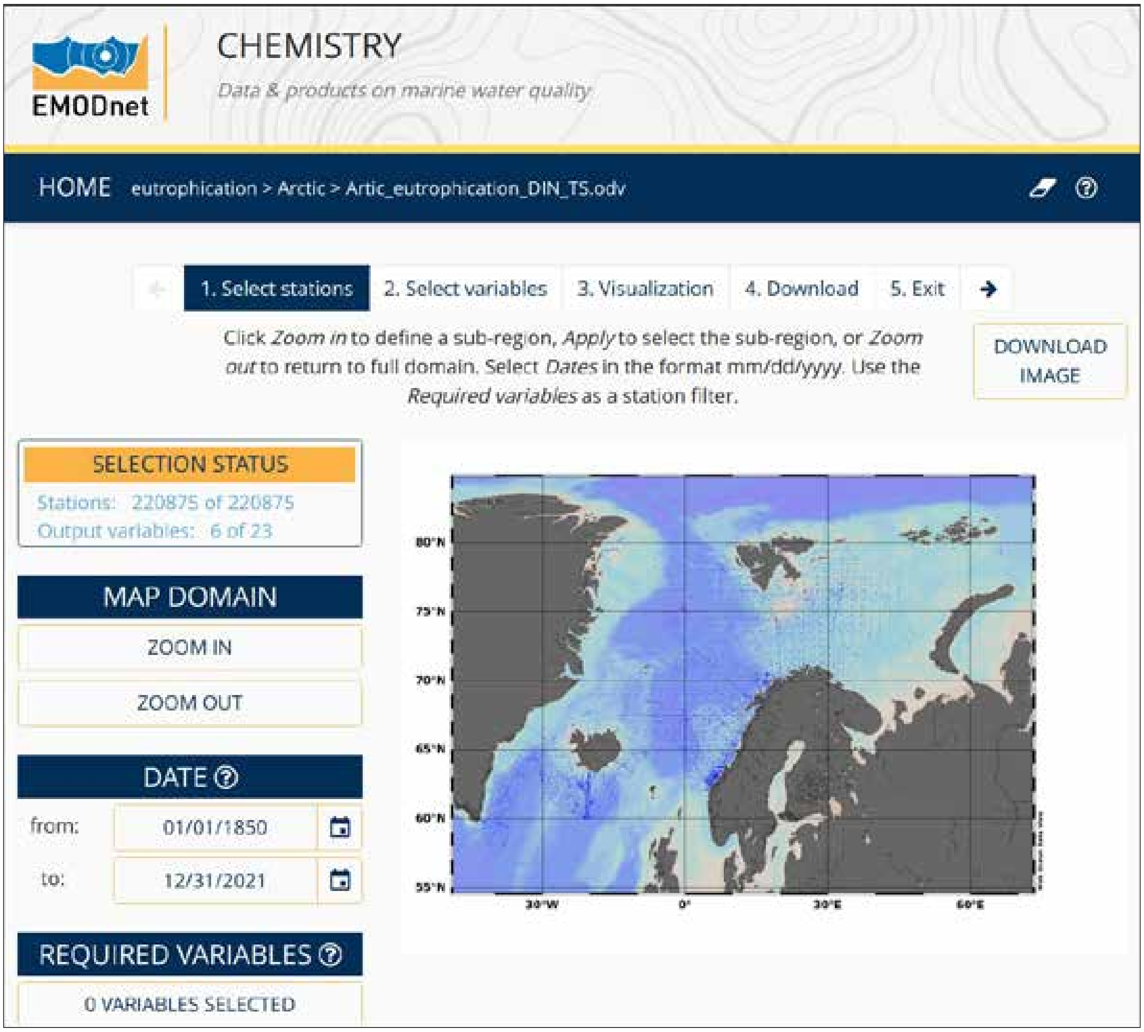


Figure 1: The EMODnet Chemistry data extractor

webODV deployment for the community

webODV is free software and we provide different packages for the deployment. Here we focus on the webODV docker operational suite, which provides the three mentioned webODV services out of the box and can be deployed easily on common server hardware. Minimal requirements are needed for the setup, which are a Linux server with installed Docker and Docker Compose software and a corresponding domain with SSL certificates. Setting up webODV then includes downloading the needed webODV packages and adapting a handful of configuration files for customizing and of course data in ODV format. The suite comes optionally with a user management system, including state-of-the-art security measures. Thus users have to register and login prior to using webODV.

While the above is the most simple way to get an instance of webODV running on a server, we will also provide a hierarchy of suites, from minimal development setups to the above mentioned Docker package.

Deep Learning for Supporting Ocean Data Quality Control

Serdar Demirel, Alfred-Wegener-Institute (Germany), serdar.demirel@awi.de
Sebastian Mieruch, Alfred-Wegener-Institute (Germany), sebastian.mieruch@awi.de
Reiner Schlitzer, Alfred-Wegener-Institute (Germany), reiner.schlitzer@awi.de

The frame and quality control

This project is part of a large European program, the SeaDataNet (2004-2016) and SeaDataCloud (2016-2020) initiatives, which have the aim to provide quality controlled ocean data via web services. Since 2004, more than 100 European data centers have been included, which provide rich data and meta data collections of several variables like temperature, salinity, oxygen, nitrate etc. for the global ocean with focus on the European seas.

In order to provide oceanographic data for research purposes, each sample of the measured variables (temperature, salinity, oxygen, nitrate etc.) is flagged by data providers and SeaDataNet ocean experts manually and/or visually. Additionally, the ocean experts have set up semi-automated workflows for the QC that consist of classical range and distribution checks (e.g. Simoncelli et al., 2018). Experts also use the Ocean Data View Software (Schlitzer, 2002) that allows them to label data with quality flags (QF). Every single ocean profile is scanned, while searching for outliers, anomalies and erroneous data. Since the size of the dataset is enormous (ca. 9 million profiles) and it is expected to grow significantly in the future, the automation/semi-automation of the QC is a necessity for the ocean science community.

Data and deep learning approach

Thus, the main motivation of our project is to assist QC experts by supporting the automation/semi-automation of the QC procedures. For this purpose, the state-of-the-art methods from the field of Artificial Intelligence (AI), specifically deep learning algorithms, are used. A binary classification problem is considered which is aiming at detecting outliers on measured temperature data in millions of oceanographic samples. For this reason, a Multilayer Perceptron (MLP) neural network, which is a class of feedforward artificial neural network (ANN), is designed which we name *Salacia* according to the roman goddess of sea water. To this end, in the framework of the EU SeaDataNet infrastructure, *Salacia* is trained with the already quality controlled "Mediterranean Sea - Temperature and salinity observation collection V2" (Simoncelli et al., 2015), where we use only data east of Gibraltar. For efficient training purposes, we sub-sampled the dataset to include only profiles with one to 100 *bad* flagged samples (and the rest *good*), to consider surely all bad flagged data. Finally, we removed the gross outliers by range filters (e.g. temperatures above 40 °C), and came up with 141,295 profiles containing 2,080,698 temperature samples. The most important aspects in machine learning are (**i**) the input features for the algorithm, i.e. the information about our data that is fed into the model, (**ii**) the separation of the dataset, and (**iii**) the architecture of the neural network.

- *Salacia* uses the most basic and informative features, which are also available for the QC experts, which are listed in the following for each sample: Depth, Temperature, Longitude, Latitude, Season (Month), Temperature gradient (change of temperature with depth), Temperature gradient from the sample above, and Temperature gradient from the sample below.
- The dataset is divided into four parts: Training data (55 %): to be used to train the network, Validation data (15 %): to tune the data to avoid under- and overfitting, Testing data (10 %): to tune the classification thresholds, Control data (20 %): to assess the skill of the model.
- A fully connected network of 3 hidden layers and 32 nodes each has been chosen.

Results and next steps

Similar to Simoncelli et al. (2018), the Mediterranean Sea has been divided into 16 Regions to evaluate the skill of Salacia on a regional scale. Figure 1 shows an example of the skill assessment for a region between Italy and Libya.

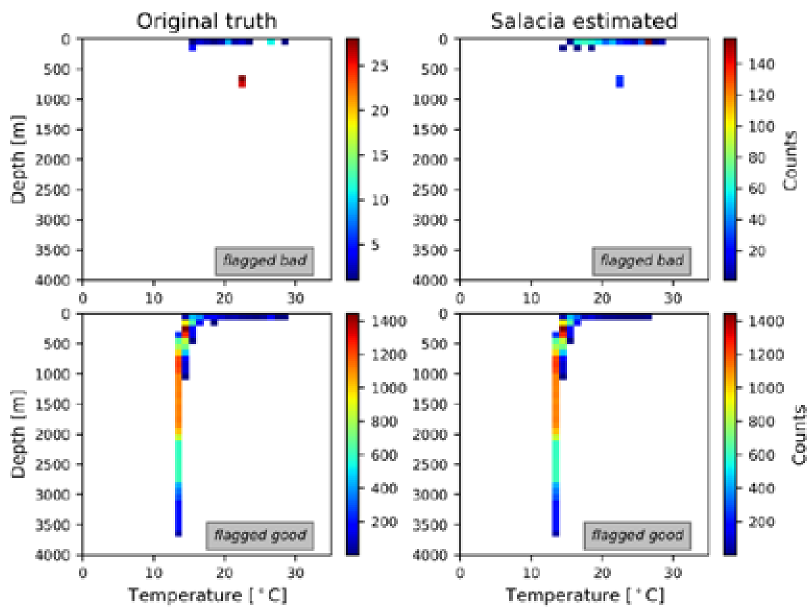


Figure 1: Temperature / Depth plots for truth (left) vs. Salacia (right). *Salacia* predicted with accuracy larger than 90 % for both *good* (green) and *bad* (red) flags.

The left side of Figure 1 shows the reference temperature measurements, flagged *bad* (top) and *good* (bottom) by the human QC experts. The right side shows the estimated flags by our algorithm *Salacia*. It is important to note that this evaluation data is "unknown" to the algorithm, i.e. it has not been used during the training process. Our evaluation reveals that *Salacia* is too sensitive in classifying data as *bad*. However, among the 646 *bad* classified samples by *Salacia*, 85 from 94 true *bad* have been found correctly (90.42 %). Regarding the *good* flagged samples, *Salacia* has found 32898 out of 33459 correctly (98.32 %). In general, we have found that the algorithm reaches high accuracies larger than 90 % in identifying *good* or *bad* data in 11 of 16 regions of the Mediterranean Sea, and for the rest of the regions, the accuracy values are oscillating between 90 and 60 % for *good* and *bad* data. Thus, Salacia could be especially useful and helpful for the

QC experts in these particular skillful regions. However, it would be recommended that the QC experts concentrate on using only the *Salacia bad* flags as a guidance and accept the *good* flags. This leads to checking only ca. 10 % of the data. Now, the crucial question is if *Salacia* can be an assistant for the QC experts by giving useful hints to potentially *bad* data on the small scale. This has to be evaluated together with the QC experts.

References

Schlitzer, R. (2002). Interactive analysis and visualization of geoscience data with ocean data view. Computers & geosciences28, 1211–1218

Simoncelli, S., Coatanoan, C., and Myroshnychenko, V. (2018). Seadatacloud temperature and salinity historical data collection for the mediterranean sea (version 1). product information document (pidoc)

Simona Simoncelli, Dick Schaap, Reiner Schlitzer (2015). Mediterranean Sea - Temperature and salinity observation collection V2. https://doi.org/10.12770/8c3bd19b-9687-429c-a232-48b10478581c

The European HF Radar node: two years distributing standardized and quality-controlled data to the major European Marine Data Portals

Lorenzo Corgnati, CNR-ISMAR (Italy), lorenzo.corgnati@sp.ismar.cnr.it
Carlo Mantovani, CNR-ISMAR (Italy), carlo.mantovani@cnr.it
Anna Rubio, AZTI Marine Research (Spain), arubio@azti.es
Emma Reyes, ICTS-SOCIB (Spain), ereyes@socib.es
Paz Rotllan, ICTS SOCIB (Spain), protllan@socib.es
Antonio Novellino, ETT SpA (Italy), antonio.novellino@ettsolutions.com
Patrick Gorringe, SMHI (Sweden), patrick.gorringe@smhi.se
Jose Luis Asensio Igoa, AZTI Marine Research (Spain), jlasensio@azti.es
Annalisa Griffa, CNR-ISMAR (Italy), annalisa.griffa@sp.ismar.cnr.it
Julien Mader, AZTI Marine Research (Spain), jmader@azti.es

High Frequency Radar (HFR) proved to be a very effective land-based remote sensing technology to monitor coastal regions all over the world due to its unprecedented capacity of mapping ocean surface currents and wave fields over wide areas with high spatial and temporal resolution. HFR main application span research, marine safety and security, pollutant monitoring, tsunami detection, fishery, navigation and renewable energy. The importance of these applications for the integrated management of the coastal zones made the HFR technology rapidly expanding in Europe (at a rate of 7 new systems per year since 2016), with over 81 HFR sites currently operating and a number in the planning stage, as shown in Figure 1.

Figure 1: Distribution of HFR systems in Europe. The operational systems are plotted in green, future installations in yellow and past deployments in red. Source: http://eurogoos.eu

In order to improve EU coordination in the management and to enhance the accessibility to HFR data for a pan-European use, the European HFR Node was established in 2018 under the coordination of the EuroGOOS HFR Task Team, as the operational asset for pushing the creation of the pan-European HFR network to a higher level. This effort benefited from the achievements of different initiatives at national and European level active since 2014 to ensure the full exploitation of the HFR technology potential in the context of the European Ocean Observing System (EOOS).

Building on the harmonization achievements (in terms of Quality Control, data format, conventions, naming and vocabularies) of the Task Team, the EU HFR Node, coordinated by AZTI, CNR-ISMAR and SOCIB acts as the focal point in EU for HFR data management and dissemination by implementing the HFR data stream (harvesting, harmonization, formatting and distribution) from the data providers towards the different EU marine data portals and global data infrastructures.

The EU HFR Node is fully operational since December 2018 in providing guidelines, free and open-source repository software and support for standardization to the HFR operators and in distributing standardized and quality-controlled HFR data towards the major European Marine Data Portals:

- near real time (within 6 hours in most cases) HFR total and radial current data to the Copernicus Marine Environment Monitoring Service In Situ Thematic Assembly Center (CMEMS-INSTAC) and EMODnet Physics, since April 2019 and March 2020, respectively;
- historical total current data with different delayed-mode reprocessing levels within CMEMS-INSTAC, EMODnet Physics and SeaDataNet data infrastructures since July 2020.

In the European framework, the EU HFR Node is now managing data from 12 HFR networks (built of 35 radar sites, representing more than 2/5 of the European Network), as shown in Figure 2, belonging to 8 countries included in 3 different ROOSes (i.e. MONGOOS, NOOS and IBIROOS), from two diverse HFR system types (i.e. Direction Finding and Phased Array), being most of them permanent installations. By end 2020, it is expected to manage 20 networks (50 radar sites).

Additionally, the EU HFR Node implements since June 2020 the integration and distribution of Global data on the aforementioned platforms, using the US network as a pilot case.

Figure 2: HFR systems distributed on CMEMS-INSTAC. Source: marineinsitu.eu/dashboard

The progress in the integration of historical and NRT HFR data achieved by the EU Node together with the HFR data providers will further boost the potential of multiplatform integration approaches for more accurate monitoring of the coastal currents, as well as the uptake of surface current coastal data by different intermediate and final users.

Novel automated quality control procedures for BGC data developed by the Copernicus Marine Service In-Situ TAC Team

Vidar S. Lien, Institute or Marine Research (Norway), vidar.lien@hi.no
Jan Even Øie Nilsen, Institute of Marine Research (Norway), jan.even.oeie.nilsen@hi.no
Håvard Vindenes, Institute of Marine Research (Norway), haavard.vindenes@hi.no
Leonidas Perivoliotis, HCMR (Greece), lperiv@hcmr.gr
Seppo Kaitala, SYKE (Finland), seppo.kaitala@ymparisto.fi
Virginie Racapé, Ifremer (France), virginie.racape@ifremer.fr

Abstract

The Copernicus Marine Service is a “one-stop-shop” providing freely available operational data on the state of the marine environment for use by marine managers, advisors, and scientists, as well as intermediate and end users in marine businesses and operations. The Copernicus Marine service offers operationally updated and state-of-the-art products that are well documented and transparent. The European Commission’s long-term commitment to the Copernicus program offers long-term visibility and stability of the Copernicus Marine products. Furthermore, Copernicus Marine offers a dedicated service desk, in addition to training sessions and workshops.

Here, we present the in situ biogeochemical (BGC) data products distributed by the Copernicus Marine Service since 2018. It offers available data of chlorophyll-*a*, oxygen, and nutrients collected across the globe. These products integrate observation aggregated from the Regional EuroGOOS consortium and Black Sea GOOS as well as from SeaDataNet2 National Data Centers (NODCs), JCOMM global systems, including bio-argo, and the Global telecommunication system (GTS) used by the Met Offices.

Through the evolution of its delayed-mode, re-processed offering of BGC data, the Copernicus Marine In-Situ Thematic Assembly Centre (INS TAC) has developed novel automated quality control procedures applied for BGC data to identify questionable data for visual inspection, thereby greatly enhancing the team’s delayed-mode quality control capabilities. Moreover, there is a potential for utilizing these procedures to improve also the near-real time quality control. The in-situ re-processed BGC product is updated two times per year and made freely available through the Copernicus Marine website. The product is delivered on NetCDF4 format compliant with the CF1.7 standard.

Introduction

The Copernicus Marine Service is the European Commission’s infrastructure for providing marine Earth Observation data and products to research, management, and businesses, among other. As a part of the Copernicus Marine Service, the INS TAC provides in-situ ocean data through two main channels: near-real time (NRT) products where observation data are provided within 24 hours post measurement, and re-processed (REP) products that consist of delayed-mode quality controlled data updated twice a year. Here, we present novel, automated quality control procedures for delayed-mode quality control of BGC data implemented for the INS TAC delayed-mode BGC REP product.

Data and products

In the production of the BGC REP product, we re-process available BGC data from 1993 and up to present date and divide the data into three main categories: chlorophyll-*a*, oxygen, and nutrients. Currently, the nutrients dataset consists of measurements of nitrate, silicate, and phosphate. The measurements include both discrete samples from bottle data, as well as sensor data. Moreover, measurements come from a wide variety of observation platforms, including CTD, bio-argo, gliders and ferryboxes, among other. Furthermore, the data originates from a large host of data providers.

Quality control procedures

Most of the data have been subject to delayed-mode quality control by the different data providers. However, the quality control is inconsistent between providers and the procedures applied may be unknown to the INS TAC partners that channel the data to Copernicus Marine. The INS TAC quality control procedures need to take these matters into account and make sure that data of high or low quality are flagged accordingly. Therefore, we apply consistent, well-documented quality control procedures for all data going into the BGC REP product, including visual inspection of data of questionable quality. To reduce the need for visual inspection and thereby increase the efficiency of the quality control process, we have developed automated quality control based on statistical testing. Moreover, where possible, we have in addition applied tests that check the data against physical constraints. Data that pass the tests are flagged as good data, while data that fail any of the tests are visually inspected before being flagged as good or bad data.

For chlorophyll-*a* a purely statistical approach has been chosen. The world ocean is divided into coastal and pelagic regions, and each region is divided into upper and deeper ocean. Then, the 99th percentile for chl-*a* concentration is computed for each region individually and used as upper boundary for accepting data, i.e., any data point outside the 99th percentile of any given region is visually checked before being flagged as either ‘1 – good’ or ‘4 – bad’. For oxygen, on the other hand, the concentrations are compared with the calculated maximum saturation of oxygen in seawater (allowing for over-saturation in the upper ocean), in addition to a regional range test based on a statistical approach. Any data points that are above 100% saturation (but allowing for over-saturation in the surface layer) are rejected and flagged as ‘4 – bad’ data. Moreover, any data points that falls outside the regional fixed ranges are visually inspected before having its final quality flag decided. For nutrients, a profile test is applied in addition to the statistical range test. The profile test identifies all profiles where the concentration in the surface layer exceed the concentration at intermediate depths (Figure 1), and all profiles that are marked are visually inspected before the final quality flag is applied. Note that the profile test is only applied in pelagic (non-coastal) regions to avoid potential impact of runoff from land, yet, as Figure 1 shows, impact from coastal waters may still occur. We find that for some variables in some regions, e.g., nutrients in the Mediterranean Sea, a substantial part of the data are flagged during the regional range test and to a lesser extent during the profile test. Moreover, generally more measurements of nitrate and silicate obtained by CTD are flagged compared to measurements from bottle data. This result was found in both the range test and the profile test.

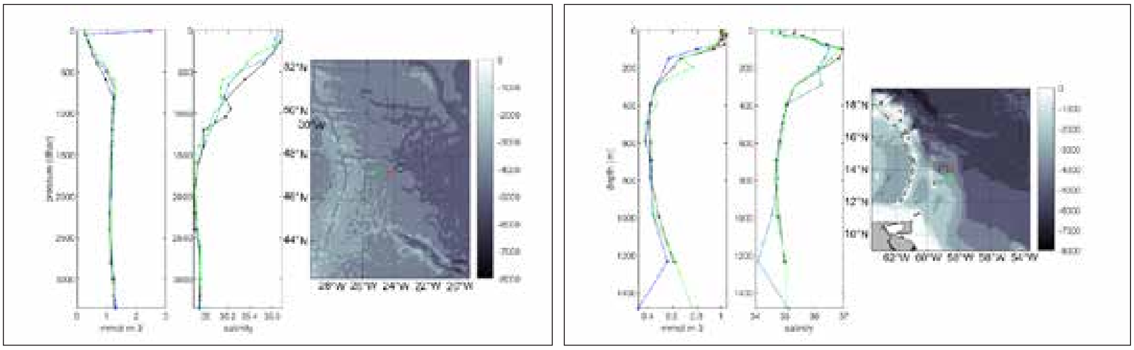


Figure 1: Examples of the profile test for nutrients (phosphate). Left: surface value marked (far left) and flagged as bad data (surface value larger than subsurface values; salinity profile to the right). Right: surface values are marked (left profile) but flagged as good data after visual inspection (influence of coastal water; salinity profile to the right).

Summary & Conclusion

We present automated quality control procedures for identifying BGC data that require visual inspection for delayed-mode quality control. The proposed automation procedures greatly reduce the required resources for delayed-mode quality control, and also provide opportunity for further improvements of near-real time quality control of BGC data.

Blue-Cloud: Developing a marine thematic EOSC cloud to explore and demonstrate the potential of cloud based open science in the domain of ocean sustainability

Sara Pittonet, Trust-IT Srl (Italy), s.pittonet@trust-itservices.com
Sara Garavelli, CSC (Finland), sara.garavelli@csc.fi
Dick M. A. Schaap, MARIS (Netherlands), dick@maris.nl
Pasquale Pagano, CNR (Italy), pasquale.pagano@isti.cnr.it
Anton Ellenbroek, FAO (Italy), anton.ellenbroek@fao.org
Kate Larkin, SSBE (Belgium), kate.larkin@seascapebelgium.be
Gilbert Maudire, Ifremer (France), Gilbert.Maudire@ifremer.fr
Alain Arnaud, Mercator Ocean (France), alain.arnaud@mercator-ocean.fr

The **European Open Science Cloud (EOSC)** is an initiative launched by the European Commission in 2016, as part of the **European Cloud Initiative**. EOSC aims to provide a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and scientific disciplines, leveraging and federating the existing data infrastructures.

Following its launch several Calls have been published and several projects have been granted for developing (parts of) the EOSC. For the marine domain a dedicated call was launched as part of 'The Future of Seas and Oceans Flagship Initiative', combining interests of developing a thematic marine EOSC cloud and serving the Blue Economy, Marine Environment and Marine Knowledge agendas.

The winning H2020 **Blue-Cloud project** is dedicated to marine data management and its aims are:

- To build and demonstrate a Pilot Blue Cloud by combining distributed marine data resources, computing platforms, and analytical services
- To develop services for supporting research to better understand & manage the many aspects of ocean sustainability
- To develop and validate a number of demonstrators of relevance for marine societal challenges
- To formulate a roadmap for expansion and sustainability of the Blue Cloud infrastructure and services.

The project will federate leading European marine data management infrastructures (SeaDataNet, EurOBIS, Euro-Argo, Argo GDAC, EMODnet, ELIXIR-ENA, EuroBioImaging, CMEMS, C3S, and ICOS- Marine), and horizontal e-infrastructures (EUDAT, DIAS, D4Science) to capitalise on what exists already and to develop and deploy the Blue Cloud.

Figure 1: Blue-Cloud federated infrastructures

The federation will be at the levels of data resources, computing resources and analytical service resources. A Blue Cloud data discovery and access service will be developed to facilitate sharing with users of multi-disciplinary datasets. A Blue Cloud Virtual Research Environment (VRE) will be established to facilitate that computing and analytical services can be shared and combined for specific applications.

This innovation potential will be explored and unlocked by developing five real-life demonstrators addressing societal challenges in the domains of genomics, fishery, aquaculture, biodiversity and environment. The demonstrators will showcase how Blue-Cloud can support ocean science research therefore contributing to the vision of the **United Nations Decade of Ocean Science for Sustainable Development**.

The modular architecture of the VRE will allow scalability and sustainability for near-future expansions, such as connecting additional infrastructures, implementing more and advanced blue analytical services, configuring more dedicated Virtual Labs, and targeting more (groups of) users. During the project, a roadmap to 2030 for expansion and sustainability of the Blue-Cloud federated infrastructures and services mobilising input and support of all major stakeholders, also beyond Europe, will be produced. This roadmap will set the basis for a global Blue-Cloud and will be a first step towards the establishment of a transparent and accessible ocean (a '**Digital Twin of the Ocean**') in support of the Green Deal priorities.

The presentation will describe the vision of the Blue-Cloud framework and the role that Blue Cloud will have in supporting the European Open Science Cloud and the Digital Twin of Ocean establishment.

The Blue-Cloud data discovery and access service (to find and retrieve data sets from a diversified array of key marine data infrastructures dealing with physics, biology, biodiversity, chemistry, and bio genomics), the Blue-Cloud VRE (to facilitate collaborative research using a variety of data sets and analytical tools, complemented by generic services such as sub-setting, pre-processing, harmonizing, publishing and visualization) and the technical architecture of Blue-Cloud will be presented via 5 real- life use-cases to demonstrate the impact that such innovation can have on science and society.

EMODnet Ingestion and safe-keeping of marine data

Sissy Iona, HCMR (Greece), sissy@hnodc.hcmr.gr
Dick M.A. Schaap, MARIS (Netherlands), dick@maris.nl

Access to marine data is a key issue for the EU Marine Strategy Framework Directive and the EU Marine Knowledge 2020 agenda and includes 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.

There is a wealth of marine data collected in Europe by public and private users, such as governmental organisations carrying out environmental monitoring, academic researchers studying the status of and fluctuations in the marine environment, private companies planning and building marine infrastructures, such as pipelines and wind farms, and even citizens science initiatives. In recent years, EMODnet has made huge progress in facilitating access to data from many sources. However, numerous data sets still remain hidden or unusable.

The '**EMODnet Ingestion and safe-keeping of marine data'** project, started mid-2016, tackles these problems by reaching out to data holders, explaining the benefits of sharing their data and offering a support service to assist them in releasing their data for subsequent processing, quality control, long term storage, and possible inclusion in EMODnet data products.

The activities are undertaken by a large European network that is geographically anchored in the countries bordering all European marine basins, and covers all EMODnet data themes. The EMODnet Data Ingestion ambassadors are representatives of national and regional marine and oceanographic data repositories and experts in marine data management. The network includes the coordinators of the EMODnet thematic projects. They work together in pan-European marine data management infrastructures such as SeaDataCloud, EurOBIS and EGDI, and in international organisations such as IODE, ICES, EuroGeoSurveys, EuroGOOS, and IHO.

Their activities are facilitated by the EMODnet Data Ingestion portal which encourages data providers to share marine data, gives marine data management guidance information, and provides a range of services such as:

- submission service for easy ingestion of marine data packages;
- view submissions service to oversee submitted data sets 'as is' and further elaborated as entries in major European marine data management infrastructures such as SeaDataNet, EurOBIS, and others which feed into EMODnet and CMEMS;
- data wanted service to post requests for specific data types.

Submission forms with data packages are assigned to qualified data centres from the EMODnet Ingestion network depending on the country of the data provider and the type of EMODnet theme.

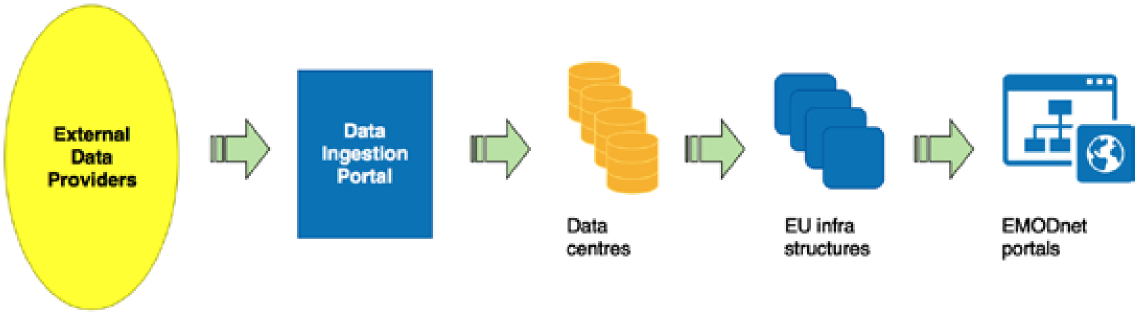


Figure 1: Flow of data sets from external data providers to EMODnet portals

After an initial period of developing the portal and services, followed by promoting and marketing to potential marine data holders, a steady inflow of submissions has been achieved. Currently, there have been more than 700 data packages submitted of which circa 600 have been completed and published ''as-is''. Of those, at present more than 250 submissions have been elaborated by assigned data centres to common formats and are made available by EMODnet thematic portals and as input for EMODnet data products.

Furthermore, a (semi-)automatic coupling was established between SEANOE and EMODnet Ingestion. SEANOE, operated by IFREMER, is a SeaDataNet service, inviting European scientists to publish their scientific papers and associated data collections in return for a DOI which will facilitate their wider citation. The coupling facilitates that (selected) scientific submissions from SEANOE are harvested by EMODnet Ingestion for further metadata completion, publishing 'as-is', and elaboration of data sets for inclusion and publishing in national, European and EMODnet portals.

EMODnet Ingestion not only concerns archived marine data sets, but also Near Real Time (NRT) and even Real Time (RT) operational oceanography data that are collected by fixed and moving platforms such as fixed stations, moorings, buoys, tide gauges, surface drifters, ferryboxes, argo floats, gliders, HF radars and other platforms. Together with EMODnet Physics, operators of operational oceanography networks and platforms are motivated and given guidance for making their datasets part of the European oceanography data exchange as managed by Copernicus CMEMS-INSTAC, EuroGOOS, and SeaDataNet, which are pillars under EMODnet Physics. In addition, EMODnet Ingestion and EMODnet Physics promote the uptake and adoption of Sensor Web Enablement (SWE) standards for operational oceanography data streams. Therefore, using the SeaDataNet SWE toolkit as developed by 52North, a SWE pilot has been set-up and is maintained. This pilot gives discovery and access to data streams from real time oceanographic monitoring systems, covering a range of operators and platforms, and allowing direct standardised access to selected data types from selected monitoring instruments.

EMSO ERIC Data Services: managing distributed data through an ERDDAP federation

Antoine Queric, Ifremer (France), antoine.queric@ifremer.fr
Rob Thomas, Marine Institute (Ireland), rob.thomas@marine.ie
Maurice Libes, CNRS (France), maurice.libes@osupytheas.fr
Enoc Martinez, Universitat Politècnica de Catalunya (Spain), enoc.martinez@upc.edu
Claudia Fratianni, INGV (Italy), claudia.fratianni@ingv.it
Tania Morales, PLOCAN (Spain), tania.morales@plocan.eu
Helen Snaith, BODC NOC (UK), h.snaith@bodc.ac.uk
Maria Sotiropoulou, HCMR (Greece), marsot@hcmr.gr
Sylvie Van Iseghem, Ifremer (France), sylvie.van.iseghem@ifremer.fr
Raluca Radulescu, GeoEcoMar (Romania), raluca.radulescu@geoecomar.ro
Paulo José Relvas de Almeida, UAlg (Portugal), prelvas@ualg.pt
Raul Bardaji, UTM-CSIC (Spain), bardaji@utm.csic.es
Ivan Rodero, EMSO ERIC CMO (Italy), ivan.rodero@emso-eu.org

Bringing together distributed data

EMSO is a consortium of partners sharing a common strategic framework of scientific facilities (data, instruments, computing and storage capacity). Formally, it is a European Research Infrastructure Consortium (ERIC), a legal framework created for pan-European large-scale research infrastructures. EMSO ERIC consists of a system of regional facilities/observatories placed at key sites around Europe, from the North East Atlantic, through the Mediterranean, to the Black Sea. The observatories are platforms equipped with multiple sensors, placed through the water column and on the seafloor. These observatories constantly measure different biogeochemical and physical parameters that address natural hazards, climate change and marine ecosystems.

Central to the EMSO ERIC mission is the collection, curation and provision of high-quality oceanographic measurements for the assessment of long-term trends. EMSO ERIC regional facilities collect a variety of data spanning oceanographic measurements through video and acoustic data types. In developing an integrated data management ecosystem, EMSO ERIC faced the common challenges where data sources that are of different volume, velocity and variety are hosted across many partners. EMSO ERIC is committed to ensuring that datasets fulfil the FAIR principles of being Findable, Accessible, Interoperable, and Reusable. The primary goal of this ecosystem is to deliver data and products from the aggregation of sources from the regional facilities in a reliable and integrated manner. It facilitates the user with a single access point to all EMSO ERIC observatories through harmonization processes. It also offers tools that enable users to easily find and access data assets, including data portals, application programming interfaces (APIs), dashboards, and a virtual research environment. This presentation outlines experiences for delivering a federated ERDDAP system, which complements other data access mechanisms of the EMSO ERIC data management ecosystem.

EMSO ERIC ERDDAP federation

ERDDAP is a data server created by NOAA in the United States that provides a simple, consistent way to serve data on the web. ERDDAP is free and open source. It uses Apache-like licenses, so it can be adapted or enhanced to fit a user's requirements. Users can download subsets of gridded and tabular scientific datasets in common file formats and make graphs and maps, which can be embedded in web pages and can be configured to update with the latest data available. In addition to a web interface, ERDDAP also provides a RESTful API that allows users to programmatically interact with the data using scripting languages such as Python, R or Matlab, and can be used to provide data to dashboards and other applications. ERDDAP is well suited to the distributed data requirements of EMSO ERIC as data do not need to be transformed from the local storage format of choice (e.g. flat files, relational database, noSQL database) to one "master" format in order to be served. This means EMSO ERIC is not forcing data architecture decisions onto each partner who may not fit with their existing architecture. ERDDAP can be set up to serve data from the storage structures already in place at each organization and it automatically aggregates files of the same XML model within a dataset. This is a useful feature for users interested in a subset of a time-series who don't want to have to stitch together a series of files themselves before working with the data. ERDDAP also facilitates interoperability and data reuse since it is able to take a variety of formats as input and output them into user preferred file formats.

Architecture and implementation

In order to provide a single ERDDAP endpoint for EMSO ERIC end-users, a number of architecture choices had to be evaluated. The relative benefits and costs for three solutions were considered: EMSO ERIC data harvested centrally and served from one single ERDDAP server; each partner serves data through their own ERDDAP server and references to other servers; a distributed/federated network of ERDDAP servers. The ERDDAP federation model was agreed to be the optimum solution, and it was decided to investigate the ERDDAP federation functionality for its implementation, i.e., ERDDAP provides the ability to reference datasets served from other ERDDAP servers. The configuration is based on a simple URL that leads to the desired dataset hosted by a remote ERDDAP server. The central EMSO ERIC ERDDAP server then provides those datasets as if they are locally hosted (Figure 1).

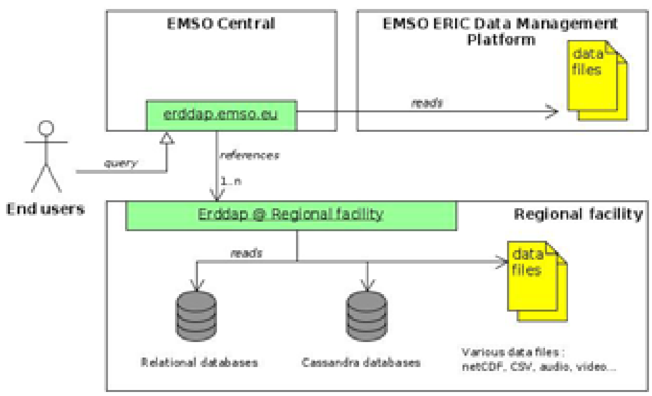


Figure 1: EMSO central ERDDAP server (http://erddap.emso.eu) references remote datasets

Whereas some EMSO ERIC partners already have experience setting up and delivering data through ERDDAP servers, not all EMSO ERIC regional facilities are currently participating in the ERDDAP federation. Undergoing efforts include the deployment of dedicated ERDDAP servers at EMSO ERIC regional facilities and the integration of datasets from the EMSO ERIC data management platform into ERDDAP (see Figure 1). It gives the user the appearance that all data are being sourced from one single location. A further goal is to provide end-users with meaningful examples for the usage of the datasets within the EMSO ERIC virtual research environment (https://jupyter.emso.eu/). Default queries are configured to show the users what they can do with the data for each dataset.

Application of community standards and formats

While ERDDAP provides a technical architecture to achieve the Findable and Accessible components of the FAIR principles, the content still needs to be well managed and marked up in order to achieve Interoperability and Reusability. Ensuring consistent metadata markup from community vocabularies and populating the metadata (or attributes) of community standards, for example, CF, OceanSITES, and SeaDataNet, is an ongoing process. Rather than bespoke code and processes being maintained locally to produce a variety of community formats, the open-source nature of ERDDAP opens possibilities for EMSO ERIC, as well as other communities across the marine domain, to contribute enhancements to future releases of ERDDAP to deliver these formats.

OceanTeacher Global Academy: IOC's Capacity Development Tool

Claudia Delgado, UNESCO/IOC/IODE (Belgium), c.delgado@unesco.org
Greg Reed, UNESCO/IOC/IODE (Belgium), g.reed@unesco.org
Peter Pissierssens, UNESCO/IOC/IODE (Belgium), p.pissierssens@unesco.org

Background and overall purpose

Capacity development is an essential tenet of IOC's mission: it enables all Member States to participate in and benefit from ocean research and services that are vital to sustainable development and human welfare on the planet. The OceanTeacher Global Academy Project (2015-2019) built on the legacy of decades of training delivered by IODE and its predecessor project, the OceanTeacher Academy. A main development of OTGA was the change in approach from a single Training Centre at the IOC Project Office for IODE in Oostende, Belgium, to a network of Regional Training Centres (RTCs) around the world in Colombia, China, India, Iran, Kenya, Malaysia, Mozambique and Senegal. This new approach enabled the use of the technical capacity already existing in the regions (teachers, experts, facilities) as well as a relevant increase in the number of courses organized and people trained in/from the regions. Additionally, it made possible to deliver courses in languages other than English, namely Spanish, Portuguese and French. The use of ICT was promoted by enabling the possibility of invited experts to contribute to courses remotely and linking classrooms following the same course (using videoconference), as well as further use of the OceanTeacher e-Learning Platform. The OTGA project successfully delivered over 220 courses and instructed over 3,300 learners from 134 Member States. The OceanTeacher e-Learning Platform has currently more than to 6,600 registered users globally.

Figure 1 a) and b): a) Trainees at different RTCs during training courses;
b) network of Regional Training Centres (2015 -2019)

OTGA Methodology and Approach

IOC's OceanTeacher Global Academy (OTGA) is a global network of Regional Training Centres delivering training on ocean sciences, services and marine and information data management (including marine biodiversity data and ocean best practices) using the OceanTeacher e-Learning

Platform. The OceanTeacher e-Learning Platform enables the sharing of standardised, quality training contents in a coordinated framework, whilst allowing the use of different languages as well as local/regional case studies. It is a comprehensive web-based training platform that supports classroom training (face-to- face), blended training (combining classroom and distance learning), and online (distance) learning.

Courses cover a range of topics related to the IOC programs, contributing to the IOC Mandate and the implementation of the IOC Capacity Development Strategy, enabling equitable participation of all IOC Member States and IOC Programs. Furthermore, OTGA supports the organisation and hosts training courses for other marine sciences related international organisations and projects.

OceanTeacher Quality Management

In April 2018, the IOC Project Office for IODE, host of the OceanTeacher Global Academy, achieved ISO 29990 certification as a Learning services provider for non-formal education and training, and was accredited by the Belgian Accreditation Body (BELAC) having satisfied the requirements of the International Standard. This certification is a recognition of the quality of learning opportunities offered by OTGA, through the IOC Project Office for IODE, and the high standard of quality learning services delivered that can support all IOC programmes in providing specialized training.

OTGA: the way ahead

A new phase of OTGA was approved for funding late 2019. The new project, due to start mid 2020, will build on the legacy of OTGA and include new initiatives and challenges now in place, for example, the

2030 Agenda and its SDGs and the UN Decade of Ocean Science for Sustainable Development. New, ready to deliver course topics will be made available online. The network of Regional Training Centres will be joined by Specialized Training Centres that can deliver very specific training topics. Training topics will also include tools that can help Member States achieving the SDGs as well as emerging topics such as Ocean Acidification and Blue Carbon, *inter alia*. OTGA-2 will build partnerships with other national and international training organizations with the aim of (i) expanding global awareness of learning opportunities, (ii) increasing learning content, (iii) developing new cross-cutting learning services, and (iv) developing and implementing quality standards for delivery of learning services. OTGA-2 will foster collaborations beyond UNESCO/IOC to position itself as the training platform for ocean-related topics within the UN and beyond.

A collaborative approach to improving access to UK marine data

Charlotte Miskin-Hymas, Marine Environmental Data and Information Network, British Oceanographic Data Centre (UK), charmisk@bodc.ac.uk
Jake Kuyer, eftec economics for the environment, London (UK), eftec@eftec.co.uk
Rob Daniel, eftec economics for the environment, London (UK), eftec@eftec.co.uk
Nicola Dewey, ABPmer, Southampton (UK), enquiries@abpmer.co.uk
Sean Gaffney, Marine Environmental Data and Information Network, British Oceanographic Data Centre (UK), cfpo@bodc.ac.uk
Robin McCandliss, Marine Environmental Data and Information Network, British Oceanographic Data Centre (UK), cfpo@bodc.ac.uk
Clare Postlethwaite, Marine Environmental Data and Information Network, British Oceanographic Data Centre (UK), cfpo@bodc.ac.uk

The Marine Environmental Data and Information Network (MEDIN) was formed in 2008 and aims to be the hub for UK marine data. It promotes sharing, re-use and improved access to that data. MEDIN is an open partnership of over 50 organisations, with representatives from government departments and agencies, research institutions and private companies. Data are accessible via an online portal, which contains information about 15,000 marine datasets from over 600 UK organisations. MEDIN are sponsored by a range of UK organisations and provides secure long-term management of marine data sets at specialist data centres.

Data are managed and delivered by a network of specialist data archive centres. These include the UK Hydrographic Office; the Met Office; the British Oceanographic Data Centre; the British Geological Survey; DASSH; Marine Scotland; CEFAS; the Archaeology Data Service; Historic Environment Scotland; and the Royal Commission on the Ancient and Historical Monuments of Wales. Over the last 12 years, MEDIN has significantly improved access to UK marine data. Free data management workshops are held across the UK to increase marine metadata knowledge, facilitating widespread adoption of an agreed set of common standards for their marine data, to enable easy discovery and re-use of data. The network provides expertise on best practice data management for marine data.

The benefits provided by marine data infrastructures, such as MEDIN, are often listed but, to date, have been relatively difficult to attribute economic value to. In order to address this, in 2019, a Cost Benefit Analysis (CBA) of MEDIN's services was conducted by experienced environmental economists, eftec, supported by marine consultants, ABPmer. The aim of the cost benefit analysis was to obtain robust quantitative and qualitative data around the benefits that MEDIN provides to its users through its portal and other services, comparing these against the financial and other costs associated with MEDIN. An online survey was circulated throughout the UK marine community during August-September 2019 focusing on four main benefits (the first three of which were quantified in monetary terms in the analysis):

1) Reduced time searching for existing marine data due to the MEDIN portal

2) Avoiding duplication of marine data gathering and savings from primary research efforts due to data uploaded by others to MEDIN along with standardized metadata.
3) Time savings for organisations in managing their own data and external data they hold due to the MEDIN porta, Data Archive Centres, data guidelines and workshops.
4) Improved decision-making due to greater availability of marine data.

There are two main costs of MEDIN, which these four benefits were compared to;

1) User time costs involved with the upload of data and metadata relating to time spent learning the MEDIN metadata standard and applying this knowledge to their own data.
2) The financial costs of running MEDIN e.g. employment of the MEDIN core team and other associated overheads.

The results of this cost benefit analysis show that the benefits of MEDIN far outweigh the costs. With a benefit to cost ratio (BCR) of approximately 8, this finding highlights the high value and cost effectiveness of MEDIN. The three monetized benefits sum to £59.7m over the 10-year appraisal period which far outweighs the total costs of £7.3m, giving a net present value of £52.4m. The most significant quantified benefit is the savings from improvements in organisations' own data management resulting from MEDIN guidelines and training. Both savings from searching for existing marine data and from primary data gathering are also substantial, and overall the contribution from all three benefits is significant. The results of this study are in line with similar ones undertaken internationally in recent years. A review of the economic effects of marine spatial data infrastructures (MSDIs) by Griffin et al., (2017) found that investing in MSDIs delivers benefit to cost ratios of between 2:1 and 18:1, with a mean of 7:1. While the costs and benefits assessed in these studies do not overlap perfectly with those detailed in the MEDIN CBA, they support the results found in this study and help support the conclusion that MEDIN and similar marine data sharing infrastructure, can and do provide far greater benefits to their users than the associated running costs.

References

Edward Griffin, Andy Coote and Joep Crompvoets (2017): A marine spatial data infrastructure in New Zealand: a systematic review on the cost-benefits, Journal of Spatial Science, DOI: 10.1080/14498596.2017.1372227

IMARDIS - A Marine Data Infrastructure Serving the Needs of the Welsh Marine Sector

David Mills, Bangor University (Wales), d.mills@bangor.ac.uk
Thomas Prebble, Bangor University (Wales), t.prebbles@bangor.ac.uk
Guy Walker-Springett, Bangor University (Wales), g.r.w.springett@bangor.ac.uk
Cathy Blakey, Bangor University (Wales), cathy.blakey@bangor.ac.uk
Gwyn Roberts, Bangor University (Wales), gwyn.roberts@bangor.ac.uk
Sudha Susarita, Bangor University (Wales), s.susarita@bangor.ac.uk
Martin Austin, Bangor University (Wales), m.austin@bangor.ac.uk

The step change in our ability to measure and simulate the marine environment can lead to improved decision making based on an improved evidence base underpinned with data. Building the data infrastructure to capitalise on the next generation of data intensive science is beyond the capacity of many commercial or public funded bodies and limits its potential impact across multiple societal benefit areas. Against this background, IMARDIS (Integrated Marine Data and Information System) has been developed as part of the £17M SEACAMS2 (www.seacams.ac.uk) project at Bangor University (BU) part funded by ERDF (European Regional Development Funds) through the Welsh government. With an initial focus on the Welsh Ocean Renewable Energy (ORE) sector IMARDIS is already in demand from other government and commercial users to deliver services that support improved decision making and further improve understanding of marine systems.

SEACAMS2 collaborative research between BU and the ORE sector encompasses multi-scale and multidisciplinary (physical, chemical, biological and ecological) observations from intertidal, shallow coastal and offshore domains. The datasets gathered are large, complex and require rapid processing to enable timely delivery of quality assured data and information required by stakeholders. Furthermore, SEACAMS has a legal and contractual requirement to share publicly funded data assets.

A review of existing data services within the UK showed they could not meet the demands of our users in terms of data access speed, access to raw and processed data, very high spatial resolution data requirements or that would retrieve data on a machine-to-machine basis for further processing. IMARDIS architecture is based on a series of services, each capable of operating semi-independently and accessible through a RESTful JSON based Application Programming Interface (API) layer. The services support a range of key functions including security and authentication, metadata management, metadata catalogue, file upload and download, tabular data manipulation, point cloud (or generic raster) data processing and time-series data manipulation. The services are implemented in Java and Python and deployed within the Amazon Web Services cloud infrastructure. This is scalable in terms of storage capacity and throughput. The publicly available API allows a range of services to be delivered as required by end-users. The data discovery and download service has been implemented and is accessible via the recently launched IMARDIS portal (portal.imardis.org). A service to ingest and publish real-time data to the web has also been implemented as part of the SEACAMS2 coastal observatory initiative.

Through these services, IMARDIS is beginning to address the requirements of three main societal benefit areas: decarbonising the UK economy, environmental policy implementation and maritime safety.

The UK Crown Estate has designated two demonstration zones for the development of wave energy (Pembrokeshire, south Wales) and tidal stream energy industries (Anglesey, north Wales). IMARDIS was specifically developed to streamline industry access to SEACAMS2 quality assured data derived from the extensive observational and modelling programmes. Data collection was driven by the ORE sectors need to address a range of critical requirements: policy compliance, fulfilling licencing requirements, Environmental Impact Assessments, resource assessment and site selection. For example, high-resolution seabed bathymetry is a key requirement for site selection for the ORE industry and has been prioritised in the initial phase of IMARDIS. A screenshot (Fig. 1) of the new portal shows the results of a typical search returning results that include seabed bathymetry.

As part of Welsh governments coastal monitoring strategy real-time and delayed mode observations are carried out at key coastal locations of meteorological variables and sediment transport processes. In response to a request from National Resources Wales (NRW) we have used IMARDIS to ingest NRW data and publish to the web in real-time. Data are used to support policy compliance, flood management and managed realignment and to improve understanding of the impact of extreme events (e.g. storms) and longer-term changes (e.g. wind direction) on coastal morphology in dynamic systems such as dune evolution. Real-time meteorological data published to the web also informs decisions made by beach goers and water sports participants and addresses the requirement of government to deliver 'public good' services.

Cemex UK Ops Ltd operate a pier on the North Wales coast that has been instrumented by BU with meteorological and sub-surface sensors (temperature, pressure, salinity, suspended load, chlorophyll). IMARDIS ingests and publishes real-time data on tidal height, wind speed and direction at the pier. This enables the operators of a bulk cargo carrier to make better informed decisions regarding berthing on the pier leading to increased safety and operational efficiencies. Generic forecasts for this specific site had previously proven unreliable due to the proximity of the pier to steep cliffs resulting in inaccurate predictions of sea state. While the maritime operations benefitted primarily from specific real-time meteorological and water depth observations the water quality parameters informed large scale studies of marine environmental status and coastal dynamics.

IMARDIS is now delivering on its promise to extract value from data and address the needs of business, government and communities. A measure of the success of IMARDIS is that it is a data infrastructure that data users already want to use. Next steps include ingestion of SEACAMS2 legacy data and the development of new analytical capabilities to serve the needs of the low carbon energy sector as part of the new Smart Efficient Energy Centre (seec.bangor.ac.uk), funded by the EU through the Welsh European Funding Office.

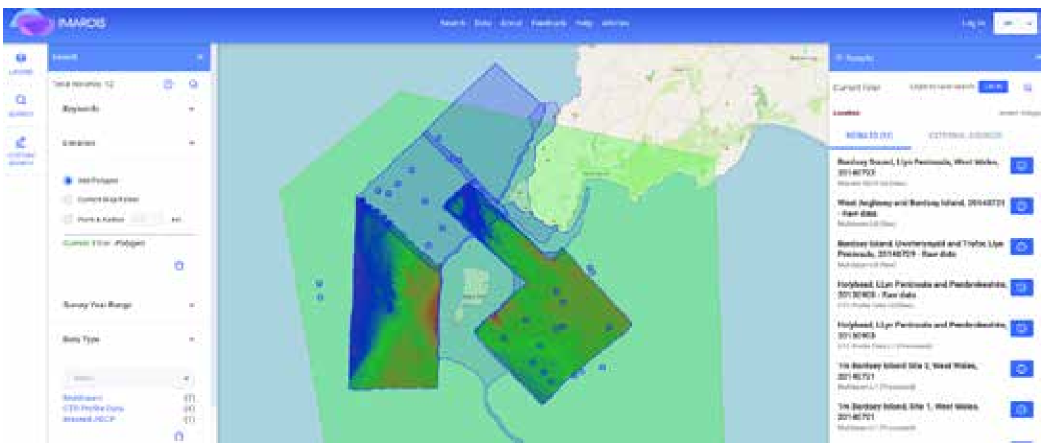


Figure 1: IMARDIS portal showing an example of the results of a search for available SEACAMS2 data in south Wales

Applying the RDF DataCube model to power data visualization and exploration dashboards for the Irish Wave and Weather Buoy Networks

Rob Thomas, Marine Institute (Ireland), rob.thomas@marine.ie
Donnchadh Ó Foghlú, Marine Institute (Ireland), donnchadh.ofoghlu@marine.ie
Kieran Lyons, Marine Institute (Ireland), kieran.lyons@marine.ie
Will Meaney, Marine Institute (Ireland), will.meaney@marine.ie

Motivation

Public administration frequently publishes Open Data in an ad-hoc manner. Yet society expects data- driven public services, not raw data. In order to ensure that public services address their needs society should be involved the design and production of such services. Within the software development world, these continuous interactions are termed "co-design" and "co-production".

There are challenges for the producers and consumers of open data, for example are the data small or large volume, how frequently are they updated, who is permitted access, how easy are they to access, etc. Within this context, data literacy (in terms of knowledge, experience and ability) of users is often overlooked when providing solutions.

OpenGovIntelligence (OGI) project

The Horizon 2020 funded OpenGovIntelligence project (OGI) planned to support and develop approaches to co-design and service co-production by providing software tools to aid decision making and better manage the complexities and precise nature of high quality statistical data. OGI sought to prove the effectiveness of the OGI approach through a diverse selection of pilot projects in six countries. The aim of the pilots was to develop services at both national and local levels to tackle challenges within society and Public Administration. The central model for data provision around which the OGI project planned to build tool was the RDF Data Cube model[1].

The RDF Data Cube model provides a vocabulary for publishing multi-dimensional data, such as statistics, on the web using the W3C RDF (Resource Description Framework) standard. The Data Cube model is compatible with the cube model on which the SDMX (Statistical Data and Metadata eXchange), an ISO standard for exchanging and sharing statistical data and metadata among organizations such as government statistics offices, is founded.

OGI Irish Pilot data dashboard - Irish Wave and Weather Buoy Networks

The Irish Wave and Weather Buoy Networks return metocean data at 5-60 minute intervals from 9 locations in the seas around Ireland. Outside of the Earth Sciences an example use case for these data is in supporting Blue Economy development and growth (e.g. renewable energy device

1 https://www.w3.org/TR/vocab-data-cube/

development). Access to wind and wave energy data are valuable information in supporting decision making by the marine renewables community, specifically in selecting suitable locations in coastal and offshore zones for deployment as well as informing the design or selection of wave energy converters for use in these areas. While the full resolution and range of data collected are valuable, summary statistics (e.g. by season) are key to designing devices optimized to exploit the prevailing conditions for a location or to determine extreme event frequency to ensure structures are designed for such conditions.

The Marine Institute, as the operator of the buoy platforms, in partnership with the OGI project has published daily summary data from the Irish Wave and Weather Buoys using the RDF DataCube model.

These daily statistics are available as Linked Data via a SPARQL endpoint API making these data semantically interoperable and machine readable. This API underpins a pilot dashboard for data exploration and visualization (Fig. 1). The dashboard[2] presents the user with the ability to explore the data and derive plots for the historic summary data, while interactively subsetting from the full resolution data behind the statistics.

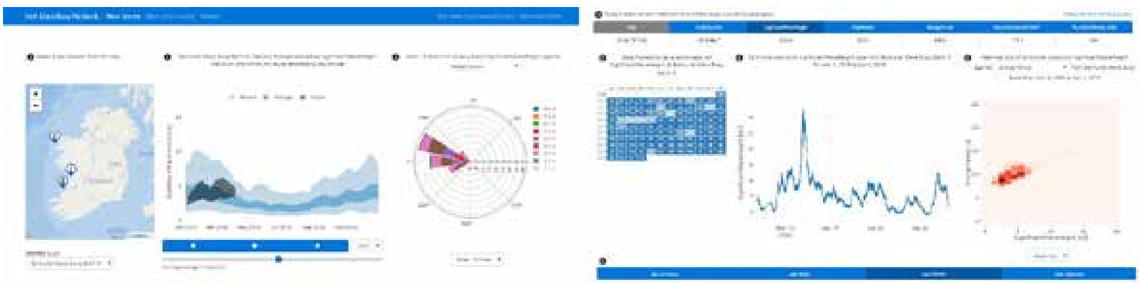


Figure 1: Irish Data Buoy Network data visualisation and exploration dashboard.

The pilot dashboard draws together summary statistics data from the SPARQL endpoint and complements these data with real-time data (using the Internet of Things MQTT protocol) and full resolution historic data from an instance of the NOAA ERDDAP data broker with a GraphQL API. GraphQL was developed by Facebook before being publicly released in 2015. Key features being the efficiency in returning results in the same structure as specified by the query and a query flexibility not available from REST services. Publishing environmental data with these technologies makes accessing environmental data available to developers outside those with Earth Science involvement and effectively lowers the entry bar for usage by providing data in self describing schemas for the RDF DataCubes and GraphQL API.

This presentation will highlight the benefits achieved by joining the project as a “pilot partner” and lesson learnt from the process. As well as discussing the technologies deployed and user feedback on from an end-user survey of the dashboard. The presentation will discuss how the dashboard is one of a range of access points that aim to cover the range of users’ data literacy and interest while minimising the challenges faced by those wishing to consume data from the Irish Weather and Wave Buoy Network.

2 https://vis.marine.ie/dashboards/

The French research vessels management: an opportunity for harmonized data

Cécile Pertuisot, Ifremer/IRSI/SISMER (France), cecile.pertuisot@ifremer.fr
Frederic Merceur, Ifremer/IRSI/ISI (France), frederic.merceur@ifremer.fr
Nolwenn Danioux, Ifremer/IRSI/SISMER (France), nolwenn.danioux@ifremer.fr
Goulwen Peltier, Ifremer/DFO (France), goulwen.peltier@ifremer.fr
Loic Petit De La Villéon, Ifremer/IRSI/SISMER (France), loic.petit.de.la.villeon@ifremer.fr

The French research vessel fleet has been selected as one of the Very Large Research Infrastructures (TGIR). Until end of 2017, the French research vessels were operated by each owner of the ships (the French research institutes Ifremer, IPEV, IRD and CNRS). Since the 1st January of 2020, 11 research vessels, mobile equipments, 4 underwater vehicles and 1 submarine are effectively gathered within the French Oceanographic Fleet (FOF). The French Ministry of Research has entrusted Ifremer to manage the whole research fleet including open ocean and coastal vessels, station ships, underwater gears and mobile equipment such as seismic acquisition systems and deep sea coring.

The TGIR Fleet is a tool that covers all fields of oceanography (marine geosciences, physical and biological oceanography, bio-geochemistry and chemistry of the oceans, paleoclimatology, etc).

The actors involved in the French research fleet activities are the chief scientists, the persons in charge of the scheduling and processing of the cruises, the ship operator Genavir in charge of achieving the cruises and SISMER which - as the French NODC - is in charge of the final repository of the data and of the data distribution.

The landscape would be incomplete without taking into account the European marine data infrastructure SeaDataNet and the currently attached project SeaDataCloud. SeaDatanet is in charge of managing common vocabularies describing, among others, vessels, geographical locations, parameters and instruments.

Considering the number of stakeholders, the challenge is to ensure the transmission of the information reported by each stakeholder related to each stage in the life of a cruise. But it is also a question of providing the entire scientific community with all the information related to the oceanographic cruises and the data from the Fleet through a single national access portal.

A new global cruise management tool.

Designing a global management tool required, first, bringing together the fleet managers and the SISMER data scientists in order to:

- Define, describe and orchestrate each stage in the life of a cruise.
- Draw up an exhaustive list of all available gear that can be operated on each vessel and map them with standardised vocabulary lists.
- Define the role of each stakeholder in order to enable each of them to act at his or her level.

Under the coordination of DFO, Ifremer/IRSI developed the Campaign Management System

(SGC - Système de Gestion des Campagnes) in 2018. The SGC is a workflow, managed by a software, which allows the overall management of an oceanographic cruise from calls for tenders to the cruise valorization. The various stages of the SGC are as follows: scientific campaign application following calls for tenders, evaluation of campaigns by external experts and national commitees, scheduling of selected campaigns, instructions, process and preparation of cruises, production of reports and records at the end of the cruises, and finally valorization of each scientific campain for its post- evaluation by national commitees.

The SGC is based on the common vocabularies defined in SeaDataNet. In particular, it produces the Cruise Summary Reports (CSR), as part of the cruise report, which feed the descriptions of each of the cruises listed in the SISMER database and are then shared at European level.

IBTS 2020 (January, February 2020) is the first cruise that has gone through all the steps of the SGC workflow, from the tender to the cruise report. The valorization will follow soon.

The cruise catalogue: a real hub for the access to all the information related to the fleet.

SISMER has been responsible for a long time for archiving (guarantee of no loss of data) and banking (guarantee of description and access to data that have passed quality control) data acquired on board Ifremer research vessels.

In order to extend its practices to other vessels, Ifremer has focused on installing Ifremer standard acquisition units on each vessel of the fleet (except 2 for geographical reasons) to facilitate the automation of its database indexing procedures. Thanks to an efficient partnership with its ship operator (Genavir), all the data acquired on board each platform are now transmitted to SISMER within 2 months after the end of the cruise.

The data are converted into standard formats and undergo visual and/or automatic quality checks to meet the needs of cataloguing and dissemination. Like the SGC, the conversion and description tools are based on the common SeaDataNet vocabulary repositories.

Today, the SISMER database contains more than 9300 cruises, the oldest of which dates back to 1913. It successfully feeds the new single access portal developed in 2018: the campaign catalogue (https://campagnes.flotteoceanographique.fr/). This cataloguing and data access tool is a real hub of access to all information related to the oceanographic cruises since it contains the CSR, the Digital Object Identifier (DOI), the associated publications and reports, and the description of the acquired data, the sampling operations, the dives, the moorings and the videos.

The unification of the French research vessels within the FOF has been an opportunity to unify the management of French oceanographic cruises and their associated data. The use of common vocabulary repositories established at the European level is an essential step to contribute to the application of the FAIR principles -Findability, Accessibility, Interoperability and Reusability- to the French data.

The presentation will describe in details the whole information and data flow system (SGC Système de gestion des campagnes / Cruise Management System) that has been developed to ensure a comprehensive tracking of cruise information and data circulation. It will also describe how it is included in a wider perimeter for sharing data and cruise summmary reports (CSR) through a single national access portal.

SatBałtyk System- modern tool for monitoring and research of the Baltic Sea

Mirosława Ostrowska, Institute of Oceanology Polish Academy of Sciences, (Poland) ostra@iopan.pl
Joanna Stoń-Egiert, Institute of Oceanology Polish Academy of Sciences, (Poland) aston@iopan.pl
Agnieszka Zdun, Institute of Oceanology Polish Academy of Sciences, (Poland) zdun@iopan.pl

The comprehensive, precise and high quality information on the state of environment should underpin not only scientific analyses, but all decisions regarding the economy, transport, recreation and signing of the international conventions between the countries bordering the Baltic Sea. The SatBałtyk System presents on website maps of spatial distribution of values of almost one hundred features of the Baltic Sea environment including the coastal zone as well as the state and optical properties of the atmosphere. This provides up-to-date, reliable data that foster accurate diagnosis of the state of the environment and allows observation of changes occurring in the Baltic ecosystem. This system was developed and deployed in 2015 by the Consortium associating four scientific institutions: the Institute of Oceanology PAN in Sopot –the coordinator, the University of Gdańsk, the Pomeranian Academy in Słupsk and the University of Szczecin.

SatBałtyk System description.

The SatBałtyk System is a satellite-based operational platform for monitoring of the Baltic Sea. There are no doubts that satellite observations that can cover large sea area in a single instant could be a perfect tool to study the marine environment. Nevertheless efficient and systematic day-to-day monitoring of the entire Baltic will provide reliable information only if we use more sources of data. For this reason the SatBałtyk System combines three types of data: satellite data used for day-to-day monitoring of large sea areas, models data which utilises hydrodynamic and ecohydrodynamic models describing phenomena taking place in the marine environment and point data acquired using traditional oceanographic measurement techniques.

To ensure the reliability of provided parameters a very complex information flow system was organized (Fig1). Important role is played by in situ measurements from the systems conducting continuous monitoring (buoys, shore stations, or other measurements platforms with research instruments) and research vessels. Traditional methods of marine ecosystem study are by their nature limited in time and space, but they accurately describe the local situation. This attribute makes them very valuable for developing satellite algorithms for the current determination of physical, chemical and biological characteristics, or validate and verify the accuracy of characteristics provided by numerical models coupled in SatBałtyk System.

Figure 1: Scheme of the SatBałtyk System infrastructure

Using models and algorithms allow estimation of marine environment parameters not only on the surface layer but also at different depths in water column and taking into account their spectral characteristics. They were also used to transgress the limitations of remote sensing methods, existing for example due to the cloud cover over the investigated area, or when data from visible and infrared domain are unavailable. An advanced and innovative methodology of merging the satellite and modelled data allows to deliver a live assessment of the current state of the Baltic Sea environment, even when satellite data acquisition could not be made.

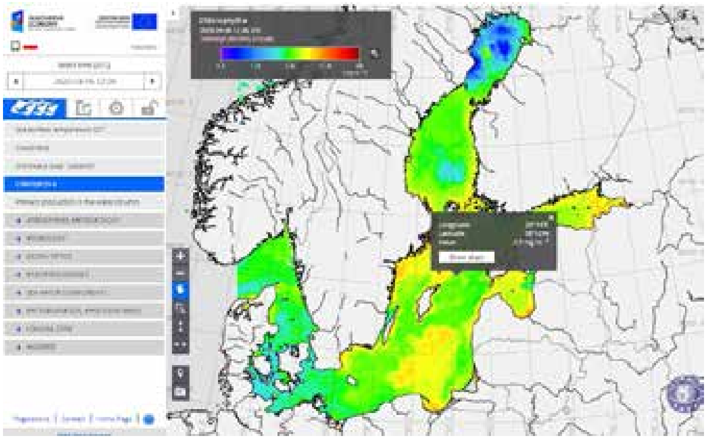


Figure 2: SatBałtyk System website http://satbaltyk.iopan.gda.pl/

The parameters currently available in the System have been divided into eight task-oriented groups (as presented in Fig. 2). Apart from the typical oceanographic characteristics of the marine environment, like sea surface temperature, the system determines specialised parameters describing complex natural processes, for example the quantum yield of photosynthesis. Many of these characteristics are unique, and their values are not determined operationally by any other comparable system.

The System's functionalities allow browsing the value distribution maps of all provided parameter in near real time throughout the entire Baltic Sea area and reading these values for each pixel with a 1 km side. Authorised users can download the numerical values of the parameters in various formats. It is also possible to analyse the variability in time of each of the parameters available in the SatBałtyk System. It enables not only to track long-term changes in the marine environment, but also the comprehensive analysis of the processes running in the Baltic ecosystem. The SatBałtyk System offers not only current NRT information but also historical data (since 2010) are available together with a short forecast of modelled parameters.

Summary.

The Satbałtyk System has became the state-of-the-art research tool, that meets the requirements of present-day oceanography, especially in tracking changes of the marine environment, resulting either from progressive eutrophication or as the effects of climate change. Such set of data describing the Baltic ecosystem allows observing changes in many of its characteristics to much greater extent, than it was possible previously on the basis of data collected during many research cruises. It contributes significantly to the development of knowledge and enables accurate diagnosis of the condition and observations of the change occurring in the Baltic environment. Data provided by the SatBałtyk System will be available soon also on the eCUDO.pl platform.

HarmoNIA project: web application for data visualisation

Damir Ivankovi, Institute of Oceanography and Fisheries (Croatia), ivankovic@izor.hr
Dalibor Jelavi, Instititute of Oceanography and Fisheries (Croatia), jelavic@izor.hr
Maria Eugenia Molina Jack, OGS Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), mmolinajack@inogs.it
Marina Lipizer, OGS Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (Italy), mlipizer@inogs.it

HarmoNIA project

Countries sharing a marine region or sub-region should adopt a common approach to environmental monitoring, good environmental status definition and assessment. In the Adriatic – Ionian marine sub- regions, the level of coherence and consistency of several EU and regional environmental policies, particularly in the assessment of contamination from hazardous substances, needs a common implementation. In the framework of enhancing the capacity to tackle environmental vulnerability and safeguard ecosystem services at transnational scale, the objective of HarmoNIA is twofold:

- to share best practices to support the harmonized implementation of marine environmental directives in the ADRION region
- to strengthen the network of data infrastructures to facilitate access and re-use of marine data among countries bordering the Adriatic – Ionian Seas.

In this framework, the Interreg project HarmoNIA has focused on improving available information related to pollution by hazardous substances in the ADRION area and on implementing tools to support pollution assessment and response.

Introduction

Data-sets about hazardous substances in sediment, biota and water column were prepared using the EU initiative EMODnet for the management and supply of fragmented marine data, and in the framework of HarmoNIA project. Data was extracted from EMODNet, validated and transform into csv tables. After that data was loaded into database and additional validated for duplicates and invalid values (metadata with no data, only bottom depth present). Data-sets cover Adriatic – Ionian Seas and the time frame is 1980-2017. These data derive from 10 different institutions. Data were collected in

2152 stations, sampled over 4282 times producing a final number of 95231 data values which are referred to 510 different parameters divided into 22 groups. All data are quality flagged according to a shared approach and quality flags can be used to filter data to visualise. Data-sets contain some data with access restrictions (by negotiation or academic - 6010 out of 101953). Those data are not shown as single values but are used for statistics calculations.

A dedicated web application (https://vrtlac.izor.hr/ords/harmonia/) developed in the framework of the project HarmoNIA shows station locations and graphical representations of data. Users can filter data by: year, project, institution, cruise, parameter group and specific parameter. Data filter is adoptive, that means that changing each category, values in all other categories are re-calculated with values according to the new criteria. For example if user choose particular years, all other categories are restricted according to data available in those years (cruises from that year, parameters...). This approach helps users to quickly gain insight into which data are available.

After setting filters and defining data subset, graphical data visualisation helps to understand the spatial/temporal distribution of contaminant concentration. In addition, basic statistics for selected subset of data are available: number of stations and values, minimun and maximum values and stations where these values are present, average, standard deviation and variance.

Methodology

Relational database used for web application is Oracle 19.3 standard edition 2 database. As application server Tomcat 9 is used. Oracle locator is used for storage of spatial data. Various JavaScript frameworks are used for advanced browser side data manipulations and visualisations (JQuery, BackGrid, Highcharts and Google Maps API v3). JavaScript is used to pair graph elements and station markers (Figure 1). Column graph is used when visualised subset contain only one value per station, and lines with negative y axis (depth), when there are values from different depths.

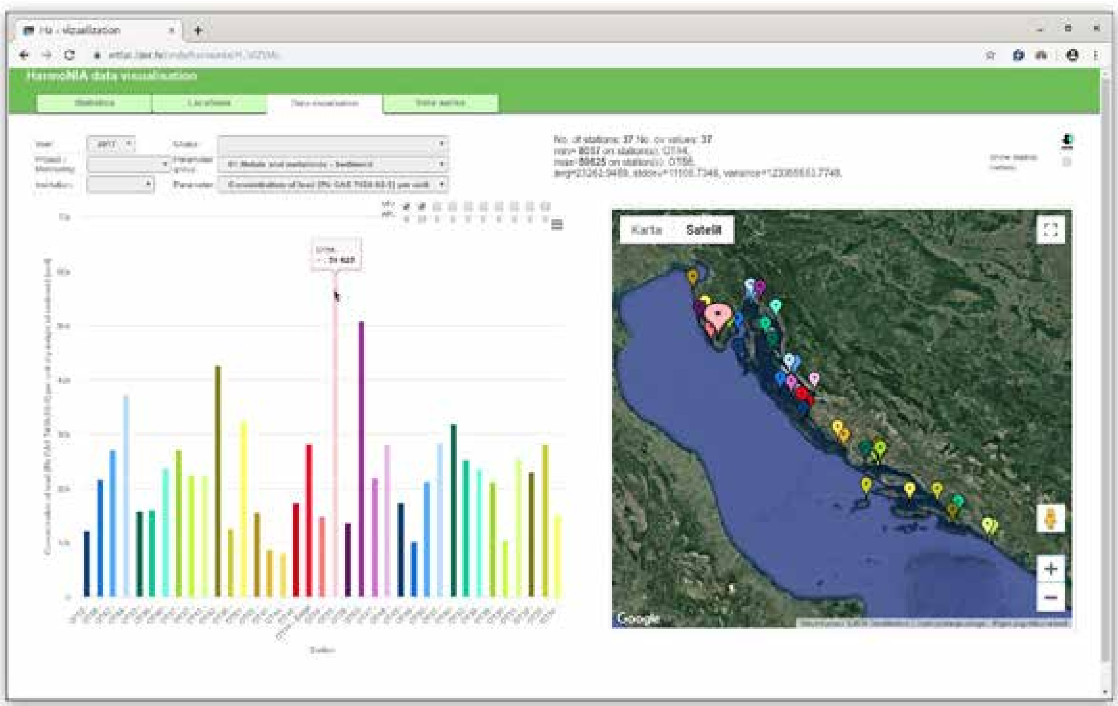


Figure 1: Application interface with highlighted paired elements for data visualisation (column and enlarged station marker)

Conclusion

Oceanographic research and monitoring are very relevant and expensive. It is important that data produced by research are used more than once, and in the proper way. Good visualizations help to better evaluate the state of marine environment. One of the tasks of HarmoNIA project was to harmonize methodologies used by different institutions. Additionally, this web application shows data heterogeneity, and lack of constant and coordinated monitoring efforts of hazardous substances in the Adriatic-Ionian region. Together with other project outputs, web application will help to address the needs of future research and monitoring.

Acknowledgment

The work was supported by INTERREG Adrion Project HarmoNIA. We acknowledge the contribution of the whole HarmoNIA partnership for sharing data and information.

PLOCAN multi-platform observatory data infrastructure

Tania Morales, PLOCAN (Spain), tania.morales@plocan.eu
Ruben Marrero, PLOCAN (Spain), ruben.marrero.ext@plocan.eu
Eduardo Caudet, PLOCAN (Spain), eduardo.caudet.ext@plocan.eu
Eric Delory, PLOCAN (Spain), eric.delory@plocan.eu
Jose Joaquin Brito, PLOCAN (Spain), joaquin.brito@plocan.eu

Introduction

The Oceanic Platform of Canary Islands (PLOCAN) is a multipurpose technical-scientific service infrastructure that provides support for research, technological development and innovation in the marine and maritime sectors, available to public and private users. As part of its activities PLOCAN runs an observatory consisting of four components that contribute to a broad range of coastal and regional sampling missions and services; (a) Coastal observing system: coastal component, expandable, permanent observing system, that will also serve as instrumentation test-bed, (b) Open-ocean deep water observatory: regional component, located 60 miles north of the archipelago in 3670m water depth (a.k.a. ESTOC site), (c) Offshore platform observing system: coastal, located on the platform physical structure, observes seabed, water column, ocean surface and atmosphere and (d) Mobile observing system: coastal, regional and global observation missions including the characterization of the Atlantic Ocean.

Multi-platform Observatory

A wide range of Observation Platforms are managed by PLOCAN's Observatory in order to provide a continuous and real-time in-situ monitoring of the ocean. These platforms can be both fixed (oceanographic moorings, weather stations, and HF radar) or mobile (autonomous underwater vehicles, unmanned surface vehicles, remotely operated vehicles and drifters), providing information about the ocean surface and/or the water column. Different sensors are placed in PLOCAN Observation Platforms allowing access to physical (temperature, salinity, currents, etc.), biochemical (ph, nutrients, carbon dioxide, chlorophyll-a, etc.) and meteorological data (air temperature, wind speed and direction, barometric pressure, etc.). Those platforms provide both engineering and scientific data in near-real-time with its own particular format and communication flow which leads to an unmanageable scenario when developing information systems. This is a common challenge shared by all ocean observatories which work with different observing systems and its solution becomes even more complex for small institutions like PLOCAN.

International initiatives and standards

As in an increasing number of regional, national, European and international infrastructures, PLOCAN is engaged in the implementation of FAIR principles. Data and metadata standards have been in place for a subset of existing and emerging observing networks, e.g. netCDF OceanSITES profile for fixed open-ocean observatories or EGO netCDF profile for gliders. However, a stronger effort in adopting common strategies and agreed protocols is still needed, working in that direction there are initiatives such as EMSO-ERIC for fixed observatories, OceanGliders and the GROOM II project for gliders or EuroGOOS HFR Task Team for HF radars. In addition, data harvesters such as The European Marine Observation and Data Network (EMODnet) and umbrella initiatives like ENVRI (via ENVRIFAIR) can be instrumental in bridging interoperability requirements across networks and set best practices. The current landscape implies that the observatory infrastructure should be flexible and prepared for adaptations, and the data need to be formatted and served to be delivered for different uses. **PLOCAN data infrastructure**

Taking into account the standards already adopted by the different communities and networks (gliders, fixed observatories, HF Radar, etc), their best practices and recommendations, PLOCAN has designed a common infrastructure, Figure 1, to manage and distribute the data produced in its observatory. A specific platform-driven data management has been developed in order to unify data flow and harmonize the different sources for posterior common treatment, which allows common visualization, statistical processing or automation on operational decision making process among others.

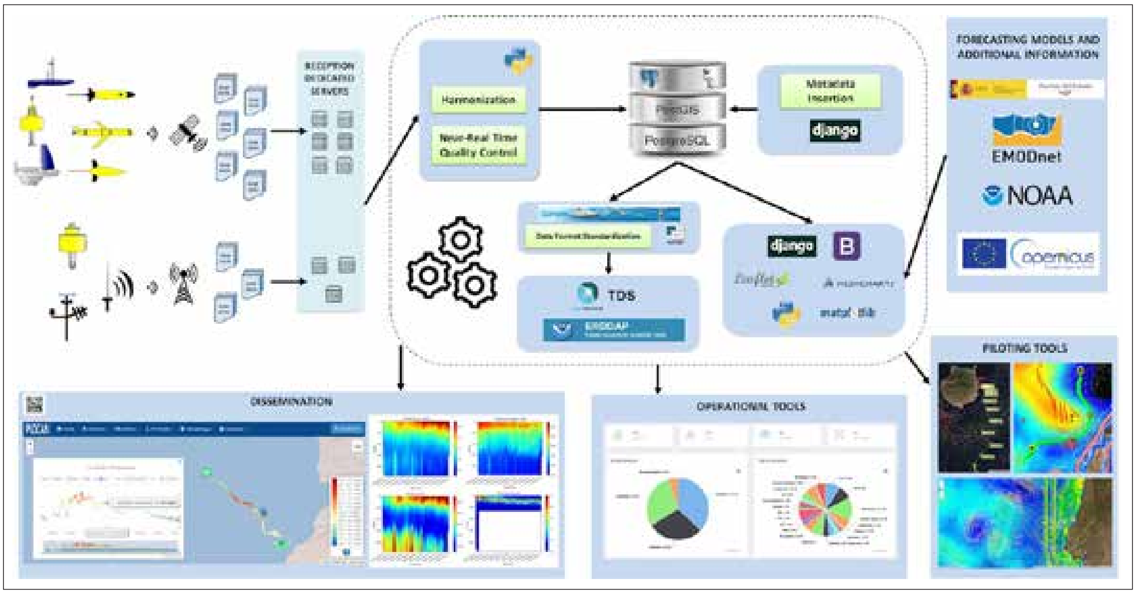


Figure 1: PLOCAN's observatory data infrastructure, dissemination and operational web tools.

Visualization tools have been built, on top of this harmonized data, with two main objectives: disseminating the observation activities (for both scientific and non-specialized audience) and providing visual tools in order to take better operational decisions. Regarding dissemination, dynamic maps and interactive data plots have been integrated in a unique web data portal that leads to a better understanding of the scientific collected data, which are also linked to its corresponding metadata. Furthermore, for operational purposes and with the aim of facilitating PLOCAN's autonomous marine vehicles fleet piloting, engineering information has been integrated in a web based operational map tool with other information sources such as forecasting models or near-real-time observations from HF radars, buoys, moorings or remote sensing platforms.

Among forecasting products it is worth highlighting the physical ones that include currents and waves as parameters which are of great importance for operational purposes. Depending on the scale, there exist different products. For instance, physical parameters at a global or regional (North-Atlantic and Maracaronesian area) scale are obtained from Copernicus Marine Environment Monitoring Service (CMEMS) and NOAA Operational Model Archive and Distribution System (NOMADS), whereas Spanish and Canary Islands local scale products are obtained from Puertos del Estado. From CMEMS it is also possible to obtain near-real-time remote sensing observation data and has a vast marine data catalog for European regions from where it is possible to obtain bathymetric products or vessel density maps. Building a common data infrastructure for PLOCAN observatory allows a more efficient use of the observation platforms and equipment available. In addition, using open data already available thanks to other institutions and initiatives in conjunction with PLOCAN observatory data, leads to a deeper understanding of the ocean.

POSTERS

SINDBAD: An Innovative Med Sea Situation Awareness Tool

Antonio Novellino, ETT SpA (Italy), antonio.novellino@ettsolutions.com
Marco Alba, ETT SpA (Italy), marco.alba@ettsolutions.com
Luca Bonofiglio, ETT SpA (Italy), luca.bonofiglio@ettsolutions.com
Paolo Brotto, PM_TEN Srl (Italy), paolo.brotto@pm10-ambiente.it
Maura Mazzarello, OnAir Srl (Italy), maura.mazzarello@onairweb.com
Giovanni Besio, University of Genoa (Italy), giovanni.besio@unige.it
Andrea Mazzino, University of Genova (Italy), andrea.mazzino@unige.it
Lorenzo Corgnati, National Research Council (Italy), lorenzo.corgnati@sp.ismar.cnr.it
Carlo Mantovani, National Research Council (Italy), carlo.mantovani@cnr.it
Maristella Berta, National Research Council (Italy), maristella.berta@sp.ismar.cnr.it
Ennio Ottaviani, OnAir Srl (Italy), ennio.ottaviani@onairweb.com

I. INTRODUCTION

Sea Situational Awareness (SSA) is the capability to provide present and future information on the state of weather and sea conditions. SSA is needed as basic information for any kind of sea-related and maritime activity. Lacking or missing of SSA rises the degree of unpredictability and incertitude for maritime working sectors and hence has an economic impact on the whole sea industry. Key aspects which make SSA highly relevant for the sea sector and blue economy are:

1. it gives important information to improve the navigation safety;
2. it can be used to reduce energy consumption with cost-saving effects;
3. it can help to improve the on board comfort.

In this paper we are going to present the developed tool and achieved results.

II. THE SINDBAD PROJECT

A group of italian SMEs (On AIR, XEDUM and PM_TEN), in collaboration with scientific community (University of Genoa and Italian National Research Council), have developed an ICT Service Infrastructure which produces highly detailed and reliable weather-marine forecasts, together with a web-based Decision Support System (DSS) able to provide customized navigation operational suggestions for conducting a boat and avoiding any kind of risk ensuring the best degree of comfort.

This work was supported by the Operational Regional Program ERDF of Liguria and included in the SINDBAD project which focused on leisure and boating navigation in the Ligurian Sea (Italy). SINDBAD consists of 4 modules:

1. Atmospheric - collects data on temperature, wind, rain, pressure
2. Sea Weather - provides information on wave height and direction
3. Comfort - calculates the ‘sickness index’ based on sea conditions and boat size
4. Drift - with forecasts at 2, 4, 6 hours from the starting point

The DSS integrates heterogeneous weather/sea forecasts and boat seakeeping data to provide specific tasks that may be of immediate utility for users, like weather-marine forecasts at different high spatial and temporal resolutions, the generation of comfort and risk maps and the delivery of boat-specific alerts and advices for navgation.

Figure 1. Sindbad's Mapviewer

The SINDBAD project output is a tool which can be easily adapted to target other of possible applications. The ultimate goal is to provide yachtsmen, fishermen, port authorities etc. with affordable and boat-specific highly detailed and personalized SSA information.

A Business Model based on free products and personalized purchased products could be applied for economic exploitations.

III. RESULTS

From April 2020 to nowaday, the SINDBAD App reached ~4.331 users, received ~10.312 pageviews (with high peeks during the vacation period) and more than 8.000 sessions. A User Experience Assessment has been carried out. Users satisfaction has been ranged between 7 and 8 out of 10, considering parmaters, such as: data visualisation quality, interfaces usability, interaction capability, etc. New evolutions of the system are planned for further improvements.

The new user interface for the development of hydrometeorological support based on integrated data

Evgenii Viazilov, RIHMI-WDC (Russia), vjaz@meteo.ru
Nikolay Mikhailov, RIHMI-WDC (Russia), nodc@meteo.ru
Denis Melnikov, RIHMI-WDC (Russia), melnikov@meteo.ru
Vasiliy Smolyanitskiy, AARI (Russia), vms@aari.ru

For solving applied tasks (AT) separate data set or database are usually used. One of the most important approaches for organizing complex processing of distributed and heterogeneous data is their preliminary integration. The general scheme of the functioning of the hydrometeorological support system of enterprises includes the following steps: configuring of information resources (IR), tools, representations; search (filtering) data; calculation of statistical characteristics and calculation of new parameters; representations of parameters values.

The configuring is intended to include in the AT profile of the necessary data; of services for calculating trends, anomalies of one or another parameters; of extrema in the selected area or over a period of time; of average values based on a time series at points or regions based on tuned IRs.

If the IR represents a wider region, then by tuning for the geographical area, parameters, spatial and temporal resolution scales, type of representation (points, polygons, trajectories, profiles) for each AT appropriate subsets of data are selecting or filtering.

Based on the available IRs using services, statistical characteristics (averages, extrema, trends, anomalies, and others) and indicators are calculated. For the observed, calculated and prognostic values of the parameters, dangerous levels are indicating by using their local threshold values. For implementation these functions, a toolbar is used - this is a set of buttons and drop-down lists for performing actions with selected data. Examples of tools are the weather monitor for every station, the meteogram - a graph of changes in parameters along any trajectory, a forecast at a point or region, calculation of trends, anomalies in a point, along a trajectory or for region. Depending on the type of IR, the buttons appear on the toolbar: animation for the appointed period, calculation of new indicators, viewing graphs, meteogram, report editor, subscription on data, and metadata on IR.

The form of representation of observed, calculated and forecasted parameters depends from:

- type of data marked in the metadata on the IR (point, grid, profile, catalog of objects);
- type of platform (fixed point - it is possible to generalize data for a time series; dynamic platform - changing parameters along a trajectory is built);
- presence of geographical coordinates (it is possible to build the spatial distribution of parameter values in the form for points or isolines);
- availability of date, time (it can build a time series), for example, for observations in a fixed point can plot graphs.

The “Graphs” section may include the construction of several figures of the time course at different points for comparison. Using the catalog of spatial objects with coordinates and time can be search satellite images. The “Reports” section is forming based on of the including of the necessary maps, graphs, and tables and the corresponding explanatory text. To accelerate the receipt of individual ready-made results, “Hot keys” are used, with which it can get a list of current natural disasters; look at local threshold values of indicators; show available reports. The scheme of the AT functioning shown in Figure 1.

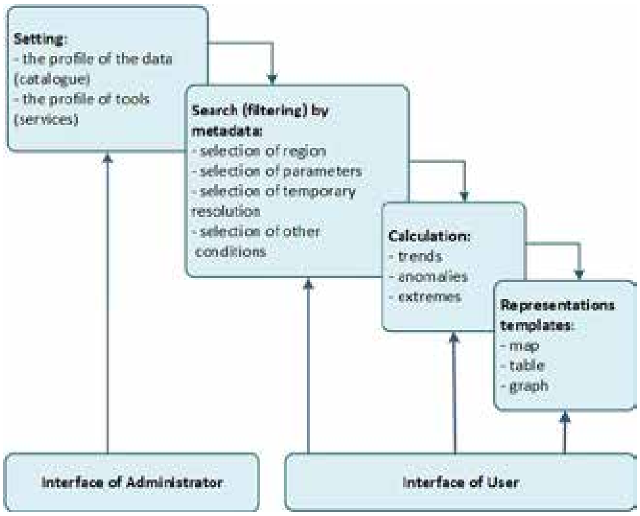


Figure 1: General scheme of the AT functioning

The following AT identified disasters monitoring, climate assessment, climate monitoring, support of long-term forecasting, and transport operations. The «Climate monitoring» AT (http://portal.intaros.meteo.ru/portal/intaros/services/) is intended for obtaining climate change indicators in the Arctic and individual seas at various temporary resolution (month, season, year, multi-year summaries) and of estimates climate change. The AT includes a diagnosis of climate change in the Arctic based on data of coastal and synoptic stations, observations in the open sea; identification of extremes. The parameters presented in the IRs include monthly and annual average, minimum, max values of atmospheric pressure, air temperature, wind speed, precipitation, wave height, permafrost active layer, river flow, sea ice, and sea level. The indicators obtained based on basic IRs include anomalies, trends of the parameters; monthly values of anomalies. Panels includes a selection of parameters as radio-button, a group of check boxes for selecting the sea. The panel “selection temporary resolution” includes a radio-button and slider. It needs to select a specific month (season or year), period for multi-year values. Tools include calculating trends, anomalies, extrema at a point and for a region.

For the first time, it is proposed not only to visualize a particular IR (array or database), but to enable the user to configure his task to the necessary resources from those available in the integrated database. A new method is presenting for universal process of processing and visualization of environmental data in various forms (maps, graphs, tables, and reports) and providing consumers with access to them in the «self-service» mode.

The presented approach develops the use of heterogeneous, distributed and integrated data for solving applied problems related to monitoring the hydrometeorological situations, with preparing results for analyzing the current situation, with calculating indicators and decisions support.

The work financially supported by the Ministry of Science and Higher Education of the Russian Federation, a unique project identifier RFMEFI61618X0103.

Southern Ocean data: A community effort to build a data ecosystem

Patrick Gorringe, Swedish Meterological Hidrological Institute (Sweden), Patrick.Gorringe@smhi.se
Pip Bricher, Institute for Marine and Antarctic Studies, University of Tasmania (Australia), phillippa.bricher@utas.edu.au
Marco Alba, ETT SpA (Italy), marco.alba@ettsolutions.com
Luca Bonofiglio, ETT SpA (Italy), luca.bonofiglio@ettsolutions.com
Antonio Novellino, ETT SpA (Italy), antonio.novellino@ettsolutions.com

I. INTRODUCTION

Data sharing is at the heart of SOOS. It is widely understood that Southern Ocean data is expensive and logistically difficult to collect and that there is never enough of it. However, there is more out there than individual researchers might expect, and we are working to make it all accessible from one place. The aim is that we, as a research community, develop the systems that will allow scientists to maximise the scientific value of observations collected in the Southern Ocean, no matter who collected them, or when, or why. At the same time, SOOS wants to protect researchers' rights to get first use out of data that they collected. In this paper, we share our progress, lessons, and plans for the future, and invite you to join us as we develop our data ecosystem.

II. SOUTHERN OCEAN OBSERVING SYSTEM

SOOS is an international initiative of the Scientific Committee on Antarctic Research (SCAR) and the Scientific Committee on Oceanic Research (SCOR). Developed over many years, SOOS was officially launched at the end of 2011 with the opening of the International Project Office, hosted by the Institute for Marine and Antarctic Studies (IMAS), and the Australian Research Council's Antarctic Gateway Partnership at the University of Tasmania, Australia. Since then, SOOS has built a network of stakeholders and contributors, all working together to achieve the community-defined mission and objectives.

The Southern Ocean Observing System bridges oceanographic and polar science programs and is one of the most intensely internationally integrated scientific communities, thanks to the logistical challenges of conducting science in these remote waters. The users of Southern Ocean data are therefore highly heterogeneous in terms of their needs and expertise.

Serving such a diverse research community requires data management systems that are flexible and focus on integration of existing data products, rather than trying to duplicate existing work. The Southern Ocean Observing System is working with both the science and data communities to design an ecosystem of data management tools, catalogues, and systems for polar oceanographic research.

What makes this system unique, compared with other data aggregators, is that SOOSmap serves an unusual community, with data coming from both a large number of nations and different scientific disciplines, despite being restricted to one geographic region. Instead, most similar portals either serve a limited number of observation types or take data from a limited number of contributing research programs.

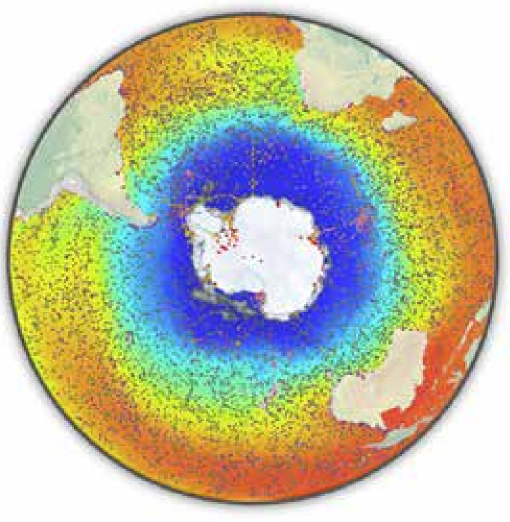


Figure 1: SOOSMAP

Key components of the SOOS data vision include data and metadata discovery tools, fieldwork coordination tools, and linking data collections with analysis tools.

Since data is coming from different sources with different formats, we process the data in order to be able to use ERDDAP as a unique Data Server for all the data we are serving. Since ERDDAP supports a wide range of data types as sources, we are able to maintain the original data in the original format and have an homogeneous point of access for retrieving the data.

Just as important are the people and communities who develop, link, and use these components. We will share, indeed, our experiences in working with EMODnet Physics to develop SOOSmap - a portal to explore, graph, and download curated spatial datasets of key observations.

SOOSmap relies on the EMODnet Physics backend, with a polar-projected map interface. In SOOSmap we added the possibility to manage different types of layers other than platforms, as data layer or background layer. Open Geospatial Consortium standards are supported plus the addition of the geojson standard. In terms of content, SOOSmap has brought in biological data on a range of topics, including birds, plankton, krill, and other key species. In terms of physical data, SOOSmap has brought in key new data sources, including the MEOP seals, which collect CTD data in areas where sea ice interferes with Argo-style floats; global CTD casts from PANGAEA, which is a major host of this data type; and the Saildrone data, which is a good example of a collaboration between EMODnet and a private company. Finally, Data are stored on a redundant Data Server and services are monitored real time for availability with Nagios (Network Monitoring Software)

III. RESULTS

In 2019, SOOSmap received ~8000 page views per month, with ~2500 near-real time data files and 11 long-term repository datasets downloaded each month. While it is difficult to fairly compare the metrics with other Antarctic data repositories, it appears to receive approximately 80 times as many views as the SOOS metadata portal on NASA's Earth Data search tool. For this community, SOOSmap has made it much more enticing to explore and download data. We will keep focusing on the challenges and need for a federated metadata search tool to improve access to a much broader range of polar oceanographic data than can be served directly through SOOSmap.

A pilot study to provide large-volume, user oriented marine spatial data of a federal institution online

Romina Ihde, Federal Waterways Engineering and Research Institute (Germany), romina.ihde@baw.de
Janina Freund, Federal Waterways Engineering and Research Institute (Germany), janina.freund@baw.de
Robert Hagen, Federal Waterways Engineering and Research Institute (Germany), robert.hagen@baw.de
Andreas Plüß, Federal Waterways Engineering and Research Institute (Germany), andreas.pluess@baw.de
Frank Kösters, Federal Waterways Engineering and Research Institute (Germany), frank.koesters@baw.de

Data of federal institutions may offer a valuable source for scientific or commercial applications. Unfortunately, these data are not always well known or easily accessible. Therefore the German Federal Ministry of Transport and Digital Infrastructure (Bundesministerium für Verkehr und digitale Infrastruktur, BMVI) funded the mobility research initiative (mFUND). Within this initiative the Federal Waterways Engineering and Research Institute (Bundesanstalt für Wasserbau, BAW) created the EasyGSH-DB project to provide user-oriented, synoptic reference data for the thematic fields of geomorphology, sedimentology and hydrodynamics. These data cover a large area, the German Bight, and a large time span (years 1996 to 2015). The purpose is to provide 20 years of consistently generated marine data in a high temporal and spatial resolution for the German Bight, which can be visualized and openly shared. To obtain a high degree of user orientation, possible applications for the data have been identified in a structured stakeholder involvement.

A web-based presentation is chosen and the portal for EasyGSH-DB was designed with a focus on the following four main goals:

- discovering and sharing of project related data and information
- providing technical documentation e. g a model validation and metadata
- free download of data
- quick and visually appealing visualisation

Figure 1 shows the structure of the website which covers information about the project, an interactive viewer for visualization and exploration of data, access to data as services and download.

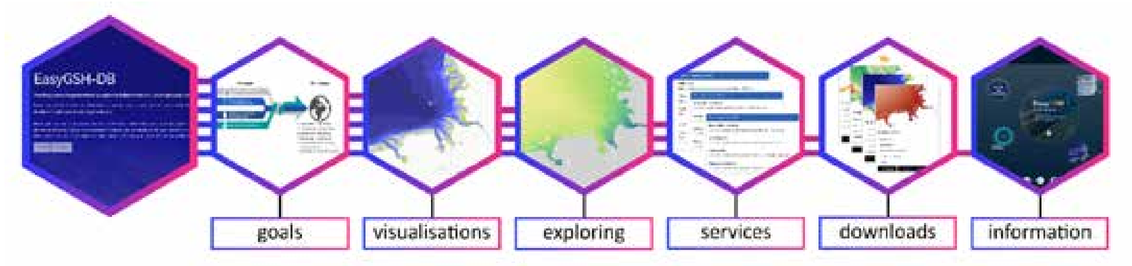


Figure 1: Short presentation of website content.

The technical set-up was chosen in a way to obtain a sustainable platform with the opportunity for easy updates, maintenance and use by stakeholders.

The data products are highly complex with different spatial resolutions and data formats e. g. resolution of bathymetry (10x10 m), hydrodynamic analyses (100x100 m) and numerical model output (1x1 km, 20 minute time steps) with vector, raster or NetCDF data formats. The contents of data and metadata for one year are highly diverse in complexity and this multiplied by temporal data coverage of twenty years, required separated data storage, services and visualisation.

Rasterized model data from the North Sea model is stored in a THREDDS Server (Thematic Real-time Environmental Distributed Data Services). From there data and metadata can be queried as Web-Map- Services (WMS) or Web-Coverage-Service (WCS) and used from web clients for viewers to support key features such as data comparison or animation.

For the three main topics geomorphology, sedimentology and hydrodynamics, data is stored in two databases depending on the data format. The databases are divided in vector and raster data. They share the same metadata with specification on their format, process steps and lineage. A Geoserver is used to generate the WMS, Web-Feature-Services (WFS) and WCS, which are published on the website and used from the Terrajs Viewer, where users may request, discover or directly download the data according to their field of research.

Using a Catalog Service for the Web (CSW) directly from the metadata repository and having all UUIDs, links or other connections partly automatically updated allows publishing of all metadata separately or together with the data. Through CSW harvesting, metadata information systems like the mCloud, govdata or GDI-DE have access to the data information and download.

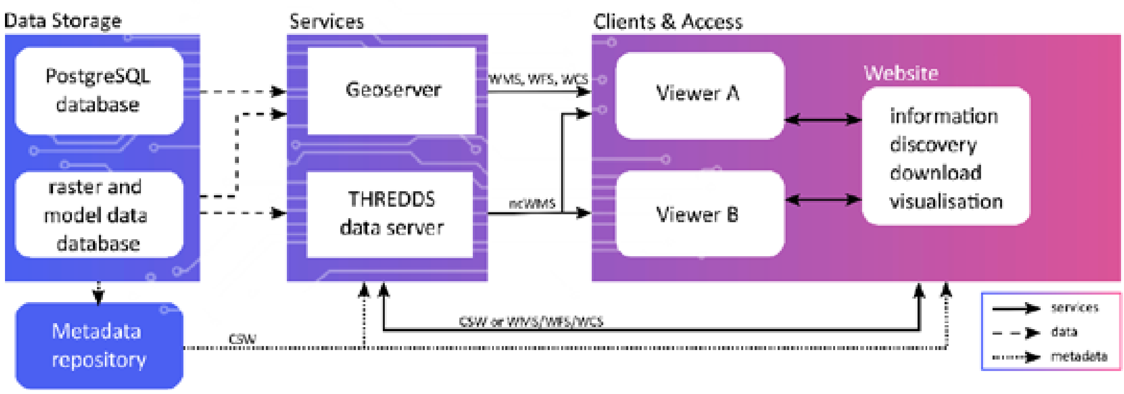


Figure 2: Basic overview about the technical infrastructure implemented from EasyGSH-DB with service, data and metadata flows.

Various tools and software were used to determine the best-practice for building a website to present the different kinds of data. The key features which are:

- easy to find downloads with different data formats,
- download of time-series for a location,
- query results for metadata with the possibility of a separate download,
- visualisation over a free and easy to use viewer,
- animation of model data and
- technical documentions for e.g. the North Sea model, validation and project status.

As a result, the EasyGSH-DB website joins the requirements from stakeholder interviews with the infrastructure of a project and offers a variety of services and visualisation tools with different features. The incorporation of different stakeholder interests, topics and features in an intuitive website is not a simple task. Maybe EasyGSH-DB can be used as an example for projects and how to overcome these challenges in the future.

Toolboxes for data management in marine environmental monitoring

Örjan Bäck, SMHI (Sweden), Orjan.Back@smhi.se
Arnold Andreasson, SMHI (Sweden), Arnold.Andreasson@smhi.se
Johannes Johansson, SMHI (Sweden), Johannes.Johansson@smhi.se
Magnus Wenzer, SMHI (Sweden), magnus.Wenzer@smhi.se
Lotta Fyrberg, SMHI (Sweden), Lotta.Fyrberg@smhi.se

Users of environmental monitoring data will always need access to highest quality of data. Advances in technology can meet these demands via technical solutions. The Swedish Meteorological and Hydrological Institute (SMHI) is the Swedish National Oceanographic Data Centre and has developed advanced systems (toolboxes) for data management and perform state-of-the-art quality controls (QC) of marine data. Specifically, the data management handles our data flows, ranging from collecting, organizing, storing data on local servers and packaging into zip archives. The QC includes outlier and range checks and comparison to historical and regional data. In addition, the toolboxes are also used for data archiving, analysis and basic/routine reports (Fig. 1).

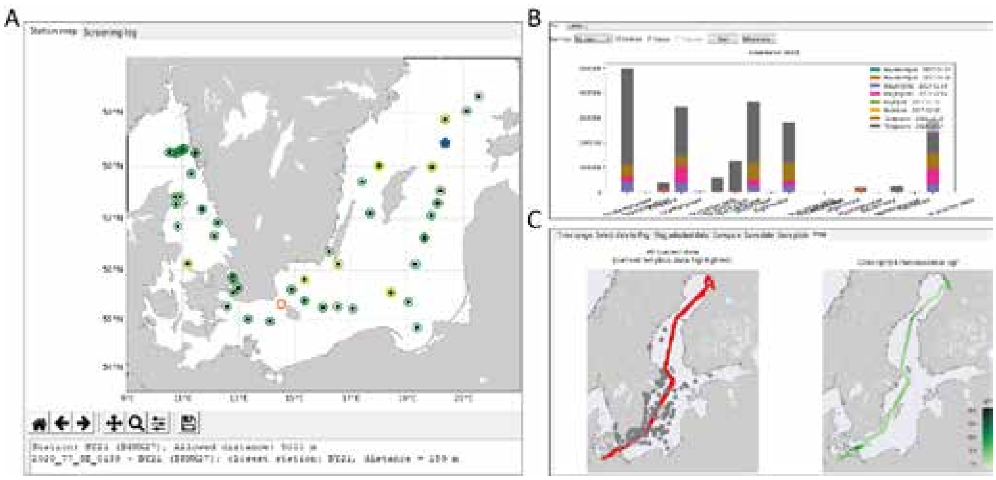


Figure 1: Graphical interfaces of the toolboxes developed by SMHI; SHARKtoolbox i.e systems for data management, quality control, data archiving and basis for reports (panel A), PLANKTONtoolbox i.e systems for data management and analysis as well as advanced counting and biovolume calculations of plankton samples (panel B), and GISMOtoolbox i.e systems for data management and quality controls of in situ ocean data from ferryboxes and fixed platforms (panel C).

All toolboxes can handle data analysis from national and regional reported physical/chemical and marine biological data. The toolboxes automate data management and reduce the amount of manual work (and human errors) to ensure higher data quality. In particular, the toolboxes combine an automatic QC system with a visual interface that interactively involves the user. Thus, the systems allow for machine and human interaction to find and resolve advanced QC issues, automate data management and perform in-depth data analysis. The toolboxes were developed using open source code and are released with an open source MIT license and are available both for Microsoft, Linux and Apple users.

Using metadata quality to filter datasets in data portals

Susanne Feistel, Leibniz Institute for Baltic Sea Research Warnemünde (Germany), susanne.feistel@io-warnemuende.de
Ulrike Kleeberg, Helmholtz-Zentrum Geesthacht (Germany), ulrike.kleeberg@hzg.de
Jörn Kohlus, Nationalpark Wattenmeer, (Germany), joern.kohlus@lkn.landsh.de
Romina Ihde, Federal Waterways Engineering and Research Institute (Germany), romina.ihde@baw.de
Carsten Schirnick, GEOMAR Helmholtz Centre for Ocean Research Kiel (Germany), cschirnick@geomar.de
Stefanie Schumacher, Alfred Wegener Institute (Germany), stefanie.schumacher@awi.de
Susanne Tamm, Federal Maritime and Hydrographic Agency (Germany), susanne.tamm@bsh.de

Sharing geoinformation and data across communities is becoming more important, which is enabled by web services. The Open Data initiative (e.g. data.gov) promoted by public agencies and research institutions enhances this development. However, poor data quality with insufficient quality information may be hindering the acceptance and re-usage of the data by the scientific community in the future. The mandatory metadata within ISO 191** is not sufficient for a comprehensive quality assessment or long-term usability of any dataset. Documentation and quality information is mostly optional and not easily found or filtered. It might be stored within the metadata as information in LI_Lineage, DQ_DataQuality, or even within the dataset itself, possibly coded. Ultimately, in most cases, quality information is not directly accessible for interested users.

On the other side, for the data creator, the documentation of quality information can get time-consuming and overwhelming, especially if required all at once in the reporting phase of a project. Furthermore, quality information gets significantly more complex if more than one dataset is considered as in data products such as maps or modeling results.

We present a more practical approach to get a sufficient and standardized quality assessment (quality flag), dynamically generated from the entire quality information of the metadata XML file. While the ISO standard offers a complex array of optional fields to provide quality information, we recommend a manageable number of fields to be filled (Fig. 1). As prepared in (Feistel et al. 2020, in prep.) there is specific information a user needs first and foremost to evaluate the usefulness of a dataset.

Figure 1: Create accessible quality information and record it in recommended fields of the metadata standards of ISO 191**.

Our task group provides a set of tools to assess quality information recorded in the recommended fields of the ISO standard. The first is a web-based form to manually assess a single dataset, the second is a program for syntactically analyzing multiple dynamically generated XML in ISO 19139. To make the quality assessment compatible and comparable between platforms and quality flag schemes, as well as machine-friendly, we propose a coded summary string in the pattern of the scheme plus a flag, e.g. "SDN::1" (equal to quality flag "1" within the quality flag scheme of "SeaDataNet") to be put in the ISO field "DQ_StandaloneQualityInformation" (Fig. 2).

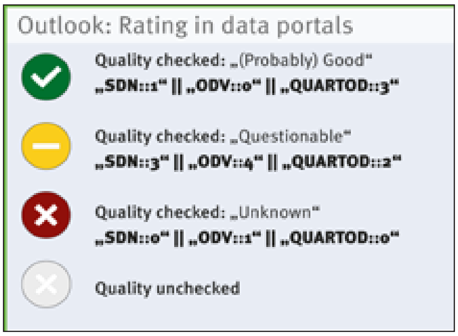


Figure 2: Proposed quality ratings through automatic assessment of metadata quality via the ISO XML checker.

Our toolbox is aimed at data creators to help provide meaningful metadata and at data portals to search for sufficiently documented data.

For data creators, metadata templates can be prepared in advance. This is especially useful for repeating data collection or creation procedures within projects, as well as long-term data series. Known instruments and methods can be inserted in the recommended fields and require little further effort. Thus, the compilation of metadata can be simplified for individual scientists.

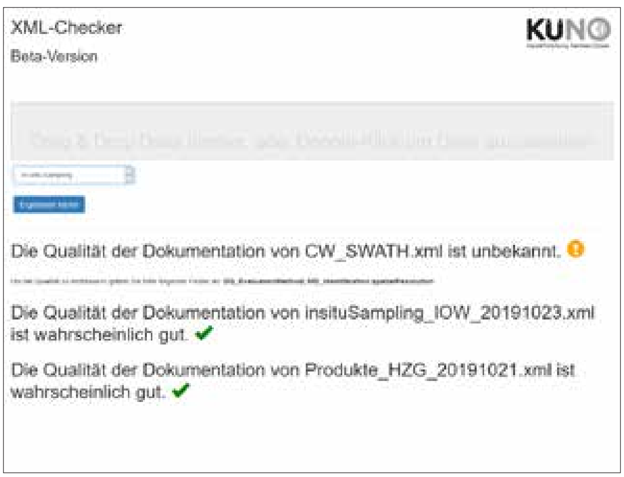


Figure 3: Drag & drop a XML-file in the checker to create a standardized metadata quality assessment.

In data portals the primary advantage of using an automatic assessment of metadata quality via the ISO XML checker is in getting a standardized quality flag based on a common quality flag scheme (Fig. 3). A time consuming individual control is not necessary. The quality control mechanism supports users in finding data relevant for their own work more efficiently. Last but not least it allows for a quality rating and filtering of any search results (Fig. 2).

Ocean Data View goes Online

Reiner Schlitzer, Alfred-Wegener-Institute (Germany), reiner.schlitzer@awi.de
Sebastian Mieruch, Alfred-Wegener-Institute (Germany), sebastian.mieruch@awi.de

Ocean Data View – desktop versus online versions

Ocean Data View (ODV; https://odv.awi.de/) is a popular software package for the analysis, exploration and visualization of oceanographic and other environmental data. ODV has more than 80,000 registered ODV users, and, based on recent software download logs, more than 10,000 active users. ODV-created graphics material is regularly published in textbooks as well as scientific journals, including high-level journals, such as *Nature*, *Science*, *PNAS* and *PlosOne*. ODV plays a fundamental role in the SeaDataNet (SDN), SeaDataCloud (SDC) and EMODnet projects, and is employed for data import, parameter aggregation and harmonization as well as for data quality control. The resulting large aggregated data collections are shared as ODV collections.

In the classical desktop configuration, a user installs the ODV software on the computer and then copies the datasets to be worked on to the same machine. The computer can be connected to the Internet, but this is not a requirement. ODV runs successfully in remote locations, with poor or no Internet connection at all. This configuration is ideal when working with private datasets and sharing with colleagues or communities is not intended or not a priority.

When working with large community datasets, however, such as the World Ocean Atlas (https://www.nodc.noaa.gov/OC5/woa18/), GO-SHIP and CLIVAR global ocean repeat hydrographic data (https://cchdo.ucsd.edu/), or aggregated regional SeaDataNet collections (https://www.seadatanet.org/Products/Aggregated-datasets), the desktop ODV configuration has drawbacks. Usually, community datasets are very large and use up significant amounts of precious disk space. Scarsity of storage space, especially on laptops, often sets limits to the selection of datasets that can be analyzed and worked with simultaneously. Making copies of the datasets on every client computer not only wastes storage, these copies also become outdated and need to be refreshed, whenever new versions of the datasets are released.

To overcome these drawbacks, we have developed and implemented webODV, a second, online version of ODV. webODV works in a client/server configuration as shown in Figure 1.

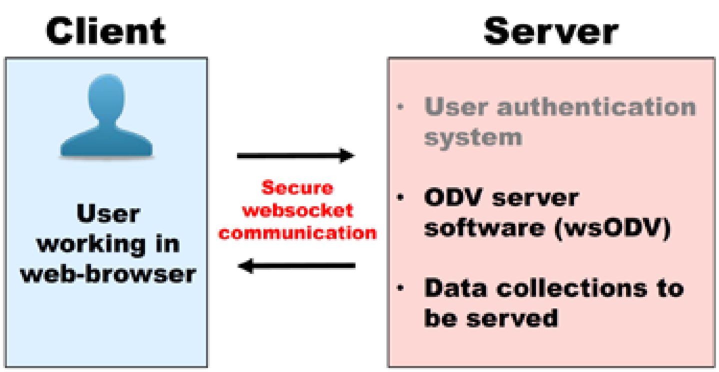


Figure 1: webODV configuration

webODV Configuration

In the webODV configuration (Fig. 1), datasets as well as the ODV software reside and run on a dedicated server machine, not on the client computer. All that is needed on the client computer is an up-to-date web browser (Chrome, Safari, Edge, Firefox). The client communicates with the server over the Internet using secure websocket communication. The ODV running on the server (wsODV) is a special version of ODV containing an embedded websocket server in an otherwise fully functional ODV application. Websockets are a relatively new web technology providing a permanent, bi-directional connection with very low latency. This is crucial for a responsive, interactive user impression. The actual client/server communication uses a set of JSON-encoded request/reply commands. Both, the special server-side wsODV application as well as the command documentation are public. This allows interested parties to develop and establish webODV webservices of their own.

ODV-online

We are developing a number of web services using the webODV configuration above. One specific service called *ODV-online* aims to provide the look-and-feel and functionality of the desktop ODV in the browser window (Fig. 2). Here the browser window resembles the ODV desktop application window consisting of a canvas with station map and data windows, and metadata, sample data and isosurface value lists on the right side. As in the desktop ODV, left mouse clicks on stations or data points select these items. Right-clicks bring up context menus providing the familiar ODV functionality. Users can download image files of the entire canvas or individual windows and can export the data of the current station set or of individual data windows. The ODV-online interface can also be used to perform data quality control in the same way as the desktop ODV. The *ODV-online* JavaScript code is available to interested parties.

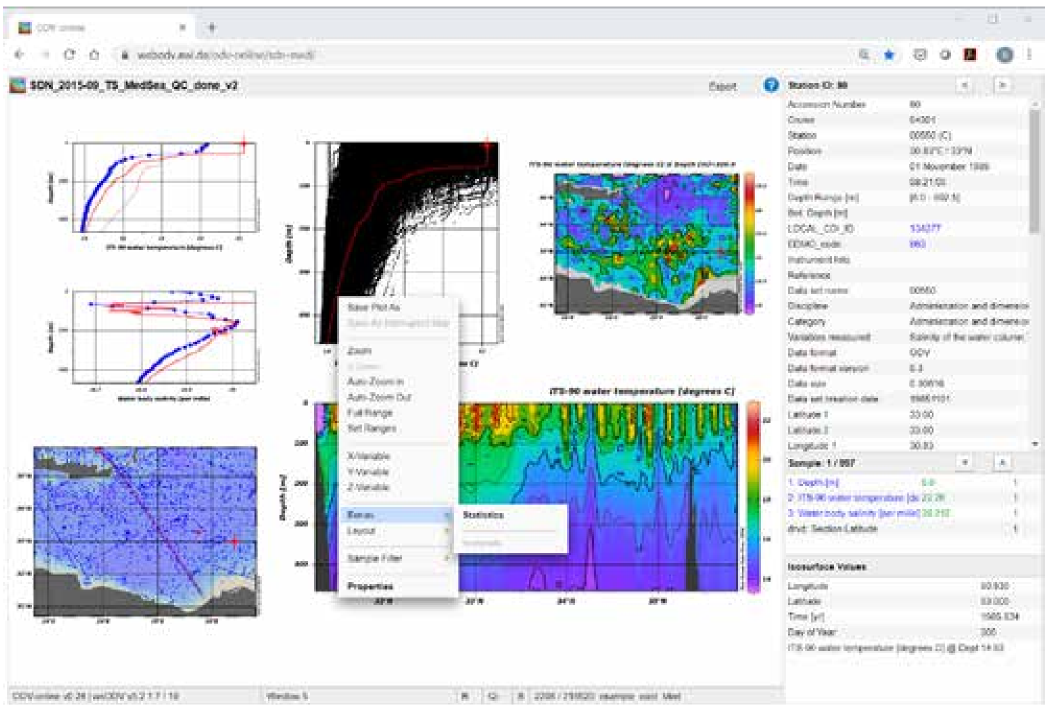


Figure 2 : ODV-online browser interface

An all-in-one web tool to apply CTD quality control, format data, and generate metadata under SeaDataNet criteria

Pablo Otero, Instituto Español de Oceanografía, Centro Oceanográfico de Vigo (Spain), pablo.otero@ieo.es
Gonzalo González-Nuevo, Instituto Español de Oceanografía, Centro Oceanográfico de Vigo (Spain), gonzalo.gonzaleznuevo@ieo.es
Elena Tel, Instituto Español de Oceanografía, Servicios Centrales, Madrid (Spain), elena.tel@ieo.es

I. INTRODUCTION

Professional data centres of the Pan-European region active in marine data collection, in particular National Oceanographic Data Centres (NODCs), routinely share diverse datasets through SeaDataNet, a distributed Marine Data Infrastructure. This network provides on-line integrated databases of standardized quality.

Sharing marine data through SeaDataNet is an optimal way of ensuring FAIR principles: Findable, Accessible, Interoperable and Reusable. However, submission of data to this infrastructure demands a set of technical tasks that cover quality control processing, adoption of common vocabularies, implement file format standards and preparation of associated metadata. Although common software tools are made available to NODCs to facilitate data and metadata preparation (see https://www.seadatanet.org/Software) these tasks continuous to be complex and time-consuming. For example, the technician must use MIKADO software to generate INSPIRE-complaint metadata and adopt SeaDataNet Common Vocabularies; NEMO software to transform the original file to a common data transport format; and OCTOPUS software to check the compliance of a file. Additional software must be used to quality check and properly flag each individual record.

The Instituto Español de Oceanografía (IEO) acts as NODC and submits diverse datasets collected by the Spanish oceanographic fleet to SeaDataNet infrastructure. Until the present, more than 35,000

CTD vertical profiles have been processed and submitted. Taking into account that this kind of data usually follows the same format and involves similar processing, a web application has been developed to perform all these tasks straightforward. The aim is to save processing time and to reduce human interaction that could lead to errors or lack of uniformity in data.

II. METHODOLOGY

A Python application (http://ctdcheck.ieo.es/) has been coded in combination with Flask, a lightweight WSGI web application framework, to get the app up and running. Parts of the code are based on the CoTeDe[1] and Python-CTD[2] packages. The source code is available at https://github.com/PabloOtero/CTDChecker.

The software parses a set of cnv files and homogenise variables (names, column order, precision, etc.). Different quality test are applied to each vertical profile. Most of the tests simply reproduce the procedure recommended by GTSPP[3] and SeaDataNet manual[2]. For example, the application checks if the date is valid, the station is at sea, records are in the expected range (globally and regionally), compares with climatology, detects spikes and performs a density inversion test among others. Depending on the success during the test, a flag is assigned to each individual record and also an overall flag to each parameter and to the entire profile. The application also plots temperature and salinity profiles and their comparison with climatologic values. Bad or suspicious points are highlighted.

Processed records are formatted to the SeaDataNet MEDATLAS auto-descriptive ASCII format. Finally, Common Data Index (CDI) metadata are generated, one per vertical profile.

As input, the software requires a Cruise Summary Report (CSR) metadata file associated to the dataset. In this way, software extracts main info from the cruise (e.g., ID, principal investigator, vessel, dates, location, project…) and ensures consistency when generating CDI metadata and correct inclusion of this information in the header of the output data file.

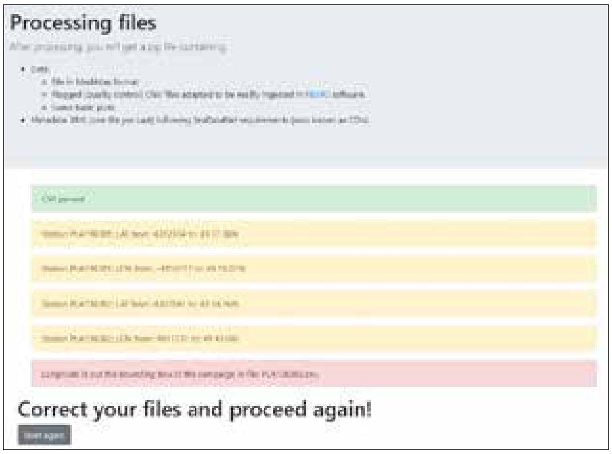


Figure 1: Screenshot example of an output of the application, showing some warnings (geographical coordinates have been reformatted and therefore requires further revision) and error (location is out the bounding box of the campaign informed in the CSR metadata file).

III. CONCLUSIONS

An all-in-one web tool to apply quaility control to CTD vertical profiles, reformat data, and generate metadata under SeaDataNet criteria has been created. The objective of this tool is saving time and reduce human errors. The reason to implement a web-based instead a desktop application is to avoid any installation process, skip upgrades and use cross-platform. The application always requires the previous existence of a CSR file in the SeaDataNet infrastructure with the aim of create completely coherent metadata (CDIs) and full-descriptive header files.

REFERENCES

[1] Castelao, G. P., (2020). A Framework to Quality Control Oceanographic Data. Journal of Open Source Software, 5(48), 2063, https://doi.org/10.21105/joss.02063

[2] Fernandes, F. (2014, August 25). python-ctd v0.2.1 (Version v0.2.1). http://doi.org/10.5281/zenodo.11396.

[3] UNESCO–IOC, 2010. GTSPP Real-Time Quality Control Manual. First revised edition. UNESCO–IOC. United Nations Educational, Scientific and Cultural Organization 7, Place de Fontenoy, 75352, Paris 07 SP. IOC/2010/MG/22Rev.

[4] SeaDataNet, 2010. Data Data Quality Control Procedures. Version 2.0 May 2010. Available at https://www.seadatanet.org

Data integration of the Russian segment of the integrated Arctic observing system

Kristina Belova, RIHMI-WDC (Russian Federation), christy@meteo.ru
Nikolai Mikhailov, RIHMI-WDC (Russian Federation), nodc@meteo.ru

Data are the raw characteristics of information processed and control by computers. Information is the presentation of data in a context that has business value to people. Information is needed by scientific, design and industrial enterprises, government agencies and other participants in maritime activities. Information is especially critical for activities in the Arctic. It is necessary to integrate heterogeneous and distributed data from the Arctic observation networks, databases and systems, to produce relevant information, and to disseminate this information in support regional maritime activities. This problem is being addressed under the INTAROS Project of the HORIZON-2020 Program and the Research Programme of the Ministry of Science and Higher Education of Russia through the creation of an integrated Arctic observation system (iAOS).

Russian segment and node iAOS node

The Russian segment of iAOS includes a network of the Arctic data sources, which is the part of the Unified State System of Information on the World Ocean (ESIMO, www.esimo.ru), and the Russian iAOS node – a software and hardware system that contains subsystems for data integration, for data processing and for provision of services. The ESIMO software and other means have been used as the infrastructure of the node. A multi-component Java-oriented development and operation environment is used as the basic software. Applied software consists of applications and SOAP/ RESTful web services implementing business processes and visual user interfaces of the node subsystems. In the context of computing hardware, a node is a cluster of servers containing virtual machines with applied software components.

Data integration of the Russian segment of iAOS

There are three main approaches to data integration: (1) Consolidation, when data are extracted from multiple sources and integrated into a centralized data warehouse. The technology is called ETL (Extract-Transform-Load); (2) Federalization based on the use of brokers - software applications for transparent access to distributed and heterogeneous data sources, retrieving and transforming data into a common data model, delivering data to a given point in an accepted exchange format. An example of this approach is enterprise information integration (EII) technology; (3) Data distribution based on the transfer of data from the data source to the receiving application. Enterprise application

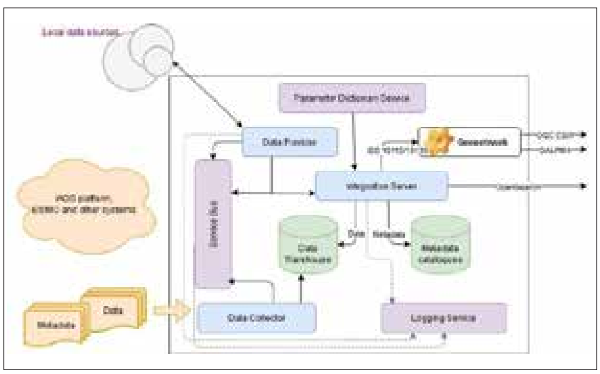


Figure 1 : Structure of the data integration subsystem of the RUSSIAN node of iAOS

integration (EAI) technology is an example of this approach. Data integration of the Russian segment of iAOS is based on a combination of EAI technology and ETL technology. The data integration subsystem (Figure) of the node includes the following components:

- Data Collector, which performs access to data and metadata of the iSNA platform and ESIMO, their transformation if required and inclusion of data messages and flows into the node processes;
- Data Provider is an integration broker. It performs access to data sources using software wrappers, extracting and transforming data into the format of a transport data file (formats supported are NetCDF (2, 3), CSV), and also forms the metadata in the ISO 19115 standard / 19139 and interaction with the Integration Server;
- Integration Server is the data integration adjuster. It maintains a registry of data providers and data sources, can add new objects or delete existing ones. It also collects data and metadata provided by the Data Providers, loads them into the file Data Warehouse, and provides data and metadata for the implementation of internal business processes of the node and for external usage and systems;
- Data Warehouse, implemented as a cache system of transport data files and metadata files for subsequent sharing;
- Service Bus in combination with Logging Service implements uniform, flexible and documented communication between the components listed above. The Parameters Dictionary Service is designed to present information about parameters to consumers in a unified form understandable to humans.

As a result of data integration, the designated data of sources (named as information resources of the node) and metadata are loaded into the file Data Warehouse and the PostgreSQL integrated database. Data Warehouses are replenished automatically and their contents are used to produce new products and services. Note that over 30 percent of the node's information resources are updated in real time, for example, hydrometeorological observation data updated every 3 hours or forecast information updated twice a day. For their integration, special metadata are used that set the schedule and rules for processing of events and organizing the exchange of messages, metadata and data between the Data Providers and the Integration Server. The Node Service Bus plays a significant role in maintaining the node's Data Warehouses in real time. Metadata are an important element of data integration. Several consolidated metadata repositories are used. Examples are the catalog of information resources of the node in the ISO 19115 metadata standard, WMO Core Profile 1.3 and CKAN, available via the OAI-PMH protocols, OpenSearch, the OGC CSW geoservices catalog, the SOAP / RESTful API specifications for web services. The catalogs are available at http://core.intaros.meteo.ru:8081/geonetwork/ and http://opendata.intaros.meteo.ru.

Development prospects

Data integration of the Russian iAOS segment is based on a combination of consolidation (ETL technology) and data federation (EAI technology) of sources, as well as the widespread use of metadata and service bus tools. This makes it possible to perceive heterogeneous and physically distributed data as a single information space. It is planned to use the means and technologies of the node as a universal node of a distributed system in the context of the tasks of iAOS, ESIMO and the WMO Arctic Distributed Climate Center. The development of data integration tools is associated with the expansion of the use of EAI technology, as well as the study of the capabilities of the EII technology and its testing.

The Marine Data Exchange: the challenges and opportunities of standardising survey data for enhanced visualisations and insights

Chelsea Bradbury, The Crown Estate (UK), Chelsea.bradbury@thecrownestate.co.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 uniquely placed to look at the bigger picture across a breadth of offshore sectors. The Marine Data Exchange (MDE) is one tool used to address key consenting and planning challenges through better data sharing and encouraging evidence-based decisions.

Developed in 2013 as the first of its kind, the MDE manages offshore survey data collected throughout the lifetime of offshore wind farms, from feasibility studies through to decommissioning or repower. The Crown Estate work closely with offshore developers and operators to ensure that survey data held on the MDE meets the MEDIN Data standards and can be shared with the public as soon as it is reasonable to do so.

Throughout the years, the MDE data holding has developed significantly to having almost 200TB of survey data from over 50 offshore projects across England, Wales and Northern Ireland; over 2,000 survey campaigns and covering over 15 survey themes, from geophysical data to marine mammal surveys. In order to generate data-driven and value-adding insights from the variety of data collected at great expense by offshore developers, the next development phase for the MDE is to enable the standardisation and automation of data, to deliver trend analysis, meaningful insight and enhanced visualisations for a wide variety of stakeholders and users.

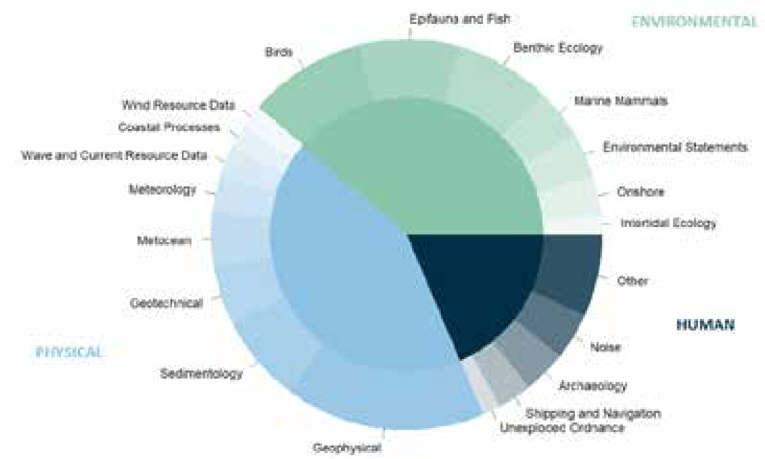


Figure 1: Survey data themes held on the Marine Data Exchange

Nevertheless, 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. This includes ensuring that the data held on the MDE meets the MEDIN Data Guideline standards and is defined by the MEDIN Metadata Discovery Standard; ensuring the discoverability and reusability of the data. However, in order to strike the right balance with the industry, there are no requirements in terms of the data format and so the MDE holds a variety of data types, collected and stored in slightly different ways.

Almost 60% of all data on the MDE is publicly available for anyone to freely download. Therefore, the MDE also disseminates data and information to a wide variety of users, including offshore developers and their consultants, academics, environmental NGOs, government bodies and the general public. Whilst the majority of users of the public website are visiting in order to find and download data, some users simply want to see what offshore wind companies are doing to measure impacts. For all of these user groups, standardisation can hugely benefit the visualisation of data; facilitating data discovery and educating many on not only the variety of data being collected by the offshore wind sector in the UK, but perhaps also unlocking the trends and insights held by the data.

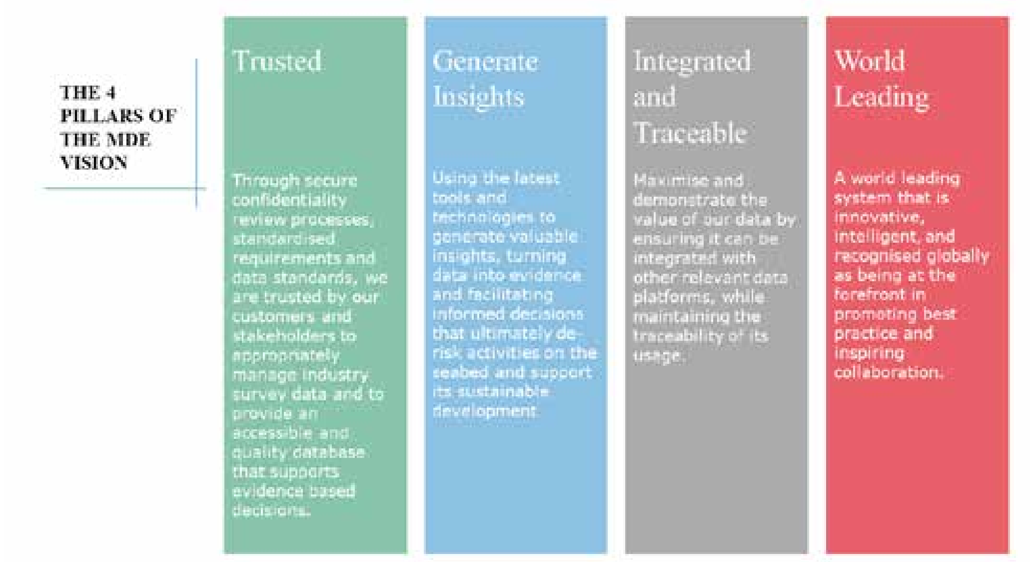


Figure 2: The Marine Data Exchange Vision

Survey data and evidence play a key role in the ongoing sustainable growth of the offshore wind industry. As a key tool in encouraging data sharing and evidence-based decisions, we are committed to improving the MDE so that it can better support the consenting and planning process. As we embark on an exciting journey to develop a new design-led MDE platform that supports our vision (figure 2), the ability to readily visualise, analyse and interpret the data will be critical. To ensure that the value of the data is realised, The Crown Estate aspire to standardise the variety of survey data that sits on the MDE using the latest tools and technologies, and the existing knowledge that sits within the marine data community.

As we embark on this journey, The Crown Estate would like to share some of the challenges, opportunities and lessons learnt.

Filling data gaps through interpolation: innovative analysis tools for oceanography

Charles Troupin, ULiège/GHER (Belgium), c.troupin@uliege.be
Aida Alvera-Azcárate, ULiège/GHER (Belgium), a.alvera@uliege.be
Alexander Barth, ULiège/GHER (Belgium), a.barth@uliege.be
Matjaž Ličer, National Institute of Biology (Slovenia), matjaz.licer@nib.si
Dimitry Van der Zande, Royal Belgian Institute of Natural Sciences (Belgium), dvanderzande@naturalsciences.be
Jean-Marie Beckers, ULiège/GHER (Belgium), jm.beckers@uliege.be

Spatial interpolation of in situ and remote-sensing observations has been performed for decades in oceanography, with many applications ranging from simple data visualisation to initialisation of numerical models. The large amount of data to process, the presence of physical boundaries (coastlines, land) and processes (advection, diffusion), combined with the variety of spatial and temporal scales all contribute to making the interpolation task not trivial. Furthermore, while satellite observations usually provide a better coverage than in situ data at the sea surface, sensors working in the optical and infrared bands are affected by clouds, which obscure part of the ocean underneath. We present here three open-source software tools that have in common to provide a gridded field as an output and to take into account the numerous specificity of the oceanographic data.

Tool	Meaning	Source code (language)
DIVAnd	Data-Interpolating Variational Analysis - n-dimensional	https://github.com/gher-ulg/DIVAnd.jl (Julia)
DINEOF	Data-Interpolating Empirical Orthogonal Functions	https://github.com/aida-alvera/DINEOF (Fortran)
DINCAE	Data-Interpolating Convolutional Auto-Encoder	https://github.com/gher-ulg/DINCAE (Python, Julia)

DIVAnd is an analysis tool designed for the generation of gridded data products from in situ observations. DIVAnd extends the 2D capabilities of the DIVA tool and allows the interpolation of observations on curvilinear, orthogonal grids in an arbitrary high dimensional space by minimizing a cost function. This cost function penalizes the deviation from the observations, the deviation from a first guess and abruptly varying fields based on a given correlation length (potentially varying in space and time). Physical constraints can be added to this cost function, such as an advection constraint, diffusion or source terms. One major advantage of the method is that it naturally decouples basins that are not connected and where water masses often have very different properties.

DIVAnd was rewritten from scratch using generalized mathematical formulation, in the programming language Julia which allows a high-level programming style but is compiled to machine code for better performance. There are also functions implemented to directly query online the following databases: World Ocean Database, Copernicus Marine Environment Monitoring Service (CMEMS) and EMODnet Physics. The bathymetries, needed to delimit the domains of interpolation can be extracted from EMODnet Bathymetry or from GEBCO.

Application: new practical application in the context of SeaDataCloud for surface currents acquired high-frequency radar data, EMODnet Physics data and EMODnet Biology data with initial test of DIVAnd coupled with a neural network will be given.

DINEOF

DINEOF is an EOF-based method to fill in missing data from geophysical fields, typically because of the presence of clouds. The method has been applied on a wide variety of parameters, including sea surface temperature (SST), sea surface salinity (SSS), chlorophyll concentration or suspended particulate matter (SPM). DINEOF also allows the analysis of 2 or more variables jointly through a multivariate approach, for instance SST and wind. The consistency between successive fields is enhanced by filtering the temporal covariance matrix. Recent applications have been dealing with very-high resolution images from Sentinel-2 missions. This gave rise to new issues such as the effect of the cloud shadows, which were not affecting the measured variables at lower resolution. An advanced algorithm has been designed to detect and remove the cloud shadows prior to the analysis with DINEOF. Another aspect under development is the analysis of observations of a common variable but from sensors with different spatial and temporal resolutions, for instance Sentinel-2, Sentinel-3 and SEVIRI. As an application, a reconstruction of SPM in the North Sea will be presented.

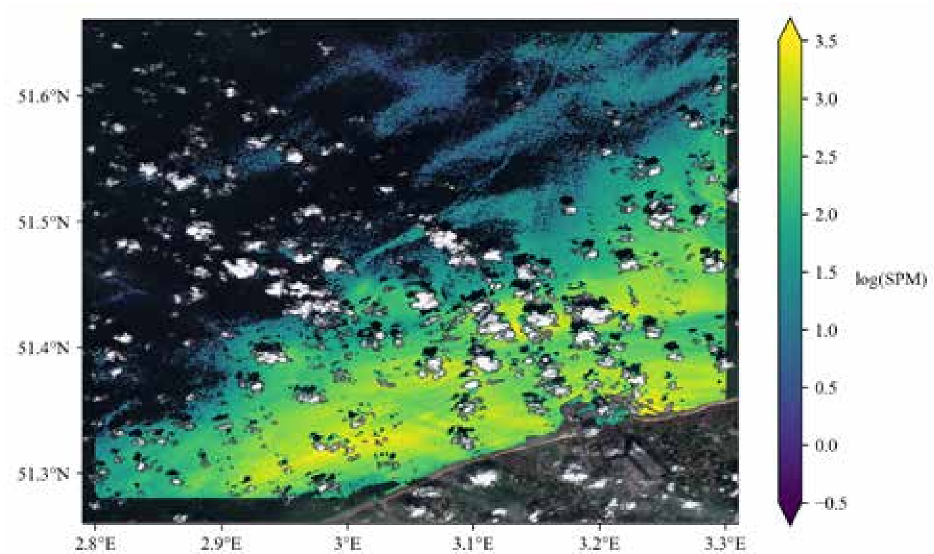


Figure 1: A typical Sentinel-2 SPM field, displaying small scale features and affected by clouds and their shadow.

DINCAE

DINCAE is a method to reconstruct missing data using a neural network. A neural network with the structure of a convolutional auto-encoder is developed to reconstruct the missing data based on the available cloud-free pixels in satellite images. Contrary to standard image reconstruction with neural networks, this application requires a method to handle missing data (or data with variable accuracy) in the training phase. DINCAE introduces a consistent approach which uses the satellite data and its expected error variance as input and provides the reconstructed field along with its expected error variance as output. The neural network is trained by maximizing the likelihood of the observed value. The approach is so far tested to 25-year time-series of Advanced Very High Resolution Radiometer (AVHRR) SST data and compared to DINEOF. The reconstruction error of both approaches is computed using cross-validation and in situ observations from the World Ocean Database. DINCAE results have lower error, while showing higher variability than the DINEOF reconstruction.

Web application and database for collecting and managing fisheries data

Damir Ivanković, Instit. of Oceanography and Fisheries (Croatia), ivankovic@izor.hr
Dalibor Jelavić, Instit. of Oceanography and Fisheries (Croatia), jelavic@izor.hr

Introduction

On the Institute of Oceanography and Fisheries web application based on relational database is developed for the purpose of managing and validation of fisheries data. Application also provide outputs needed for preparing various reports. Application manage data from different fishing gear / fishing method (bottom trawl, cast net, drift net, etc.) and for commercial and non commercial species. Interface have fields for filtering data on the left side, and it is organized hierarchically with tabs. Tabs are: metadata about fishing including coordinates, weight distribution, length distribution laboratory analysis, map and length-weight distribution (Figure 1). Data are inserted directly from web browser with build in semiautomatic validation.

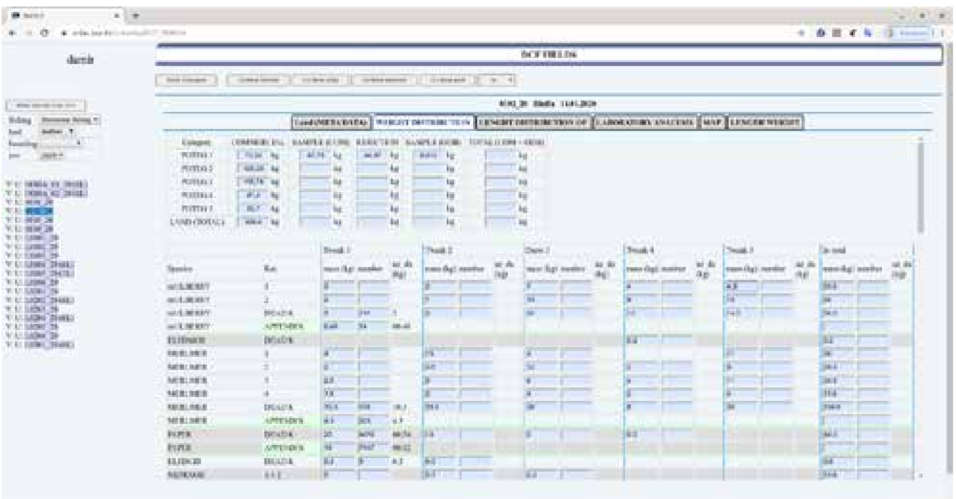


Figure 1: Application interface with active weight distribution tab

Development

Excel spreadsheets where used before usage of relational database for storing fisheries data. First version of application has multiple stage loading process:

- exporting multiple worksheet excel spreadsheet into one csv file using macro
- parsing csv file and loading into temporary structure
- correcting structural (depedencies) errors like wrong name of ship or species
- transponing data into permanent and related database tables

This process where used to insert data from already inserted excel files. For normal data insertion in second stage of application development, direct web interface was developed. Database structure was also improved to can manage different fishing methods and tools data. Catch is generally divided into categories and to commercial and non commercial catch. Side catch is also managed.

Results

As output results from application and database various outputs were created (Figure 2). Data quality was increased by two major methods. Direct input have consistency checks and limitations (no free text for species or ship) with automatic sums checks (total weight of sample vs. sum of weight distribution). Also all outputs were available at any time (not need to collect all data and then perform analysis) what in case of spotted irregularity makes space for organizing additional data collection for some area or time period. Instant visualisations showing length distribution and length-weight helps in the identifying and correction of input errors. Database is used for data management of national fisheries monitoring and for European Commission - Data Collection Framework reporting. Outputs are prepared according reporting demands in tabular form (downloadable in csv format) and also in graphical forms.

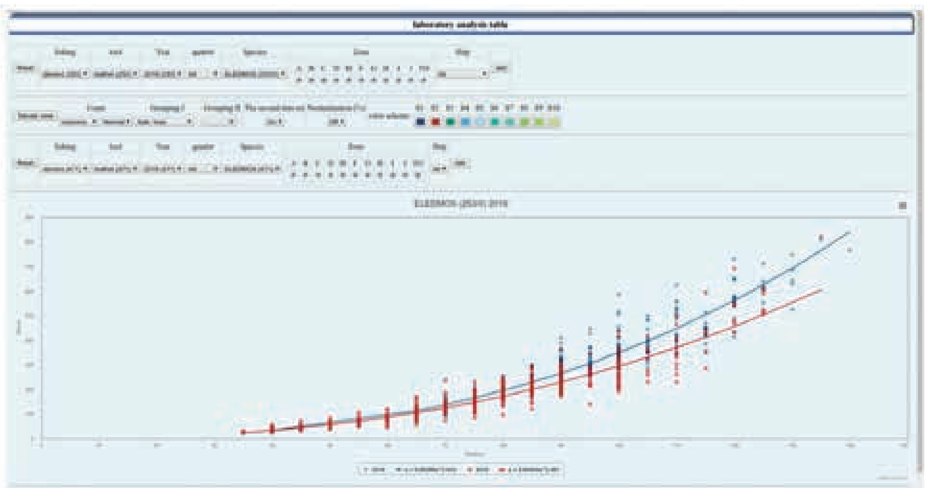


Figure 2: comparation of two yearly data sets and length-weight relationship using in the standard allometric equation W=aL^b (y=ax^b)

Conclusion

Biological data are often hardest for transposing into some predefined structure. Tight collaboration between fisheries expert, data manager and IT developer is needed. Database procedures manages methodology for calculation of estimated total stock based on avaliable measured samples. As result we have improved data quality and semi automatically prepared data for reporting. Visualisation of some trends and fish distribution (Figure 3) make possible for fish expert to better plan and manage such valuable resource as fish stock.

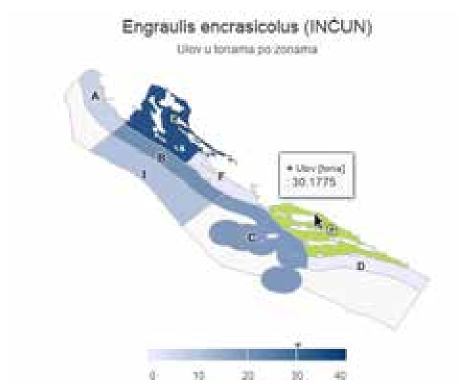


Figure 3: distribution of species within zones

Citizen science and crowdsourcing in the field of marine scientific research – the MaDCrow project

Paolo Diviacco, OGS National Institute of Oceanography and Applied Geophysics (Italy), pdiviacco@inogs.it
Antonio Nadali, Transpobank (Italy), nadali@tbk.it
Massimiliano Iurcev, OGS National Institute of Oceanography and Applied Geophysics (Italy), miurcev@inogs.it
Rodrigo Carbajales, OGS National Institute of Oceanography and Applied Geophysics (Italy), rcarbajales@inogs.it
Francesca Malfatti, OGS National Institute of Oceanography and Applied Geophysics (Italy), fmalfatti@inogs.it
Lorenzo Grio, Transpobank (Italy), grio@tbk.it
Alessandro Busato, OGS National Institute of Oceanography and Applied Geophysics (Italy), abusato@inogs.it
Alessandro Pavan, OGS National Institute of Oceanography and Applied Geophysics (Italy), apavan@inogs.it
Massimiliano Nolich, University of Trieste (Italy), mnolich@units.it

Introduction

MaDCrow (Marine Data Crowdsourcing) is a marine technology research and development project co-funded by the European Regional Development Fund (ERDF), that aims to create an innovative technological infrastructure for the acquisition, integration and dissemination of data on the marine ecosystem of the Gulf of Trieste based on a citizen science paradigm.

Citizen science is the process whereby citizens are involved in science as researchers (Kruger & Shannon, 2000) and in this perspective one very important issue is the participatory attitude of volunteers (Conrad & Hilchey, 2011).

The project aims, at the same time, to develop the tools to allow such perspective to take place and to increase public awareness of environmental issues and in particular of climate changes as drawn within goal 13.1 of the UN Sustainable Development Goals, “Improve education, awareness-raising and human and institutional capacity on climate change mitigation, adaptation, impact reduction and early warning”.

The activities of the project are linked also to the priorities of the POR (Programma Operativo Regionale), within its target «Investments towards a growth in employment» 2014-2020 Axis 1 – “To strengthen research and development growth”, Activity 1.3 – “To support R&D collaborative activities for the development of new sustainable technologies, new products and services”.

In this perspective an aspect that is also particularly important within MaDCrow is to support open innovation between public institutions and private companies, basing it on environmental awareness and blue economy. In this perspective MadCrow aimed also at creating new competences in the field of marine technologies for already existing small and medium enterprises (SME).

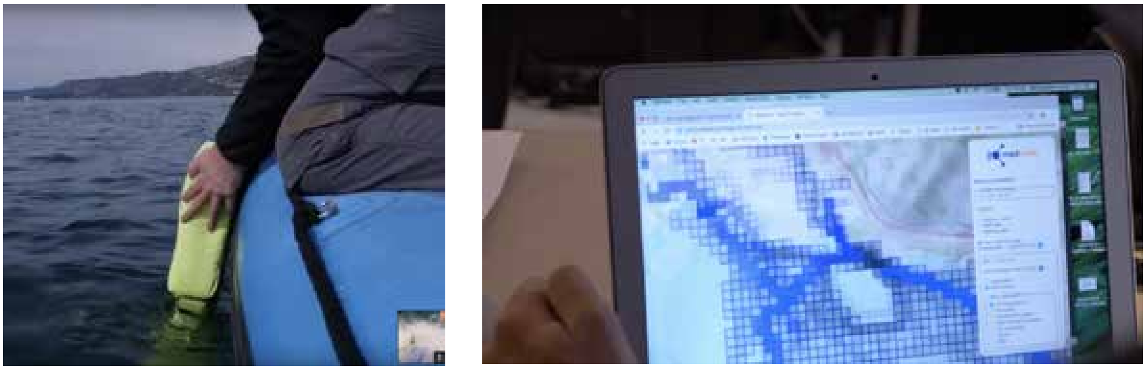


Figure 1: Marine data acquisition and transmission allows near real time data visualization on the MadCrow platform

The system

The MadCrow project developed all the tools that cover all the phases from acquisition, to processing to data visualization and access. A specific sensing and transmission box has been developed with innovative materials that can be easily deployed on small private boats (be it a recreation boat, a fisherman ship or else) without interfering with the normal activities of the vessel. The sensing box contains pH, temperature, dissolved Oxygen and salinity sensors. Others, such as for example turbidity will be soon included. The acquisition system is based on Arduino system that takes care also of GPS data geo-referencing and time-stamping. Both the sensor box and the acquisition system are packed in a specifically developed bag that is deployed attaching it to the hull of a ship through adjustable straps. There, the sensors are directly in contact with the sea water while the electronics are closed within a waterproof case. The sensors can be placed approximately at depths ranging from 50 cm to 1 meter, but of course the movements of the ship can change this and can sometimes even bring the sensors outside the seawater. The system therefore has limitations in its use in relation to sea conditions. It must be said also that volunteers hosting the system probably will not go at sea in conditions that are not compatible with the acquisition system. Anomalous recordings that come from these types of problems are common and suggest the need of a careful phase of quality control on the raw data. In addition, precision and accuracy of sensors worsen when vessels travel at velocities larger than 9-10 knots. A first step in data quality control is therefore checking the vessel position and speed. Issue can be due to errors in the GPS or the actual speed while both can be handled essentially in the same way: when positioning show velocities that exceed the acceptable speed, these measurements are automatically removed. A second quality check consist in controlling the range of measurement, when this show values that are outside the expected range they are automatically removed. Further quality check steps are performed superimposing a grid that allows to bin values pertaining to the same cell. The cells have dimensions of 200x200 metres per one hour. Measurements corresponding to the same cell are statistically averaged and the output is assigned to the cell as the representative value. Of course, statistics works best if there is a sufficient number of points in each cell. This is often not the case since depends very much on how many platforms acquire data at the same time; an issue that is intrinsic in any citizen-science based project. Further validation is under development, in order to check data upon what is available using traditional method. After validation, data are made available, almost in real-time, through standard OGC web services (WFS, WMS) and through a dedicated portal that allows to map data geographically and in time.

We are currently working on the integration of MadCrow with other data sharing initiatives such as for example SeaDataNet or EMODnet starting from the standardization of web services that have been implemented already in the project.

First tests at sea

MadCrow was aiming mainly at developing a prototype (TRL 6), while the outcomes obtained went further, resulting in a fully operative infrastructure. This was tested in several test surveys, with 3 platforms simultaneously and showed to be robust and resilient even in complicated sea conditions. We tested the system on different types of boats, with different success; smaller boats performed better that larger ones which suggested that also other methods of deployment should be considered. Although MadCrow was mainly targeting at a relatively small area such as the Gulf of Trieste (Northern Adriatic Sea) the infrastructure can be easily scaled to larger and different areas.

Volunteers are provided at no cost with the acquisition system while the rest of the infrastructure is maintained by the Italian National Institute of Oceanography and Applied Geophysics – OGS and the private company Transpobank.

Being a project aiming at citizen-science, after having developed the technologies behind the infrastructure, we will now tackle the issue of building a fleet of volunteers. In this, two paths will be followed and namely (I) a commercial approach: where some additional serviced will be made available only to specific end users, and (II) a completely voluntary based approach, where ethical rewarding mechanisms will be tested. All raw data will be made available to the scientific community.

Conclusions

Madcrow demonstrated non only the feasibility but also the vast possibilities that the citizen science approach can have from the perspective of providing researchers with large quantities of data covering large geographical areas with very low costs.

MadCrow web site: http://www.madcrow.it

References

Conrad, C. and Hilchey, K. (2011) A review of citizen science and community-based environmental monitoring: issues and opportunities. Environ. Monit. Assess. 176, 273–291

Kruger, L. E., & Shannon, M. A. (2000). Getting to know ourselves and our places through participation in civic social assessment.Society and Natural Resources, 13,461–478.

Improving connectivity with distributed partners in the Australian Ocean Data Network

Sebastien Mancini, Integrated Marine Observing System, University of Tasmania (Australia), sebastien.mancini@utas.edu.au
Mark Rehbein, The Australian Institute of Marine Science (Australia), m.rehbein@aims.gov.au
Dave Watts, CSIRO Oceans and Atmosphere (Australia), dave.watts@csiro.au
Johnathan Kool, The Australian Antarctic Data Centre (Australia), johnathan.kool@awe.gov.au
Andrew Carroll, Geoscience Australia (Australia), Andrew.carroll@ga.gov.au
Miles Nicholls, The Atlas of Living Australia (Australia), miles.nicholls@csiro.au

The Australian Ocean Data Network (AODN) is an interoperable online network of marine and climate data resources. It is a collaboration between six major Australian Commonwealth agencies, the Integrated Marine Observing System (IMOS), and a growing list of universities, state government offices and others organisations in Australia, New-Zealand and the Pacific.

AODN data collections cover a large geographic area (from coast to open ocean, from equator to Antarctica), a wide range of observed parameters (physical, chemical, biological) and are obtained using a variety of platforms and technologies (e.g. ships, autonomous floats and vehicles, gliders, moorings, satellites, animal tags, coastal radar). The end users include researchers, students, managers, policy makers, consultants, sailors and fishers. All data products contributed to the AODN are made freely and openly available to the public via the AODN Portal (https://portal.aodn.org.au/).

In 2016, the AODN Portal underwent major improvements in order to focus its functionality on searching dataset collections through its existing metadata catalogue, visualising data on a map using Web Map Service (WMS) and providing access to data through multiple download services in a single user interface. At the launch of the revamped AODN Portal, almost all dataset collections available to users were collected by the IMOS program.

Soon after, IMAS (Institute for Marine and Antarctic Studies) and NIWA (National Institute of Water and Atmospheric Research) became the two major external contributors of dataset collections to the AODN Portal. This early success was explained by their choices of infrastructure. They implemented a complete copy of the AODN software stack making the integration much simpler.

More recently, the Marine Research Data Cloud project, funded by the Australian Research Data Commons (ARDC), enabled the AODN team to establish infrastructure connections and data integration with the following major oceanographic data providers in Australia:

- The Australian Insitute of Marine Science (AIMS)
- Geoscience Australia (GA)
- The Australian Antarctic Data Centre (AADC)
- The Atlas of Living Australia (ALA)
- The CSIRO Oceans and Atmosphere Data centre

The new dataset collections made available on the AODN Portal are key datasets with existing public access and interest.

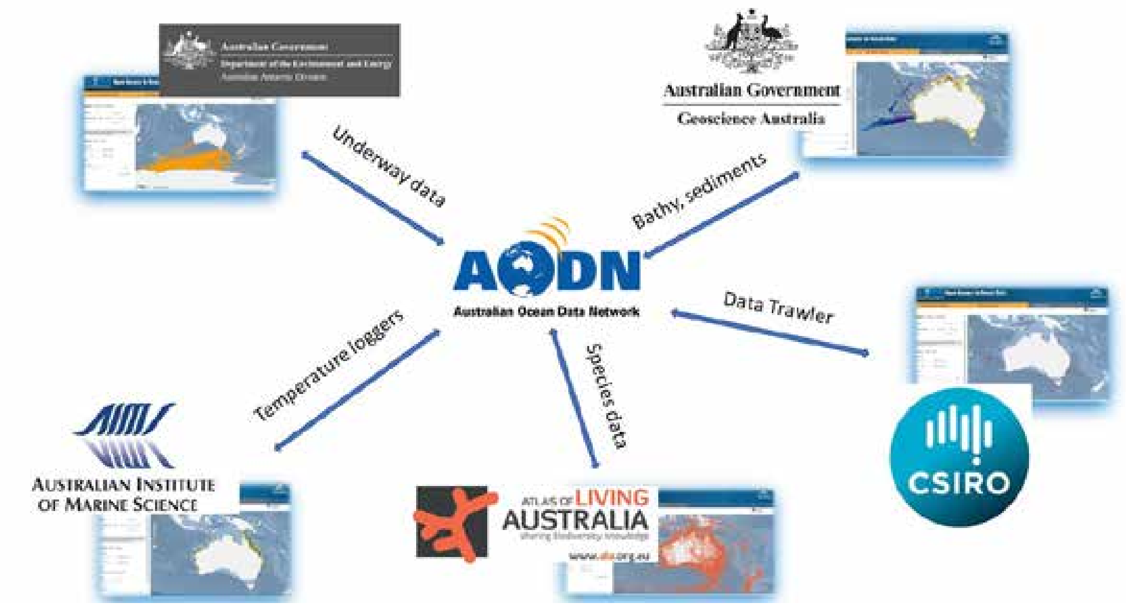


Figure 1: an overview of the new dataset collections published on the AODN Portal

These five organisations have been targeted to demonstrate that the integration of data provided through multiple types of infrastructure is possible. Each organisation currently uses different tools to publish metadata records, Web Map Services and download services (e.g. Web Feature Service or Web Processing Service). Upgrades to the existing infrastructure of each organisation and new developments to the AODN Portal were required to allow the AODN Portal to connect to existing web services and integrate new dataset collections.

The AODN Portal now provides more than 250 dataset collections to end users. Half of these dataset collections represents data collected by IMOS while the other half is made available by the different partners of the network.

Wider adoption of published controlled vocabularies, harmonisation of data services provided by partners, improved workflows to configure and test new dataset collections pre-publication and securing on-going funding to expand on these initiatives are some of the key challenges that need to be overcome in order to expand the marine data resources made available on the Australian Ocean Data Network (AODN).

The demonstrated data integration and improved connectivity with its distributed partner dataset collections provides a proof of concept of the truly distributed data network the AODN has aspired to for so long.

Pilot for accessing distributed marine data using the OneData concept

Peter Thijsse, MARIS, Nootdorp (The Netherlands), peter@maris.nl
Dick Schaap, MARIS, Nootdorp, (The Netherlands), dick@maris.nl
Gergely Sipos, EGI Foundation, Amsterdam (The Netherlands), gergely.sipos@egi.eu

There are many challenges, also in the SeaDataNet community, in the domain of access to distributed data. A large portion of the challenges is related to the various aspects of data storage management. For example, having access to files locally available, synchronous file transfers, proprietary log formats, and log locations. OneData (adopted by EGI) offers many advanced features related to data storage management. Therefore, MARIS has spent effort as part of the EOSC-HUB project on working out several use cases with the OneData concept. As users two different groups can be distinguished: Developers facing the challenge with more and more, and larger datafiles. End-users benefitting from the faster response times and more direct access to large data collections.

About Onedata

With Onedata, users can access, store, process and publish data using a global data storage backend provided by computing centers and storage providers worldwide. Onedata focuses on instant, transparent access to distributed data sets, without unnecessary staging and migration, allowing access to the data directly from your local computer or worker node. The most important concepts of the Onedata platform are:

- Spaces - distributed virtual volumes, where users can organize their data
- Providers - entities who support user spaces with actual storage resources exposed via Oneprovider services
- Zones - federations of providers, which enable creation of closed or interconnected communities, managed by Onezone services.

More background information of Onedata can be found at www.onedata.org. For the pilot these concepts have been used, configured and investigated.

SeaDataNet practice

In the SeaDataNet practice (just like in other infrastructures) developers are confronted many times with situations that data sets are stored at different locations while we want to undertake central processing. For instance, there is great interest in so-called BioGeoChemical (BGC) data sets as these provide input for determining indicators about the quality of marine waters and as such are very relevant for the Marine Strategy Framework Directive of the EU which aims at establishing Good Environmental Status (GES). Through its engagement with EMODnet Chemistry, SeaDataNet is actively supporting Regional Sea Conventions, EU DG Environment, and European Environment Agency (EEA) in compiling and providing harmonised and validated data collections for eutrophication and contaminants which are derived from the BGC data as gathered by the SeaDataNet data centres. Moreover, SeaDataNet has established cooperations with Copernicus CMEMS as well as with Euro-Argo to work together on mutual data exchanges and on improving and innovating quality control and processing of large BGC data collections for various purposes, including MSFD. Access to the data files, as well as controlling quality and processing the distributed datasets, currently have performance issues and existing solutions might have to be replaced at some point with new concepts.

Pilot content

In order to investigate the potential of Onedata solutions for SeaDataNet purposes, during 2020 a test configuration has been set up using OneData in combination with Cassandra and Elasticsearch. OneData has been configured to give access to a Onedata “Space” to data of a number of data providers on the cloud, each provided with BGC data collections in the SeaDataNet ODV format. Data is collected via a OneData “Zone”, cached and stored in a Cassandra open source NoSQL database with wide column store, which allows high searching performance on large data sets with many numbers. Elasticsearch has been configured on top in order to optimize free text search on the metadata of the data sets to facilitate fast and precise subsetting of data collections from the master collection.

During the session we will present the conclusions with key insights, examples how this could be used and user perspectives, as gained during development.

Web service for storing and processing sea water data measured in situ concurrently with satellite survey

Nikita Kniazev, Space Research Institute of RAS (Russia), nkkniazev@gmail.com
Dmitriy Elizarov, Space Research Institute of RAS (Russia), dimaelik71@gmail.com
Olga Lavrova, Space Research Institute of RAS (Russia), olavrova@iki.rssi.ru
Ksenia Nazirova, Space Research Institute of RAS (Russia), knazirova@gmail.com
Evgeny Krayushkin, Space Research Institute of RAS (Russia), box_evk@mail.ru

The report discusses the Web service developed by Space Research Institute RAS (IKI RAS) for storing and processing sea water parameters measured *in situ* concurrently with satellite survey. Every year, the IKI RAS team conducts field measurements in the coastal zones of the Black, Baltic and Azov Seas focusing on retrieving information on the 3D structure of hydrophysical processes reflected in quasi-synchronous satellite data. The main processes of interest are: coastal currents, submesoscale eddies, internal waves, river and lagoon plumes.

The measurements are made from small boats using a set of oceanographic instruments, of which the key ones are an RBR-concerto CTD probe equipped with turbidity and chlorophyll *a* concentration (CHL-a) meters and an Acoustic Doppler Current Profiler (ADCP). The CTD sampling rate is 6 Hz, one down-and-up cycle can yield a dataset of over 1500 measurements. As a rule, 30-40 stations are made on ship route. For preliminary data processing, a software module is created implementing processing and sorting of the obtained dataset according to user-defined parameters. The output file contains the list of stations (measurement points) with coordinates and the array of filtered downcast and upcast data. All the data, preliminary and processed, are uploaded to the web service for further processing.

The web service backend is written in the Laravel framework. The main components of the system are presented in Fig. 1. User access to the system tools is realized via a graphics interface developed in the VueJS framework. The interface is very simple and easy to use.

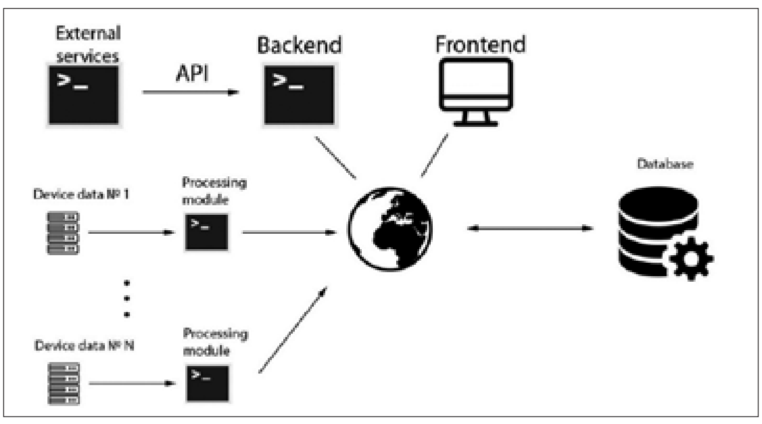


Figure 1. System architecture

As exemplified by CTD data, the database table contains values of station number and coordinates and measured depth, temperature, salinity, sound velocity, turbidity and CHL-a. During a day of work, over 300 thousand values are accumulated in the table. Similar tables are compiled for ADCP data. For individual expeditions, a separate table is created with numbers and dates of specific measurements as well as an array of local archives that is associated with particular expedition. In the system, on the basis of the uploaded data, profiles of physical parameters at the stations are drawn taking into account changes in depth and route distance. Examples of such graphs are shown in Fig. 2.

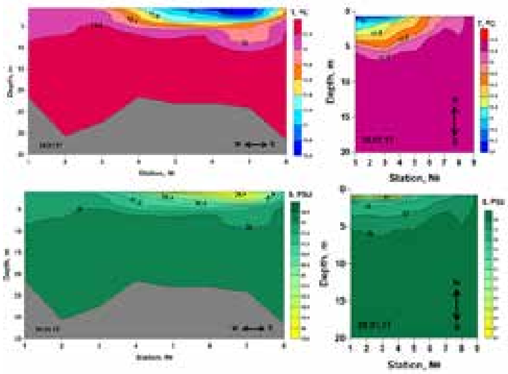


Figure 2. CTD data from one day of measurement

Today, data processing programs are written and incorporated into the system. The basis is the algorithm for filtering and sorting CTD data recorded in Excel format (.xlsx). This format was chosen because of the oceanographic equipment software that exports the raw data in a tabular Excel format (Ruskin, WinRiver). The algorithm for creating new files with processed data and station coordinates (by default absent in CTD logs) is written in Python using OpenPyxl and XlsxWriter third-party libraries. Primary data processing includes removing unnecessary columns, searching for the beginning and end of stations without measurements, which are responsible for calibrating the probe, and averaging the data over the selected parameters. The processed data is stored in the service database, where it is used for subsequent visualization of the results. For this, a cartographic interface displaying selected stations is added to the web service. It is possible to perform joint analysis of measurement data and shipboard meteorological data (Fig. 3).

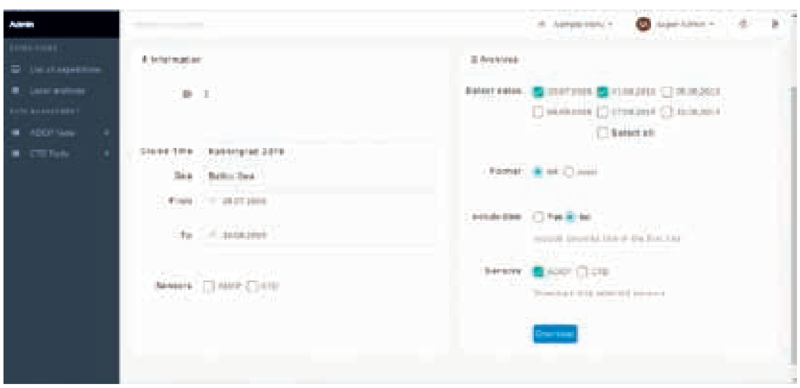


Figure 3: Service interface. Data selection

The joint use of in-situ measurements and remote sensing data in existing scientific works is used as an auxiliary method for various qualitative descriptive work.

That's why the principal upgrade direction of the web service is the development of an interface with the See the Sea information system (STS). Created and maintained by IKI RAS, STS is intended for studying processes and phenomena in oceans and seas using all available satellite data (Loupian *et all*., 2018 ; Lavrova *et all*. 2019). The data stored in the Web service should flawlessly be supplied to STS for display in its cartographic interface and joint analysis with satellite data.

Nowadays, the web service is used by users of the IKI RAS, but soon the project is planned to be published for the general public, where any scientist, which engage in oceanological research, can store and processing primary observation data in prepared instrument.

The development of the See the Sea information system is supported by theme "Monitoring", registration no. 01.20.0.2.00164. The information system is functioning with use of the data from the "IKI Monitoring" Center for Collective Use.

References

Lavrova O.Yu., Mityagina M.I., Uvarov I.A., Loupian E.A. (2019) ***Current capabilities and experience of using the See the Sea information system for studying and monitoring phenomena and processes on the sea surface***. Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa, 16(3), 266-287.

Loupian E., Lavrova, O. Kashnizky A., Uvarov I. (2018). ***"See The Sea" - new opportunities for distributed collaboration aimed at solution of oceanographic problems using remote sensing***. Bollettino di Geofisica Teorica ed Applicata, 2018, Vol. 59, Supl. 1, pp. 91-93.

UTM-CSIC Data Service Architecture: from acquisition on-board to final dissemination

Juan Luis Ruiz, UTM-CSIC (Spain), jlruiz@utm.csic.es
Susana Diez, UTM-CSIC (Spain), sdiez@utm.csic.es
Xavier Rayo, UTM-CSIC (Spain), xrayo@utm.csic.es
Guillermo Garriga, UTM-CSIC (Spain), ggarriga@utm.csic.es

Overview

The Marine Technology Unit of the Spanish Research Council (UTM-CSIC) is responsible for the management of oceanographic research vessels. Within the UTM, the Data Service is in charge of managing spatial data acquired onboard oceanographic cruises and make these data as interoperable as possible in order to disseminate them so that they can be used and reused.

To achieve this goal an information architecture has been developed to control all the processes of data workflow described below. The data model can be consulted under http://data.utm.csic.es/lib/datamodel/

1. Data acquisition and events registration on-board

Underway data acquired by different vessel systems and sensors (position, heading, vessel speed, depth, weather, water temperature, salinity and other water physical parameters) are sent to the Continuous Oceanographic Data Acquisition system (CODA) which integrates them. The result is sent in real time to the next phase of the system, the Data Hub, and also in non-real time so that there is no loss of data in case there is a problem with the real time.

The Eurofleets Automatic Reporting System software (EARS) allows scientists and technicians working on research vessels to register events that take place during an oceanographic cruise, key aspect for the generation of metadata. Theses events are merged with some of the data acquired in CODA and sent in non-real time to the Data Hub.

2. Data Hub and web services on-shore

On-shore UTM Data Hub concentrates data and metadata of CSIC fleet through satellite connections using different communication protocols in real time (UDP) and non-real time (email, rsync ...) with the aim of feeding dissemination services and tools. In addition, the hub harvests data and metadata from other centers in order to be standardized and disseminated to portals of international projects such as Eurofleets or Seadatanet.

This information is stored in files offered through two types of web services, feeding end user applications. Those related to metadata, such as *getCruise* to retrieve cruise information and *getEvent* to retrieve events registered using EARS, and those involved in the management of time series of continuous acquisition parameters, such as *getLast* with the last registered datagram, *getSerie* that offers all the records between two dates and *getPos* that offers the position for a given time.

3. UTM-CSIC Data Portal. Data dissemination I

The UTM-CSIC Data Portal gathers a series of tools and services for data and metadata dissemination:

Metadata catalogue: Catalogue of Oceanographic Cruises (with more than 600 cruises) and related datasets, carried out since 1991 on research vessels managed by CSIC. The basic metadata of each oceanographic cruise is completed with documentation (cruise plan and report) and online resources such as DOI citation, cruise map and plot and download of underway data.

Geoportal: Web portal with an interactive display that allow users to navigate through a map to search and query datasets. Users are allowed to add their own data and connect to external WMS services in order to create their own maps. The query and search results are also linked to the metadata catalogue.

Quality Control Panel and Underway Plot: Web application to represent graphically the underway datasets acquired during Oceanographic cruises. Data graphics and maps are merged interactively and the user can navigate using the map or the graphics. Data quality control can be done from this application from both our served data or from local file from users.

Real time panel: Panel showing real time underway data acquired by different vessel systems and sensors managed via CODA (position, heading, vessel speed, depth, weather, water temperature, salinity and other water physical parameters).

SOS Client: Similar to the underway plot, in this case the SOS client shows historical series from underway data following the standards of SOS services.

Data Download: All raw data acquired during oceanographic cruises are available for downloading through the Data Download portal by registration, according to the privilege of each user. In case of restricted data, a formal request is required and cases are studied independently according to the current convention.

4. European Data portals. Data dissemination II

The UTM is a National Oceanographic Data Center (NODC) of the Distributed European Sea Data Infrastructure ***SeaDataNet*** (SDN). *Cruise Summary Reports (CSR*) are provided through a dedicated catalogue to be harvested by SDN and *Common Data Index (CDI*) through the Replication Manager and Import Manager, both developed in SeaDataCloud H2020 project. The related unrestricted data are available in the cloud to be downloaded through the SDN Portal.

During Transnational Access **Eurofleets** cruises on board the RV Sarmiento de Gamboa the Data Service provides the "European Virtual Infrastructure in Ocean Research" Eurofleets portal (EVIOR), with underway data and events information, sailing tracks and current position and metadata.

Scalable and high performance infrastructure for ocean data discovery and visualization

Glenn Judeau, IFREMER (France), Glenn.Judeau@ifremer.fr
Jérôme Detoc, IFREMER (France), Jerome.Detoc@ifremer.fr
Charlie Andre, IFREMER (France), Charlie.Andre@ifremer.fr
Léo Bruvry-Lagadec, IFREMER (France), Leo.Bruvry.Lagadec@ifremer.fr

This contribution presents a new data management and processing system that can handle massive amounts of oceanographic data as, for instance, held in the Coriolis database. This is achieved by combining established and widely used components in a clever way. Users will also benefit from a wide range of visualization styles. The Coriolis "in-situ" dataset is historically stored in Oracle and represents terabytes of data. While the dataset grows, reaching billions of measures, Oracle has shown limitations to address innovative use-cases.

IFREMER has built a Big Data solution to face modern challenges, ensuring sustainability of the dataset in the future and responding to both use-case workloads: Online Analytical Processing (OLAP) and Online Transaction Processing (OLTP). Those use case include interactive and complex metadata search-engine, sub-second data plotting, robust and high performance sub-setting and innovative Copernicus diffusion with large NetCDF4 files.

The solution implements three innovative data engines *(fig. 1)*:

- Spark, a scalable distributed processing engine
- Cassandra, a scalable high performance storage and access engine for data
- Elasticsearch, a scalable high performance metadata storage and powerful search engine

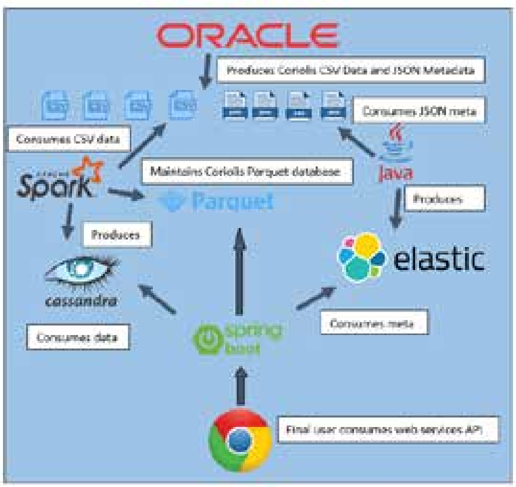
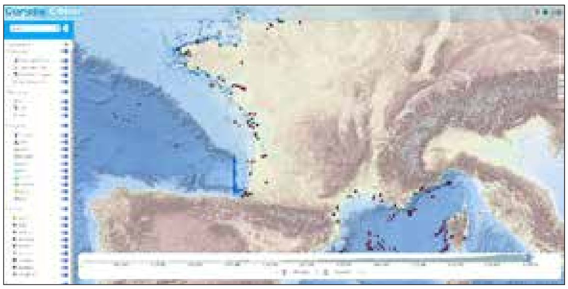


Figure 1: Infrastructure principles

Each data engines deserve its own purpose to give best performances for the user's needs. In order to provide services to different endpoints, *IFREMER* set up Java REST APIs for each:

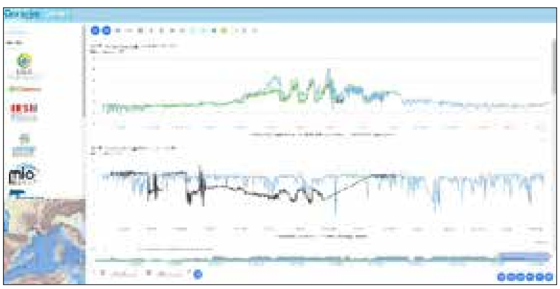
- Data-Discovery API to research using provided facets through Elasticsearch, returning JSON formatted metadata.
- Data-Plot API to get data-plots for profiles, time series, trajectories. Returning JSON format. Because returning thousands or millions data for visualization is time consuming and not relevant most of the time, a down-sampling algorithm has been implemented to the API: Largest Triangle Three Buckets (LTTB).
- Sub-setting API to download data in CSV or NetCDF4 format. User can be notified by email and data is pushed on *IFREMER*'s shared storage.

The use of API eases the ability to call services through different web portal for different purposes *(fig. 2)*:

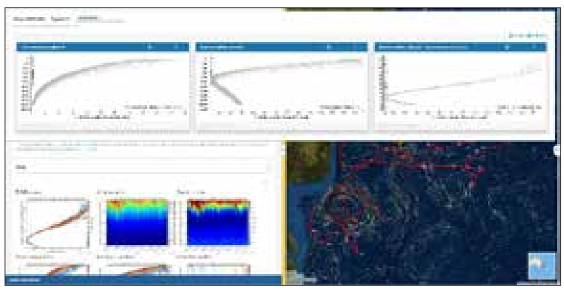


Coriolis-Côtier (data research)

Argo fleet monitoring (data research)



Coriolis-Côtier (data visualization)



Argo fleet monitoring (data visualization)

Figure 2: https://data.coriolis-cotier.org/and https://fleetmonitoring.euro-argo.eu/

MORSE, Marine Organisms and Resources Storage system, towards a global system to ensure the traceability of all Ifremer biological samples

Jerome Detoc, Ifremer (France), jerome.detoc@ifremer.fr
Sylvie Van Iseghem, Ifremer (France), Sylvie.Van.Iseghem@ifremer.fr
Florian Nivert, Ifremer (France), Florian.Nivert@ifremer.fr
Glenn Judeau, Ifremer (France), Glenn.Judeau@ifremer.fr
Catherine Borremans, Ifremer (France), Catherine.Borremans@ifremer.fr

Ifremer has started to implement the information system MORSE (Marine Organisms and Resources Storage system). It aims to ensure the traceability of all Ifremer biological samples as required by the international regulations APA (*Access and Benefit Sharing)* and CITES (*Convention on International Trade in Endangered Species of Wild Fauna and Flora*).

MORSE provides an identity card of each Ifremer biological sample which is also required to serve the scientific objectives of valorisation and publication (Figure 1). The information stored for each sample cover the entire sample life: acquisition metadata, sample type and processing, storage information, responsible(s), identification, administrative data.

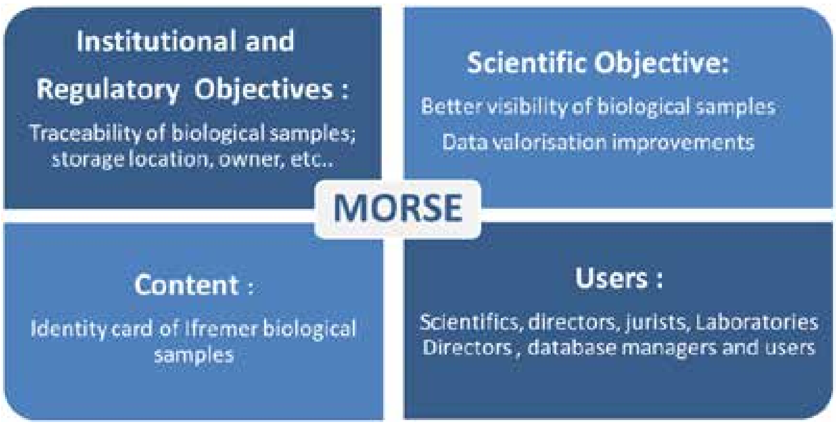


Figure 1: MORSE principles

Data localisation:

IFREMER owns within all its laboratories thousands of biological samples. They are referenced in many ways:

- Locally with dedicated software as LabCollector or Excel
- Locally and manually, without dedicated software
- Centrally, integrated in various and specific databases and information systems.

The information systems managed by SISMER (the French NODC hosted in Ifremer) and concerned by MORSE deal with deep-sea environment data and geoscience data (BIGOOD), fisheries data (Harmonie) and coastal environment data (Quadrige 2).

MORSE functionalities and structure:

MORSE is a unique tool that will answer the following needs:

- Aggregation of biological sample information
- Interoperability with standardised referential
- Visualization of the required information on biological samples (including physical storage, sampling)
- Description of the sample filiation, duplication, modification, etc.
- Unique internal code generating an unique label
- Labels generation before, during and after samples collection
- Valorisation of biological samples using DOI.

In addition to scientific data management systems, Morse interoperates with administrative systems (Corail, SAP, SGC) and data portals (Seanoe, Sextant). Technically, **MORSE** is composed by

- A dynamic and contextual dashboard built on the Angular 8 framework to access directly to the desired information.
- A micro-service to display/print on demand the QR Code and metadata labels adapted to the sample container dimension.
- A micro-service to direct, for each sample or group of samples, to the associated DOI landing page.
- An administration dashboard built on the Angular 8 framework to control the interfaces with thematic information systems, with administrative systems and with common referentials.

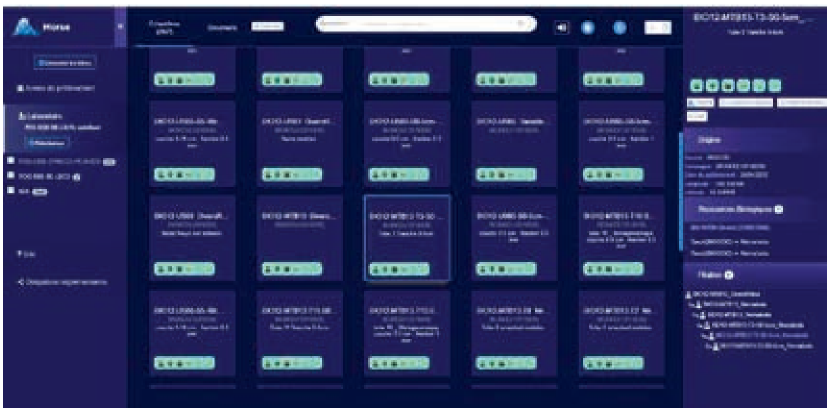


Figure 2: Screenshot of the user dashboard

Remaining questions and challenges

To get the MORSE system fully operational, many challenges remain. Crucial points will be to link this Ifremer System to European referential and standards, DISSCO for example, to attribute DOI to biological samples and last but not least to involve users with a wide variety of profiles.

Data To Product Thematic Services Integration into e-JERICO

Miguel Charcos, SOCIB (Spain), mcharcos@socib.es
Francoise Pearlman, IEEE (France), jsp@sprintmail.com
Jay Pearlman, IEEE (France), jay.pearlman@fourbridges.org
Juan Gabriel Fernández, SOCIB (Spain), jfernandez@socib.es
Lorenzo Corgnati, CNR (Italy), lorenzo.corgnati@sp.ismar.cnr.it
Lauri Laakso, FMI (Finland) and North-West University (South Africa), Lauri.Laakso@fmi.fi
Milla Johansson, FMI (Finland) and North-West University (South Africa), milla.johansson@fmi.fi
Taavi Liblik, TalTech (Estonia), taavi.liblik@taltech.ee
Julien Mader, AZTI (Spain), jmader@azti.es
Carlo Mantovani, CNR (Italy), carlo.mantovani@cnr.it
Simone Marini, CNR (Italy), simone.marini@sp.ismar.cnr.it
Marc Picheral, LOV (France), marc.picheral@obs-vlfr.fr
Emma Reyes, SOCIB (Spain), ereyes@socib.es
Anna Rubio, AZTI (Spain), arubio@azti.es
Inmaculada Ruiz, SOCIB (Spain), iruiz@socib.es
Jukka Seppala, SYKE (Finland), jukka.seppala@ymparisto.fi
Lars Stemmann, LOV (France), stemmann@obs-vlfr.fr
Timo Tamminen, SYKE (Finland), timo.tamminen@ymparisto.fi
Joaquín Tintoré, SOCIB (Spain) and IMEDEA (Spain), jtintore@socib.es

Introduction

The Joint European Research Infrastructure network for Coastal Observatory (JERICO) is a Pan European coastal marine observatory network that integrates a large diversity of resources including data from multiple types of observatories. This diversity of data and platforms offers an assortment of assets to support the creation of high-quality and diverse data products. In the new phase of the project, namely JERICO-S3, the implementation of a virtual infrastructure (called e-JERICO) will allow access to the information from assets, including data, that are distributed among all internal and external stakeholders. It will also provide a full range of access to documents, tools, software and ocean best practices (OBPs). JERICO-S3 aims, among other objectives, to demonstrate the benefits of an e-Infrastructure for JERICO (e-JERICO) by designing and implementing the first elements of this virtual environment. Data To Product Thematic Services (D2PTS) are practical cases that respond to specific needs of the coastal community despite a diversity of data and platforms by providing added value products of different kinds. Four D2PTS will support the pilot phase and provide these capabilities of e-JERICO during the operation phase. They will provide advanced products, software capabilities and services, for specific thematic interests in the Integrated Regional Sites (IRS) and Pilot SuperSites (PSS) around the coast of Europe.

The **HF-Radar (HFR) D2PTS** will provide physical oceanography products to fill the gap in water surface current data products addressing the need for a comprehensive understanding of ocean surface flows. It offers advanced analysis software and services for HFR. A pilot application will be undertaken in Iberian Peninsula IRS and NW-Mediterranean PSS. The **glider D2PTS** will contribute with estimations types and transport monitoring of sea water masses by combining

Biogeochemical (BGC) and physical parameters. These advanced products will fill a gap in physical ocean transport analysis. A pilot application will be undertaken in the Gulf of Finland (GoF) and NW-Mediterranean PSSs. The **BGC D2PTS** will provide advanced data products (e.g.. HAB situation and remote sensing "sea-truth") based on combined multiplatform NRT data, demonstrating the capabilities of coordinated transnational observations. It will respond to the needs of the integration of multiplatforms observations. A pilot application will be undertaken in the GoF PSS. The JERICO-**EcoTaxa D2PTS** will provide new insights in the biological field by facilitating the study of coastal plankton monitoring products from optical and imaging sensors. It responds to a need to bring together the biological community to a joint effort of analysing biological images. A pilot application will be undertaken in NW-MED, Gulf of Finland, Channel and NorthSea PSSs.

D2PTS Integration

From a general perspective, e-JERICO (see Figure 1) harvests and connects information of different assets (i.e. data, documents, tools, software, etc) in a knowledge-based catalog of interconnected resources from distributed infrastructures such as data aggregators (EMODnet, SeaDatanet, CMEMS), Sensor Observation Services (SOS), documentation repositories (OBPS, OceanDocs) and software repositories. Thematic services will be integrated into e-JERICO in different ways depending on their level of maturity. The D2PTS represent various scenarios of integration that will demonstrate the entire range of capabilities of the e-JERICO infrastructure.

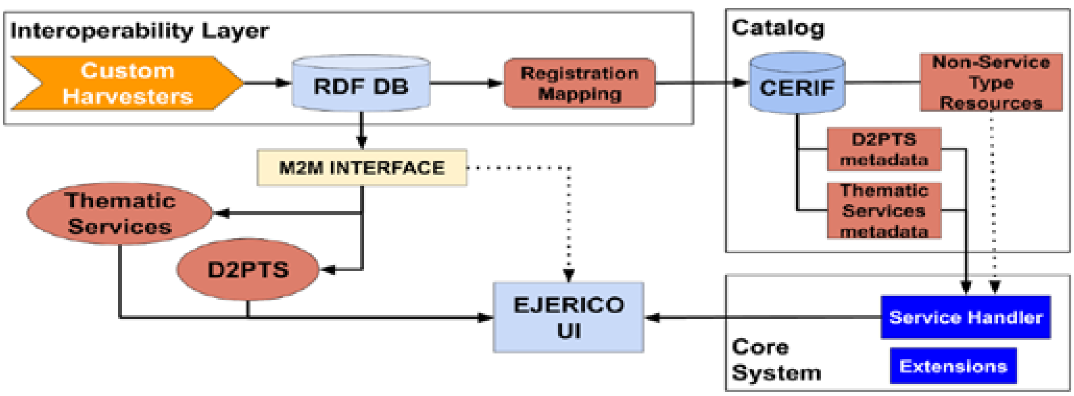


Figure 1: Interaction flow of D2PTS that are integrated in e-JERICO

The integration of services provides the D2PTS users the opportunity to take advantage of the information of the JERICO community that is represented in the resource catalog. In a basic scenario, users will be able to access all the information to run the D2PTS including the necessary support such as manuals, best practices and portal information. A more advanced level of integration will allow thematic services to make use of the information of the resource catalog to support the service. For example, the BGC D2PTS can collect the list of platforms of the region to improve collaboration and the search of multiplatform data. Integration of a service could even allow seamless and remote execution of processing components alone or workflows of concatenated services. In this poster, we will examine the way these diverse resources are handled in e-JERICO for the four D2PTS.

PHIDIAS: boosting the use of cloud services to benefit marine data management, services and processing

Charles Troupin, GHER/ULiège (Belgium), ctroupin@uliege.be
Julie Abergas-Arteza, Trust-IT Services, Pisa (Italy), j.arteza@trust-itservices.com
Alexander Barth, GHER/ULiège (Belgium), a.barth@uliege.be
Frederic Huynh, IRD (France), frederic.huynh@ird.fr
Jukka Seppälä, SYKE (Finland), jukka.seppala@ymparisto.fi
Seppo Kaitala, SYKE (Finland), Seppo.Kaitala@ymparisto.fi
Aleksi Kallio, CSC (Finland), aleksi.kallio@csc.fi
Cecile Nys, IFREMER, Plouzané (France), Cecile.Nys@ifremer.fr
Francesco Osimanti, Trust-IT Services, Pisa (Italy), f.osimanti@trust-itservices.com
Peter Thijsse, MARIS (The Netherlands), peter@maris.nl
Gilbert Maudire, IFREMER, Plouzané (France), Gilbert.Maudire@ifremer.fr
Boris Dintrans, CINES, Montpellier (France), boris.dintrans@cines.fr

The project and its goals

PHIDIAS - Prototype of HPC/Data Infrastructure for On-demand Services - addresses the development and concrete realisation of a set of High-Performance Computing (HPC) based on interdisciplinary services and tools to exploit large satellite datasets of public European interest provided by:

- Satellite observation of Earth,
- FAIR access to processed datasets,
- Value-added services through large data storage capability and high-bandwidth networks across Europe.

PHIDIAS aims at creating access services to increase the HPC and data capacities of the European Data Infrastructure in the context of the 'Connecting European Framework (CEF)' on open data. The PHIDIAS project seeks for – and establishes – links with the scientific, engineering and industrial communities that may be interested in the project's results and outcomes, in order to initiate and fulfil high-level exchanges at different levels.

The Ocean use case

Observing the ocean is challenging: missions at sea are costly, different scales of processes interact, and the conditions are constantly changing. This is why scientists say that „*a measurement not made today is lost forever*". For these reasons, it is fundamental to properly store both the data and metadata, so that access to them can be guaranteed for the widest community, in line with the FAIR principles: Findable, Accessible, Interoperable and Reusable. The PHIDIAS Ocean use case is focused on 3 aspects:

1. **Improvement of long-term stewardship of marine *in situ* data**. The SEANOE service allows users to upload, archive and publish their data, including the processed data via HPC, to which a permanent identifier (DOI) is assigned so the dataset can be cited and

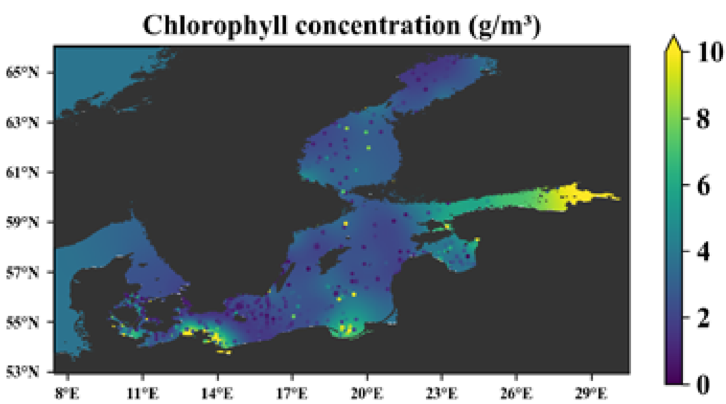
referenced. Efforts will be articulated around the scalability, the exchanges between data centres in charge of related data types and the protection of long-time archives. The *long-tail* data, referring to data not acquired routinely but during scientific missions or specific events, are of particular interest. Usually they cannot make their way so easily to the data centers, as there is no automated procedure.

2. **Improvement of data storage for services to users**. The goal is to provide users with (1) fast and interoperable access to data from multiple sources, for visualization and subsetting purposes; (2) parallel processing capabilities within dedicated high-performance computing, using, for example, Jupyter notebooks or the PANGEO software ecosystem.
3. **Marine data processing workflows for on-demand processing**. The objective is that users can access data, software tools and computing resources in a seamless way to create added-value products, for example quality-controlled, merged datasets or gridded fields.

Key results

The efforts of the pilot use cases in PHIDIAS strive to improve the activities of the researchers and specialists on different aspects:

- **Data publication:** with the improved capabilities of SEANOE, they will be able to seamlessly upload large datasets, ensure their long-term archiving (also of the processed datasets) and publish them following standards, best practices and recommendations from Data Management groups. This will also enhance the ingestion of *long-tail* data, which in turn will be made available to a larger community.
- **Data access:** thanks to fast access to the most recent data collections obtained from different sources and providers (Euro-ARGO, SeaDataNet, EMODnet, CMEMS, imaging flow cytometer), users will be able to demonstrate as pilot operations such as sub-setting (based on regions, parameters), quality-control, visualisation or spatial interpolation.
- **Data processing:** the deployment of cutting-edge tools such as DIVAnd (spatial-temporal interpolation, https://github.com/gher-ulg/DIVAnd.jl) in an HPC environment will allow scientists and experts to perform spatio-temporal interpolation of large datasets. In particular, this use case will be in the North Atlantic Ocean and the Baltic Sea, which represents 10 million observations for a total of approx. 250 GBytes. The final product will consist of an inter-comparison of satellite data and in-situ data of sea surface salinity, including Inspire-compliant online services for data visualization and access.



SOURCE software's reprocessing and merging of different sea temperature and salinity time series data collections from SeaDataCloud and CMEMS

Paolo Oliveri, Istituto Nazionale di Geofisica e Vulcanologia (Italy), paolo.oliveri@ingv.it
Simona Simoncelli, Istituto Nazionale di Geofisica e Vulcanologia (Italy), simona.simoncelli@ingv.it
Pierluigi di Pietro, Istituto Nazionale di Geofisica e Vulcanologia (Italy), pierluigi.dipietro@ingv.it
Gelsomina Mattia, Istituto Nazionale di Geofisica e Vulcanologia (Italy), gelsomina.mattia@ingv.it

Several European, regional and global marine data, and data product initiatives provide access to marine data (Martín Míguez et al., 2019). CMEMS and SeaDataNet initiatives aggregate, archive, quality assess, update and distribute the data with a different scope (https://www.seadatanet.org/About-us). The CMEMS in situ TAC gathers both Near Real Time (NRT) and Delayed Mode (DM) observations to serve the EU Monitoring and Forecasting Centers, while the SeaDataNet diffuse network of National Oceanographic Data Centers mainly aggregate DM marine data coming from observing systems and research programs. This implies that the integration of the their DM data collections is of paramount importance to assure the usage of the largest data sets for scientific applications.

The data originating from a specific data provider may be available only in one of the two infrastructures or their availability could be fragmented in terms of temporal and spatial data coverage and resolution or in terms of the distributed parameters. This may result in partial data availability, with great impact for the researchers who wants to access as much data as possible.

The objective of this work is to analyze the mooring time series data from CMEMS in situ TAC and SeaDataNet, compare and integrate them when necessary in order to obtain the most complete data set for temperature and salinity over the Mediterranean Sea.

Due to the different data management procedures, it was necessary to first convert all the data in an unique common data format (netCDF). Then it has been applied an accurate quality control analysis, a duplicates detection and process.

This data collation and reprocessing requires specific skills on how to manipulate the data, that final users might not have or it would be very costly for them.

INGV team developed a software called **SOURCE** (Sea Observations Utility for Reprocessing, Calibration and Evaluation) written in Python. SOURCE handles the pre-processing of the data coming from the different infrastructures, their merging in one final database and the post-processing by filtering out the duplicates. The final data quality assurance and consistency process consists of several tools that follow the ocean best practices guidelines.

The software was already ready to handle the CMEMS in situ TAC Mediterranean NRT data set with the main target to quickly acquire and publish data into an operational chain already in place. Then a first data merging test has been performed between SeaDataCloud V2 and CMEMS temperature and salinity data collections.

The SeaDataCloud time series data collection is a Global Ocean soon-to-be-published dataset that could represent a reference for the users. It has been released in binary, user friendly Ocean Data View format, originally implemented for vertical profiles, but successively adapted for time series. The pre-processing of those data was very difficult in order to concatenate and aggregate the different time series belonging to the same platform through likeness of ID parameter strings, organizing metadata, harmonizing time units, filtering the data by area of interest or instrument type, producing information on the original SeaDataNet QC scheme and producing log files to track the problems encountered during the processing steps (i.e. missing time, depth, data, wrong Q/C variables).

The reconstructed SeaDataCloud time series, divided by parameter and organized in a fit for purpose dataset, gave the possibility to start the integration process with the already pre-processed CMEMS time series in order to obtain the final dataset and maximize the data coverage. Statistics have been computed per each platform in order to further QC the temperature and salinity time series and discard the data anomalies.

The data has been processed and published through a new INGV data access and visualization service that allows to visual inspect the content of the obtained aggregated database and its statistics. Moreover the INGV web service allows to visualize the synthetic time series extracted from INGV model analyses and reanalyses at each mooring location. This functionality could also be expanded to the available CMEMS prediction model products.

This work allows a continuous monitoring of the coastal environment through the all available existing moorings using both observational and model data displaying high quality data but it represent a continuous calibration/validation tool to assess the skill of the existing models This is the base from which to start to develop new services and applications to serve a variety of different users that perform operations at sea, integrated coastal monitoring systems, early-warning system for coastal environmental protection and preservation.

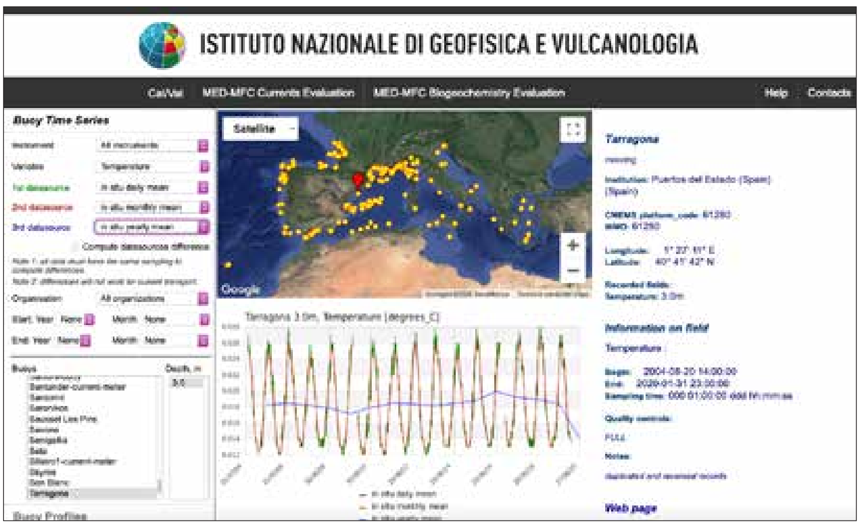


Figure 1. Snapshot of the new INGV service portal for visualization and evaluation of processed in situ observations from fixed platforms.

Martín Míguez B, Novellino A, Vinci M, Claus S, Calewaert J-B, Vallius H, Schmitt T, Pititto A, Giorgetti A, Askew N, Iona S, Schaap D, Pinardi N, Harpham Q, Kater BJ, Populus J, She J, Palazov AV, McMeel O, Oset P, Lear D, Manzella GMR, Gorringe P, Simoncelli S, Larkin K, Holdsworth N, Arvanitidis CD, Molina Jack ME, Chaves Montero MM, Herman PMJ and Hernandez F (2019) The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe. Front. Mar. Sci. 6:313. doi: 10.3389/fmars.2019.00313

Mapping Arctic Observing Systems and In Situ Data Collections

Frode Monsen, Nansen Environmental and Remote Sensing Center (Norway), frode.monsen@nersc.no
Tor I. Olaussen, Nansen Environmental and Remote Sensing Center (Norway),
Hanne Sagen, Nansen Environmental and Remote Sensing Center (Norway), hanne.sagen@nersc.no
Torill Hamre, Nansen Environmental and Remote Sensing Center (Norway), torill.hamre@nersc.no
Florian Geyer, Nansen Environmental and Remote Sensing Center (Norway), florian.geyer@nersc.no
Kjetil Lygre, Nansen Environmental and Remote Sensing Center (Norway), kjetil.lygre@nersc.no
Roberta Pirazzini, Finnish Meteorological Institute (Finland), roberta.pirazzini@fmi.fi
and the INTAROS consortium

Abstract

There is a great need for better information on the state and changes in the Arctic marine environment. Climate change in the Arctic is significant and will have far-reaching consequences for marine life and sustainable industrial development in the Arctic. It produces a lot of data from polar field satellites, but these are limited to the surface of sea, land and ice. Direct measurements (in situ) of the marine environment in Arctic waters are absolutely necessary to estimate the state and changes in the marine environment. Since collecting in situ data in polar regions most often funded by research projects, much of the data becomes time-limited and without a general plan. The H2020 INTAROS project has conducted a survey of Arctic in situ observing systems, in situ and remote sensing data collections (M. Tjernstrøm et al., 2019). Based on the questionnaires from this survey we have developed a user friendly web-based system for collecting and maintaining information about in situ observing systems and data collections in a project funded by the Norwegian Ministry of Climate and Environment. This system, ARCMAP, will be used to maintain and extend the INTAROS survey of Arctic observation systems and data collections, as a contribution to GEO and SAON.

The arcmap survey application

ARCMAP is a web application for collecting and updating information about in situ observation systems. It allows the user to create and edit detailed descriptions of observation systems that are, have been or will be deployed to measure various environmental parameters in the Arctic. This information on observation systems is organized in a database, enabling easy retrieval and updating, and facilitating analysis of the capacity for environmental monitoring in a given area and time period.

The application is developed at NERSC using open source frameworks for web applications, including Python Django 2 and the wq.io online survey framework, and packaging the application in a Docker container that is deployed on a server at the Center. We have extended the wq.io framework to provide the needed functionality for ARCMAP, among others, allowing multiple types of geographic objects to represent the location and extent of an observation system composed of several subsystems. The survey is now presented as a series of forms that can be created and updated separately. This allows users to complete the system description in multiple sessions, and only reopen those parts that need updating when there is new information about their systems (Figure 1).

Extraction of indicators and statistics from the ARCMAP database

ARCMAP stores the survey results in a relational database. The collected system description can be easily retrieved and combined to extract information about, among others, which parameters are being measured in different regions and time periods, where collected data are stored, and which application areas the systems are designed to serve (Figure 1). The extracted information is then presented using the Matplotlib open source plotting library.

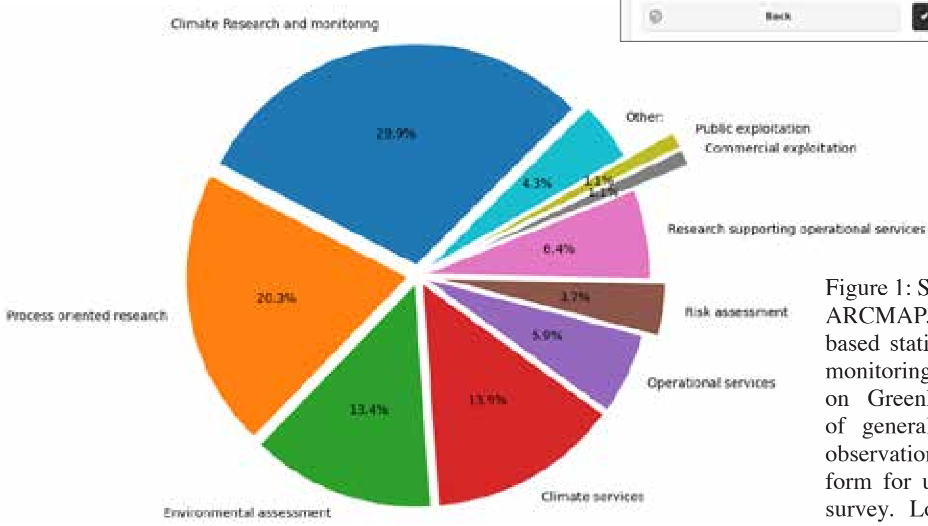


Figure 1: Selected screen shots from ARCMAP. Upper: Locations of land based stations in an environmental monitoring network for glaciers on Greenland. Middle: Example of general information about an observation system and the input form for updating this part of the survey. Lower: Application areas served by the surveyed observation systems.

Acknowledgement: This work has received funding from the Norwegian Ministry of Climate and Environment and the Nansen Environmental and Remote Sensing Center and is a contribution to the project INTAROS. INTAROS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727890.

References

Tjernstrøm, Michael; et al. 2019. Deliverable 2.10. Synthesis of gap analysis and exploitation of the existing Arctic observing systems. INTAROS project 2019. https://intaros.nersc.no/sites/intaros.nersc.no/files/D2.10_INTAROS_Synthesis_v9.0.pdf

The NextGEOSS Cold Region pilot: Improved discoverability and access to polar data

Torill Hamre, Nansen Environmental and Remote Sensing Center (Norway), torill.hamre@nersc.no
Bente Lilja Bye, BLB AS (Norway), bentelil@hotmail.com
Markus Fiebig, Norwegian Institute for Air Research (Norway), markus.fiebig@nilu.no

Abstract

GEO Cold Regions coordinates global efforts to provide Earth Observation (EO) products and services to science, decision- and policy-makers with a vested interest in the cryosphere (in particular) and the environment (in general) of polar regions and mountain areas around the world. The NextGEOSS Cold Regions Pilot focuses on three areas: (1) the Arctic/Svalbard region, (2) Antarctica, and (3) the Himalayan glaciers, linking together satellite and in situ data from the targeted regions, including the atmospheric, marine, and terrestrial domains, and making them available in the NextGEOSS Data Hub and the NextGEOSS Cold Regions Community Portal. The pilot liaise with ongoing initiatives such as SIOS, GEOCRI, WMO GCW, as well as national programs in Antarctica. The products resulting from the Cold Regions pilot can be used to develop Information Services for the Cold Regions Initiative, using the NextGEOSS system and making use of existing interoperability standards. This presentation will introduce the first version of the Cold Regions Community Portal, aiming to make polar data relevant for Cold Regions more easily discoverable and accessible for users.

NextGEOSS Cold Regions Community Portal

The NextGEOSS Cold Regions Community Portal provides access to a wide range of polar data from satellite, in situ and models. It harvests metadata from a series of data providers and provides a joint entry point to data originating from all of these sources. The users can search for free text, geographic area, time periods, provider's name, and multiple other criteria (Figure 1). A summary of the metadata for all datasets fulfilling these criteria is presented, allowing the user to drill down further into the details of particular datasets. Having identified a dataset of interest, a user can then follow the data access links to retrieve the data.

In addition to data search, the Community Portal provides descriptions of a selection of the products offered (Figure 2) including plots of product examples and links to further information. This aims to inform the users further to enable them to decide whether various products can be useful for a specific purpose. The Community Portal will also link to educational material and webinars on other relevant topics and products for the Cold Regions Initiative Community, including resources on how to the use the NextGEOSS Data Hub and Platform.

The Community Portal has been established by means of a portal framework developed by the Norwegian Meteorological Institute. This framework is used in several other data portals, e.g. in the the SIOS Data Management System and the YOPP Data Portal, to harvest metadata from distributed data centres and harmonising the metadata into a common structure. The aim is to customise this framework to harvest metadata for data of relevance to Cold Regions from the NextGEOSS Data Hub into the Community Portal.

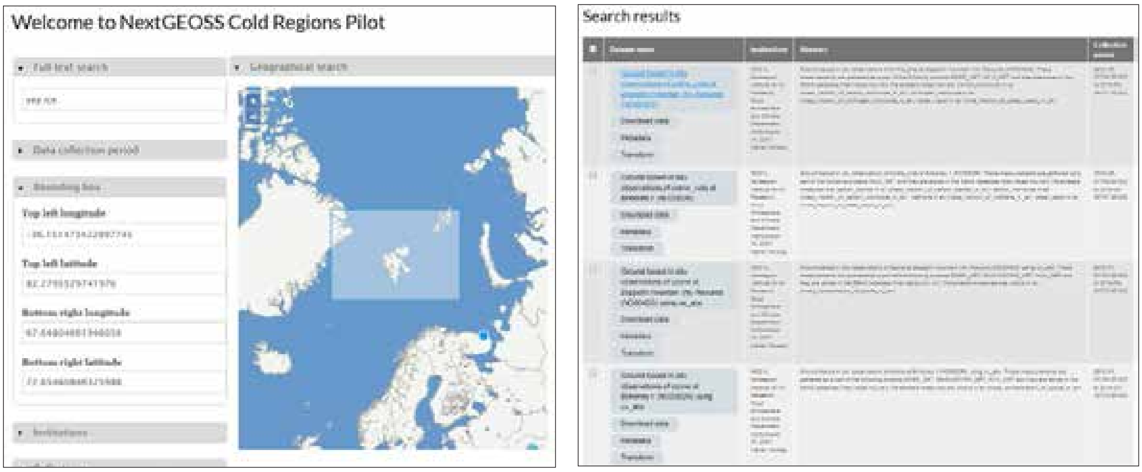


Figure 1: The NextGEOSS Cold Regions Community Portal, home page (left) and search results (right).

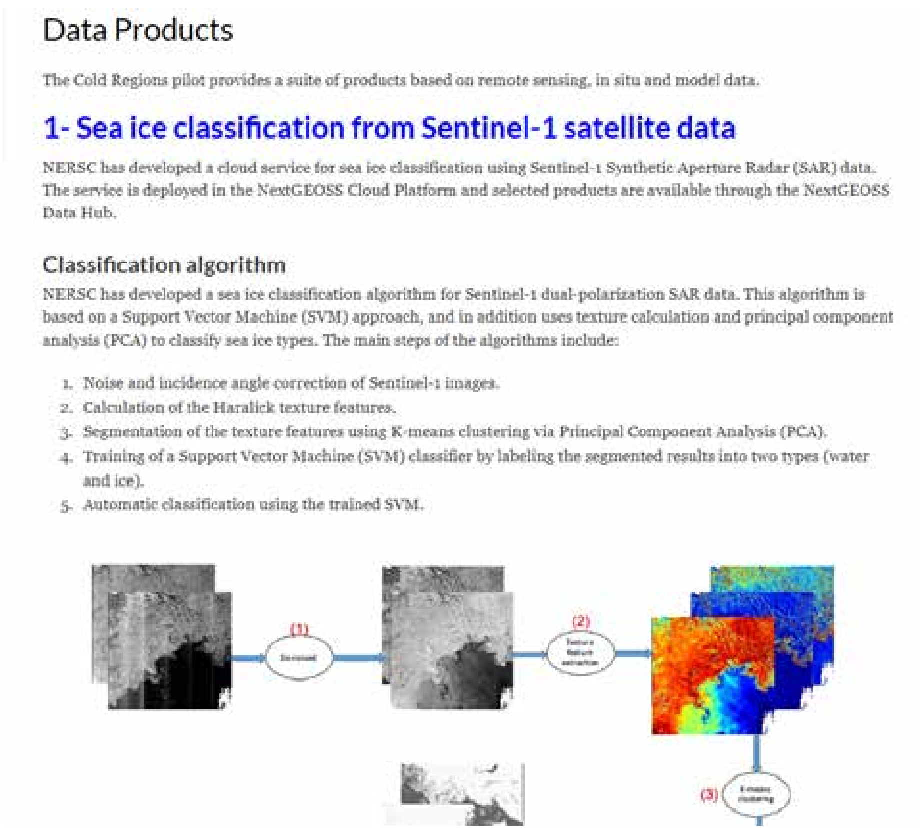


Figure 2: Product descriptions in the NextGEOSS Cold Regions community portal.

SESSION TECH:

Technical developments for marine information and data management

- Standards for data interoperability (e.g. INSPIRE, vocabularies, data formats, Sensor Web Enablement, Semantic Web, Linked data)
- Metadata/Data FAIRness
- Quality assurance
- Data citation, DOI, PID and data versioning
- Monitoring and tuning
- Best practice
- Data centre/repository accreditation/certification (e.g. IODE Quality Management Framework, CoreTrustSeal, etc.)

ORAL PRESENTATIONS

How to stop re-inventing the wheel: a data management case study

Jen Thomas, Swiss Polar Institute (Switzerland), jenny.thomas@epfl.ch
Marco Alba, EMODnet Physics (Italy), marco.alba@ettsolutions.com
Eric Bouillet, Swiss Data Science Center (Switzerland), eric.bouillet@epfl.ch
Antonio Novellino, EMODnet Physics (Italy), antonio.novellino@ettsolutions.com
Carles Pina Estany, Swiss Polar Institute (Switzerland), carles.pinaestany@epfl.ch
Michele Volpi, Swiss Data Science Center (Switzerland), michele.volpi@sdsc.ethz.ch

Introduction

Over the past three years, the Swiss Polar Institute (SPI) has adopted and integrated a number of existing tools and services to manage data from a recent expedition in order to assist scientists in making them openly accessible. SPI aimed for secure data storage, well-documented, discoverable and citable datasets and recording of dataset provenance.

Here we put into context the data to be managed, explain the use of tools chosen to implement the data management workflow, as well as notable successes and challenges of putting this into practice.

The data in question

Initial data to be managed resulted from the Antarctic Circumnavigation Expedition Walton and Thomas (2018) (ACE). Data and metadata were managed during the expedition using **Django**, **Python** and **MySQL** Pina Estany and Thomas (2019); tools that allowed rapid development of a relational database to record metadata, monitor backup of files, and produce simple reports or displays of the data collected. Almost 30,000 samples and over 20 TB of data were recorded, resulting in around 200 datasets from a range of natural science disciplines (Craig-Wood, 2020) https://zenodo.org/communities/spi-ace.

Selection of tools and services

Rclone, an open-source tool, is used to transfer raw data files to the object storage and perform consistency checks on data files.

SPI was able to adopt **RENKU** as a tool for recording the provenance of datasets through a collaboration (ACE-DATA: Delivering Added value To Antarctica) between SPI, the Paul Scherrer Institute and Swiss Data Science Centre. This open platform addresses the issue of data versioning and ensures reproducibility by capturing what was done to the data, by whom and when.

Zenodo offers long-term archiving of published datasets and was chosen for ACE dataset publication because citation and dataset re-use metrics are facilitated using a Digital Object Identifier (DOI).

EMODnet Physics incorporates already-published data from Zenodo into its well-established repository, displaying it through a user-friendly map-based portal as well as adding value to the data through visualisations. Data are described using machine-readable metadata schemas from **Frictionless Data** and validated by goodtables-py prior to publication in Zenodo. In addition to a Github repository with Github actions to provide this information for already-published datasets, this provides machine-readable metadata to EMODnet Physics.

Facilitating the discovery of data from ACE is important to ensure that it reaches its potential for maximum re-use and can be incorporated into further regional and discipline-specific studies. **RENKU**, **Zenodo** and **EMODnet Physics** all contribute to this aim: **RENKU** exposes data-derived insights to a perhaps unforeseen audience; text-based search of **Zenodo** allows discovery and through their use of schema.org and adoption of the DataCite Metadata Schema, also opens up the data to potential re-use through other repositories such as Mendeley Data and Google dataset search; and the **EMODnet** Physics map-based portal displays the data to more domain-specific users as well as those interested in the geographic region through the SOOS map.

Internally, SPI uses **Python** as the language of choice for writing utility scripts that bring some parts of this workflow together. **Github** is used to manage, document and finally publish releases of the code through its link with **Zenodo**. Whilst dataset documentation is always maintained alongside the data, project and data management documentation is written in **Github** and a **C4Science** wiki.

	rclone	MySQL	Django	Python	Gitlab	Github	c4science	Frictionless Data	RENKU	Zenodo	EMODnet Physics
Project management and documentation						■	■				
Code management and publication						■				■	
Data file management	■	■	■	■	■				■		
Data provenance				■	■				■		
Data description								■	■	■	
Data publication						■			■	■	■
Data discovery						■		■	■	■	■
Data visualisation				■							■

Table 1. Overview of tools and services, and the roles they play within data management at the Swiss Polar Institute.

Can we continue to "reduce, reuse and recycle" software tools and services in this world of plenty?

The adopted tools meet the needs of the SPI for managing current and future datasets. Researchers, institutes, data managers and others are always looking to existing tools to meet their needs. Use of the open-source technologies described here, is in line with the principles of open science such as reproducibility and transparency, setting an example for requiring open access to datasets.

The most pressing challenge is the lack of integration between these tools and services. As the number of datasets grows, it is becoming more evident that connecting each tool with custom-developed software to form an automated and continuous workflow, would greatly improve reliability, robustness and scalability.

We will continue to look for existing tools to adopt that meet the needs of what we require, but cannot shy away from building something new if it would be more suitable. In short, it is possible to stop reinventing the wheel, adopt existing software tools and services, collaborate and integrate.

References

Walton, D.W.H and J. Thomas. (2018). Cruise Report - Antarctic Circumnavigation Expedition (ACE) 20th December 2016 - 19th March 2017 (Version 1.0). Zenodo. doi: 10.5281/zenodo.1443511

Pina Estany, C and J. Thomas. (2019). Swiss-Polar-Institute/science-cruise-data-management v0.1.0 (Version 0.1.0). Zenodo. doi: 10.5281/zenodo.3360649

Craig-Wood, Nick. (2020). Rclone. https://rclone.org/

Swiss Data Science Center. RENKU. https://datascience.ch/renku/

Thomas, J., S. Landwehr, M. Volpi and J. Schmale. (2019). ACE-DATA: Antarctic Circumnavigation Expedition: Delivering Added value To Antarctica (Version 1.0). Zenodo. doi: 10.5281/zenodo.2587954

Zenodo. (2009-2017). CERN. https://zenodo.org

EMODnet Physics project. European Marine Observation and Data Network. www.emodnet-physics.eu/map

Fowler, D., Barratt, J. and Walsh, P. (2018) 'Frictionless Data: Making Research Data Quality Visible', International Journal of Digital Curation, 12(2), pp. 274–285. doi: 10.2218/ijdc.v12i2.577

Frictionless Data. goodtables-py. Available at: https://github.com/frictionlessdata/goodtables-py

schema.org. (2020). https://schema.org

DataCite Metadata Working Group. (2019). DataCite Metadata Schema for the Publication and Citation of Research Data. Version 4.3. DataCite e.V. doi: 10.14454/f2wp-s162

SOOSmap. Southern Ocean Observing System. (2020). www.soosmap.aq

A modular approach to cataloguing marine science data

Adam Leadbetter, Marine Institute (Ireland), adam.leadbetter@marine.ie
Will Meaney, Marine Institute (Ireland), will.meaney@marine.ie
Elizabeth Tray, Marine Institute (Ireland), elizabeth.tray@marine.ie
Andrew Conway, Marine Institute (Ireland), andrew.conway@marine.ie
Sarah Flynn, Marine Institute (Ireland), sarah.flynn@marine.ie
Tara Keena, Marine Institute (Ireland), tara.keena@marine.ie
Caoimhín Kelly, Marine Institute (Ireland), caoimhin.kelly@marine.ie
Rob Thomas, Marine Institute (Ireland), rob.thomas@marine.ie

The ability to access and search metadata for marine science data is a key requirement for answering fundamental principles of data management (making data Findable, Accessible, Interoperable and Reusable) (Wilkinson, *et al.,* 2016). In order to address the requirements of the Findable aspect of FAIR data, a dataset needs to be described by rich metadata in a searchable resource and the dataset must be assigned a clearly labelled persistent, unique identifier. Additional requirements include meeting domain-specific, community defined standards and legislative requirements placed on data publishers, for example maintaining semantic heterogeneity in cross-disciplinary domains, such as Marine Spatial Planning (MSP).

Therefore, in the sphere of oceanographic data management, the need for a modular approach to data cataloguing which is designed to meet a number of requirements has been demonstrated. According to Friddell, *et al*., 2014, modularity is required in order to represent datasets, projects or programmes and other polar data resources within the catalogue system in other cross-disciplinary topics.

This modular approach has been adopted by the Marine Institute Ireland (Leadbetter, *et al.,* 2020) in developing a data cataloguing system to meet the needs of legislative requirements including the European Spatial Data Infrastructure (INSPIRE) and the Marine Spatial Planning directive. The data catalogue described here makes use of a metadata model focused on oceanographic-domain. It comprises a number of key classes which will be described in detail in the paper, but which include:

- Dataset - combine many different parameters, collected at multiple times and locations, using different instruments
- Dataset Collection - provides a link between a Dataset Collection Activity and a Dataset, as well as linking to the Device(s) used to sample the environment for a given range of parameters. An example of a Dataset Collection may be the Conductivity-Temperature-Depth profiles taken on a research vessel survey allowing the individual sensors to be connected to the activity and the calibration of those sensors to be connected with the associated measurements.
- Dataset Collection Activity - a specialised dataset to cover such activities as research vessel cruises; or the deployments of moored buoys at specific locations for given time periods
- Platform - an entity from which observations may be made, such as a research vessel or a satellite
- Programme - represents a formally recognized scientific effort receiving significant funding, requiring large scale coordination

- Device - aimed at providing enough metadata for a given instance of an instrument to provide a skeleton SensorML record
- Organisation - captures the details of research institutes, data holding centres, monitoring agencies, governmental and private organisations, that are in one way or another engaged in oceanographic and marine research activities, data & information management and/or data acquisition activities

The data model (figure 1) makes extensive use of controlled vocabularies to ensure both consistency and interoperability in the content of attribute fields for the Classes outlined above.

The data model has been implemented in a module for the Drupal open-source web content management system, where ISO19115/19139 compliant XML records can be exported, and made available. Publically available metadata XML records are harvested by a GeoNetwork, which comprises the Marine Institute's public facing data catalogue available at http://data.marine.ie. Over 400 ISO19115/19139 compliant XML datasets are published on the Marine Institute's data catalogue.

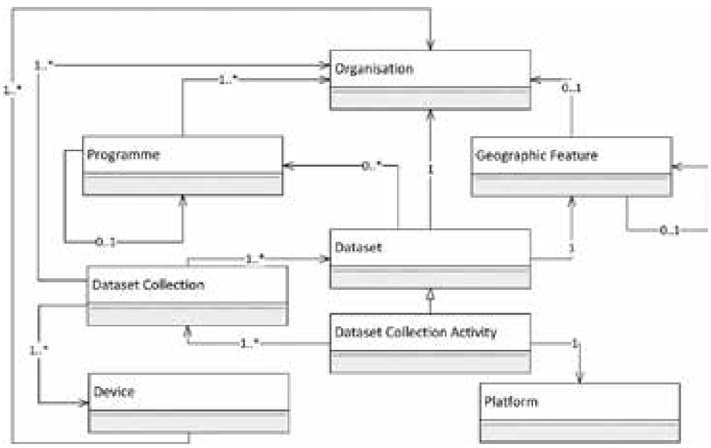


Figure 1: A high-level overview of the data model used in the modular data catalogue approach. The overall class structure is shown in the Unified Modelling Language

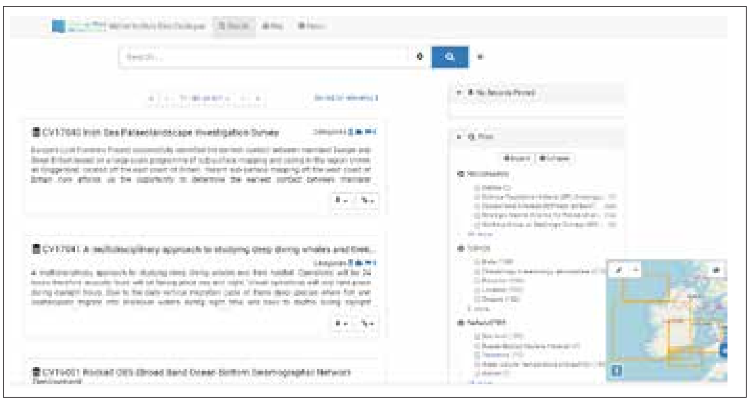


Figure 2: Marine Institute's Metadata Catalogue available at data.marine.ie

References

FRIDDELL JE, LEDREW EF, VINCENT WF (2014) The Polar Data Catalogue: best practices for sharing and archiving Canada's polar data. Data Science Journal: IFPDA-01

LEADBETTER, A., MEANEY, W., TRAY, E., CONWAY, A., FLYNN, S., KEENA, T. KELLY, C. AND THOMAS, R. (2020). A modular approach to cataloguing marine science data. *Earth Science Informatics.* https://doi.org/10.1007/s12145-020-00445-w

WILKINSON, M., DUMONTIER, M., AALBERSBERG, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. https://doi.org/10.1038/sdata.2016

FAIR Semantics and the NVS

Alexandra Kokkinaki, British Oceanographic Data Centre, National Oceanography Centre (UK), alexk@bodc.ac.uk
Gwenaëlle Moncoiffé, British Oceanographic Data Centre, National Oceanography Centre (UK), gmon@bodc.ac.uk
Yann Le Franc, e-Science Data Factory (France), ylefranc@esciencefactory.com

The FAIR principles provide guidelines for the publication of digital resources such as datasets, code, workflows, and research objects aiming at making them Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2018). Amongst them, the I of the FAIR promotes interoperability and more specifically principle I2 suggests that metadata should use vocabularies that themselves follow the FAIR principles. Recently, FAIRsFAIR[1] project officially published a first iteration of recommendations for making vocabularies FAIR (Le Franc et al., 2020). These recommendations include 17 general recommendations aligned with the different FAIR Principles and 10 Best Practice recommendations. The main objective of these recommendations is to provide a set of guidelines for creating a harmonised and interoperable semantic landscape easing the use and reuse of semantic artefacts from multiple different scientific domains.

The NERC Vocabulary Server (NVS) is an operational service, managed as a shared resource by the British Oceanographic Data Centre (BODC) according to rigorous content governance principles (Moincoffe and Kokkinaki, 2018). Since the 17 general FAIRsFAIR recommendations will impact terminology providers we volunteered in examining them from an operational terminology service point of view and in this presentation we will share our experience and findings.

The initiative behind this assessment stems from our commitment to serve the marine community with up-to-date and FAIR semantics. Since its inception NVS has undergone incremental enhancements in its model including publishing the versioning of its concepts, the provenance of mappings, allowing interaction via github, being listed in registries like fairsharing.org. In this work, we wish to evaluate the level of compliance of NVS to the FAIR principles according to this first set of recommendations as depicted in Figure 1. Our goal is to help shape the FAIR Semantics recommendations through the lens of a pragmatic approach applied to an existing, well established and operational terminology service.

For this purpose, we are working in partnership with the FAIRsFAIR project that drafted the recommendations, within an international context involving initiatives such as the GO FAIR Inter implementation network[2], the fairsharing.org[3] community, the Research Data Alliance (RDA) Vocabulary Semantic Services Interest Group (VSSIG)[4] and other terminology providers. In order to coordinate these various initiatives, a Task Group (TG) is being set up under the umbrella of the RDA VSSIG to focus on the evaluation of the FAIR Semantics recommendations with

1 https://www.fairsfair.eu/
2 https://www.go-fair.org/implementation-networks/overview/go-inter/
3 https://www.FAIRsharing.org
4 https://www.rd-alliance.org/group/vocabulary-services-interest-group

respect to semantic artefact services (i.e. repositories and registries). The TG will be used as a platform for collaboration and discussions on the topic.

The first step involves a crude analysis of the compliance of NVS to the 17 general recommendations, planning to be extended to the rest of semantic artefact services (SAS) involved in the RDA dedicated TG. This simple approach will allow the analysis of the practical implementation of these recommendations within each of the SAS. This first analysis will help uncover the diversity of implementations of each of the recommendations and reveal possible commonalities which should become, for instance, practical recommendations for other SAS. In the meantime, we are analysing the recommendations that are not fulfilled by any repositories and we will propose either a common solution for addressing this recommendation whenever possible or simply refine and clarify the recommendations. We hope our experience can benefit other terminology providers who are currently trying to evaluate and improve the FAIRness of their services and their content.

Rec#	Recommendation	NVS	Other Terminology Service	Comment
P-Rec.1	Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefacts, their content and their versions	Yes/No	Yes/No	
P-Rec.2	Use Globally Unique, Persistent and Resolvable Identifier for Semantic Artefact Metadata Record	Yes/No	Yes/No	
P-Rec.3	Use a common minimum metadata schema to describe semantic artefacts and their content	Yes/No	Yes/No	
P-Rec.4	Publish the Semantic Artefact and its content in a semantic repository	Yes/No	Yes/No	
P-Rec.5	Semantic repositories should offer a common API to access Semantic Artefacts and their content in various serializations for both use/reuse and indexation by any search engines	Yes/No	Yes/No	
P-Rec.6	Build semantic artefacts' search engines that operate across different semantic repositories	Yes/No	Yes/No	
P-Rec. 7	Repositories should offer a secure protocol and user access control functionalities	Yes/No	Yes/No	
P-Rec. 8	Define human and machine-readable persistency policies for semantic artefacts metadata	Yes/No	Yes/No	
P-Rec. 9	Semantic artefacts should be represented using common serialization formats, e.g. Semantic Web and Linked Data standards	Yes/No	Yes/No	
P-Rec.10	Use a Foundational Ontology to align semantic artefacts	Yes/No	Yes/No	
P-Rec.11	Use a standardized language for describing semantic artefacts	Yes/No	Yes/No	
P-Rec.12	Semantic mappings between the different elements of semantic artefacts should use machine-readable formats based on W3C standards	Yes/No	Yes/No	
P-Rec.13	Crosswalks, mappings and bridging between semantic artefacts should be documented, published and curated	Yes/No	Yes/No	
P-Rec.14	Use standard vocabularies to describe semantic artefacts	Yes/No	Yes/No	
P-Rec.15	Make the references to the reused third-party semantic artefacts explicit	Yes/No	Yes/No	
P-Rec.16	The semantic artefact should be clearly licenced for machines and humans	Yes/No	Yes/No	
P-Rec.17	Provenance should be clear for both humans and machines documented, published and curated	Yes/No	Yes/No	

Figure 1: Initial form to evaluate terminology services alongside the 17 recommendations

References

Wilkinson, Mark & Sansone, Susanna-Assunta & Schultes, Erik & Doorn, Peter & Bonino da Silva Santos, Luiz Olavo & Dumontier, Michel. (2018). A design framework and exemplar metrics for FAIRness. Scientific Data. 5. 180118. 10.1038/sdata.2018.118.

Le Franc, Yann, Parland-von Essen, Jessica, Bonino, Luiz, Lehväslaiho, Heikki, Coen, Gerard, & Staiger, Christine. (2020). D2.2 FAIR Semantics: First recommendations (Version 1.0). FAIRsFAIR. 10.5281/zenodo.3707985.

Moncoiffé, Gwenaëlle and Kokkinaki, Alexandra (2018). The BODC Parameter Usage Vocabulary (PUV) semantic model exposed. Bollettino di Geofisica Teorica ed Applicata. 59, 36. Proceedings of IMDIS 2018

Semantic interoperability of operational parameter terminologies in marine sciences

Gwenaëlle Moncoiffé, BODC (UK), gmon@bodc.ac.uk
Alexandra Kokkinaki, BODC (UK), alexk@bodc.ac.uk
Alison Pamment, UKRI STFC (UK), alison.pamment@stfc.ac.uk
Neil Holdsworth, ICES (Denmark), NeilH@ices.dk
Hans Mose Jensen, ICES (Denmark), hans.jensen@ices.dk
Adam Shepherd, BCO-DMO (US), ashepherd@whoi.edu
Adam Leadbetter, Marine Institute (Ireland), adam.leadbetter@marine.ie
Rob Thomas, Marine Institute (Ireland), rob.thomas@marine.ie

In October 2019 a new Research Data Alliance's working group, InteroperAble Descriptions of Observable Property Terminology (I-ADOPT WG)[1], launched its 18-month work plan with the ultimate goal to deliver a semantic interoperability framework for observable property terminologies. The goal is ambitious but the group contains a broad range of scientific, data and semantic experts, a very committed core group, and a growing membership of users and developers of terminologies within a broad spectrum of scientific domains centered mainly but not exclusively on environmental sciences including terrestrial ecology, marine, atmospheric and earth sciences, as well as related human activity disciplines like agriculture, forestry. While the ultimate goal is semantic interoperability across domains, it is clear that this work will also benefit semantic interoperability within domains wherever there are overlaps between established resources.

In the case of marine sciences, interoperability between data sources has inadvertently been hindered by difficulties in aligning long established parameter naming schemes such as the Climate and Forecast (CF) Standard Names used by the Climate and Forecast community and users of the netCDF format, the BODC P01 Parameter Usage Vocabulary (PUV) adopted by the SeaDataNet (SDN) community and close partners, and the ICES parameter dictionary used by many monitoring government agencies and laboratories. Each of these terminology resources grows in response to requests from their user communities based on their own local semantic rules and constraints.

Table 1 shows examples of how the labels used in CF, PUV and ICES parameter vocabularies can be decomposed into useful atomic components if one knows their structure or "grammar". However these components are not always in perfect alignment and human interpretation is necessary in order to achieve an accurate mapping between the terminologies. For example, the quantity term "concentration" has to be inferred from the units in the ICES vocabulary; then "mole concentration" used in CF is broader than the quantity "concentration" used in PUV and ICES and an equivalence between the two would require access to the units; we also have important qualifier such as "dissolved" which is an attribute of the analyte in CF and ICES but a phase of the matrix in the PUV.

1 www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg

Community	Parameter vocabulary	Parameter label	Quantity term	Chemical substance term	Matrix/medium term
CF	Standard names	mole_concentration_of_ dissolved_molecular_ oxygen_in_sea_water[2]	mole concentration	dissolved molecular oxygen	sea water
SDN	PUV	Concentration of oxygen {O2 CAS 7782-44-7} per unit volume of the water body [dissolved plus reactive particulate phase][3]	concentration	oxygen	water body [dissolved plus reactive particulate phase]
ICES	<multiple>	Dissolved oxygen[4]	<inferred from units>	dissolved oxygen	water

Table 1: Examples of terminology used by the CF, PUV, and ICES vocabularies for dissolved oxygen concentrations and comparison of selected atomic component terms used for the quantity measured, for the chemical substance and for the matrix or medium

While manual matching of terms across vocabularies has been achieved for small vocabulary subsets, mainly driven by relevance to particular projects (such as the mapping of ICES and SDN vocabularies for contaminants for EMODnet Chemistry), this is not a scalable solution for complete vocabularies covering many types of observations. The increasing pressure to provide fast, Findable, Accessible, Interoperable, Reusable, machine readable and user friendly access to data from a growing number of sources and for a growing number of applications has made the need to improve semantic interoperability of scientific terminologies more acute.

The semantic interoperability framework that will be delivered as part of the I-ADOPT collaborative work will provide a common method to systematically express or represent observable properties. In the marine domain, example mappings from the BODC P01 PUV to the Complex Properties Model were prepared originally by Leadbetter and Vodden (2016). These alignments will need to be validated against the I-ADOPT agreed framework. They will then be operationalised and implemented using Linked Data principles to provide a trusted, reproducible, and harmonised framework to build semantic translation and brokering services. Such services will be needed in projects such as EnvriFAIR in order to access data across multiple marine observation networks and research infrastructures, and relate them to broader concepts like e.g. Essential Ocean and Climate Variables or environmental assessments indicators or Sustainable Development Goals criteria. Such a semantic framework will also help standardise the associations between fine granularity terminologies and higher level ontologies and semantic models such as Semantic Sensor Network Ontology, SensorML, schema.org.

This presentation aims to promote the work of the RDA I-ADOPT working group, highlight the benefits to the marine community, and stimulate collaboration and contributions.

References

Leadbetter, A. & Vodden, P. (2016) Semantic linking of complex properties, monitoring processes and facilities in web-based representations of the environment, International Journal of Digital Earth, 9:3, 300-324, DOI: 10.1080/17538947.2015.1033483

2 vocab.nerc.ac.uk/collection/P07/current/CF14N29/
3 vocab.nerc.ac.uk/collection/P01/current/DOXYZZXX/
4 https://vocab.ices.dk/?CodeID=33506

Increasing FAIRness of marine data within ENVRI-FAIR

Peter Thijsse, MARIS, Nootdorp (The Netherlands), peter@maris.nl
Thierry Carval, IFREMER, Plouzané (France), Thierry.Carval@ifremer.fr
Dick Schaap, MARIS, Nootdorp, (The Netherlands), dick@maris.nl
Valerie Harscoat, IFREMER, Plouzané (France), Valerie.Harscoat@ifremer.fr
Sylvie Pouliquen, IFREMER, Plouzané (France), Sylvie.Pouliquen@ifremer.fr
Ivan Rodero, EMSO ERIC, Rome (Italy), ivan.rodero@emso-eu.org
Benjamin Pfeil, University of Bergen, Bergen (Norway), benjamin.pfeil@uib.no
Katrina Exter, VLIZ, Oostende (Belgium), katrina.exter@vliz.be
ENVRI-FAIR WP9 partners, IFREMER, Plouzané (France), envrifair-wp9-coordination@ifremer.fr

Humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data. Machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) is key, but how to achieve this? Starting from an RDA initiative in 2016 guidelines have been drafted - The FAIR principles - to improve the findability, accessibility, interoperability, and reuse of digital assets [*GoFair.org*]. Several metrics have been defined to "score" FAIRness against the principles, although the main concept of FAIR is the process to improve, more than a quantitative score card.

The ENVRI-FAIR project is engaging Research Infrastructures (RIs) in the environmental domain covering the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems. The overarching goal of ENVRI-FAIR is that all participating research infrastructures (RIs) will improve their FAIRness and become ready for machine-to-machine access, e.g. for services in the European Open Science Cloud (EOSC).

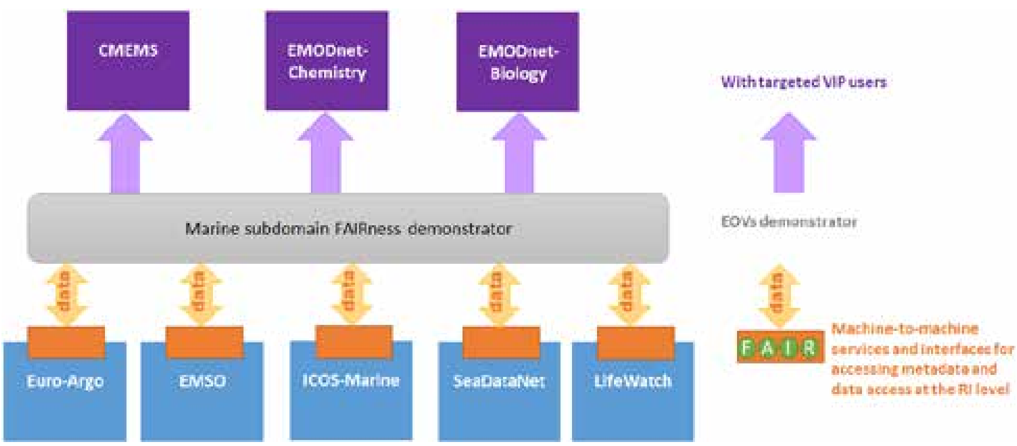


Figure 1: FAIRness roadmap for the Marine subdomain within ENVRI-FAIR

The Marine subdomain work package (WP9) of ENVRI-FAIR has a focus on (see figure 1): Euro-Argo-ERIC, EMSO-ERIC, ICOS (Marine) and LifeWatch (Marine), as RIs listed on the ESFRI roadmap, as well as SeaDataNet as European marine data management infrastructure. The overall aim is to analyse the FAIRness of each of these RIs and to implement within each RI the necessary actions to improve its FAIRness. The latter is critical for the Marine subdomain as it will provide a coherent base for developing the integrated services systems required by a broad variety of research, regulatory and operational communities. Climate change and anthropogenic impacts are among the key issues that already affect and concern European and worldwide societies, but there are others that also have an enormous socio-economic impact (natural hazards, water quality, and acidification). Therefore, "Essential Ocean Variables" (EOV) as identified by the Global Ocean Observing System (GOOS) will provide guidance to the WP9 activities for setting priorities at data type level.

Considering the ENVRI-FAIR challenge of multiple RIs and multiple subdomains, the agreed way forward is that the FAIR principles will be implemented within each RI to improve RI FAIRness at three levels: 1) to better serve its own users; 2) to facilitate the development of cross RI services at Marine subdomain level; and 3) to facilitate the development of cross subdomain services at the level of ENVRI-FAIR cluster. The approach is bottom-up: respecting the autonomy of RIs concerning requirements and solutions, however in close and regular interaction with experts in ENVRI-F AIR about common standards, training, common implementation options at environmental domain level, and shared implementation activities with other subdomains.

An analysis process on FAIRness in the Marine RIs has been undertaken by the partners in the marine subdomain to come to a list of priorities to enhance RI data FAIRness. This list of priorities for enhancing their FAIRness has been drafted in a roadmap of actions and an implementation plan that is on its way till the end of 2020. The next step within ENVRI-FAIR project will be to setup an "EOV" demonstrator in the form of a metadata broker for the Marine subdomain (figure 1) that aims to serve SeaDataNet, CMEMS and EMODnet, as integrators and processors of Marine EOV data accessed via the RI machine-to-machines services.

Delivering Quality Marine Data and Services: the IODE Quality Management Framework

Greg Reed, UNESCO/IOC/IODE (Belgium), g.reed@unesco.org
Rob Thomas, Marine Institute (Ireland), rob.thomas@marine.ie

Marine data are fundamental to our understanding of the processes that control the environment. These data are a key requirement for effective strategic decision making and play an important role in promoting the development of economic activities and underpin many of our activities. Reliable marine data are key inputs to the efficient management of marine resources and it is essential that accurate data be accessible in a timely manner to facilitate informed decision making. The value of such data increases when they are provided by organizations that have established and adhere to quality management principles.

Quality management is the process for ensuring that all the activities necessary to design, develop, and deliver a product or service meet the requirements of the end-user. Quality management focuses not only on product and service quality, but also on the means to achieve it. Users of marine data, products and services are increasingly calling for quality management systems to be put in place to provide a level of confidence in quality.

The International Oceanographic Data and Information Exchange (IODE) programme of the Intergovernmental Oceanographic Commission (IOC) of UNESCO maintains a global network of more than 100 National Oceanographic Data Centres (NODC) and Associate Data Units (ADU) responsible for the collection, quality control, archive, and online publication of many millions of ocean and marine observations which are made available to Member States.

The IODE has long held the view that there is a need for a quality management framework to ensure that NODCs and ADUs are established and operate according to defined principles, including adherence to agreed standards and the requirements of the IOC Oceanographic Data Exchange Policy. This will ensure NODCs and ADUs are able to provide data of known quality to meet the requirements of a broad community of users.

The IODE has implemented a Quality Management Framework (IODE-QMF) to provide the overall strategy, advice and guidance for NODCs and ADUs to design and implement quality management systems for the successful delivery of marine and related data, products and services. The IODE-QMF addresses the implementation of quality management systems that will assure the quality of final data, products and services.

The IODE has implemented an accreditation process whereby NODCs and ADUs must satisfy a minimum set of requirements to ensure compliance with IODE standards and to establish a mechanism to regularly monitor and assess the quality of their data and services. IODE has been accredited as a WDS Network Member and is certified against the Network Member Catalogue of Criteria developed and managed by the WDS-SC. Network Members are bodies representing groups of data stewardship organizations, some of which may or may not be WDS Regular Members. IODE contributes to the WDS through its network of NODCs and ADUs and the IODE accreditation process will ensure these centres can demonstrate their capability to meet the WDS certification requirements.

In response to the IOC-IODE guidance and to the requirements of funding agencies, the Marine Institute of Ireland included “Quality” as a goal in its Data Strategy (2017-2020), with a target of achieving the IOC-IODE accreditation as the NODC for Ireland. The Marine Institute was awarded IOC-IODE accreditation in February 2019. In its submission to IOCE-IODE, Ireland’s National Marine Data Centre (hosted by the Marine Institute) included a Data Management Quality Management Framework (DM-QMF) model; a manual detailing this model and how it is implemented across the scientific and environmental data producing areas of the Marine Institute; and, at a more practical level, an implementation pack consisting of a number of templates to assist in the compilation of the documentation required by the model and the manual (Leadbetter, et al., 2019).

This paper discusses the key steps required to implement a quality management system and the accreditation process used to ensure NODCs and ADUs can demonstrate their capabilities to provide data, products and services in compliance with established standards and procedures. The real-world experience of the Marine Institute in developing a DM-QMF is presented along with a discussion of the ongoing implementation progress within the organisation 18-24 months after the DM-QMF was initially piloted.

References

Leadbetter, A., Carr, R., Flynn, S. et al. Implementation of a Data Management Quality Management Framework at the Marine Institute, Ireland. Earth Sci Inform (2019). https://doi.org/10.1007/s12145-019-00432-w

Evolving the UNESCO/IOC Ocean Best Practices System: preparing methods for the oceans' digital ecosystem

Pier Luigi Buttigieg, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (Germany), pier.buttigieg@awi.de
Cora Hörstmann, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (Germany), cora.hoerstmann@awi.de
Pauline Simpson, International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission. (Belgium) p.simpson@unesco.org
Jay S Pearlman, IEEE (France) jay.pearlman@ieee.org

Introduction

Best practices are one of the most valuable outcomes of collective human behaviour, encapsulating the culmination of years - perhaps centuries - of methodological development. Despite this, many best practices, along with the methods that precede them, are lost when projects end, generations turn over, or simply when URLs fail to resolve. Addressing this, the Intergovernmental Oceanographic Commission of UNESCO has deployed an Ocean Best Practices System (OBPS; Buttigieg et al., 2018; Pearlman et al., 2019). Here, we briefly describe subsequent developments in the system's technology (Buttigieg et al., 2019), outlining an ongoing co-development process to support the UN Decade of Ocean Science for Sustainable Development (Ryabinin, 2019).

Advancing metadata, structured documentation, and version control

Actionable metadata is at the heart of the OBPS. A revised set of mandatory fields and optional fields allows the OBPS' technologies to interlink content across its entire corpus, seeding new potentials in technical and human interoperability. Free-text entry fields (for e.g. SDG codes, EOV names, and spatial descriptors) are being substituted by those referencing external community vocabularies and ontologies to increase interoperability, in line with the vision of the Ocean Data and Information System. Key examples (see Buttigieg et al 2019) include 1) UN Environment's SDG Interface Ontology, 2) NERC Vocabulary server resources, and 3) the Environment Ontology which includes semantic references for the Essential Ocean Variables (EOVs). We constantly seek additional terminology resources, and invite the community to suggest new high-quality, marine-focused terminologies.

While metadata is essential, document text (with multi-lingual range) is the OBPS' primary resource. Parsing this corpus with natural language processing (NLP) is hindered by inconsistent document structure. In response, Hörstmann et al. (2020) have published guidance for future submissions to allow improved processing. More machine-friendly templates have been created (e.g. for sensors), with elements illustrated in Figure 1. Tabulated information will be used to auto-populate OBPS metadata, while prose will be augmented through semantic tagging.

Preserving the version history of a method, as it evolves into a best practice, is essential. With the recent upgrade of the OBPS' archiving software, we are implementing automated and user-controlled item-level versioning (Hörstmann et al., 2020). This will upgrade the current, manual approach and allow users to trace method evolution and maturity through multiple rounds of review.

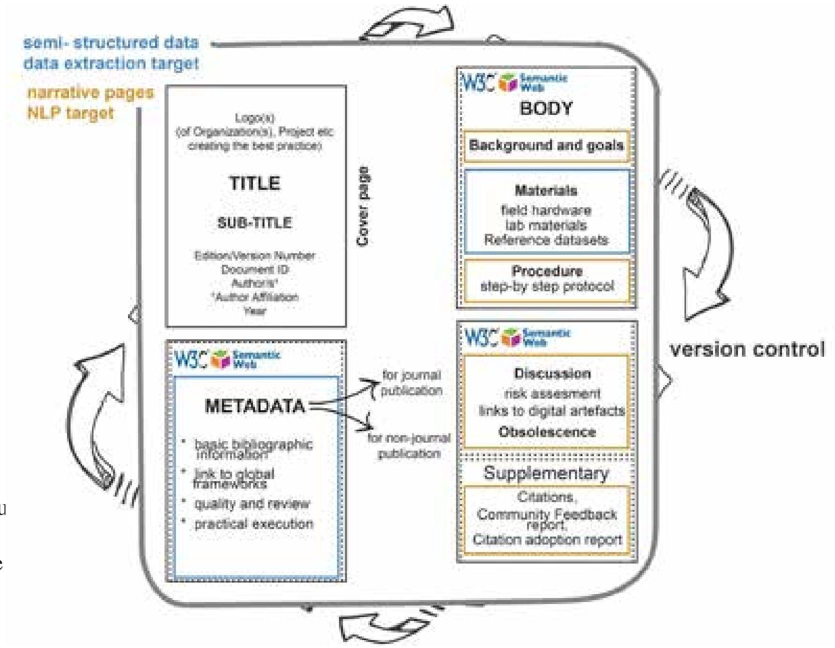


Figure 1: Illustration of the OBPS template (modified after Hörstmann et al. 2020). Blue boxes: document parts where semi-structured metadata are expected and mined. Yellow boxes: narrative content subject to text mining. Semantic web technologies using reference terminologies are used throughou to enhance interoperability. Version control systems preserve corpus history.

Seeding a new federation of interoperable document stores

The OBPS' open codebase is intended to support reuse by all to create interoperable document stores. As an example, the EU Horizon 2020 project CAPARDUS is creating an "Arctic Common Practices System" (ACPS) based on an Arctic Collection within the OBPS. An open co-development process will help meet the needs of Arctic stakeholders in their rapidly changing context.

Conclusion & outlook

The OBPS will continue to evolve through community input and partnerships in projects such as JERICO S3, EuroSea, OceanObs RCN and CAPARDUS. We invite ocean technologists, informaticians, and software developers to join our implementation team in advancing this system into a gold standard for handling methods and best practices in ocean sciences and applications. We also welcome the technical expertise of this community in identifying and endorsing methods as best practices to help guide the ocean community in the use of high-quality digital resources.

References

Buttigieg PL, Simpson P, Pearlman J et al. (2018) Technologies for a FAIRer Use of Ocean Best Practices. IMDIS 2018: International Conference on Marine Data and Information Systems. Barcelona, Spain: November 5-7.

Buttigieg PL, Simpson P, Caltagirone S, Pearlman JS (2019) The Ocean Best Practices System -Supporting a Transparent and Accessible Ocean. IEEE. OCEANS 2019 MTS/IEEE. Seattle, USA: October 27-31.

Hörstmann C., Buttigieg PL, Simpson P, et al. (2020) A Best Practice for Developing Best Practices in Ocean Observation (BP4BP): Supporting Methodological Evolution through Actionable Documentation. Paris, France, UNESCO. Intergovernmental Oceanographic Commission Manuals and Guides No. 84. (IOC/2020/MG/84).

Pearlman JS, Bushnell M, Coppola L et al. (2019) Evolving and sustaining ocean best practices and standards for the next decade. *Front Mar Sci.6*: 277

Ryabinin V, Barbière J, Haugan P, et al. (2019) The UN Decade of Ocean Science for Sustainable Development. *Front Mar Sci. 6*:470.

Digital twin of the North Sea, A visual geo tool with computer modelling to support stakeholder engagement, to aid decision makers and support citizen involvement.

Joan Staeb, marine information and data centre IHM (Netherlands), joan.staeb@rws.nl
Igor Mayer, BUAS (Netherlands), mayer.i@buas.nl
Yvonne Koldenhof, MARIN (Netherlands), y.koldenhof@marin.nl
Fedor Baart, Deltares (Netherlands), fedor.baart@deltares.nl
Petra Jeurissen, Rijkswaterstaat (Netherlands), petra.jeurissen@rws.nl
Peter Thijsse, MARIS (Netherlands), peter@maris.nl

Introduction

The North Sea is very busy and many users like fishery, wind farms, sea traffic, nature, military and recreation are requesting space. This leads to immense pressure on marine spatial planning. The project Digital twin of the North Sea (Digitwin) gives insight into the real data and provides easy modelling in an easy to use platform to assist in discussions with stakeholders and facilitate decision makers.

Digishape (digishape.nl) is a consortium of governments, businesses and research institutes that is promoting the use of digital techniques in the Dutch water sector. We experiment with open data in projects using new data and advanced IT techniques.

Scope

The project created a ready-to-use platform bringing together new technologies such as open source data and mapping services, cloud computing, gaming and virtual reality. This will bring a better understanding of the pressures all stakeholders experience, and will lead to better decision making

Methodology

The platform consists of a landing page where the three tools can be accessed. The Browser is the easiest and most familiar entry for all users. Depending on interest maps can be opened to see the different space claims in the Dutch North Sea (fig 1.).

Much effort was put into finding the best open data and maps and into presenting them in a user-friendly way. Models include a levelized cost of energy model for windfarms, a shipping safety model a fishery model and an ecological model.

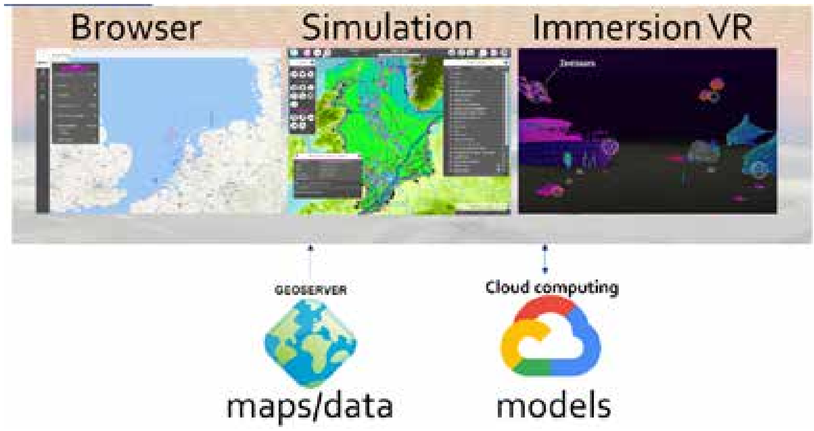


Figure 1: Architecture of the Digitwin. Maps and models feed the 3 user interfaces. A user can make changes in each of the user interfaces by e.g. building wind farms or changing shipping routes. The effects are calculated in the cloud and immediately visualised in the 3 user interfaces.

Results

Starting from the present situation the user is invited to claim new sea space e.g. for shipping or the building of a wind farm. It is likely that conflicts will arise with already existing uses of the sea. It is quite possible that there is already an oil platform or military use in that area. The platform takes all those maps and data from a central geoserver.

Next calculations and simulations are done using Google's high-performance cloud computing. The costs for energy production in new windfarms are calculated using state of the art and well accepted models from the Dutch wind industry. Shipping costs and safety is calculated using state of the art but simplified models from the Dutch institute Maritime Research Institute (MARIN). The platform is designed in such a way that different models and maps can be activated depending on the demands of the users.

A next level is introduced by the simulation platform. The same maps and computer models are available and give the same results. But in the simulation platform stakeholder processes can be played by multiple users. Discussions will emerge and the players must find solutions together or in competition. Conflicts with existing use of the sea cannot be simply neglected as is possible in the browser, and timing of decisions becomes crucial. The simulation tool really brings interaction at a higher level.

Finally, the virtual reality (VR) tool allows the user to immerse in the world of the sea. Citizens and politicians only rarely go out on the sea and have difficulty in understanding the challenges. VR helps you to see the outcome of decisions that are made. With 3D glasses on your head you can be the captain of a big oil tanker or even a sea bird trying to find its way in the turmoil of shipping lanes and windfarms.

Conclusions

An easy-to-use digital twin platform of the North Sea was created successfully. The platform can accommodate any map or model. So, the question is merely which maps and models you choose for your discussion or decision making. The Digitwin can be used in simple stand-alone way or in a more complex multiuser gaming mode. Depending on your question simple or more complex cloud computing models can be activated to evaluate scenarios and facilitate discussions.

Digital twins are still very new for many users. For further acceptance and effective use of digital twins, the maps and data must be reliable, and the models used must gain acceptance. Finally, more people should get acquainted with those tools by just playing with them and by using different VR tools and games that are now available for everyone.

Enhancing the technical architecture of the Unified State Ocean Information System (ESIMO) through the use of a cloud platform and digital technologies

Sergey Belov, RIHMI-WDC (Russian Federation), belov@meteo.ru
Nikolai Mikhailov, RIHMI-WDC (Russian Federation), nodc@meteo.ru
Alexey Kozlovcev, RIHMI-WDC (Russian Federation), akozlovcev@meteo.ru

The Unified State Ocean Information System (ESIMO) is a state interagency information system aimed at solving following problems: the integration of information about the World ocean state, received from the information systems of the federal bodies; provision of comprehensive information about the situation in the oceans to the state authorities of the Russian Federation, as well as to legal entities and individuals engaged in maritime activities; information exchange with international systems.

As of 2020, ESIMO is integrating more than 250 databases on more than 450 environment parameters in the World Ocean (more than 3,500 so-called "information resources") provided by over 20 departmental information systems. The total amount of data from the ESIMO information resources is about 15 terabytes. The system is composed of a global node (Moscow, Roshydromet), two regional nodes (Far East regional node in Vladivostok and northwestern node in Saint Petersburg responsible for the Arctic region), several specialized nodes in ministries, and departmental nodes around the country in data provider organizations.

ESIMO is a distributed system that provides metadata and data exchange between the nodes of a unified system. ESIMO applies following interoperability standards - ISO 19115 for metadata, NetCDF 2/3 for data, SOAP, REST, OGC, and SPARQL for the web services. Later developments include the wider user of Semantic Web standards such as OWL and RDF.

The hardware and software of ESIMO was formed in 2011–2013 under the federal target program "World Ocean". At present, about 30 percent of the equipment has failed completely; periodic failures in existing facilities are registered. Many efforts are required to maintain human resources in the field, capable of providing technical support. The most relevant requirements for the modernization of ESIMO in technical terms are:

- use of modern digital technologies (cloud computing, big data, etc.) and last generation software and hardware at the data centers operating in the Russian Federation to ensure the sustainable operation of the system;
- wider adoption and use of interoperability standards;
- use of open / free and shareware software products, dominantly developed and maintained in Russia;
- updating of middleware and specialized software;
- deployment of information security tools in accordance with the requirements of the Federal Service for Technical and Export Control of Russia;

The composition and quality of modernized hardware and software should be sufficient to ensure the operation of a unified system, including: a) information interaction with more than 350 departmental databases and ensure their availability; b) data integration and maintenance of the consolidated information base of at least 4,500–5,000 data sets; c) information services for at least 100 users/per moment, more than 5000 calls daily, data delivery to end-users in the amount of 350 GB or more per month.

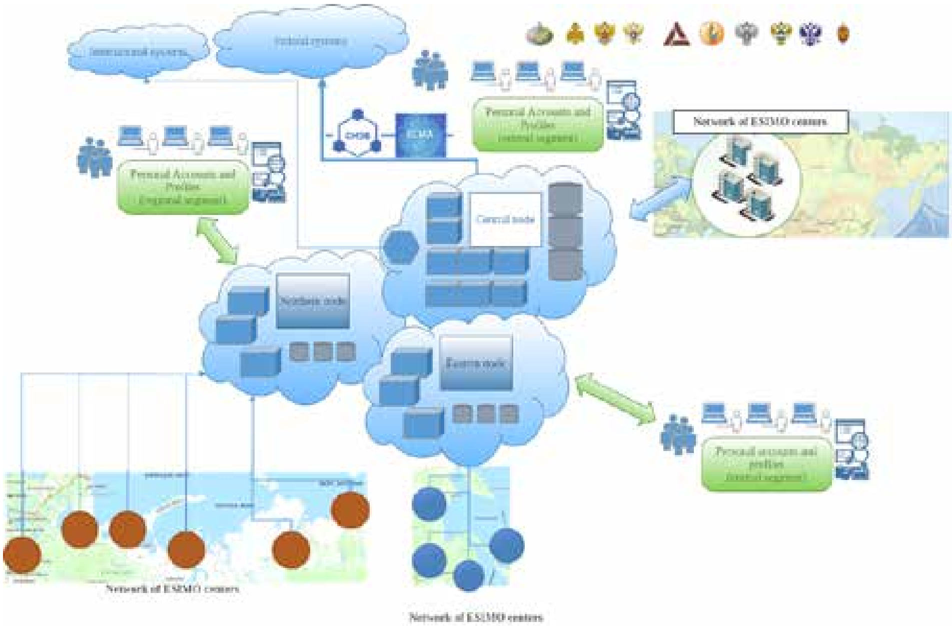


Figure 1: Generalized architecture of the modernized ESIMO

The concept of modernization of ESIMO is based on the transformation of a unified system as an interdepartmental state system of information on the situation in the oceans into an interdisciplinary digital platform focused on integrated information support of the main types of maritime activities in the Russian Federation. From an information point of view, a digital platform is an integrated distributed repository of observations, information products and services in the field of the situation in the World Ocean, to which effective access of information consumers is provided.

The digital platform of ESIMO will support a variety of communication interfaces between information providers and consumers. It will also enhance interoperability arrangements (new services and protocols), including semantic interoperability through the wider use of Semantic Web and Linked Open Data techniques. Services will be provided to consumers in a "one-stop-shop" manner. Personalization of ESIMO data and services is foreseen as: from end-user profiles (user groups) of "workplace" type on stationary / mobile devices to a digital platform (in fact, a profile of an ESIMO digital platform) for an enterprise or industry.

POSTERS

Maritime Spatial Planning INSPIRE data model

Andrej Abramic, ECOAQUA, University Las Palmas de Gran Canaria (Spain), andrej.abramic@ulpgc.es
Alejandro Garcia Mendoza, ECOAQUA, University Las Palmas de Gran Canaria (Spain), alejandro.garcia@ulpgc.es
Olvido Tello Antón, Instituto Español de Oceanografía (Spain), olvido.tello@ieo.es
Gerardo Bruque Carmona, Instituto Español de Oceanografía (Spain), gerardo.bruque@ieo.es
Luis Miguel Agudo, Instituto Español de Oceanografía (Spain), luismi.agudo@ieo.es
Ricardo Haroun, ECOAQUA, University Las Palmas de Gran Canaria (Spain), ricardo.haroun@ulpgc.es

The Maritime Spatial Planning (MSP) INSPIRE data model concept has been developing from 2014, applying Infrastructure for spatial information in Europe Directive 2007/2/EC (INSPIRE) data management concepts for marine planning, through the Marine Pilot project (EC Joint Research Centre 2014-2016) and continuing with the PLASMAR project (INTERREG–V 2017-2020). The results and findings delivered have been published in the paper "Maritime spatial planning supported by infrastructure for spatial information in Europe (INSPIRE)" (Abramic et al., 2018).

Currently, there are difficulties in harmonising products, visions, maps and frameworks of maritime spatial plans delivered by countries sharing the same marine (sub)region. This is mainly due to the fact that maritime plans do not use a common symbology and data structure to describe maritime activities. A solution for this issue is to apply on a marine spatial plans, INSPIRE standards for data sets, layers and portrayals.

The MarSP project was a perfect opportunity to finalise conceptual data model development and, what is more important, to test results applying it on the real use cases, developed in the Macaronesia (Azores, Madeira, Canaries) MSP process.

Initially, the INSPIRE data model for terrestrial planning (*Planned Land Use*, Figure 1) was tested to see if it could be applied for MSP. Tests pointed out that the terrestrial data model is robust, and can map MSP's, but it tends to lose detail and specific information on marine uses. To be applied for MSP, the *Planned Land Use* data model needs to be adapted for planning of the maritime activities in the marine space.

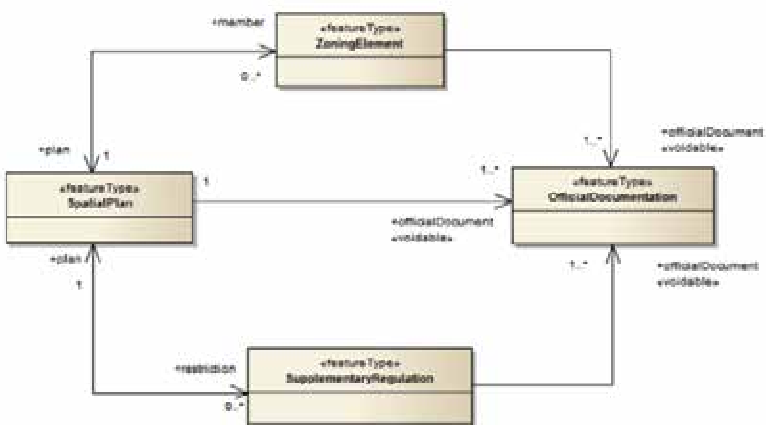


Figure 1: Overview of the INSPIRE "*Planned land use*" conceptual data model adapted for MSP. This figure is extracted from the INSPIRE UML data model repository publicly available in HTML format at http://inspire.ec.europa.eu/

Conceptual model was analyzed, adapted, applying data modeling techniques, adjusting for MSP requirements:

1. Developed conceptual MSP data model, extending *Planned Land use*, using Unified Model Language (Figure 2);
2. Extending spatial scope of the data model - from two-dimensional land planning to the three dimensions planning. Extended structure includes maritime activities within the sea surface, water column, seabed and subsoil, when *land model* consists mainly of land surface planning;
3. Developed specific maritime uses classification (including register), extending Hierarchical INSPIRE Land Use classification (HILUCS);
4. Developed MSP data model templates, using simplified and feature complex spatial data architectures. Different type of codification templates, for advanced, standard and rookie GIS users (gml, GeoPackage, Shape file, available at Canaries MSP platform);
5. Styled Layer Descriptor (color & symbol layout) for MSP, based on International Hydrographic Organization standards.
6. Data specification document v1.0 for Maritime Spatial Planning INSPIRE data model

MarSP 2nd capacity building workshop was a great opportunity to test MSP data model results. Participants were trained on how to apply MSP data model on selected use case (Madeira MSP draft), during the "hands on" session, discussing potential issues and technical solutions.

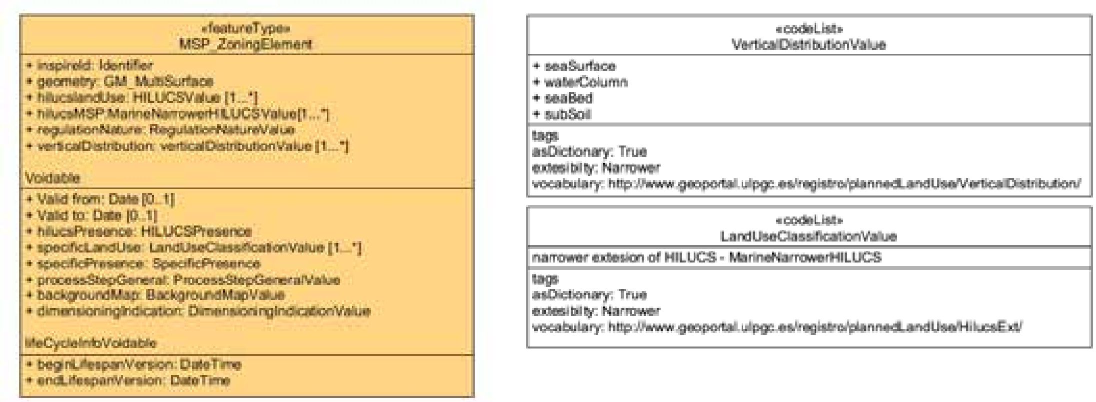


Figure 2: MSP INSPIRE conceptual data model

Cataloguing Ocean Data at Web Scale

Adam Leadbetter, Marine Institute (Ireland), adam.leadbetter@marine.ie
Andrew Conway, Marine Institute (Ireland), andrew.conway@marine.ie
Tara Keena, Marine Institute (Ireland), tara.keena@marine.ie
Will Meaney, Marine Institute (Ireland), will.meaney@marine.ie

In order to allow users of marine data to discover (or "F"ind, Wilkinson et al., 2016) the datasets which meet their needs, a data cataloguing solution is required (Leadbetter, et al., 2020). Data catalogues should then point users to accessible data download services and where possible the catalogues should use common vocabularies for interoperability and best practice metadata profiles in defined standards. Data cataloguing may take place at an institutional or national level, and for needs such as marine spatial planning or reporting under the Marine Strategy Framework Directive. The datasets described by the local catalogues may be then aggregated to a regional or an international scale.

As such, Data Catalogues are often required to feed into other, aggregated Data Catalogues. For this to be achieved, the base metadata schema of the Data Catalogue system must be mapped in a crosswalk to the metadata schema of the target Data Catalogue. This will provide syntactic interoperability and, if controlled vocabularies are used to populate the fields of the source Data Catalogue and mapped to the controlled vocabularies of the target metadata schema, semantic interoperability (Schaap and Lowry, 2010). Where the aggregation takes place in a more generic data portal, such as a data.gov portal, or uses a generic metadata profile, such as Schema.org, the crosswalk may lead to some loss of detail compared with the source metadata.

In the past, metadata validation, crosswalks and other related tasks have been achieved either through one-off calls to scripts or services or through batch runs of jobs on a schedule. However, applying principles from modern software engineering approaches to these data management tasks yields an alternative, web-scale approach to both metadata publishing and metadata engineering tasks.

Continuous Integration (CI) is a process by which a team of software developers contribute changes to a single working copy of a code base. CI is reliant on a source control system being used to manage the code base. Once code is committed to a CI pipeline, Continuous Delivery tools are used to build the software product into a deployable artifact. If the build fails, the developers will receive warnings as to why but the previous, fully built version of the software artifact will remain available. Once a build is completed, Continuous Deployment tools can seamlessly push the new software artifact to users.

The Marine Institute, Ireland has undertaken a pilot project applying this paradigm to data managers who are responsible for metadata cataloguing. The data managers are given access to a source control repository, through the GitHub platform. A data manager may commit a completed metadata record in ISO19139 XML format to a folder in the source control repository. Once the new metadata record, or records, have been pushed to the source control repository, a number of Continuous Delivery tasks are automatically started through the TravisCI platform. These tasks include: generating DataCite metadata kernels for use in minting digital object identifiers for data citation; producing HTML landing pages with Schema.org annotations (Leadbetter, et al., 2018) to allow for indexing in Google's Dataset Search; and creating Global Biodiversity Information Facility records. Continuous Deployment is achieved through the TravisCI tasks commuting back to the source code repository and general access to HTML representations of the metadata through web hosting via GitHub Pages.

The Continuous Delivery tasks have been scripted in Python, and are based around a class which has been developed to provide access to the information held within a metadata record. The various target export formats are templates within a templating framework, such as Jinja. Through Python code, the templates have access to data from instances of the metadata class.

Whilst the approach described in this paper has been undertaken as a proof-of-concept, it is anticipated that it will become operational over time and the approach could be used for future aggregations such as the Intergovernmental Oceanographic Commission's Ocean Data and Information System (http://odis.iode.org/) proposed under the UN Decade of the Ocean.

References

Leadbetter, A., Thomas, R., Shepherd, A., Fils, D. And O'Brien, K. (2018). The place of Schema.org in Linked Ocean Data.

Leadbetter, A., Meaney, W., Tray, E., Conway, A., Flynn, S., Keena, T. Kelly, C. and Thomas, R. (2020). A modular approach to cataloguing marine science data. Earth Science Informatics. https://doi.org/10.1007/s12145-020-00445-w

Schaap, D. and Lowry, R. (2010). SeaDataNet - Pan-European infrastructure for marine and ocean data management: Unified access to distributed data sets. International Journal of Digital Earth 3(Sup. 1): 50-69.

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18

Marine Hydrophysical Institute and All-Russian Research Institute of Hydrometeorological Information – World Data Center: Comparison and Development of Black Sea Databases

Eugeny Godin, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), godin_ea@mhi-ras.ru
Alexander Vorontsov, All-Russian Research Institute of Hydrometeorological Information – World Data Center (Russia), av10921@meteo.ru
Andrey Ingerov, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), ingerov@mhi-ras.ru
Elena Isaeva, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), isaeva-ea@mhi-ras.ru
Tatiyana Kasyanenko, Marine Hydrophysical Institute, Russian Academy of Sciences (Russia), kas_tanya@ mhi-ras.ru

The Black Sea is one of the most studied areas of the World Ocean by the number of oceanographic stations accomplished there. At present, there is a series of Black Sea oceanographic databases, which were developed within both national and international projects. Databases created in All-Russian Research Institute of Hydrometeorological Information – World Data Center and Marine Hydrophysical Institute, Russian Academy of Sciences, should be attributed to the most comprehensive ones. Joint work to compare the Black Sea oceanographic data arrays held in the institutions has been carried out since 2017. A preliminary analysis showed that the two arrays differ not only in total number of stations (148,329 stations for 1884–2015 and 156,980 stations for 1890–2015, respectively) but in their distribution both in time and throughout the Black Sea.

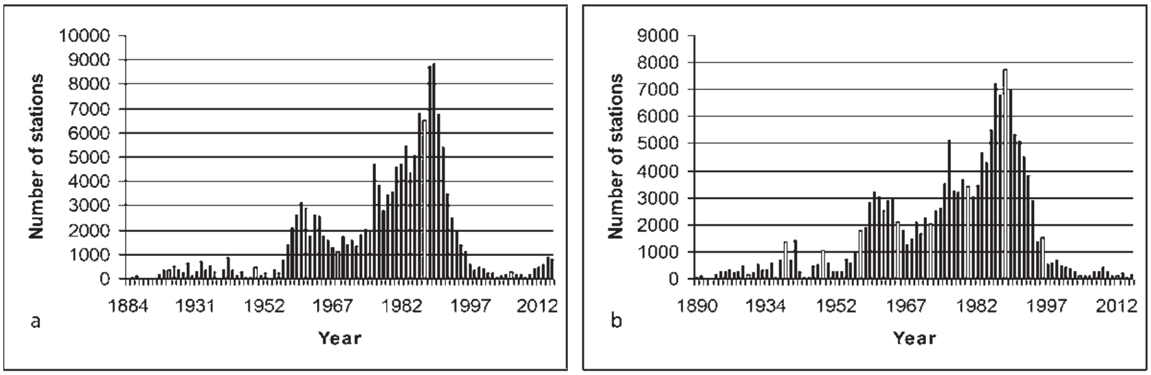


Figure 1: Distribution of oceanographic stations on years (a – All-Russian Research Institute of Hydrometeorological Information – World Data Center, b – Marine Hydrophysical Institute, Russian Academy of Sciences)

Special software was developed using Lazarus IDE in the Oceanographic Data Bank of Marine Hydrophysical Institute to optimize the procedure of comparing the two data arrays. The software allowed to look through metadata, visualize them on a map, and, based on comparing coordinates and time of stations and number of measurements, set “flags” to mark stations as “doubles” or “doubles in question”. Data marked as “doubles” or “doubles in question” (after further consideration) compose the overlapping part of the databases. It is kept as an array of coupled numbers of coinciding stations. After that, data, which do not coincide (“unique” data), are detected by exclusion. However, its use showed that the comparison of only metadata did not give sufficient information to classify stations as “unique” or “doubles”. A new version of the software was designed to eliminate the arising ambiguities. It offers more functions, including those of selecting, displaying, and comparing both metadata and data. The current version of the software provides:

- Metadata selecting both by cruises (data sets) and by spatial and temporal criteria, as well as viewing them as graphs and tables
- Displaying lists of measured parameters for a couple of stations under consideration from the compared arrays; presenting corresponding vertical profiles for a chosen parameter
- Possibility of shifting one of the arrays in time by a value determined by user (it makes the comparison easier if one array contains UTC/GMT and the other has local time), etc.

At the same time, it was found out that in the compared arrays data completeness of stations with identical time and coordinates often varies. The software allows creating a file that contains information on measured parameters, depth range, and number of levels in each profile. The information can be used to form the most complete data set for every station of that kind.

In future, forming a joint oceanographic database on the Black Sea seems to be expedient. According to preliminary estimates, the database will include above 185,000 oceanographic stations that will serve as an informational basis for analysis of climatic changes in the Black Sea and creation of various informational products using great experience in this field accumulated by the two institutions.

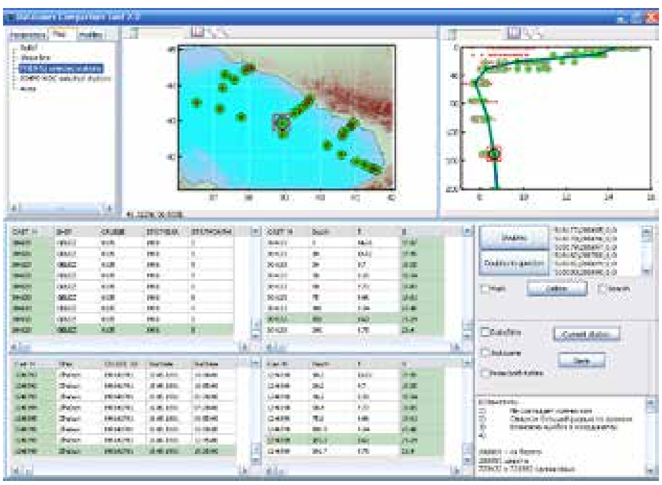


Figure 2: A screenshot of the software operation

The work is carried out in the framework of the contract on creative collaboration between Marine Hydrophysical Institute, Russian Academy of Sciences, and All-Russian Research Institute of Hydrometeorological Information – World Data Center, and within the state task on theme No. 0827-2018-0002 “Operative Oceanology”.

Data Interoperability in the French Marine Environmental Information System (SIMM)

Clémence Rabévolo, French Research Institute for Exploration of the Sea (France), clemence.rabevolo@ifremer.fr
Steven Piel, French Agency of Biodiversity (France), steven.piel@ofb.gouv.fr
Armelle Rouyer, French Research Institute for Exploration of the Sea (France), armelle.rouyer@ifremer.fr

SIMM ("Système d'Information pour le Milieu Marin") is the Marine Environmental Information System created in 2018 at the request of the French Ministry for Ecological and Inclusive Transition. SIMM was created to support public decision-making, to provide France citizens with information and to support European data policies related to the marine environment such as Regional Sea Conventions or European Directives including: MSFD (Marine Strategy Framework Directive), WFD (Water Framework Directive), and MSP (Maritime Spatial Planning).

SIMM provides a unique framework for bringing together standardized descriptions and for providing access to marine datasets processed in various contexts including observation data from monitoring centers, data products from GIS maps and models, and indicators oriented toward public policy decisions such as the indicators required by MFSD. SIMM manages both environmental and human activities data (fisheries, marine renewable energies, marine traffic, etc.).

Marine data in France are provided by many producers and managed through various databases and information systems. The SIMM portal "milieumarinfrance.fr" is a public service, which allows for sharing and retrieving, in a one place, open access marine data for scientists, environmental stakeholders and the general public.

In the European Marine Strategy Framework Directive (MSFD), significant advances have been achieved both at national and transnational levels including data policies, sharing of environmental data and in the technical implementation of distributed spatial data systems (metadata management, portal accessing distributed repositories, etc.). The creation of a new national information system offers the opportunity for integrating all current good practices in marine data management.

Data repositories: SIMM "common language"

SIMM governance has established an authority dedicated to interoperability: the Repository Administration Service (SAR). SAR's roles are to manage the SIMM repositories and to implement and develop the technical means to provide a common language within SIMM's framework, which brings together more than 20 information subsystems. SAR's goal is to enable the interoperability of databases within SIMM while ensuring that it is connected with other systems (federal, public, or international systems).

SAR is required to provide common vocabularies to describe various parameters, such as a taxonomy for species as well as physics, chemistry or biology observation parameters. SAR is careful not to re-create existing data repositories if the already existing ones match with marine environmental community needs. As such, through a detailed analysis of existing repositories in other national or international organizations, SAR chooses the one repository that best suits

SIMM needs. If no repository already exists, SAR keeps open the opportunity to create a new one. In any case, compliance with international standards is a key point for the success of the “common vocabulary” repositories integration.

Thus, SAR must ensure the interoperability of SIMM with international standards, such as the ones developed under MSFD. Therefore, SAR has been commissioned to represent France at the Technical Working-Group on Marine Data of MSFD in order to integrate their recommendations and best practices in SIMM repositories.

Data models

SAR is also in charge of data models management, which allows databases to organize their data to encourage exchanges and reuse at least at the national level. For instance, SAR applies ISO and OGC standards, and INSPIRE and IHO recommendations about data models. A modeling software such as Enterprise Architect is used in order to be able to exchange models with other standardization organizations.

Cooperation between French standardization organizations

SAR works closely with the other French standardization bodies. For example, a working group has been set up with SANDRE (equivalent of SAR for the French Fresh Water Information System, established 20 years ago) and CARET (equivalent of SAR for the French Biodiversity Information System, established in early 2020) with the aim of sharing their knowledge and experience in order to improve their repositories management. These organizations work together to put in place innovative tools, and to integrate and promote international recommendations into their practices. They cooperate to follow up on international recommendations, and to present their work and needs on repositories to international working groups. Finally, this working group is helped by the “pôle INSIDE”, which offers its expertise on interoperability standards and innovative tools and advises on good practices in this area.

A principle of subsidiarity exists among these three organizations: SAR can choose to use a repository managed by SANDRE and vice versa, and it can ask for minor changes such as adding international identifiers. If each information system has its own repository on the same objects, cross-coding tables are set up in order to be interoperable.

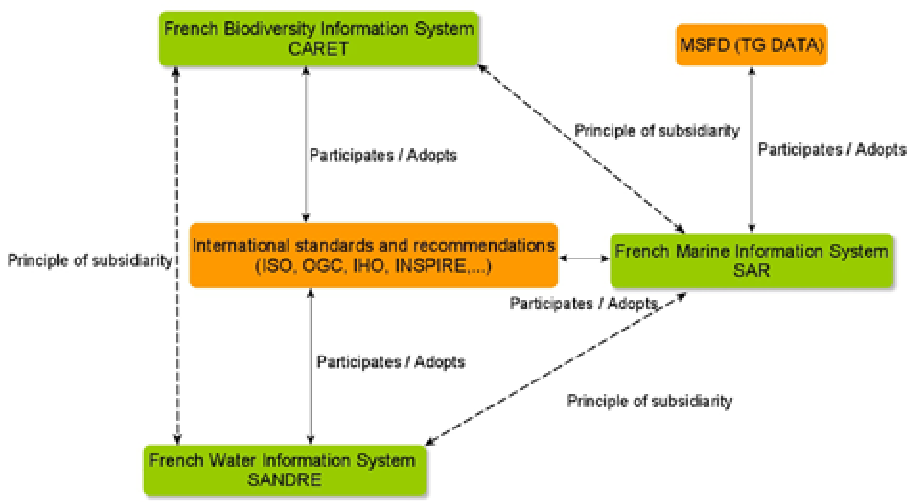


Figure 1: Links between the SAR and other standardization organizations

An open-source database model and collections management system for fish scale and otolith archives

Elizabeth Tray, Galway-Mayo Institute of Technology (Ireland), Elizabeth.Tray@gmit.ie
Adam Leadbetter, Marine Institute (Ireland), Adam.Leadbetter@Marine.ie
Will Meaney, Marine Institute (Ireland), Will.Meaney@Marine.ie
Andrew Conway, Marine Institute (Ireland), Andrew.Conway@Marine.ie
Caoimhín Kelly, Marine Institute (Ireland), Caoimhin.Kelly@Marine.ie
Elvira de Eyto, Marine Institute (Ireland), Elvira.Deeyto@marine.ie
Siobhan Moran, Marine Institute (Ireland), Siobhan.Moran@Marine.ie
Niall Ó Maoiléidigh, Marine Institute (Ireland), Niall.Omaoileidigh@marine.ie
Deirdre Brophy, Galway-Mayo Institute of Technology (Ireland), Deirdre.Brophy@gmit.ie

Biomineral samples, such as scales and otoliths (ear stones), from fish are routinely sampled for age estimation and fisheries management purposes. Fish scales and otoliths can be used to answer a diverse range of ecological research questions using both established and advanced analytical techniques. Measurements of growth marks in calcified structures provide a proxy for fish growth, and can be used to assemble extended biochronology time series and to examine relationships with climatic and fishing related drivers (Morrongiello et al. 2012). Studies have used trace elements and isotopes in the inorganic portion of biomineral structures to infer fish origin (Adey et al. 2009) and habitat usage (Brennan et al. 2019). Continuous sampling programs and opportunistic sampling of fish scales and otoliths can create invaluable, albeit irregular, archive collections. A common problem associated with biological collections, is that while sample intake grows, long-term physical storage is rarely a priority. The maintenance of scale and otolith collections can change depending on management and project-specific research needs (Rivers & Ardren, 1998). Essential fish metadata (**Figure 1**) is often recorded in a field notebook or a digital spreadsheet, which is rarely stored with the associated physical samples, and if it is, it can become separated after time. Propriety software is currently available for implementing collection management (CM) systems, but it can be expensive, dependent on ongoing investment, and limited in scalability.

Global legislative efforts to ensure that research data remains accessible have contributed to significant improvements in data quality and interoperability (Clarke & Magetts et al. 2014). At the European regional level, the INSPIRE Directive is a legislative driver for governments to construct standardized spatial data infrastructure, specifically for environmental data (Craglia & Annoni et al. 2007). To comply with INSPIRE regulations, data infrastructures must contain the flexibility to align within an Observations and Measurements (O&M) model, (Cox et al. 2016). Furthermore, other standards structures have gained widespread community usage, such as biodiversity data standards, like Darwin Core (Wieczorek et al. 2012). Recently, the FAIR (Findable Accessible Interoperable and Reusable) open-data principles (Wilkinson et al. 2016), were created through a bottom-up, community based initiative. A component of the

FAIR principals requires the usage of existing standards (e.g. INSPIRE, O&M, Darwin Core). Consequently, scientists have implemented FAIR data repositories specifically intended for sample and data management (Conze et al. 2017). However, a FAIR compliant system and model for local repositories, purposefully designed for physical ecological sample archives and their derived data, has yet to materialize.

The Marine Institute (MI) in Ireland, recently constructed an INSPIRE compliant Data Catalogue for its environmental data (Leadbetter et al. 2020). The CM system presented in this study is an extended feature of the Data Catalogue described by Leadbetter et al. 2020. The CM system was explicitly designed to house a fish scale and otolith collection, which consists of thousands of samples, dating back to 1928, from a range of geographic locations. The aim of this study is to lay a road map, and provide a toolbox, for other aquatic agencies who wish to create an 'open' digital and physical infrastructure for their own biomineral archives. The system utilizes the FAIR (Findable Accessible Interoperable and Reusable) open-data principles, and includes a physical repository, sample metadata catalogue, and image library.

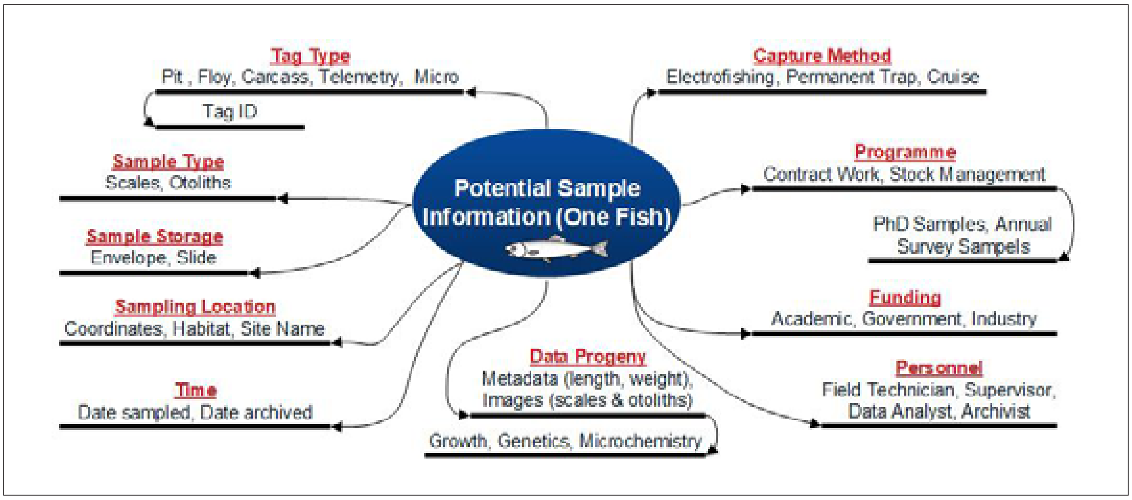


Figure 1: A mind map of the potential information associated with a fish scale or otolith sample.

References:

ADEY, E.A., BLACK, K.D., SAWYER, T., SHIMMIELD, T.M. AND TRUEMAN, C.N., 2009. Scale microchemistry as a tool to investigate the origin of wild and farmed Salmo salar. Marine Ecology Progress Series, 390: 225-235. https://doi.org/doi:10.3354/meps08161.

BRENNAN, S. R., SCHINDLER, D. E., CLINE, T. J., WALSWORTH, T. E., BUCK, G., AND FERNANDEZ, D. P., 2019. Shifting habitat mosaics and fish production across river basins. Science, 364(6442): 783–786. https://doi.org/10.1126/science.aav4313.

CLARKE, A. AND MARGETTS, H., 2014. Governments and citizens getting to know each other? Open, closed, and big data in public management reform. Policy & Internet, 6(4): 393-417. https://doi.org/10.1002/1944-2866.POI377.

CONZE, R., LORENZ, H., ULBRICHT, D., ELGER, K. AND GORGAS, T., 2017. Utilizing the International Geo Sample Number Concept in Continental Scientific Drilling During ICDP Expedition COSC-1. Data Science Journal, 16:2. http://doi.org/10.5334/dsj-2017-002.

Craglia, M. and Annoni, A., 2007. INSPIRE: An innovative approach to the development of spatial data infrastructures in Europe. Research and theory in advancing spatial data infrastructure concepts, 93-105.

Leadbetter, A., Meaney, W., Tray, E., Conway, A., Flynn, S., Keena, T., Kelly C., and Thomas, R., 2020. A modular approach to cataloguing marine science data. Earth Science Informatics. https://doi.org/10.1007/s12145-020-00445-w.

Morrongiello, J. R., Sweetman, P. C. and Thresher, R. E., 2019. Fishing constrains phenotypic responses of marine fish to climate variability. Journal of Animal Ecology, 88(11): 1645-1656. https://doi.org/10.1111/1365-2656.12999.

Rivers, P. J. and Ardren, W. R., 1998. The value of archives. Fisheries, 23(5): 6-9. https://doi.org/10.1577/1548-8446(1998)023<0006:TVOA>2.0.CO;2.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E. and Bouwman, J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3. https://doi.org/10.1038/sdata.2016.18.

Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T. and Vieglais, D., 2012. Darwin Core: an evolving community-developed biodiversity data standard. PloS one, 7(1). https://doi.org/10.1371/journal.pone.0029715.

Automatically generating ISO 19115-1:2014 metadata from relational databases

Linda Baldewein, Helmholtz-Zentrum Geesthacht (Germany), linda.baldewein@hzg.de
Dietmar Sauer, Helmholtz-Zentrum Geesthacht (Germany), dietmar.sauer@hzg.de
Ulrike Kleeberg, Helmholtz-Zentrum Geesthacht (Germany), ulrike.kleeberg@hzg.de
Housam Dibeh, Helmholtz-Zentrum Geesthacht (Germany), housam.dibeh@hzg.de
Lars Möller, Helmholtz-Zentrum Geesthacht (Germany), lars.moeller@hzg.de

Long-term storage solutions for scientific data are currently multiplying at an enormous rate as more and more journals require Digital Object Identifiers for the data used in their publications. The treatment of metadata, however, is somewhat lagging behind. Each data publisher requires different metadata fields and only few of them follow international standards. At the same time a variety of metadata catalogue systems, such as GeoNetwork, Geoportal, etc. exist. These catalogues provide standard interfaces to access, search and harvest the imported metadata, but simple tools to automatically generate the import files are missing.

At the Helmholtz Coastal Data Center (HCDC) we developed a method to extract all relevant information for the ISO 19115-1:2014 standard from different relational databases and to automatically generate XML metadata files for our datasets. Figure 1 gives an overview of the process. The relational databases contain measurement values, of for example biogeochemical samples. In addition, a lot of metadata information is stored across the various database tables, such as contact information, measurement parameters, units and methods and general information on the dataset, such as the campaign. Every table is checked for relevant metadata information. Each of the found attributes will be used to fill an ISO 19115-1:2014 element.

The metadata fields are loaded through SQL calls into the Feature Manipulation Engine (FME). FME is a software product to connect different data sources, manipulate their features and generate new results through automated workflows.

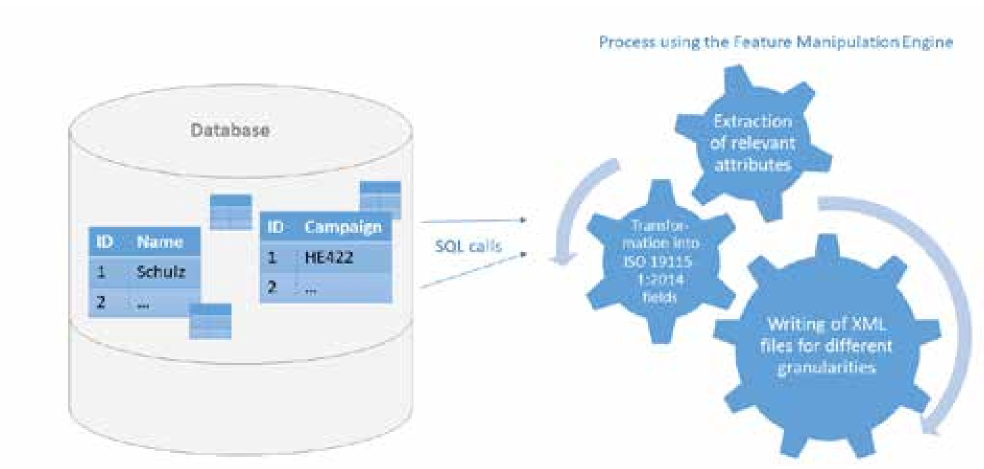


Figure 1: Generating XML files from database tables using the Feature Manipulation Engine

Using FME, the attributes gathered from the database are aggregated, transformed and mapped to the relevant metadata fields in the ISO 19115-1:2014 standard. For example, contact information is mapped to the field CI_ResponsibleParty and descriptions of parameters as well as standardized parameter names from controlled vocabularies, such as the ones provided by SeaDataNet, are mapped to MD_Keywords. The most evolved part of the process is the automated generation of the abstract of the dataset, which is part of the field CI_Citation. A generic template for all datasets is developed, that summarizes all known information of the campaigns, such as their purpose, and the involved projects. It also includes information on Digital Object Identifiers of the dataset.

As we have seen in the previous paragraphs, the metadata files are filled based on the information stored in the database. Thus, increasing the amount of such information stored is important. The sooner details about a measurement, like the coordinates of a station, are digitalized, the better the resulting metadata files will be. Since 2018 we have been generating metadata of campaigns already in the field, while the scientists are still on the research vessel. All campaigns have been equipped with high accuracy GNSS receivers and water proof tablets to electronically gather metadata. Using an app called Survey123 by Esri, the scientist fills in a questionnaire asking for example for the name of the station, the time of the sampling and the coordinates. The latter is directly received from the GNSS receiver, which is connected to the app to get the location of each sampling site with approximately one meter accuracy. The results of the questionnaire are then transferred via the mobile network from the tablet to a cloud storage. From there it is imported into the relational database, after carrying out quality checks (see Figure 2).

Figure 2: Metadata workflow from the field to the metadata catalogue

Once all metadata elements are mapped to the appropriate ISO 19115-1:2014 fields, an ISO 19139 XML file is generated by FME. This process is repeated for different granularities, such as each campaign and each project. The output XML files are stored in a folder, which is regularly harvested by the metadata catalogue system. A secondary batch process is set up to ensure that outdated metadata records are purged from the catalogue by directly deleting the old entries from the index. In our case of the metadata catalogue system Geoportal, old entries are removed from the open-source Elasticsearch instance and replaced by new information.

The automated process described above has been adapted to different relational databases, storing different data types, for example biogeochemical data and real time observational data. It has proven to be easily adjustable to new situations. After an initial setup phase, it automatically generates ISO 19115-1:2014 conform metadata files. In the future it might be possible to develop a user friendly generic tool from the described process that generates metadata files from any coastal research database.

Sextant, a Marine Spatial data infrastructure: implementation of OGC protocols for the dissemination of Marine data

Mickael Treguer, Ifremer (France), Mickael.Treguer@ifremer.fr
Catherine Satra Le Bris, Ifremer (France), Catherine.Satra@ifremer.fr
Erwann Quimbert, Ifremer (France), Erwann.Quimbert@ifremer.fr
Julien Meillon, Ifremer (France), Julien.Meillon@ifremer.fr

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, in order to manage, share and retrieve these products for its partners and the general public. Thanks to the OGC and ISO standard 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 DG research (SeaDataNet), DG Mare (EMODNET) and DG Growth (Copernicus MEMS).

The discovery service of Sextant is based on the metadata catalogue. The data description is normalized according to ISO 191XX series standards and Inspire recommendations. Access to the catalogue is provided by the standard OGC service, Catalogue Service for the Web (CSW 2.0.2).

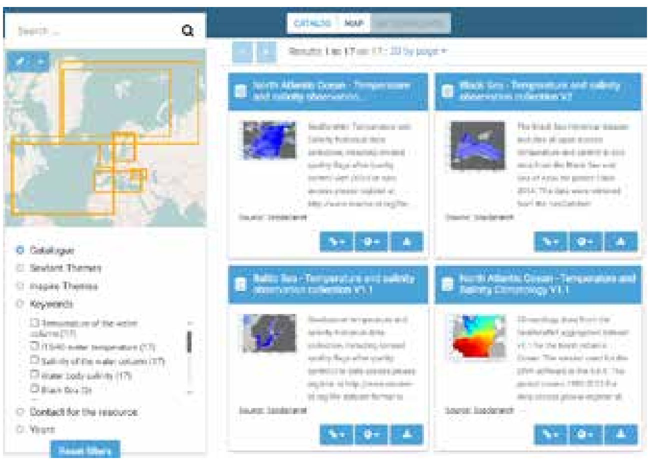


Figure 1: SeaDataNet content catalogue

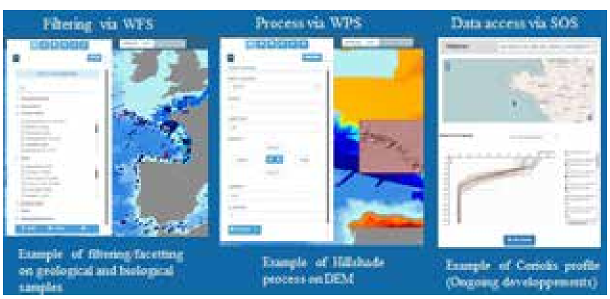


Figure 2: Data analysis

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 developped to improve the visualization, analyse and access to data, eg : data filtering, online spatial processing with WPS services and acces to sensor data with SOS services and SensorThings API.

On using a Sensor Observation Service as an INSPIRE-compliant download service

Nicolas de Ville de Goyet, Royal Belgian Institute of Natural Sciences (Belgium), ndeville@naturalsciences.be
Michael Fettweis, Royal Belgian Institute of Natural Sciences (Belgium), mfettweis@naturalsciences.be
Thomas Vandenberghe, Royal Belgian Institute of Natural Sciences (Belgium), tvandenberghe@naturalsciences.be

Due to the use of sensors, the volume of scientific data produced every day has become massive,, so there is a strong need to organize them and to set up a data infrastructure for their efficient management. Open access, FAIRness (Findable, Accessible, Interoperable and Re-usable) and INSPIRE-compliance are increasingly becoming the norm for (environmental) data management. In order to achieve complex aspirations such as FAIRness and INSPIRE-compliance a simple idea might help: "Collect Once, Use many times". Data collected today must be stored, documented and published in order to increase their knowledge extraction and to allow for an efficient re-use in the future. With the multiplication of sensor deployments in monitoring programmes, the new challenge is to publish time-series efficiently using state-of-the-art technologies.

The MOMO project (MOnitoring and MOdelling of the cohesive sediment transport and the evaluation of the effects on the marine ecosystem resulting from dredging and dumping operations) has been deploying tripod platforms consisting of ADCP and ADC sensor packages in the Belgian Part of the North Sea for the last two decades. These deployments generated considerable amounts of valuable data that are currently stored as csv-formatted text files that are not interoperable, both in terms of content and access protocol. The full potential of those data is currently safeguarded only thanks to the knowledge and expertise of the scientists involved in the project. The consequences of knowledge loss is therefore significant.

The objective of this work is to explore the merit of implementing an Open Geospatial Consortium Sensor Observation Service (OGC SOS) for the publication of FAIR INSPIRE-compliant time-series data. The INSPIRE directive must be applied to data by the end of 2020 so the question is: "can a data provider be INSPIRE-compliant using SOS without complex and heavy developments?"

The 52°North SOS implementation was chosen in this project, partly because they provide a complete open-source solution (database creation, SOS client and SOS viewer).

FAIRness and INSPIRE-compliance

To achieve a FAIR (INSPIRE) compliance a complete environment must be created with metadata, services and data. Most of the work is done by now as the official RBINS metadata catalog is online (https://metadata.naturalsciences.be) with INSPIRE-compliant metadata and services. (Meta)Data are now Findable, Accessible and Re-usable but the Interoperability still remains a challenge.

The Sensor Observation Service has been recognized as an INSPIRE-compliant download service and the 52°North implementation out-of-the-box includes the INSPIRE O&M Specialized Observation (OMSO) profile. To deliver interoperable data only two steps are needed: create the database (schema provided by the SOS client) and populate it. Pre-defined datasets are then shared using the GetObservation SOS operation in the metadata file. This approach achieves FAIRness and INSPIRE-compliance.

Achieving full INSPIRE-compliance requires also creating a view service (which shows a map). We have exposed a view on the Sensor Observation Service's database as a WMS feature (using GeoServer); this might prove an effective way to publish sensor data in the INSPIRE context (see Figure 1).

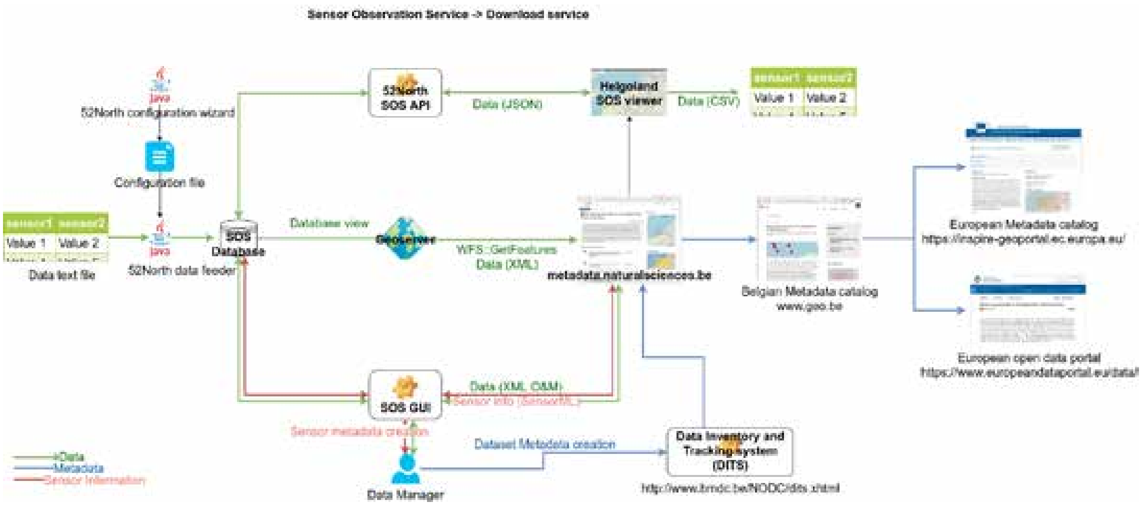


Figure 1: INSPIRE-compliant data publication workflow from text files to metadata publication.

Performance

The publication of time-series automatically comes with the burden of ensuring download service performance. The amount of data generated each year in this project is massive and providing those data in INSPIRE-compliant OMSO standard is challenging. This is due in part because interoperable data is inherently verbose.

A first attempt was to import all the project data in the database (+150e6 values from four sensors located at multiple locations). Metadata files were created for each dataset and a GetObservation request was made to retrieve the dataset data. However, each time-series contain millions of points and the big disadvantage of heavy XML encoding quickly becomes obvious. We are of the opinion that the OMSO format is not optimal for time series publication as many information elements are redundant between observations. This leads to download requests ending up in time-out errors. If one wants to publish INSPIRE time series the use of O&M time series format should be studied but it is not available in the 52°North implementation.

The second attempt was to import only the "aggregated" measurement (i.e. measurements that don't need to be associated to other measurements to have a value). The main time-series in the monitoring programme such as temperature, salinity or sediment concentration are published via the SOS service and multi-parameters sensors (more than a hundred for Acoustic Doppler Profiler sensor) are left aside. The size of the database remains much smaller and the service is able to react much more quickly. The optimal solution, at this time, is to choose for a mixed approach with SOS for standalone time-series and a files-based system for sensors that generate too many time-series. An out-of-the-band gml encoding could be used to describe the data in the files but the interoperability of this method still has to be demonstrated.

A Metadata Hierarchy for Enhanced Management of Hydro-Numerical Simulation Data

Mohammad Shafi Arif, Federal Waterways Engineering and Research Institute, (Germany), mohammad-shafi.arif@baw.de
Peter Schade, Federal Waterways Engineering and Research Institute (Germany), peter.schade@baw.de
Frank Kösters, Federal Waterways Engineering and Research Institute (Germany), frank.koesters@baw.de

In hydraulic engineering numerical simulations are commonly conducted for a better understanding of coastal and estuarine physics. Moreover, hydro-numerical simulations are carried out in the framework of environmental impact assessments, e.g. to evaluate the impact of the construction of coastal structures or the adaptation of approach channels on estuarine hydrodynamics and morphodynamics. These assessments require a large number of simulations, e.g. the model calibration for different scenarios covering natural variability. Individual model runs may differ in some minor variation of parameters only, which makes it difficult for the user to keep track of differences among models.

In terms of quality assurance a reliable documentation is therefore required but can be tedious and therefore error prone. Here the incentives of digitization and advantages of web-enabled geospatial services, as well as FAIR (Findable, Accessible, Interoperable, and Reusable) data management principles come into play with the utilization of metadata. Consistent and detailed documentation is of importance for both storage and reproduction of data as well as services such as searching, filtering and cataloguing. In spite of numerous benefits which metadata can offer, manual recording the parametrization attributes as metadata with a complex data structure is considered challenging, i.e. labor intensive and user unfriendly. Hence, the provision of an adequate data structure as well as a partial-automation for the recording is considered as a prerequisite in order to achieve reliable metadata.

Addressing the complexity for data management, the research and development project Data Management and Quality Assurance in hydraulic engineering (DMQS) has been initiated at the Federal Waterways Engineering and Research Institute (BAW). A hierarchical metadata management approach was chosen for the data organization in a simulation project, which is depicted in Figure 1 as a detailed tree diagram for a project with an integrated data and metadata organization.

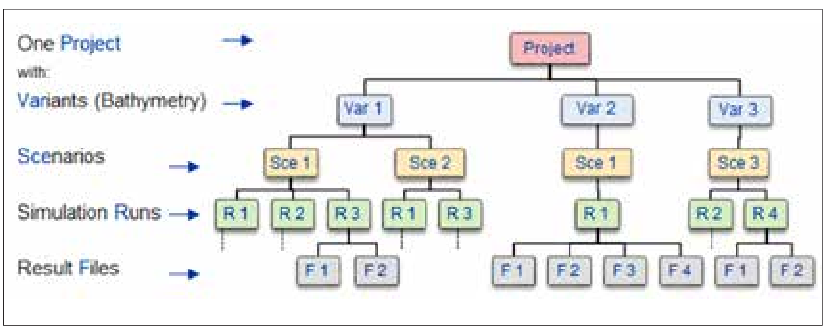


Figure 1: Metadata hierarchy in a tree structure for simulation projects.

Five levels of metadata are used with the project as the root element on top of the hierarchy. This method highly mitigates the impediments to the data management and to metadata organization in a metadata information system. The root is followed by the variants, one of the main features of hydro-numerical simulations. Variants are defining the bathymetry of the model area, consisting of a horizontal mesh and the water depth. Variants are followed by scenarios defining the parametrization e.g. the hydraulic roughness or the amount of head water. The subsequent hierarchy level is represented by the individual simulation run. Each run can generate one or more result files, which are distinguished by their Universally Unique Identifiers (UUIDs).

Aiming at a reduction of the efforts associated with metadata recording and generation, a three step approach has been chosen. At first, while starting a new project, the user is required to specify a few project related elements such as contact information just once. Pursuing this further, the user has to fill out only four meta elements related to the simulation run. At last, the rest is acquired from input data for the numerical simulations, which are mostly proprietary meta elements. For instance, the bounding box is computed by analyzing the extent of the input bathymetry, and the post processing tools protocol their lineage information so that the workflow up to the final analysis is retraceable.

All these metadata are then written and stored in Network Common Data Format (NetCDF) files as extended Climate and Forecast (CF) metadata. Later on, the NetCDF CF metadata is converted to XML (Extensible Markup Language) data based on ISO schemas, INSPIRE technical guidelines and the BAW metadata profile (GDI-BAW 1.3). The use of UUIDs represent a reliable method to identify the result files. This allows establishing the tree structure from automatically generated UUIDs on all levels.The data aggregate elements of ISO 19115 facilities the cataloguing tree as well. The UUIDs are reproducible so that later simulation runs can be linked to the same project or variant. Metadata titles of the level are concatenated to a string that is used as input to the UUID generation. Figure 2 depicts the use and application of UUIDs as parent identifiers.

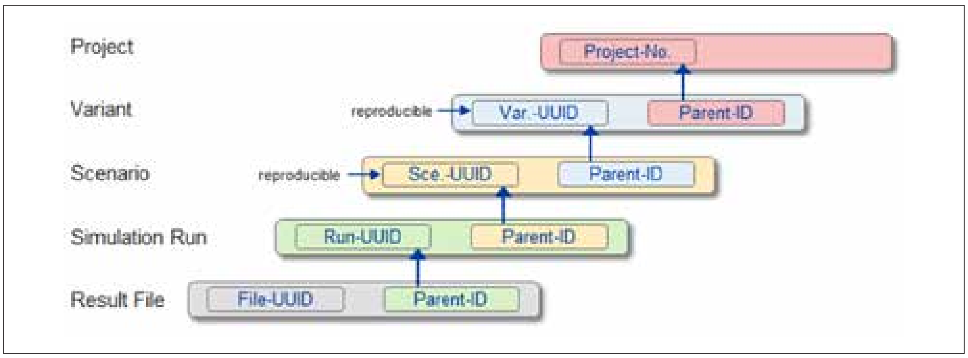


Figure 2: Application of UUIDs as parent identifiers.

To conclude, the proposed automation method has tackled the metadata generation bottleneck of a rival manual recording by using the metadata of NetCDF simulation files. Workflow automation has reduced erroneous metadata recording and enhanced reproduction as well as management of the metadata. Through the implementation of a metadata hierarchy users can better keep track of simulation results, especially of those from simulations with small differences. The convergence of CF conventions and ISO standards ensures interoperability and consistency for emerging data exchange.

Acknowledgements

The authors would like to thank Georg Carstens, Vikram Notay, Günther Lang and Shamimul Islam for their contributions towards the development of the DMQS project.

Improving data discoverability for the Antarctic Seismic Data Library System (SDLS) through SDN, ISO19115-3 and INSPIRE compliance.

Paolo Diviacco, National Institute of Oceanography and Applied Geophysics-OGS (Italy), pdiviacco@inogs.it
Alberto Viola, National Institute of Oceanography and Applied Geophysics-OGS (Italy), aviola@inogs.it
Mihai Burca, National Institute of Oceanography and Applied Geophysics-OGS (Italy), mburca@inogs.it
Alessandro Busato, National Institute of Oceanography and Applied Geophysics-OGS (Italy), abusato@inogs.it
Nikolas Potleca, National Institute of Oceanography and Applied Geophysics-OGS (Italy), npotleca@inogs.it

Introduction

The Antarctic Seismic Data Library System (SDLS) is a consolidated data dissemination initiative created in 1991 under the mandates of the Antarctic Treaty System (ATS) and the auspices of the Scientific Committee on Antarctic Research (SCAR), to provide open access to Antarctic multichannel seismic-reflection data (MCS) for use in cooperative research projects. The ATS mandates that all institutions that collect MCS data in Antarctica must submit their MCS data to the SDLS within 4 years of collection and remain in the library under SDLS guidelines until 8 years after collection, thereafter, the data switch to unrestricted use and can be requested to the SDLS for open use (Diviacco & Wardell, 2003).

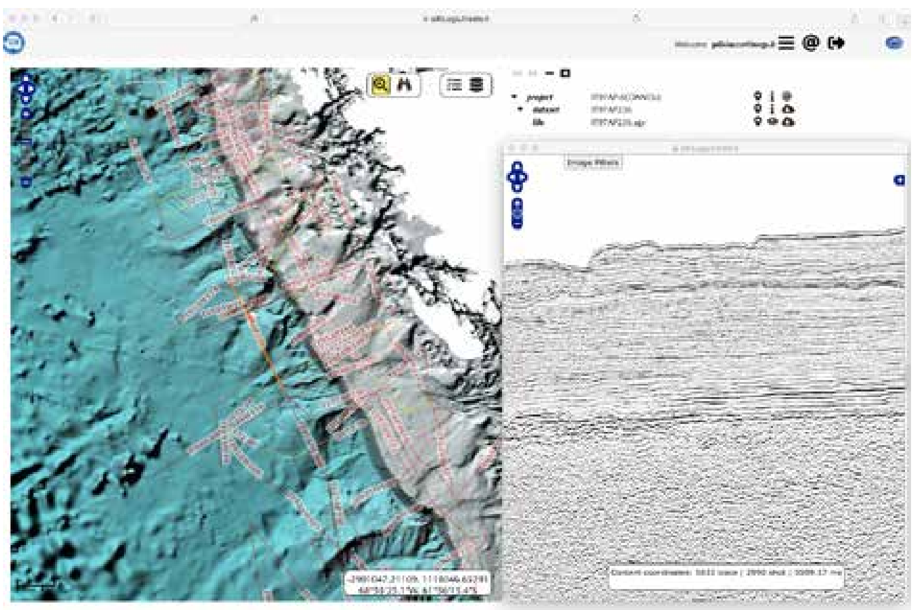


Figure 1: SDLS web portal allowing to view and interactively access all existing seismic data in Antarctica (http://sdls.ogs.trieste.it)

The web system

In order to trigger and foster as much as possible collaborative research within the Antarctic research community, the SDLS developed a web portal (http://sdls.ogs.trieste.it), hosted by OGS, that allows any data therein to be discovered, browsed, and directly accessed and downloaded. Great importance is given to compliance with the ATS legal framework and the Intellectual protection of data owners.

The SDLS web portal uses several resources and software developed within the SNAP web-based geophysical data access framework developed by OGS (Diviacco, 2005) and later integrated with the SeaDataNet infrastructure within the GeoSeas Project (Diviacco & Busato, 2013). The O&M and SensorML based metadata model used in SDLS, allows to detach general metadata from domain specific metadata. O&M and SensorML metadata in the case of SDLS is produced internally but can be prepared also using the Mikado software made available by SeaDataNet and later upload to the system. The metadata model allows therefore to embed all controlled vocabularies made available by SDN/BODC. Due to several legacy issues, unfortunately, not all services available in SNAP are available also in the SDLS. Recently, a lot of work has been done in order to overcome these limitations and to expand its interoperability with most of the international dissemination initiatives. Key to this upgrade has been the integration in the system of SeaDataNet controlled vocabularies, and the strive to reach ISO19115-3 and INSPIRE compliance. This has been attained through the introduction and integration in the SDLS of GeoNetwork v. 3.10.1.

Integration of SDLS and GeoNetwork (v. 3.10.1)

GeoNetwork is a catalogue application that provides tools for metadata editing and search functions. The metadata editor allows to generate content that complies with the ISO19115-3 metadata standard and allows to associate concepts from different thesauri to a record, in order to describe the metadata within the scope of the INSPIRE 2007/2/CE Directive. GeoNetwork allows importing external thesaurus in order to assign, as keywords, specific concepts and terms to a metadata record. In this way, through the import of the SeaDataNet common vocabularies (W3C DCAT Themes for the SeaDataNet EDMED Catalogue) it was possible to describe the SDLS metadata records with standardised terms.

Furthermore Geonetwork provides a validation system to verify the fully compliance of a record with INSPIRE requirements and the ISO19115-3 metadata standard. In this framework, the integration of the SDLS web portal with GeoNetwork allows to obtain a repository and a registry of metadata in a catalogue based on INSPIRE and ISO19115-3 standards.

Conclusions

The SDLS is the main data hub for the geophysical community working in the area of Antarctica. The portal, developed and hosted by OGS, allows researchers to access over 300.000 km of freely accessible seismic lines, and attained through the integration with GeoNetwork v.3.10.1 the possibility to be compliant with ISO19115-3, INSPIRE and SeaDataNet which vastly improved its discoverability from most of the existent international data dissemination initiatives.

SDLS web site: http://sdls.ogs.trieste.it

References

Diviacco, P., Mancini, S., Busato,A. Hoenner, X., Nitsche, F. (2018)"Tools to handle environmental concerns in marine seismic data exploration" IMDIS 2018.

Diviacco P., Busato A., (2013) "The Geo-Seas Seismic data viewer: a Tool to facilitate the control of data access" Bollettino di Geofisica Teorica ed Applicata vol.54/2 june 2013

Diviacco P. (2005) "An open source, web based, simple solution for seismic data dissemination and collaborative research", Computers and Geosciences, 2005, Vol 31/5 pp 599-605

Diviacco P., Wardell N. (2003),"Reprocessing and dissemination of Antarctic Seismic Data" Terra Antarctica reports 2003,9 133,136

Sauli, C., Diviacco,P., Busato,A., Cooper,A.,Nitsche, F., Burca,M.,Potleca, N. (2020) "The Antarctic Seismic Data Library System (SDLS): fostering collaborative research through Open Data and FAIR principles" EGU2020

Prospective FAIR bathymetry data archiving in PANGAEA - Data Publisher for Earth & Environmental Science

Daniel Damaske, MARUM / PANGAEA (Germany), ddamaske@marum.de
Janine Felden, Alfred-Wegener-Institute / PANGAEA (Germany), janine.felden@awi.de
Frank Oliver Glöckner, Alfred-Wegener-Institute / MARUM/ PANGAEA (Germany), frank.oliver.gloeckner@awi.de
Gauvain Wiemer, DAM (Germany), wiemer@allianz-meeresforschung.de
on behalf of the DAM and PANGAEA Teams

Objective

The PANGAEA data information system (https://www.pangaea.de/) is worldwide one of the leading data repositories for Earth & Environmental data hosted by the Alfred-Wegener-Institute, Helmholtz Centre for Polar and Marine Research (AWI), and the Center for Marine Environmental Sciences (MARUM), University of Bremen. Over the last 25 years, PANGAEA archived and published - supervised by scientific data curators – hundreds of thousands of datasets from all over the world from different scientific disciplines and research domains including bathymetry data.

PANGAEA already holds a few hundreds of bathymetry raw datasets containing multi-beam echosounder raw data from entire scientific cruises and as well as processed datasets consisting of vector, ASCII and raster data. The first bathymetry dataset was already stored in 2001.

Since several years, PANGAEA receives an increasing number of bathymetry data submissions, coupled with a constant growth of the data volume. Unfortunately, current data storage and download solutions in PANGAEA are laborious for searching and subsequent downstream processing of bathymetry data. Many datasets do not contain sufficient metadata in order to assure efficient long-term reuse of data. In addition, essential ancillary data, such as sound velocity profiler data (since the sound velocity has profound influence on depth computation) are not yet allocated to most bathymetry datasets stored in PANGAEA. In fact, based on an investigation of all bathymetry related datasets in PANGAEA conducted in September 2019, only eight out of 173,365 files are sound velocity profile (SVP) data files.

To cope with the increasing amount of data to be delivered to PANGAEA, standard operating procedures (SOPs) for bathymetry data, compliant with the FAIR data principles (Wilkinson et al., 2016), is going to be compulsory for all data submissions. Such consistent and efficient data storage concepts are currently developed as part of the "Underway Research Data" project, an initiative of the German Marine Research Alliance (Deutsche Allianz Meeresforschung e.V. (DAM), https://www.allianz-meeresforschung.de/). In addition, within the scope of these SOPs, bathymetry data in PANGAEA are automatically processed to generate tracklines and coverage polygons of the surveyed area including PANGAEA metadata and raster previews, which can prospectively be explored on the DAM Data Portal www.marine-data.org/. The visualized data here are also offered as OGC Web Map Services. These efforts and services can contribute to international projects such as the *Seabed 2030* Project from the Nippon Foundation and GEBCO (General Bathymetric Chart of the Ocean), or other bathymetry data collecting data portals relying mainly on pre-existing data streaming services.

Methods

One of the key challenges of bathymetric data curation is the proper amendment of the metadata description in the data curation process. Consistently adding key parameters to each single multi-beam raw data file, e.g. latitude, longitude, date and time - which can be up to thousands of data files per dataset submission - are necessary to meet international standards (e.g. ISO 19139 metadata standards). Since most data owners do not have the capacity or the means to process and provide these information during data submission, new semi-automated processes and procedures are currently developed and tested for metadata retrival, mainly relying on the open source bathymetry processing tool MB-System. Similar methods are also developed for processed (gridded) bathymetry datasets and to partially reprocess older bathymetry datasets to the new metadata standard.

Using controlled vocabularies, such as SeaDataNet (NERC Vocabulary Server NVS2.0), ensures further data consistency and interoperability of the published datasets. Furthermore, bathymetry data of the German marine research fleet (predominately of RV Sonne, RV Maria S. Merian, RV Meteor, RV Polarstern) are connected to detailed sensor descriptions in a Sensor Information System https://sensor.awi.de/ (Koppe et al., 2018), therefore providing the user with additional technical device descriptions.

These measures enhance the possibility to find, reuse and compare data and - at best - judging data quality and getting simultaneously - if available – significant ancillary data (SVP) along the data submission.

Project background

The DAM "Underway Research Data" project, spanning across different institutions, started in mid-2019. The aim of the project is to improve and unify the constant data flow from German research vessels to data repositories like PANGAEA. This comprises multibeam-echosounder and other permanently installed scientific devices and sensors following the FAIR data management aspects (Fig.1). Thus, exploiting the full potential of German research vessels as instant "underway" mobile scientific measuring platforms, provides an added value for scientific research.

The DAM is founded by the German federal government and the northern German federal states of Bremen, Hamburg, Mecklenburg-Western Pomerania, Lower Saxony and Schleswig-Holstein and is currently involving 19 German universities and non-university research institutions with a key focus in marine and climate research. The DAM contributes with its activity to the German effort of the National Research Data Infrastructure (NFDI) and especially NFDI4Earth - the NFDI Consortium for Earth System Sciences.

The "Underway Research Data" project is closely cooperating with the MareHub project of the Helmholtz Association of German Research Centers. Together they develop new visualisation technologies with the aim to offer a web-based map display of marine data of the German marine research vessel fleet, and thus facilitating the discovery of bathymetry data - including a vast package of metadata descriptions - for the national and international scientific community.

Figure 1: Data flow from research vessels to data archives, FAIR data archiving in PANGAEA and web-based visualisation of bathymetry data. *Data management is accompanied and supported by the German Marine Research Alliance (Deutsche Allianz Meeresforschung e.V., DAM)*

References

Koppe, R. , Gerchow, P. , Macario, A. , Haas, A. , Schäfer-Neth, C. , Rehmcke, S. , Walter, A. , Düde, T. , Weidinger, P. , Schäfer, A. and Pfeiffenberger, H. (2018): SENSOR.awi.de: Management of heterogeneous platforms and sensors , RDA 11th Plenary, Berlin, 21 March 2018 - 23 March 2018

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data 3,* 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Data management in Eurofleets+: the whole picture

Thomas Vandenberghe, Royal Belgian Institute for Natural Sciences (Belgium), tvandenberghe@naturalsciences.be
Susana Diez Tagarro, Consejo Superior de Investigaciones Científicas (Spain), sdiez@utm.csic.es
Dick Schaap, MARIS B.V. (The Netherlands), dick@maris.nl
Guillaume Clodic, IFREMER (France), Guillaume.Clodic@ifremer.fr
Juan Luis Ruiz, Consejo Superior de Investigaciones Científicas (Spain), jlruiz@utm.csic.es
Hong Minh Le, Royal Belgian Institute for Natural Sciences (Belgium), hmle@naturalsciences.be
Yvan Stojanov, Royal Belgian Institute for Natural Sciences (Belgium), ystojanov@naturalsciences.be
Christian Autermann, 52°North (Germany), c.autermann@52north.org
Simon Jirka, 52°North (Germany), s.jirka@52north.org

Eurofleets+ is a consortium of 42 research vessel operators aiming to provide access to ship-time for high-quality marine campaigns, including equipment and remote sampling access. From the start, the project has given data management a central place. This approach acknowledges the important drivers of efficient data management: a) broad acquisition by means of a data management plan, b) adequate transformation by software agents and c) integrating the exchange technology used by data repositories such as SeaDataCloud, all three designed to work together.

Eurofleets+ (EF+) is a 4-years H2020-funded project, and is currently in its second year. At this moment, no cruises have yet departed. For the cruise and dataset metadata funded by Eurofleets 2 (2013-2017), it has not always been apparent what their funding context was, let alone that a centralized view on the generated datasets was possible. For the Eurofleets+ proposal, the gaps in achieving this have been filled. For a better synergy with other aspects of the project, they have been separated into multiple work packages. Compared to Eurofleets 2, included in the description of work are a) the procurement of a data management plan (DMP) as a mandatory evaluation criterion, to assure data provision, and b) the assignment of dedicated data management organisations to assist principal investigators and vessel operators, to ensure the follow-up of the DMP and the data dissemination of EF+ cruises.

An additional reason to enforce DMPs is that it is a requirement of any H2020 project. Both the project and each individual cruise have a DMP. The cruise DMPs are managed on a forked DMP Roadmap web application (created by the UK Digital Curation Centre and the University of California Curation Center) and contains a number of questions adapted for EF+ from the H2020 Open Research Data Pilot.

The DMP website (http://dmp.ef-ears.eu) also provides the data management guidelines. These guidelines state the data workflow, from acquisition to dissemination. A distinction is made between en-route data and manual data. 'Manual' data (sample-derived) will be posted by the Principal Investigator on the EMODnet Data Ingestion Platform and data managed by three reference data centres, i.e. HCMR, OGS and BMDC. These will take care of the actual dissemination and promotion of both en-route and manual data by publishing the corresponding metadata in global directories (SeaDataNet and thence to EurOBIS, EMODnet, GEOSS, IOC-IODE portal) but also on a dedicated EF+ dataset catalogue, providing persistent links (DOIs) to the actual data, accessible through the project website and the "European Virtual Infrastructure in Ocean Research" portal (EVIOR). Specific attention is paid to 1) meteorological data, 2) "Essential Ocean Variables" (e.g. sea temperature, salinity, currents, oxygen, nutrients, carbon, plankton biomass,…), 3) 3.5 kHz or Chirp light seismic; and 4) multi-beam bathymetry, as these are underrepresented and have a high potential.

The main software agent is the Eurofleets Automated Reporting System (EARS) which provides software and services for en-route data acquisition, records cruise and event metadata, and transforms this metadata into the necessary European and global marine data standards. An optimized EARS „v2.5" will be distributed to vessel operators for use during the first few 2021 cruises (All 2020 cruises have been postponed due to the COVID-19 crisis). Version v3.0 is under development and will be released in April 2021.

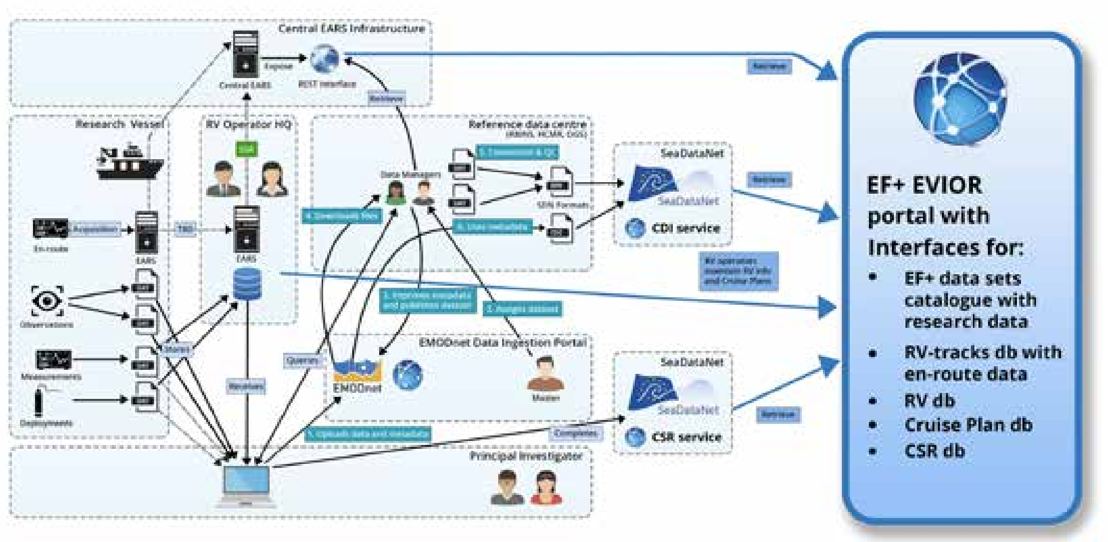


Figure 1: The complete project data workflow

The new EARS server distribution is based on docker and available on GitHub together with guidelines on installation. A docker image is a lightweight software package that combines the actual software plus the server environment. It is to be installed on the vessel, relies on TechSAS for data acquisition, and stores the metadata and data in the local EARS database, on the vessel. The metadata is transferred to shore by EARS, whenever the connection allows this, but the data transfer remains the responsibility of the vessel operator. The end goal is to let each RV operator have a 52°North Sensor Observation Service (SOS) installed on-shore as a central interoperability hub for acquisition data and event metadata. For operators without the possibility to install this SOS, a central SOS set up at CSIC will remain as a central datahub. The tracks and data of three Research Vessels are currently available on this datahub.

The EVIOR data portal connects to the SOS (GetObservation) to display the cruise tracks and primary en-route data (navigation, meteorology and thermosalinometry) as soon as it is made available by the vessel operator. The EVIOR portal and the operator's SOS will be the main interface for the data managers of the three reference data centres to retrieve the en-route data.

The EARS manual event database has been redesigned for EARS3, in order to contain all elements of a Cruise Summary Report, including references to P02, C77, the gml track and a summary of measurements. The java libraries to produce CSRs and Sensor ML are available on GitHub. In order to be able to express the device events in SensorML, a new BODC vocabulary (W11) has been created.

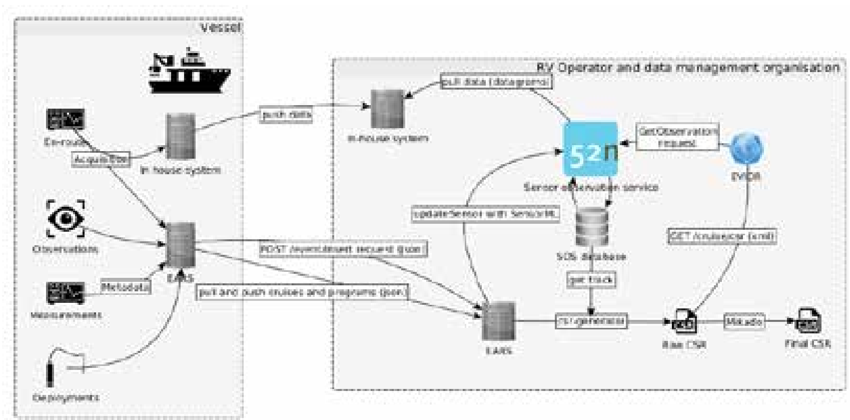


Figure 2: technical outline of the (meta)data flow

Application of metadata standard for data management in marine geological field

Jiyoon Lee, KESTI (KOREA), jiyoon4550@kesti.co.kr

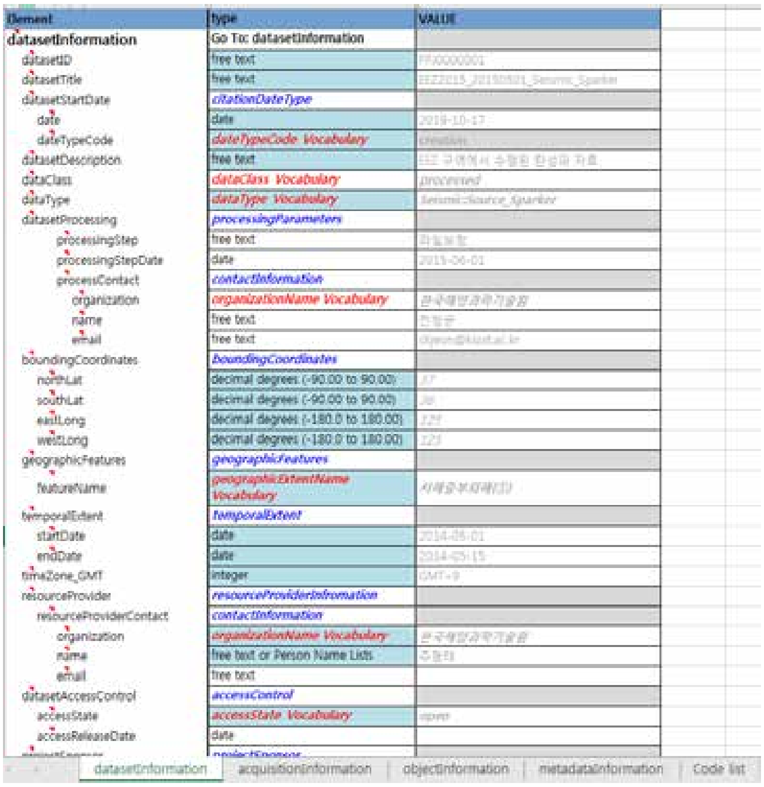
Certain standards are required to collect and distribute marine data (data and metadata) surveyed and produced by various institutions and researchers. ISO 191115/19139, the international standard for geographic metadata, is used to improve consistency and interoperability and to enable standardized data and metadata to be distributed to users.

ISO 191115/19139 does not provide acquisition information related to 'equipment and platform'. In addition, there are no metadata items that can reflect the characteristics of each sector data (physical, chemical, biological, geological, etc.), and there are no metadata items that can reflect the characteristics of individual data files included in the data set. To complement these shortcomings, institutions are developing new metadata standards or extending ISO 19115 to distribute them within the community.

In this study, we reviewed the metadata standards in the marine geography field commonly used in foreign countries, and defined the metadata standards to represent the meta information stored in seismic binary files constituting equipment and platform information and data sets. In addition, a metadata registration form was developed and the system was configured to automatically match the files and metadata that make up the data set when uploading the data set.

The applied metadata standards are largely divided into four sections. It consists of "datasetInformation" that can write general content of a data set, "questionInformation" related to data set acquisition information, "objectInformation" that can write the contents of each file that constitutes a data set, and "metadataInformation" that contains meta data information. In particular, in "objectinformation", where information of each individual file can be entered, the characteristics of each file can be used as a search keyword by marking the metadata with the user. In addition, manageable words such as equipment and sea area names were added to the code list in order to increase convenience and search utilization when writing metadatan(Figure 1).

Researchers register their own research data in the system so that they can manage seismic files in units of projects and cruise. At this time, if detailed information about the registered file is collectively entered in the metadata form (csv file) and uploaded, each file and the metadata are matched and displayed in XML format (Figure 2).



Element	type	VALUE
datasetInformation	Go To: datasetInformation	
datasetID	free text	[illegible]
datasetTitle	free text	EEZ2015_20150521_Seismic_Sparker
datasetStartDate	*citationDateType*	
date	date	2019-10-17
dateTypeCode	*dateTypeCode Vocabulary*	creation
datasetDescription	free text	[illegible]
dataClass	*dataClass Vocabulary*	*processed*
dataType	*dataType Vocabulary*	*Seismic:Source_Sparker*
datasetProcessing	*processingParameters*	
processingStep	free text	[illegible]
processingStepDate	date	2015-06-01
processContact	*contactInformation*	
organization	*organizationName Vocabulary*	[illegible]
name	free text	[illegible]
email	free text	[illegible]
boundingCoordinates	*boundingCoordinates*	
northLat	decimal degrees (-90.00 to 90.00)	*37*
southLat	decimal degrees (-90.00 to 90.00)	[illegible]
eastLong	decimal degrees (-180.0 to 180.00)	[illegible]
westLong	decimal degrees (-180.0 to 180.00)	[illegible]
geographicFeatures	*geographicFeatures*	
featureName	*geographicExtentName Vocabulary*	[illegible]
temporalExtent	*temporalExtent*	
startDate	date	2014-05-01
endDate	date	2014-05-15
timeZone_GMT	integer	GMT+9
resourceProvider	*resourceProviderInfromation*	
resourceProviderContact	*contactInformation*	
organization	*organizationName Vocabulary*	[illegible]
name	free text or Person Name Lists	[illegible]
email	free text	
datasetAccessControl	*accessControl*	
accessState	*accessState Vocabulary*	open
accessReleaseDate	date	

datasetInformation | acquisitionInformation | objectInformation | metadataInformation | Code list

Figure 1: Developed marine geological metadata standard registration form

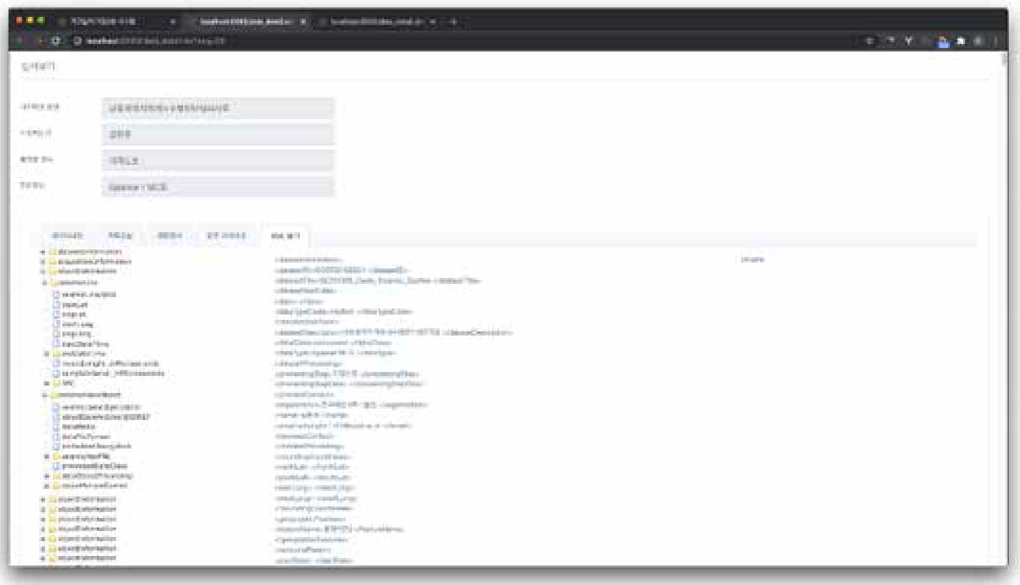


Figure 2 **:** Detailed view of applied metadata

An enabling semantic pipeline for fisheries observation data to improve sustainable management of fisheries resources

Aileen Kennedy, NUI Galway (Ireland), a.kennedy28@nuigalway.ie
David Currie, Marine Institute (Ireland), david.currie@marine.ie
Enda Howley, NUI Galway (Ireland), enda.howley@nuigalway.ie
Jim Duggan, NUI Galway (Ireland), jim.duggan@nuigalway.ie

Abstract

Fisheries scientists typically collect and analyse data such as commercial landings, fishing effort and fleet capacity - there is also a requirement for biological sampling data and spatio-temporal data in relation to vessel positioning. These data sources provide large quantities of heterogeneous data. As commercial fisheries data is often private it has tended not to follow interoperable standards and, as such, is more difficult to integrate. This work aims to use semantic web techniques to integrate and analyse heterogeneous marine and fisheries data sources. The Observation and Measurement (O&M) Ontology is used to provide a generic, non-domain specific ontology that can be used to allow interoperability between the different data sources to provide a standardised framework. The aim of the work is to create a semantic data management infrastructure that integrates fisheries observational data at a national level to aid decision support systems. Once proven at a national level, the generic pipeline can be scaled and migrated to a regional level. The integrated data could provide a data platform for machine learning prediction and forecasting techniques to aid in the sustainable management of fisheries resources.

Context

O&M is a domain neutral international standard information model. The scope of O&M includes *in-situ* observations, remote sensing, *ex-situ* observations, numerical models and simulations, and forecasts. It can include any action whose result is an estimate of the property value. The INSPIRE European standards include a Species Distribution theme in which the O&M standards have been identified as being relevant however its purpose is not to directly record observations rather aggregations of such (INSPIRE, 2013). Our approach is to implement O&M to allow the integration and interoperability of fisheries data at a detailed level (Curre et al., 2016), (Kennedy et al., 2018) - we also use appropriate controlled vocabularies from sources such as ICES and NERC Vocabulary Server for this reason

System Design and Implementation

The existing data set, which is gathered according to multi-stage hierarchical sampling schemes (an example of which is shown in Figure 1), is stored on a number of heterogeneous relational databases and is composed of both fine grain spatial data and coarser grained biological sampling data. The data is typically siloed and querying across many different databases can be

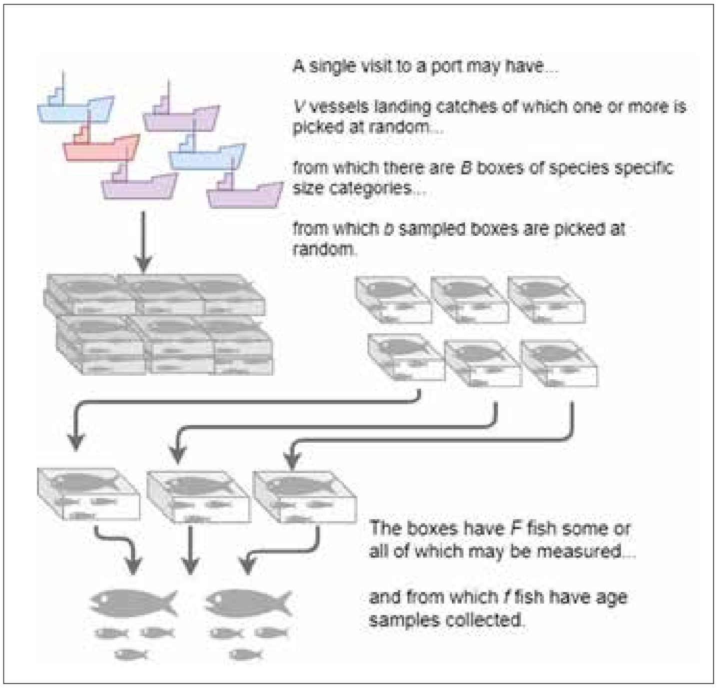


Figure 1: Sampling Process

a significant challenge. Our system extracts the data from the appropriate databases using SQL. The extracted data is transformed to RDF triple format, with each triple composed of a *subject, predicate* and *object.* The structure of the triples is determined by the structure of the ontology.

The ontology is implemented in OWL (Web Ontology Language) using classes and properties from the "oml-lite" and "samfl" ontologies to ensure that it is generic and interoperable (Cox, 2017) (Figure 2). Each physical sample can have a number of different measurements e.g. weight, length, age etc. Each measurement has its own class which stores the result of the measurement and other data such as observed property, time and procedure.

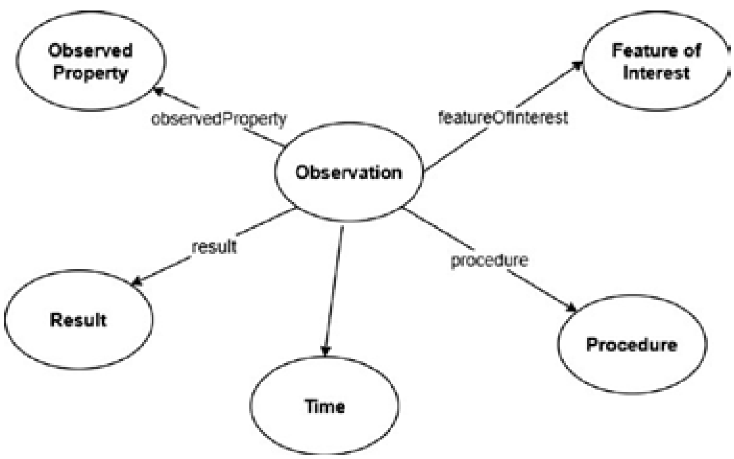


Figure 2: Simple Observation in O&M Ontology

The technical architecture includes a number of different elements: the ontology is written in OWL, using Protégé; the data, stored in relational databases, are extracted and transformed to RDF using Python; the resulting serialised RDF data is stored in a triple store and queried using SPARQL, an RDF query language.

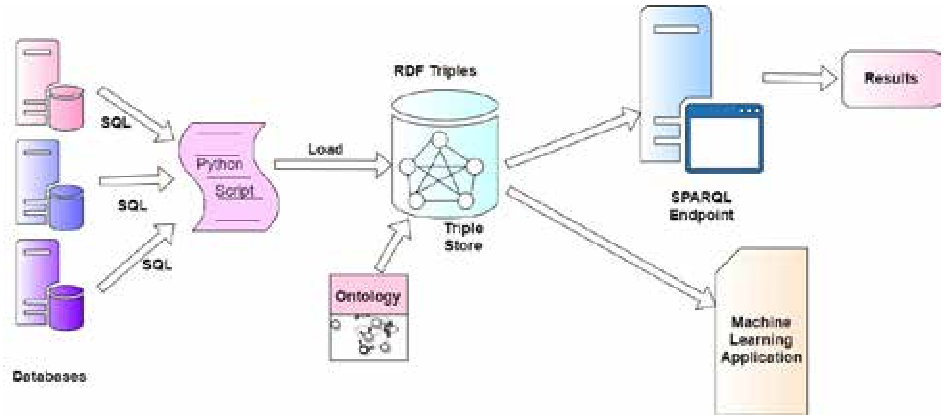


Figure 3: System Architecture

Results and Future work

We have applied the techniques discussed here in the paper to commercial fisheries sampling data and will present the outputs. The use of SPARQL to query the integrated data could enable users to write more intuitive queries rather than the complex SQL. Insight can also be gained through the use of inference on the transformed data, which would not be possible with the underlying relational databases. The pipeline integrating the data into a more standardised, interoperable format opens up opportunities for further analytics and application of machine learning techniques.

References

Cox, S.J., 2017. Ontology for Observations and Sampling Features, with Alignments to Existing Models. Semantic Web, 8(3), pp.453-470

INSPIRE Thematic Working Group Species Distribution, 2013, D2.8.III.19 INSPIRE Data Specification on Species Distribution – Technical Guidelines

Currie D., Howley E., and Duggan J. (2016) A Data Analytics Framework for Ecosystem-Based Fisheries Management. ICES Annual Science Conference 2016

Kennedy A., Currie D., Howley E., and Duggan J. (2018) Semantic Fisheries Data Integration and Analytics, IMDIS Conference 2018

A wireless network for off-shore marine research

Krzysztof Gierłowski, Gdańsk University of Technology (Poland), krzgierl@pg.edu.pl
Michał Hoeft, Gdańsk University of Technology (Poland), michal.hoeft@pg.edu.pl
Marcin Wichorowski, Institute of Oceanology PAS (Poland), wichor@iopan.pl

This paper presents functionality, the most important technology developments and potential use cases of digital communication system, developed to facilitate selected tasks commonly required during off-shore operations and marine research, including: broadband information sharing between platforms isolated from an on-shore communication infrastructure, acquisition of large data-sets from variety of off-shore data sources (e.g. fleet of drones) as well as remote access and control to the research infrastructure exploited in environment with limited communication between neighboring devices.

The shortage of well-developed, easily accessible communication infrastructure at sea makes the data acquisition and exchange tasks challenging (Woznak et al., 2017), but often limits research conditions or generating significant costs (Lighthouse). In contrast, the on-shore ubiquity of access to variability of inexpensive communication networks is driven by agile popularization of ICT support for numerous tasks – from home automation to critical infrastructure management. To change this situation the system design aimed to deliver a solution addressing the most common research communication needs and providing maximum all-purpose capabilities of deployment and ease of use feature to end-users. We have based the system design on the results of the netBaltic research project (Hoeft et al., 2018).

The building block of the system is the mobile appliance to which radio communication devices are connected with the standard Ethernet interface (Fig. 1). Any communication device can be used, if it supports IP communication, making it possible to choose a communication technology the most appropriate for the use-case, constrained to available resources (energy/ space constraints on small vessels such as RIBs). These resources could be VHF data links, popular Wi-Fi technologies and solutions dedicated for marine environment. It has been verified that, with high quality Wi-Fi hardware, it is possible to maintain a 3 Mb/s link between vessels about 10 km distance.

Figure 1: Deployment of marine data sharing appliance and picture of a portable, battery powered prototype

The appliance is capable of automatically utilizing its connected radio communication devices to detect similar units operating in the communication range and establishing cryptographically secured data transmission links to these units. Linked devices automatically create a self-organized mesh network capable of data exchange not only between direct neighbors, but between any pair of devices as well, as long as any connecting path traversing other devices can be found.

Administrator is able to control access to the system easily, by granting interaction only between the authenticated devices of dedicated group, each of them uniquely identified with an automatically generated cryptographic certificate. The process is transparent to users and allows creation of well-protected private systems assuring that connecting new device to the group is an easy task.

Figure 2: Common use-cases of marine data sharing system
(long-range communication links – purple lines; short-range, energy efficient links – blue lines)

One of the tested services is distributed file repository. The system utilizes this communication capability, combined with Delay Tolerant Networking (DTN) mechanisms, to provide set of shared file repositories synchronized between vessels even if data links between vessels are available only temporary – data in repositories are synchronized as soon as the vessel reconnects with the system (Fig. 2 - A). This service proves usefulness of the system during multi-vessel research activities - for example to facilitate collaboration between groups of researchers on different vessels, global monitoring of measurements conducted by the fleet of platforms, and to foster fast dissemination of updated measurement plans. It can also be useful for hosting on vessel-on-route an up-to-date replica of the data repository hosted on-shore (Fig. 2 - B).

The second service offered by the system is a reliable data acquisition capability, specifically designed to operate in limited-connectivity environment. Gathered information is organized into protected data packages, which are forwarded between devices towards a selected destination, where it is aggregated (Fig. 2 - C). Packages can wait extensive time periods for opportunity to be transmitted, which makes this service especially useful for isolated sets of measuring devices. In such case the vessel acting as a carrier has to enter a communication range of any device in the interconnected set and gather data to be transferred further.

The third service, command and control of research infrastructure, operates in a similar way – a configuration package is delivered to a remote device connected to elements of research infrastructure (located on a ship, buoy, drifter, etc.). The status of this operation is returned in encrypted package to commander.

Users can connect to the appliance by built-in Wi-Fi access or it can be connected to on-board Ethernet (Fig. 1). The ability to select different radio options make the appliance suitable for a wide range of platforms: small drifters, RIBs, buoys and large vessels, while the set of services provided to end users fits the requirements of the off-shore research.

This work has been partially supported by Grants: PBS3/A3/20/2015 and POPC.03.03.00-00-0008/16.

References

J. Wozniak, K. Gierlowski and M. Hoeft, Broadband communication solutions for maritime ITSs: Wider and faster deployment of new e-navigation services, *2017 15th International Conference on ITS Telecommunications (ITST),* Warsaw, 2017, pp. 1-11.

Lighthouse, Maritime research and innovation for the future, https://www.lighthouse.nu/sites/www.lighthouse.nu/files/attachments/160630_eng_lighthouse_programomraden_webb.pdf

M. Hoeft, K. Gierlowski, J. Rak, J. Wozniak, NetBaltic System-Heterogenous Wireless Network for Maritime Communications, 25(2), 2018, pp. 14-26.

Integration of Biology Sensor Outputs in the European Marine Observation and Data Network

Véronique Créach, Cefas (UK), veronique.creach@cefas.co.uk
Luis Felipe Artigas, CNRS-LOG (France), felipe.artigas@univ-littoral.fr
Alain Lefebvre, Ifremer (France), Alain.Lefebvre@ifremer.fr
Markus Lindh, SMHI (Sweden), markus.lindh@smhi.se
Fabien Lombard, CNRS-LOV (France), lombard@obs-vlfr.fr
Klas Ove Möller, HZG (Germany), klas.moeller@hzg.de
Lennert Schepers, VLIZ (Belgium), lennert.schepers@vliz.be
Jukka Seppälä, SYKE (Finland), jukka.seppala@ymparisto.fi
Melilotus Thyssen, CNRS-MIO (France), melilotus.thyssen@mio.osupytheas.fr

JERICO-S3 is a new building-block in the "Joint European Research Infrastructure network of Coastal Observatories" concept. It has been initiated in 2011 and aims at addressing the challenge of observing the complexity and high variability of coastal and shelf seas from local to Pan-European level. An important target of ***JERICO-S3*** is to provide the research community with continuous, interoperable and more valuable coastal data, coupling physical, biogeochemical and biological information into an ecosystem approach, to support the EU Water Framework and Marine Strategy Framework Directives, Regional Seas conventions such as OSPAR and HELCOM, and to provide operational marine services of high societal value.

From Sensors to service access

Challenges involving data processing, quality control, standardisation are often hidden behind the exciting step of developing and applying new technologies. However, good data management is the key step which allows scientists to find easily the data, trust them and used them according to FAIR (Findable, Accessible, Interoperable, Reusable) principles in innovative approaches for environmental studies. During the last 15 years, the development of a new generation of biological sensors has drastically changed the studies of marine plankton from lab bench work to *in situ* and real-time observations. The new instruments have also demonstrated a great versatility. Many studies have already demonstrated installation in buoys, scientific vessels, and commercial ships, and that measurements can be made at high frequency. Target organisms, from bacteria to microphytoplankton and zooplankton, can now be optically characterized and/or photographed and archived. Consequently, scientists are now facing the difficulty to handle a large amount of data which need to be processed rapidly and harmonised before being stored in databases and then made accessible to different scientific/environmental management communities. One of the aims of JERICO-RI is to provide a framework for the data flow following the FAIR principles and to increase their visibility using the European Data Infrastructures. In order to draw up best practices in data management, we encourage data users and experts operating the sensors to (a) develop standardised protocol descriptions and minimal technical metadata elements for effective re-use;

(b) identify and extend the appropriate vocabularies; (c) identify tools for data integration and platforms for trust-worthy long-term archival; (d) map sensor-specific formats to standardized data formats to be ingested by European data infrastructures; (e) discuss the correct scale for meaningful spatial and temporal data aggregation. According to the size of the user community, the readiness of the technologies, and the relevance of information for the ecosystem system approach in the monitoring of coastal seas, three main types of biological sensors have been selected for data implementation in *JERICO-S3*:

- plankton imagery: different cameras and types of *in situ* or inflow machines, image acquisition and analytical tools have been improved over the last few years for addressing the biodiversity of plankton communities. Confronted with a huge number of images, platforms such as EcoTaxa have been developed to sort and annotate taxonomically the organisms using automated classification algorithms based on random forest or deep learning approaches.
- multispectral fluorometry: it had addressed the bulk phytoplankton community response, based on their pigment composition, to multi-wavelength excitation, rather than taxonomy. The technology is now more and more regularly implemented on board of research vessels and in FerryBox systems. Data produced has the same frequency as classical single excitation fluorescence sensor. If some approaches have been developed for automated estimation of pigment groups (from scientist or companies), the data have not yet been hosted by a European data infrastructure, and data format and vocabulary still need to be better defined.
- on-line automated flow cytometry: this single-cell/particle inflow technology becomes more and more popular for optically defining phytoplankton functional groups over the whole size range, due to the versatility and the autonomy of the instruments. Based on the work already done during previous European projects (DYMAPHY, JERICO-NEXT and SeaDataCloud), a first version of common vocabulary (BODC) and data format were established (SeaDataNet) and needs to be confirmed and accepted by a growing scientific community.

As these three technologies reach today different stages of data management, JERICO-S3 aims at providing the support to comply with international standards, to establish control quality procedures, to harmonise the data and build tools to help the data flow towards the data platforms before inviting the partners of the consortium to apply the different procedures to their biological data collected in the North Sea, English Channel, Cretan Sea, north west Mediterranean Sea and Baltic Sea. Combining high frequency biological data with physical and biogeochemical data with complementary approaches (e.g: remote sensing and modelling) is the key for an integrating monitoring and a better understanding of the changes in the ecosystems.

Remote command and control capabilities for data acquisition systems provided by delay-tolerant network mechanisms

Michał Hoeft, Gdańsk University of Technology (Poland), michal.hoeft@pg.edu.pl
Krzysztof Gierłowski, Gdańsk University of Technology (Poland), krzgierl@pg.edu.pl
Marcin Wichorowski, Institute of Oceanology PAS (Poland), wichor@iopan.pl

The paper presents an assessment of a remote device reconfiguration service employing a Delay Tolerant Network (DTN) mechanisms. This service has been implemented as a part of a communication appliance dedicated to marine data transfer in off-shore and open sea areas. The service has been successfully deployed and validation test have been completed. The practical use-case has been defined as remote access to the equipment operating onboard RV "Oceania" during cruise on Southern Baltic Area. A summary of service characteristics is included, as well as simulation results of large scale deployment of the system in the south and central areas of Baltic Sea.

Continuous observations conducted at sea have a wide range of applications: enhancement of environment understanding, observations of climate changes, preparation or validation of numerical models (Visbeck, 2018). Advances in sea telemetry, observations and data acquisition offer unprecedented opportunities introduced by an increasing number of floaters or buoys equipped with new digital sensors and communication devices. On the other hand, surface waving, seawater properties and evaporation effects cause harsh conditions for radio waves propagation and make wireless communication and data collection difficult (Weller at al., 2019). Many solutions to these problems in off-shore data acquisition have been presented, however there is one important element of the process, which is frequently overlooked – the ability to deploy, verify and modify configuration of off-shore environment monitoring systems from the distance. As opposed to data collection procedures that can be performed off-line, the opportunity to command, control and reconfigure devices remotely and on-demand is highly desirable.

Selected data acquisition processes conducted on-board RV Oceania have been used as the test case for this infrastructure's capability measurements. The ongoing measurements ongoing during cruises cover Temp, Sal, Oxy, pH and pCO2 parameters. Sensors used for data acquisition should be controlled and underway data should be delivered and published in NRT mode.

Delay-Tolerant Networking (DTN) is a class of computer-network solutions, which do not require an uninterrupted network communication between the source and destination, but mobile nodes of such network can store significant amounts of data to be delivered when opportunity arises. Data packages are carried by intermediary nodes and transferred to newly encountered neighbors, eventually reaching the destination.

The proposed remote device reconfiguration mechanism was implemented as a part of an information sharing system (developed during netBaltic project (Hoeft et al., 2018) and enhanced during ePionier initiative) (ePionier iterative: http://excento.pl/e-pionier/). It is dedicated to offshore marine research, adopts DTN mechanisms to disseminate configuration updates for a wide range of devices (for example dedicated to environment monitoring). Each request is encrypted by means of authenticated encryption assuring confidentiality and authenticity of data. Thus, any attempt of a manipulation performed by intermediary nodes can be detected by the destination. DTN mechanisms operate over a uniquely adaptable self-organizing heterogeneous network – capable of transparently utilizing effectively any communication technology supporting IPv6 communication, which makes it easy to deploy in a variety of different use-cases (Woznak et al., 2017).

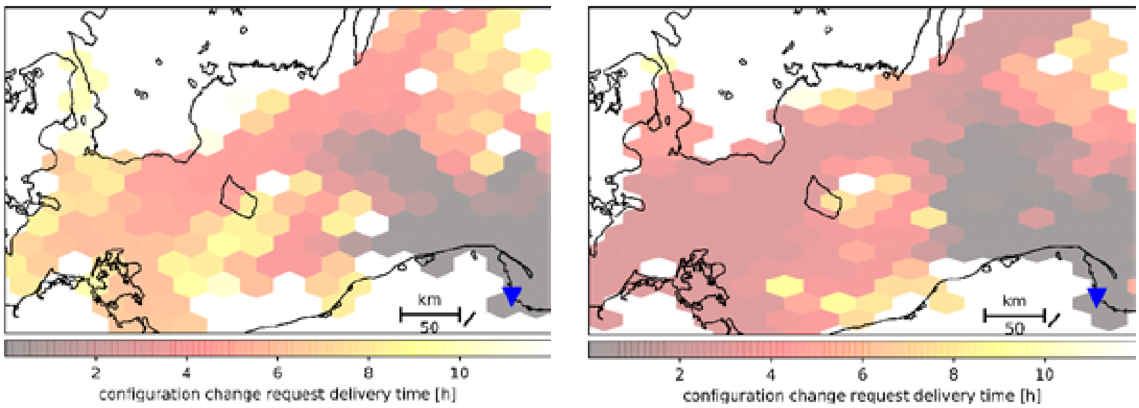


Figure 1: Visualizations of the minimum time required to deliver a re-configuration request to remote location for radio communication range: a) R_c=10km; b) R_c=15km.

Event-driven simulations of the proposed system have been conducted using vessel mobility traces obtained from a real-world AIS dataset aggregated by VTS. Based on results of the research conducted as a part of netBaltic project, it was assumed that an AIS-equipped vessel can also be equipped with a radio link offering communication in a ranges of R_c=10 km and R_c=15 km. Such communication ranges have been archived using the mentioned data sharing system in combination with off-the-shelf, industrial Wi-Fi equipment installed about 4 m over the sea-level. The presented results refer to the best-case scenario in that 100% of AIS-equipped vessels in the area participate in the system.

A reconfiguration request (marked as blue triangle at Fig. 1) is generated on-shore, in Sopot (Poland) – at the location of the Institute of Oceanology Polish Academy of Sciences. The request has been sent at the beginning of the simulation to be disseminated using DTN principles. Figure 1 illustrates the amount of time necessary to deliver the request to recipients in the particular area. Recipients in white areas did not receive the message before requested time (12 hours) for delivery void.

It can be seen that even for the shorter range of R_c=10km the coverage of the system includes most of the area in question, but the delivery time exceeds 8 h in less frequented zones. If the communication range is extended to R_c=15km, the coverage becomes almost complete and the time required to deliver a reconfiguration request tends to be lower than 6 hours.

While it is obvious that a participation of a significant number of vessels assumed in the presented scenario extend transfer capacity, we should also consider the possibility employing a number of vessels, which will modify or even plan their route to automatically deliver such a reconfiguration request to devices in a specified area. Additionally, the data sharing system allows different communication technologies to be used instead of an industrial Wi-Fi – including a number of relatively low-throughput, but long ranged and energy efficient solutions well suited for autonomous measurement devices. Combined with the fact, that the system also allows data acquisition using DTN communication, we are confident, that it presents a useful solution for offshore research.

This work has been partially supported by Grants: PBS3/A3/20/2015 and POPC.03.03.00-00-0008/16.

References

M. Visbeck, Ocean science research is key for a sustainable future, NATURE COMMUNICATIONS, (2018) 9:690

R. A. Weller, et al., The Challenge of Sustaining Ocean Observations, Frontiers in Marine Science, 2019, 6, 2019, pp. 105-123

M. Hoeft, K. Gierlowski, J. Rak, J. Wozniak, NetBaltic System-Heterogenous Wireless Network for Maritime Communications, Polish Maritime Research, 25(2), 2018, pp. 14-26.

J. Wozniak, K. Gierlowski and M. Hoeft, Broadband communication solutions for maritime ITSs: Wider and faster deployment of new e-navigation services*, ITST 2017,* Warsaw, 2017.

FAIR Data Management for Genomics Observatories

Katrina Exter, VLIZ (Belgium), katrina.exter@vliz.be
Cedric Decruw, VLIZ (Belgium), cedric.decruw@vliz.be
Georgios Kotoulas, HCMR (Greece), kotoulas@hcmr.gr
Dimitra Mavraki, HCMR (Greece), dmavraki@hcmr.gr
Matthias Obst, UGot (Sweden), matthias.obst@marine.gu.se
Christina Pavloudi, HCMR (Greece), cpavloud@hcmr.gr
Marc Portier, VLIZ (Belgium), marc.portier@vliz.be
Lennert Tyberghein, VLIZ (Belgium), lennert.tyberghein@vliz.be

Genomics Observatories (GOs) are an increasingly important resource to study the effect of climate change on marine populations. The data gathered by GOs allow one to map and track how marine populations change and how those changes relate to the local and global conditions. Such data may be used to calculate Essential Biodiversity Variables (EBVs) and can provide important information for predictive modelling of marine biodiversity.

To take full advantage of data from GOs – whoever produces the data – it is necessary that their data are FAIR: the data are findable in community catalogues, they can be accessed by human and machine, they use community standards in the data formats and the metadata vocabularies, provenance is fully documented; and ideally – but not necessarily – the data are also open access. Taking into account the complexity of GO data – with (a)biotic, genomic, geographic, and etc. parameters that need to be linked in a humanly-understandable but machine-interoperable way – it is clear that solid thinking and planning about the management of the data is essential. Done well, this allows scientists to be creative in what they are doing, by freeing them from the how of what they are doing.

VLIZ is involved in a number of GO projects, and here we explain the steps we are taking in managing the data to satisfy current and future scientific needs.

Two of these GOs – ARMS (Europe) and Ocean Sampling Day – operate (partially) under the umbrella of ASSEMBLE Plus, and we are working with EMBRC to ensure the long-term sustainability of these marine GOs.

- European ARMS programme (Matthias Obst, UGot): long-term monitoring and biodiversity assessment of invasive and indigenous hard-bottom communities. Running as collaboration of dozens of institutes, this network of *A*utonomous *R*eef *M*onitoring *S*tructures (ARMS) are deployed in the vicinity of marine stations and LTER sites in Europe and Ant/arctica for a period of 3-24 months at a time. Visual, photographic, and genetic assessments are made of the communities that settled on the ARMS units: these data will be used to track the species populations, in particular sensitive and invasive species, to map migrations, and to identify EBVs for hard-bottom fauna. ARMS units have been deployed each year since 2018 and currently have a time-series of 3 years of data.
- Ocean Sampling Day (Georgios Kotoulas, HCMR): a simultaneous sampling of the worlds oceans on the summer solstice of each year. OSD began in 2014, and continued under ASSEMBLE Plus from 2018 onwards. Following standardised sampling and sequencing protocols, these data allow for an assessment of the species populations. Packaged together with the collected (a)biotic parameters, standardised datasets will be produced that can be

compared over space and time. An extension to a monthly sampling campaign, sponsored by EMBRC, will also soon begin.

The data from these two GOs consist of

- Sequences: metabarcoding and shotgun metagenomics → analysed to produce information about the diversity of species and populations, and about the metabolic diversity of prokaryotic communities
- Images and visual assessments of the ARMS plates → analysed to produce species and biomasses/abundances
- Abiotic and biotic parameters → analysis of environmental parameters

To take full advantage of these projects, it is a must that the ARMS and OSD datasets from each station and each year can be compared, that the data can always be linked back to each sampling event and their unique samples, and that all data can be loaded into statistical codes and any variety of data-analysis workflows and virtual research environments (VREs). *But we also want that the ARMS and OSD datasets can be combined with each other – even with other data from other projects –* and this should be possible for any scientist to do, not only for those intimately involved in the projects. *In short, the data need to be FAIR*.

To this end, we are tackling the following aspects:

Data life-cycle management

- Capturing the data from the field: helping the scientists create their digital logsheets with the necessary data and metadata, and using standard vocabularies from the very beginning
- Data archiving: ensuring easy, automatic, and permanent storage of raw and processed data, permits, protocols, etc.
- (Meta)data cataloguing: automatic creation of rich metadata records for the OSD and ARMS data, with an organisation that follows the life-cycle of these projects
- Provenance: ensuring provenance is fully included as (meta)data

Data processing management

- Versioning and timestamping
- Applying workflows for data analysis, allowing for machine2machine interactions with the archives and catalogues
- Applying semantics and using controlled vocabularies
- Provenance: ensuring that the links between all and any raw or processed data, data products, scientific results, and their provenance, is done with as little human work

Engagement Management

- Capturing metrics, comments and derived results and corrections
- Writing short and sweet HowTos and cheat-sheets, to overcome the oh-too human resistance to reading documentation; and creating templates with pre-selected intelligent suggestions, to minimise the resistance to filling in forms!

Creating rich data products and data explorers

- Creating Darwin Core (OBIS Event) data products to hold all information: linking the identified species to the data from which they were identified back to the samples from which the data were obtained
- Exploring these DwC files with a user-friendly explorer: to allow the entirety of the OSD and ARMS data to be visualised, inspected, selected, grouped, cross-matched, and explored.

Quantifying quality assurance in European fisheries biological data collection

David Currie, Marine Institute (Ireland), david.currie@marine.ie
Jose Rodriguez, IEO (Spain), jose.rodriguez@ieo.es

Aim

Quality assurance of fisheries data collection is an important topic however since some of the most important fisheries data (for example commercial fisheries data) is confidential it is often not possible to directly evaluate data quality at a European level. This paper focuses on the specific operational interpretation of quality assurance in EU fisheries biological data collection however broader reviews of the topic are also available – see (STECF, 2017) for example. EU countries have a requirement to submit annual reports which describe how they collected the data required under the EU Common Fisheries Policy. One element of this report is a summary of their biological data quality assurance for each of their sampling schemes where they specify general principles, methods and tools that can provide guidance and evidence of their work. Since there is no common framework in use for biological data quality assurance it is hard to compare the information that different countries supply. The aim of this work was to define indicators which would both allow comparison between countries and allow changes to be tracked over time.

Method

It was agreed that it would not be possible to evaluate the actual data quality tools, techniques, and manuals that countries are using. This was because: i) not all information is publically available, ii) the information that is available will often be in the native language of the country, iii) even when the information is publically available in a language that is understood by the authors it would be a difficult, specialised, and time-consuming task to decide whether the techniques were appropriate for the circumstances of the country. It was therefore decided that the authors would evaluate the existence, availability, recency and pertinence of the quality documentation – with the belief being that this will have a strong correlation with the overall quality of the sampling programme.

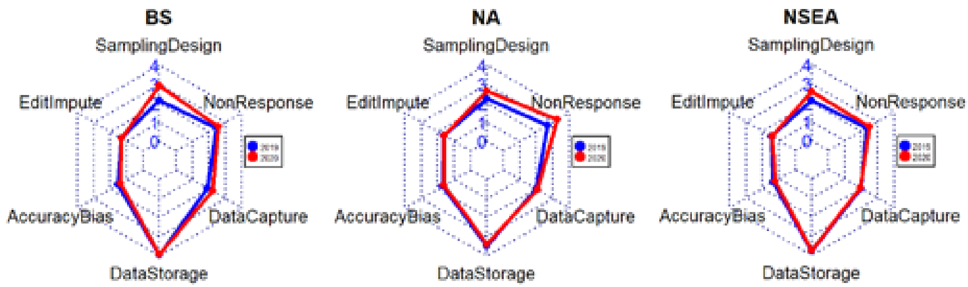
Within the annual report questions are asked on six quality assurance topics: **Sampling Design**, **Non-responses and Refusals**, **Data Capture**, **Data Storage**, **Accuracy and Bias**, and **Editing and Imputation**. For each of these topics an indicator was defined which had a range from 1 - 4 (with 1 being worst and 4 being the best). Not Applicable (NA) could also be allowed in certain cases.

To give an example of these indicators the "Sampling Design" topic has two related questions: 1) "*Is the sampling design documented?*", and 2) "*Where can documentation on sampling design be found?*". The indicator values for this topic were defined as: **Level 1**: Sampling design not documented; **Level 2**: Sampling design documented but either (i) not publically available, or (ii) the link to documentation doesn't work, or (iii) the documentation is old (from earlier than 2014); **Level 3**: Sampling design is recently documented (2014 and later) and publically available; **Level**

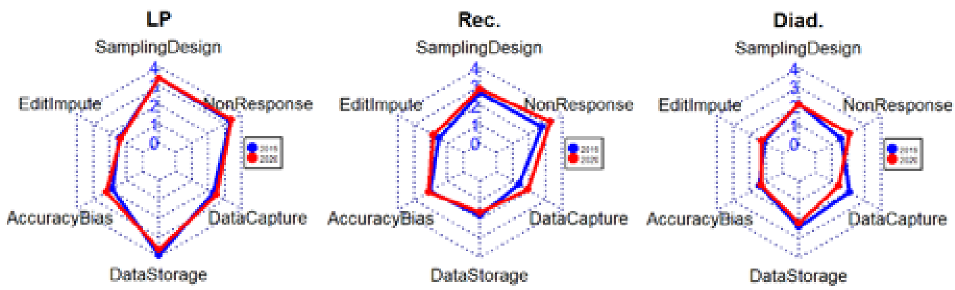
4: Sampling Design is recently documented (2014 and later) and publically available and follows good/best practices (a clear reference to guidelines established by an Expert Group or similar is provided). This indicator could be allowed to have a value of NA (not applicable) in certain circumstances, for example in the case of census data collection.

The overall method used was: i) the annual reports of each country were collated (the original annual reports can be downloaded from https://datacollection.jrc.ec.europa.eu/wp-np-ar); ii) for each row of the biological quality assurance section these indicators were assigned a score from 1 – 4; iii) each row in the collated reports was evaluated for each country which attends either the North Atlantic, North Sea & Eastern Arctic, or Baltic Sea Regional Coordination Groups (RCG); iv) the mean of these indicators for different groupings (e.g. for all countries) was then calculated. This evaluation process has been performed for 2 years (2019 and 2020) with the intention to continue in the future.

Results and discussion



It was seen that the indicators varied significantly by country (results shown in RCG, 2020) - this was because the data for each country can include submissions by different institutes, different regions (e.g. North Sea and Baltic), and different sampling schemes (e.g. commercial and recreational). A more consistent picture emerged when means were calculated for different regional groupings. It was seen that the indicators aggregated by Regional Coordination Group (RCG) had very similar values – with all showing a characteristic "shield" shape on radar plots. Radar plots are shown above for the Baltic Sea (BS), North Atlantic (NA), and North Sea & Eastern Arctic (NSEA) RCGs. The bottom point of the shield was due to the fact that their commercial fisheries data is being uploaded to the Regional Database (RDB) which gives the highest indicator value. The weakest areas are the procedures around "Editing and Imputation", "Accuracy and Bias", and "Data Capture". Radar plots for the Large Pelagic (LP), Recreational (Rec.), and Diadromous (Diad.) groups are shown below.



The Large Pelagic group had high scores because they were able to refer to internationally coordinated manuals from organisations like ICCAT. They showed a similar weakness in the areas of “Editing and Imputation”, “Accuracy and Bias”, and “Data Capture”. The Recreational data does not share the shield-shaped radar plot. Marine recreational data capture is still at the pilot stage in many countries which means many of the quality processes are not at the same maturity as the established commercial fishery sampling schemes. Recreational sampling is also more heterogeneous than commercial fisheries sampling which can make the quality procedures harder to document. Although typically data has been collected for longer time-series Diadromous programmes share some characteristics with the Recreational programmes in that there is not a single international database and that data collection is also more heterogeneous than for commercial sea fishery sampling programmes.

References

Scientific, Technical and Economic Committee for Fisheries – Quality Assurance for DCF data, 2017, doi:10.2760/680253

RCG North Atlantic, North Sea & Eastern Arctic and RCG Baltic Reports, 2020, https://datacollection.jrc.ec.europa.eu/docs/rcg

How certification process is helping SOCIB to improve Data Quality Management

Juan Gabriel Fernández, SOCIB (Spain), jfernandez@socib.es
Inmaculada Ruiz, SOCIB (Spain), iruiz@socib.es
Joaquín Tintoré, SOCIB and IMEDEA (Spain), jtintore@socib.es

Background

The Balearic Islands Coastal Ocean Observing and Forecasting System (SOCIB) is a Marine Research Infrastructure, a multi-platform and integrated ocean observing and forecasting system that provides streams of data, added value products, and forecasting services from the coast to the open ocean (Tintoré et al., 2013).

Through the different stages of the data lifecycle, SOCIB follows international ocean data management recommendations and best practices as described in UNESCO-IODE, OceanSites or EuroGOOS.

Link between Data Quality Management and accreditations

SOCIB is an established IODE Associated Data Unit (ADU) since February 2018. The IODE Committee (within the Quality Management Framework) strongly encourages to develop, implement and manage a quality management system to ensure that ADUs can prove their capabilities to provide data and services in compliance with established standards and responsibilities. Once the quality management system is operational and has stabilized, a formal application for accreditation can be made. This accreditation requires that the ADU performs a minimum set of requirements to ensure compliance with IODE standards and to establish a mechanism to regularly monitor and assess the quality of data and services.

Moreover, the "Fostering FAIR Data Practices In Europe" (FAIRsFAIR) project aims to supply practical solutions for the use of FAIR data principles. Under this framework, the SOCIB repository was selected in September 2019 as one of the Data Repositories to be supported towards achieving CoreTrustSeal certification. The process to obtain the certification follows two clear steps: (1) Self assessment based on 16 Requirements and (2) Peer review by two expert and independent reviewers under the responsibility of the CoreTrustSeal Standards and Certification Board.

The preparation procedures for these two accreditations will help SOCIB Data Center:

- to recognize the stakeholders and user needs in terms of products and services through satisfaction measurements
- to self-assess the internal procedures and activities regarding the repository
- to check all the current state of the repository documentation and resources
- to identify all the people in SOCIB that need to be involved
- to determine the strengths and weaknesses
- to understand the level of maturity needed to meet our organizational needs

- to document the gaps and major issues
- to create the steps that must be taken to reach the objective, identifying the key activities, tasks and risks
- to focus on the continuous improvement of data quality, but taking into account not disrupting the day-to-day operations of managing data

The certifications obtained (Figure 1) will guarantee the trustworthiness of the SOCIB digital repository giving confidence for stakeholders and increasing the reputation of the repository, and thus will also optimize and maximize the data sharing for contribution to global knowledge, fully in line with the challenges of the 2030 UN Decade for Ocean Sciences.

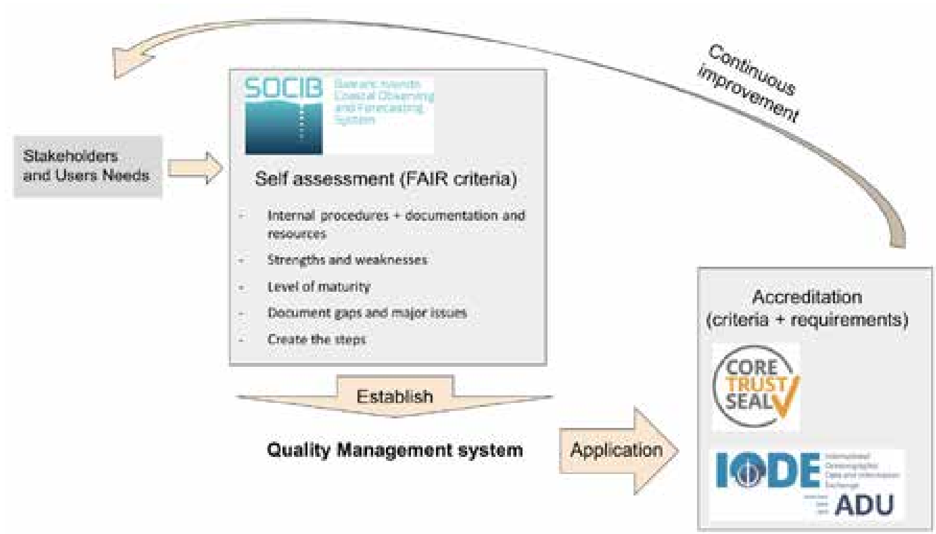


Figure 1:. Quality Management system and accreditation procedures

Conclusions

Going through a certification process is a very productive process since it will clearly help SOCIB to revise, list, update and document all activities. It encourages the whole team to identify the weaknesses and missing documentation and protocols. Finally, it is very useful in identifying where to put the efforts.

References

Dillo, Ingrid & Leeuw, Lisa. (2018). CoreTrustSeal. Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare. 71. 162. 10.31263/voebm.v71i1.1981.

Paris. Intergovernmental Oceanographic Commission of UNESCO. 2019. IODE Quality Management Framework for National Oceanographic Data Centres and Associate Data Units (Revised edition). (IOC Manuals and Guides 67, rev. ed.) 34 pp. (English) (IOC/2013/MG/67 Rev.)

Tintore, J. et al. (2013), The Balearic Islands Coastal Ocean Observing and Forecasting System Responding to Science, Technology and Society Needs, Marine Technology Society Journal, 47 (1), doi: 10.4031/MTSJ.47.1.10

SeaDataNet monitoring infrastructure near real time status updates, rich alerts, trends and insights.

Angelos lykiardopoulos, Hellenic Centre for Marine Research (Greece), angelo@hcmr.gr
Themis Zamani, National Infrastructures for Research and Technology - GRNET. (Greece), themis@admin.grnet.gr
Iordanis Michalis, Hellenic Centre for Marine Research (Greece), iordanism@hcmr.gr
Kostas Koumantaros, National Infrastructures for Research and Technology - GRNET (Greece), kkoum@admin.grnet.gr
Kostas Kagkelidis, National Infrastructures for Research and Technology - GRNET. (Greece), kaggis@admin.grnet.gr
Iona Athanasia, Hellenic Centre for Marine Research (Greece), sissy@hnodc.hcmr.gr

SeaDataNet (SDN) is a standardized system for managing the large and diverse data sets collected by the oceanographic fleets and the automatic observation systems. SDN includes national oceanographic data centres of 35 countries, active in data collection. The whole platform operates a unique virtual data management system providing integrated data sets of standardized quality on-line. The SDN infrastructure is extended in order to efficiently store, replicate and deliver the required datasets. In order to fulfil its objectives the SDN infrastructure needs to be constantly monitored for the availability and reliability of the whole system as well as for each service separately. The main scope of the monitoring service, developed during the SeaDataCloud H2020 European project, is to provide valuable and reliable services so as to ensure that the end-user has a consistent and reliable experience when interacting with the services.

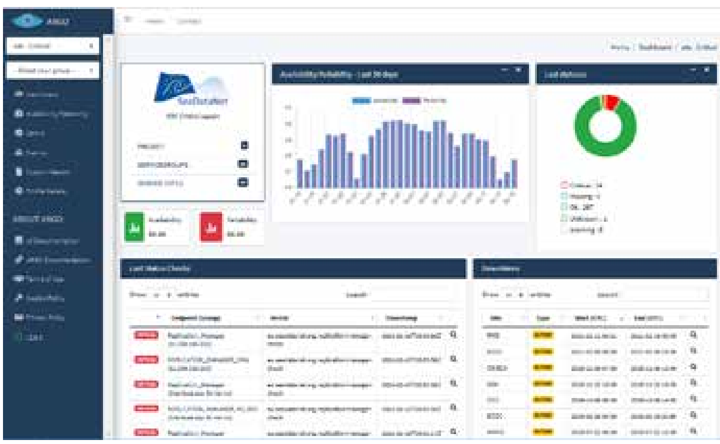


Figure 1- Main dashboard of the SeaDataNet monitoring service

The SDN Monitoring is based on ARGO, a lightweight service for Service Level Monitoring designed for medium and large sized Research Infrastructures. ARGO monitors services by emulating typical user scenarios that allows to derive the quality of service the actual user gets. It offers near real-time status updates which allow both end-users and site admins to have an overview of the services offered at any given point in time via a web user interface and via enriched email notifications, containing more information about incidents and how they affect the connected services. The rich monitoring data collected in ARGO service - data which is used for providing Availability, Reliability and status results - is actually stored in a highly flexible big-data friendly form using state-of-the-art computational pipelines and formats. This provides the ability to reuse & analyze the data in different ways such as to highlight service usage patterns and provide a number of trends and insights.

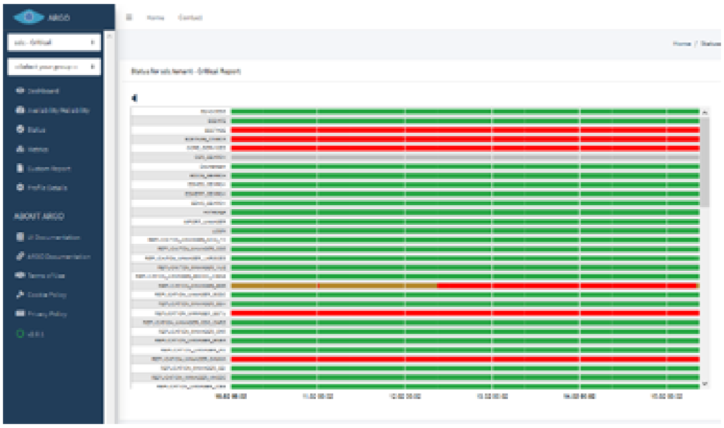


Figure 2- Status of all the SeaDataNet services

ARGO monitors the various instances of SDN services and organizes them in hierarchies and groups that reflect the relationships between them: how instances contribute to provide a user-facing service or how they organize themselves to provide high availability. Main services are the upstream and downstream services, the vocabulary services (which includes all semantic related services), the VRE (which includes all the service types and instances that contribute to the offering of the Virtual Research Environments) and the Replication Managers . In Virtual Research Environments (VRE), multiple types of services in several nodes cooperate as components to provide a unified dashboard with a plethora of functionalities to the end-users. ARGO not only monitors individually the proper function of these services but also combines them in logical groups and hierarchies to reflect and accurately monitor the reliability of their interoperability so as to ensure that the top-level service offer (VREs) works as expected.

Figure 3- Status of the SeaDataNet VRE

SESSION INFRA:

Marine environmental infrastructures for observation data

(data management and access)

- Operational oceanography metadata/data systems
- Physical and bio-chemical metadata/data systems
- Geophysical and geological metadata/data systems
- Fisheries and biological metadata/data systems
- Cross infrastructure activities (e.g. ENVRI, EuroSea...)
- Marine pollution related (contaminants, litter,...) metadata/data systems

ORAL PRESENTATIONS

A novel, spatially based, real-time software solution for the avoidance of "choke" bycatch species

Amos Barkai, OLSPS (South Africa), amos@olsps.com
Heidi Henninger, OLSPS (United States of America), heidi@olsps.com
Ronald Smolowitz, Coonamessett Farm Foundation (CFF) (United States of America), rjsmolowitz@cfarm.org

Background

In order to collect higher quality fishery-dependant data, the Northeast US limited access scallop fleet needed to develop a real-time electronic bycatch monitoring and reporting system. The objective of this system is to have the fishers report their bycatch and scallop catch real-time electronically to a central database. These data will then be anonymized and made available to the entire fleet in order to alert the fishers to areas of high bycatch CPUE areas. This would allow fishers to make informed decisions regarding where to fish in order to avoid or reduce the yellowtail flounder bycatch, and possibly a premature closing of the access area.

The Approach

In collaboration with The Coonamessett Farm Foundation Inc. (CFF), OLSPS was contracted to develop and customise its existing Olrac eLog technology to create a real-time electronic bycatch reporting system for needs of the US Scallop fleet. The Bycatch Avoidance solution was designed to change the way stakeholders collect fishery data in order to maximize catch and minimize bycatch, while increasing productivity throughout the industry. The software developed was designed to allow for data collection at sea, real-time data transfer to shore, and subsequent data transfer back to the vessel to notify fishermen of bycatch rates.

The Bycatch Avoidance solution was comprised of two software components. These were:

1. Olrac–Dynamic Data Logger (OlracDDL) – an on-board, GIS-based, vessel unit which is used to record catch and effort data and send reports to the shore.
2. Olrac–Dynamic Data Manager (OlracDDM) – A web-based reports management database, used to aggregate and analyse reports sent from the OlracDDL and transmit bycatch reports back to the fleet.

The system was designed so that fishers would only need to manually input a minimal amount of catch information in order to determine discard rates and total catch volume. Data collected by fishers on the OlracDDL included scallop catch weights, discards of several commercially important species, dredge specifications, as well as weather and tow information. In addition to these data collected, the Bycatch Avoidance solution allows for easy data collection of many other related data including of operational, biological and environmental fishing data.

After receiving data from the vessels via the OlracDDL, the OlracDDM analysed and extracted bycatch ratios for different periods of time (i.e high bycatch areas for the past day, week, or month).

These ratios produced a collection of coloured boxes identifying bycatch hot spots. These data were then anonymized and made available to the entire fleet in the form of high density CPUE bycatch maps, alerting fishers to areas of high CPUE for bycatch species (Figure 1.). This allowed fishers to make informed decisions on future fishing operations in order to avoid or reduce the occurrence of yellowtail flounder bycatch, and possibly a premature closure of the fishing areas.

Additionally, as part of the OlracDDM, a web-based interface was provided, which permitted designated users to view data sent to shore from the fishing vessels. Scientists, fishermen, regulators and other users of scallop fishery data could log in remotely to the server at CFF and access the collected data at any time.

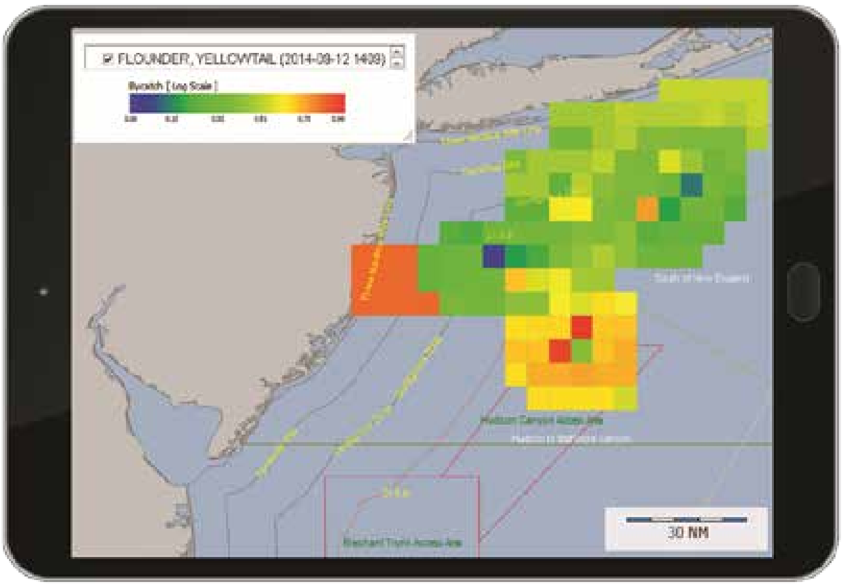


Figure 1: Example of bycatch map sent to the vessels at sea to inform fishermen of areas of low to high bycatch for all areas of scallop fishing activity

Implementation

The Olrac software was successfully installed on fifteen pilot-program vessels. Before embarking on a fishing trip, the captain and mate of each vessel were trained on how to use the software. Data for over 1095 tows was successfully recorded. Following the initial trip, CFF researchers met with the captains to debrief and receive insights on how better to optimize the software, how best to train future captains, and address issues that arose while at sea collecting data.

While the data collected and transmitted by the fifteen vessels does not accurately represent the fishing behaviour of the entire fleet, analysing the incoming data successfully demonstrated the potential for this program for both industry and management.

The Olrac bycatch avoidance system passed all field tests and is now ready for large scale deployment across the entire scallop fleet.

Data Management Architecture to enable Multinational Co-operation

Andrew Conway, Marine Institute (Ireland), andrew.conway@marine.ie
Adam Leadbetter, Marine Institute (Ireland), adam.leadbetter@marine.ie

Wastage, expending resources in a needless way, is a prevalent risk across almost every project undertaken. It is often seen as an unavoidable byproduct of undertaking any project. In particular, it has a tendency to negatively impact projects which involve more than one organisation. Communications are duplicated and parallel systems are often spun up, leading to redundancy. Data management efforts are not exempt from this, but tactics can be deployed to mitigate against some of the larger resource wasters. The INTERREG VA COMPASS project, a cross-boundary project involving organisations from Northern Ireland, the Republic of Ireland and Scotland, aimed to establish a regional network for ocean monitoring of marine protected areas, incorporating the application of tactics to minimise wastage within the project.

Instead of each partner organisation developing their own individual data management processes in isolation, with limited consultation or collaboration, the COMPASS project took the approach of agreeing on architecture at the project's outset. This architecture informed the data management choices that would be made by each partner. Key to the design of the core data architecture was a tacit agreement that all project data, systems and resources would be shared without discourse.

In terms of the local technical infrastructure used by each partner, it was decided early on that no prescription would be made to each partner on how their data would be stored (for example, SQL or THREDDS). This eliminated effort that might be wasted by any organisation having to translate data from their preferred local data management solution in order to conform to a data structure set out exclusively for a single project with finite resources and time-frame.

Once that had been decided, methods for transformation and sharing of the data were agreed upon. A federated data management system, where nodes of data are held and controlled at a remote location, was determined to be the most efficient mechanism for data sharing. The principles of federation keep a single data source for each partner's data and remove the need to keep multiple replicated sources synchronised across the project as updates occur. This was particularly beneficial when data were being added in near-real-time and also when the partners were from different nations which may have specific local legislative or community requirements for environmental data publishing (e.g. EU's INSPIRE or MEDIN standards in the UK). The system used to provide federation within the project architecture was NOAA's ERDDAP data server.

The ERRDAP data server is open source software written in Java that provides a platform through which data can be shared between partners and also published more widely. Developed out of NOAA's Monterey Laboratory, the platform-agnostic ERDDAP, or the Environmental Research Division's Data Access Program, builds upon the open-source ideals of the OPeNDAP, WCS, SOS and OBIS standards. It offers a consistent, yet an easy-to-use way of downloading or viewing scientific data in a variety of formats. Through all of this, ERDDAP also provides the ability to generate RESTful web API links in a relatively straightforward manner, which makes the process of integrating data into web-based applications simple.

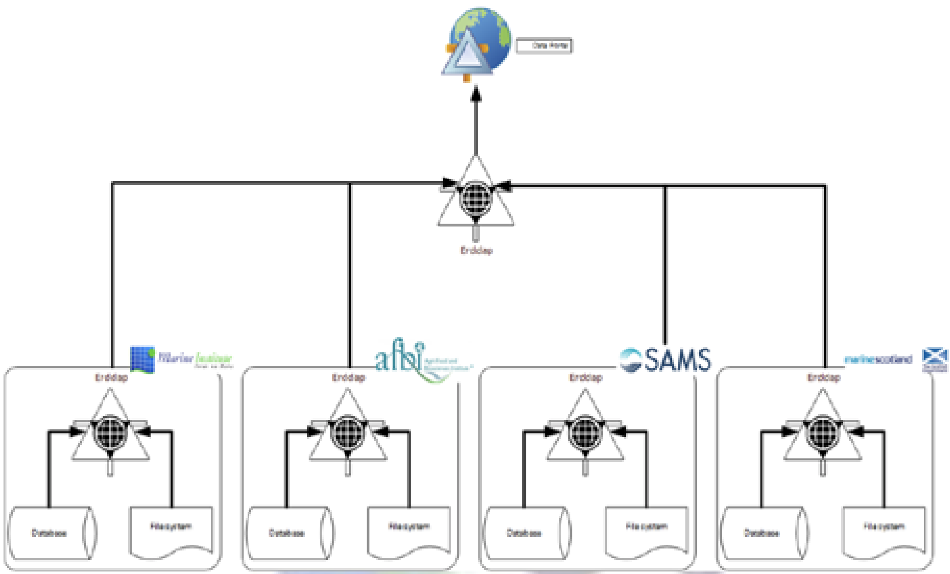


Figure 1. Anticipated high-level architecture for the COMPASS data management architecture

The use of a federation of ERDDAP servers (Figure 1), deployed at each partner location, guaranteed that each contributing organisation could maintain full control of their data while also contributing to a larger dataset, which was not dependent on any single partner to keep publicly available. Previously, this could only be attained through a single partner hosting the sum total of the project output data. This generally would involve significant duplication of data and a significant monetary cost. Further, each partner is equally involved in data publishing. Rather than one partner ring-fencing the data management/publication role, the skills obtained as well as the infrastructure deployed, are available for future use by all partners.

The establishment of these ERDDAP nodes now could have potential benefits in the future. Rather than the accepted norm of project-specific web portals for hosting data, which eventually fade into oblivion as projects age following completion, these ERDDAP servers can now act as a ready-made hub for current and future data. A beacon for publishing and a single point for data management and maintenance. Thus, the use of this data management architecture has alleviated those costs and enabled more time and resources to be spent on scientific analysis both now and hopefully into the future.

Overall, the data management architecture generated for the INTERREG VA COMPASS project, shows that through data management, multinational cooperation can be cultivated.

Unifying specialized databases for a central search portal – the HCDC approach

Linda Baldewein, Helmholtz-Zentrum Geesthacht (Germany), linda.baldewein@hzg.de
Ulrike Kleeberg, Helmholtz-Zentrum Geesthacht (Germany), ulrike.kleeberg@hzg.de
Dietmar Sauer, Helmholtz-Zentrum Geesthacht (Germany), dietmar.sauer@hzg.de
Robin Luckey, Helmholtz-Zentrum Geesthacht (Germany), robin.luckey@hzg.de
Philipp S. Sommer, Helmholtz-Zentrum Geesthacht (Germany), philipp.sommer@hzg.de
Housam Dibeh, Helmholtz-Zentrum Geesthacht (Germany), housam.dibeh@hzg.de

Coastal research is characterized by a large variety of overlapping research fields. Each of them has its own requirements for data formats, storage and retrieval solutions. For example, observational oceanographers use sensors for a continuous collection of near real-time data, while climate modelers generate terabytes of high resolution model data, oftentimes on unstructured grids. Providing access through a single portal for a wider scientific community as well as stakeholders to these different data sets is a challenging task for data managers and providers.

At the Helmholtz Coastal Data Center (HCDC) we have three main branches of data with their own storage solutions. Biogeochemical campaign data, received as either ASCII or Excel file from the author, are stored in a relational database. Metadata for biogeochemical data are available for each individual measurement. The second type, near real-time observational data stored in ASCII files, are sent via web transfer protocols directly from the sensors, on for example ferries or underwater knots, into a database set up for time series data. Related metadata are aggregated for each platform that is sending data. The third data type is model output. They are stored in NetCDF files directly at the site of the High Performance Computing system for Earth system research at our partner institute, the German Climate Computing Center (DKRZ), and the metadata are available through the data publishing platform CERA (cera-www.dkrz.de).

Before starting the journey of unifying all three data systems in a single search portal, we questioned our stakeholders, mainly colleagues from our research institute, other research centers and public authorities, and defined the required functionalities of the search portal. The users desire a portal with a search engine that "is like Google", meaning a single search field that allows them to find all data, like a one-stop shop. At the same time they want to be able to filter by time, geographic locations and a variety of other metadata fields.

Such a search process is only possible if the metadata systems are unified, because it simplifies the process compared to searching for data across three different metadata solutions with different technical and logical setups. Thus, a separate metadata database is set up, where metadata are aggregated per data source. A list of common metadata fields is created to which all three systems are mapped as well as possible. For example all three metadata systems contain parameters, measurement / file counts and the data source, e.g. a platform, model or campaign. This metadata database is continuously updated to reflect changes in the three source databases / storage locations.

The number of aggregated datasets for the combined sample, observational and model data is in the range of several millions. The use of a highly scalable full-text search engine on top of the relational database facilitates a real-time search experience. Thus, the metadata are transferred into an Elasticsearch cluster, an open-source NoSQL search engine.

To generate a "Google"-like search feeling while using filters at the same time, the intelligent criteria search has been developed. While the user is typing a search term on the responsive website, a fuzzy search is started over all keywords in the Elasticsearch cluster. The most likely matches as well as the search criterions to which they belong are displayed, independent of the database from which they originate. The user can now select the keyword from the resulting list and use it as a filter. The search results are displayed immediately and can be further refined by adding more filters (see Figure 1).

Figure 1: Unified HCDC data portal for different databases. A single search field at the top is followed by selected search criteria and the result list.

The resulting datasets are then listed to best match the searched terms. With a single click, the dataset can be added to the cart and the download can be requested. Different output file formats are offered to the user so that each specialist can continue working in their desired format. The user will receive an e-mail once the dataset is ready for downloading.

The data portal is build using state-of-the-art web frameworks and techniques. The front end is a progressive webapp based on Angular 8. It is set up as a series of services and components. The services store the application's state and provide the means of inter-component information transfer and communication with the backend. The services are being injected into predominantly stateless view components by the angular dependency injection system. The backend provides an API for searching for Metadata in the Elasticsearch cluster. It is a stand-alone, lightweight server based on ExpressJs. Its functionalities include the real-time, prefix-based, fuzzy completion suggestion for the search in the front-end, the data aggregation to provide an overview over the existing data measurements that fit the filter and a functionality for downloading these measurements in various data formats. In order to allow minimal response time albeit the multitude of requests necessary, JavaScript promises are heavily made use of for the means of asynchronous communication.

The unified HCDC data portal provides a single access point to three different data bases and storage locations for the individual user. As such, it serves as a real-life example for users and data managers how to access various subject specific databases through a single and intuitive entry point. The search field allows for filtering and downloading data across different coastal research disciplines. The data portal, using state-of-the-art web technologies, simplifies data access for all stakeholders and prevents time consuming searches across different platforms and heterogeneous user interfaces.

ODATIS: cluster for French marine data management

Valérie Harscoat, Ifremer (France), Valerie.Harscoat@ifremer.fr
Sabine Schmidt, CNRS (France), sabine.schmidt@u-bordeaux.fr
Gilbert Maudire, Ifremer (France), Gilbert.Maudire@ifremer.fr
Cécile Nys, Ifremer (France), Cecile.Nys@ifremer.fr
Joël Sudre, CNRS (France), joel.sudre@legos.obs-mip.fr
Caroline Mercier, Akka (France), externe.caroline.mercier@thales-services.fr
Gérald Dibarboure, CNES (France), Gerald.Dibarboure@cnes.fr
Frédéric Huynh, IRD (France), frederic.huynh@ird.fr

Context

The ocean is the largest compartment on Earth, and the current and expected consequences of global changes are multiple. The recent decades have seen a significative increase in the number of marine and coastal observations, both for *in situ* and remote sensing measurements. The number and variety of acquisitions now require effective tools to make such large amounts of data available to the research community. These observations must also be accessible to a wide public, from scientists to managers and citizens. The implementation of such an approach must be coordinated at least at the national level to achieve an adequate level of interoperability.

The Ocean cluster of the French national initiative for Earth observation data

DATA TERRA is the French research infrastructure for the management and processing of Earth observation data. The data produced by the national research infrastructures using satellites and terrestrial, coastal and marine observatories, are handled by four thematic clusters: AERIS (Atmosphere), ODATIS (Ocean), THEIA (Continental Surfaces) and FORM@TER (Solid Earth).

The aim of DATA TERRA, through its clusters, is to provide wide access to georeferenced data, products, software, tools and/or services on the Earth system produced by the French scientific community.

The ODATIS (Ocean DAta Information and Services) cluster, entry point to all French marine data, has the ambition to become an essential tool for the community to describe, quantify and understand the global ocean and its evolution across disciplines: physic, chemistry, biogeochemical cycles, marine ecosystems. The data managed include variables from all disciplines whatever the measurement platform used (satellites, *in situ* observatories, field campaigns, research cruises, laboratory analyses). ODATIS is therefore in charge of marine observation data and the associated elaborated data, from the coastline to the open sea and from the sea surface to the sea bottom. There are many interfaces with other domains - land/sea on the coastline, ocean/atmosphere, ocean/sub-ocean floor - and observation means are jointly shared.

The four main objectives of ODATIS are (figure 1) : (1) in situ and satellite data assembling; (2) data management and processing; (3) dissemination, promotion and support for data usage; (4) scientific expertise

To achieve those objectives, the ODATIS cluster relies on a network of Assembling Centers and Data and Service Centers (DCS) that carry out data management activites from data collection to data processing and that provide data services. It brings together six organizations: CNES, CNRS, Ifremer, IRD, Shom, as well as the network of Marine Universities. ODATIS is also part of European and international initiatives such as Copernicus, SeaDataCloud and EMODnet.

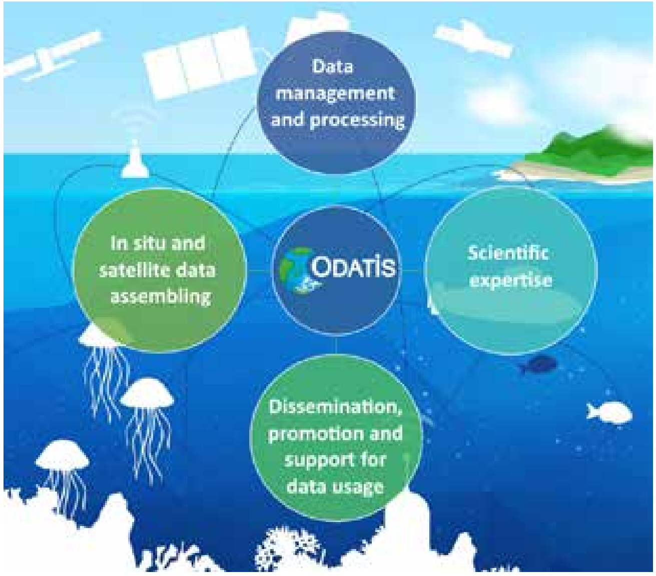


Figure 1: the objectives of ODATIS cluster

To promote and facilitate the use of ocean observations, ODATIS catalogs and faciliates access to all the data collections through its Web portal (www.odatis-ocean.fr/en). The ODATIS catalogue service provides users with several data access tools: search with selection filters, data description, viewing, and downloading (directly or via the local partner portal). As the data managed are geo-referenced by nature, the provisions and protocols relating to this type of data apply (Inspire directive, ISO 19115 family metadata, OGC interoperability protocols). The second task is to develop processing tools for handling such large amounts of data, and to generate products from observations for policy-makers, practitioners and academics.

Moreover, to define the technical orientations of the cluster, every year several technical workshops are organised (www.odatis-ocean.fr/en/activities/technical-workshops), gathering partners around presentations and round tables. Also trainings are proposed on data visualisation, analysis and processing software packages.

Finally, ODATIS relies on Scientific Expertise Consortiums (www.odatis-ocean.fr/en/activities/scientific-expertise-consortium), in order to promote and develop innovative processing methods and products for space, airborne or insitu observation of the ocean and its interfaces (atmosphere, coastline and seabed) with the other DATA TERRA clusters (AERIS, THEIA and FORM@TER).

Towards data FAIRness

The ocean is an evolving environment in a context of global changes. Each observation is valuable because it is non-reproducible: it shows a state at a specific moment and allows the constitution of time series characterizing trends and allowing to model them and understand the evolution of the ocean. Scientists must tackle these challenges in a spirit of ethics, transparency and reproducibility of results. In this context, the application of FAIR principles to Earth observation data is a top priority for ODATIS. The FAIR principles (Findable, Accessible, Interoperable, Reusable) aim at making data discoverable, accessible, interoperable and reusable. Harmonizing data management procedures, in application of the FAIR principles, is an ambition shared by all the RI DATA TERRA thematic clusters, in order to access in a transparent way to all the data of the Earth system, and in particular to those at the interfaces, such as Ocean/Atmosphere, Earth/Sea Continuum.

Gena, a cross-domain crowdsensing infrastructure

Paolo Diviacco, National Institute of Oceanography and Applied Geophysics - OGS (Italy), pdiviacco@inogs.it
Rodrigo Carbajales, National Institute of Oceanography and Applied Geophysics - OGS (Italy), rcarbajales@inogs.it
Massimiliano Iurcev, National Institute of Oceanography and Applied Geophysics - OGS (Italy), miurcev@inogs.it
Nikola Potleca, National Institute of Oceanography and Applied Geophysics - OGS (Italy), npotleca@inogs.it
Alessandro Busato, National Institute of Oceanography and Applied Geophysics - OGS (Italy), abusato@inogs.it

Introduction

The notion of Crowdsourcing is around since 2006 (Howe, 2006) when it was mainly focussed on outsourcing problem-solving by organisations or companies. Later, considering it in the perspective of the outcomes of modern Epistemology and Sociology of Science (Hanson, 1958; Polany, 1966; Diviacco, 2013), new approaches to the traditional loop of Science were made possible, extending and overlapping roles between researchers and the Society at large.

At the same time, the evolution of technologies and in particular the ubiquitous possibility to connect to the Internet led to the perspective known as IoT (Internet of Things). All these factors led to a new paradigm in which scientists can freely focus on interpreting the data, that are instead made available through an "external" infrastructure that collects, validates and disseminates observations acquired by volunteers (the crowd).

This approach has many advantages that span from the reduction of the costs of acquisition, to the possibility to explore larger areas, but also to the awareness of volunteers on the specific topic they are dealing with.

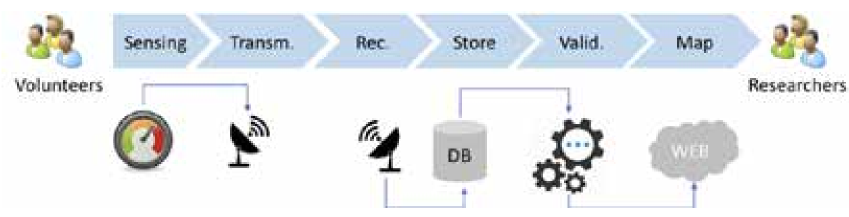


Figure 1: Scheme of the Gena cross-domain crowdsensing Infrastructure

The infrastructure.

To implement this approach OGS developed the full suite of tools that allow to follow the complete path from the acquisition, transmission, storage, integration and real time visualization of the crowdsourced data. The system has been named Gena - CROwdsensing CrOss Domain Infrastructure

The acquisition system is based on a hub box that collects data from a slot of sensors. These can be chosen, depending upon the parameter that is to be measured, among the large set of low cost sensors available on the market. Several possible configurations have already been tested for marine surveys such as for example pH, temperature, DO or salinity (Diviacco et al., 2020), while low cost sensors for CO_2, Particulate matter and other parameters have been tested for the case of atmospheric measurements (Carbajales et al. 2020).

All boxes embed a GPS device that allows geolocation and timing. Transmission is granted by a specific unit that allows to opt between GSM, WiFi or LoRaWAN technology depending on the available coverage.

Crowsensed data are collected in an InfluxDB database, which allows easy integration with TheThingsNetwork for LoRaWAN network management and directly with GSM and WiFi connections.

Validation of data

Low cost sensors are generally a synonym of low quality measures. To mitigate the impact that this has on the data, QA/QC activities can be based on one hand on the redundancy of the crowd-sensed data themselves, and on the other on their comparison with high quality measurements done in the same area (when available).

In the first case, binning data within cells that are delimited geographically and in time allows to statistically process data in order to retain an average value per cell which can be used for further visualization in gridded or contour maps. In the second case the availability of high quality measurements (generally only at very sparse points), besides the possibility to understand the error of the crowsensed data, allows to extend the high quality measurements, at least qualitatively, to the rest of the area covered during the crowdsensing survey.

Visualization and data access

Gena provides end users with a nearly real-time web interface based on OpenLayers where all data can be accessed and using standard OGC compliant web services. Server side processing and conversion scripts generate both filtered and aggregated data, by computing averages on a spatial and temporal grid. Automatic interpolation techniques like Inverse Distance Weighting or Natural Neighbours provide detailed online maps with contouring and boundary definition.

Conclusion

Gena is a complete solution for crowdsensing. It allows seamless acquisition, transmission, integration, processing and visualisation/dissemination of cross-domain crowd-sensing data, while being compliant to OGC standards. It has already been tested for marine and atmospheric studies showing the potentialities of this paradigm.

References

Carbajales, R. J., Iurcev, M., Diviacco, P. (2020), “Low cost sensors and crowd-sourced data to map air pollution in urban areas” EGU 2020.

Diviacco, P. Nadali, A., Malfatti, F., Iurcev, M., Carbajales, R., Busato, A., Pavan,A., Grio,L., Nolich M., (2020), “Citizen science and crowdsourcing in the field of marine scientific research – the MaDCrow project” EGU 2020.

Diviacco, P. (2013). Addressing Conflicting Cognitive Models in Collaborative e-Research: A Case Study in Exploration Geophysics. In Collaborative and Distributed e-Research: Innovations in Technologies, Strategies and Applications, Hershey, PA: IGI Global press.

Hanson, N. R. (1958) Patterns of Discovery. Cambridge, UK: Cambridge University Press.

Howe. J., The rise of crowdsourcing, Wired, July 2006, https://www.wired.com/2006/06/crowds/ [retrieved 7/4/2020]

Polanyi, M. (1966). The tacit dimension. New York, NY: Anchor Day Books.

River data management for coastal oceanography

Antonio Novellino, ETT SpA (Italy), antonio.novellino@ettsolutions.com
Flavio Santos, MARETEC, Istituto Superior Tecnico, Universidade de Lisboa (Portugal), flavio.t.santos@tecnico.ulisboa.pt
Estrella Olmedo, Barcelona Expert Centre (Spain), olmedo@icm.csic.es
Marco Alba, ETT SpA (Italy), marco.alba@ettsolutions.com
Luca Bonofiglio, ETT SpA (Italy), luca.bonofiglio@ettsolutions.com
Patrick Gorringe, Swedish Meteorological and Hydrological Institute, SMHI (Sweden), Patrick.Gorringe@smhi.se
Francisco Campuzano, MARETEC, Istituto Superior Tecnico, Universidade de Lisboa (Portugal), campuzanofj.maretec@tecnico.ulisboa.pt

I. INTRODUCTION

Rivers are the major pathways for material fluxes from land to sea. They supply nutrients and sediments to marine ecosystems, whereas riverine freshwater supplies largely affect the coastal and offshore circulation of oceanic systems. Unfortunately, most of the river data is fragmented and not accessible. In this paper, we describe the results from a joint European cross-cutting action between EMODnet Physics and the LAMBDA project for making river data available, accessible, downloadable and usable. In the framework of EMODnet Physics and SeaDataNet NODCs have started designing the workflow for long term river data (for oceanography) management.

a. EMODnet Physics

EMODnet Physics (www.emodnet-physics.eu) is one of the seven domain-specific portals of the European Marine Observation and Data Network (EMODnet). EMODnet-Physics map portal (www.emodnet-physics.eu/map) provides a single point of access to in situ datasets, products and their metadata.

b. EMODnet Data Ingestion

The EMODnet Data Ingestion portal facilitates data producers and managers to ingest their marine datasets for further processing and publishing as open data. In particular, it contributes to identify new data sources and define the management flow of this new typology of data. Its role is central also for providing a standard codification for formats; conventions and modalities of dataset management, making the system easily accessible form other programs, such as Copernicus.

c. CMEMS SE - LAMBDA

The LAMBDA project (http://www.cmems-lambda.eu/) aims to improve the CMEMS MFCs thermohaline circulation in coastal areas by a better characterisation of the land-marine boundary conditions, with special regard to the salinity fields, through exploring the capacities of watershed numerical modelling and its coupling to mesoscale regional ocean models.

II. FIND, ACCESS, INTEROPERATE, REUSE DATA AND PRODUCTS

Usually, river data are managed by local environmental agencies and despite the fact that some are making data (water level along the river) accessible, the fragmentation of the sources and the variety of the data formats and naming conventions limit the usability of the data. To overcome this harmonization of both data transport format and naming conventions are needed and to set up a single point of access for this federated network of sources. The EMODnet Physics-LAMBDA working group have defined common standards to manage the data flow and make it accessible and visible.

a. Data conventions and data management

Transport file is NetCDF v3.6 (and v.4.0), data is stored in a data server according data age. CF convention/SeaDataNet P09 are used for parameters. Data access and services

b. Data access and service

EMODnet Physics has developed a dedicated page for each river station and provides the user with metadata and the latest data chart. Features to download data and search for longer and older time series are also offered. Once integrated in EMODnet, data are made available under the EMODnet Physics M2M interfaces

Figure 1: Example of river data plot

The LAMBDA project provides the user with reanalysis river data (river flow and temperature) from 1992 to 2020 and forecast. This dataset is helpful for CMEMS MFCs reanalysis simulations. The project also provides remote salinity SMOS products focussing on coastal river inputs.

The possibility to access harmonized river flow data from a single point is enabling new possibilities and new features in operational services and river management projects. While LAMBDA is collaborating with the CMEMS MFCs in order to run specific model sensitivity tests focused on evaluating the potential impact of using LAMBDA products in model applications analogous to the ones used in their operational forecast services, other projects e.g. HazRunoff (http://www.hazrunoff.eu/) are adopting the data stream for early warning & detection, follow-up, and early response to different or combined types of flooding and hazmat pollution in inland and transitional waters.

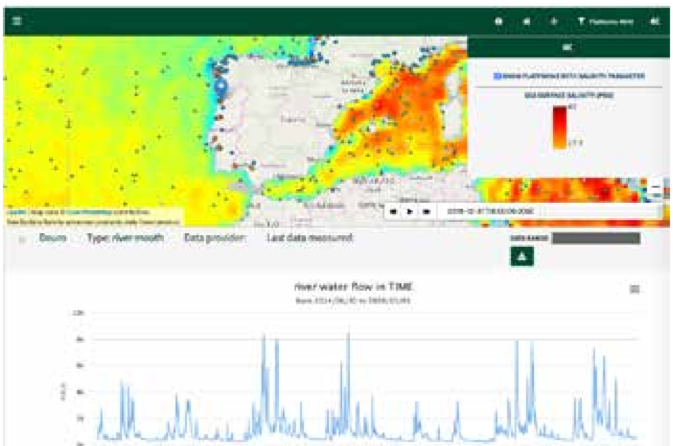


Figure 2: Capture of LAMBDA project data portal (http://www.cmems-lambda.eu/mapviewer/)

III. CONCLUSIONS

The need for river data management for coastal applications and coastal oceanography represents a key emerging topic and users welcome an easy and innovative service supporting their daily activities. Results will be presented on progress made in making river data available, accessible, downloadable and usable.

Welcome to the now: How to handle EOV Inorganic Carbon data in respect to the Sustainable Development Goals and the FAIR Data Management Principles

Benjamin Pfeil, University of Bergen (Norway), benjamin.pfeil@uib.no
Steve Jones, University of Bergen (Norway), steve.jones@uib.no
Maren K. Karlsen, University of Bergen (Norway), maren.karlsen@uib.no,
Camilla Stegen Landa, University of Bergen (Norway, camilla.landa@uib.no
Rocio Castano Primo, University of Bergen (Norway), rocio.primo@uib.no

Some data simply does not meet the criteria for mainstream (often physical) oceanographic data but the demands on how to make it available, perform higher level QC, achieve near real-time (NRT) data products are rising in times of the Agenda 2030 of the United Nations and their Sustainable Development Goals (SDG) where one target (14.3) addresses the Essential Ocean Variable Inorganic Carbon - on top data has to follow the FAIR data management principles. How do adopt new data flows, implement new tools and procedures into established scientific communities?

EOV Inorganic Carbon observations collected from instruments at sea are typically processed by individual PIs before being submitted to data centres and other data archives. Often this work is done on an ad hoc basis using unpublished self-built software and published in unique formats. This leads to lacks in data flow and data availability in respect to SDG target submissions and NRT data availability. Inconsistent data treatment and delivery, lacks of reproducibility are hindering/ impeding the Interoperability and Reusability of the FAIR principles since much work is needed to convert data formats while effective re-use of the data is challenging with lacking enriched metadata and details about data processing. If measurements are processed using open source, fully documented standard tools, all data can be traced fully back to source and reassessed if necessary.

The European Research Infrastructure ICOS (Integrated Carbon Observation System) aims at increasing the fundamental understanding of the marine, atmospheric and ecosystem carbon cycle, it's underlying processes and verify the effectiveness of policies aiming to reduce greenhouse gas emissions. Within the marine part of ICOS, the Ocean Thematic Centre is developing QuinCe, an browser-based tool for uploading, processing, quality control and publication of data from underway pCO_2 systems on ships, moorings and SailDrones. Data from the instruments can be uploaded directly in any text format, where it will be standardised and processed using algorithms approved by the community. PIs can perform full quality control of the data following SoPs and best practices, which is recorded and then sent to the ICOS Carbon Portal and SOCAT (Surface Ocean CO2 Atlas) project for publication where it is used for decision making and informs the annual Global Carbon Budgets of the Global Carbon Project and can be submitted to the SDG 14.3 target: "average marine acidity measured at an agreed suite of representative sampling stations".

Where data is transmitted directly from ship, mooring or SailDrone to shore, QuinCe processes, quality controls and publishes Near Real Time data to the ICOS Carbon Portal and to Copernicus Marine Environmental Services (CMEMS) as soon as it is received with no human intervention, greatly reducing the time from measurement to data availability and is the baseline for the CMEMS Carbon NRT data product.

All quality control decisions recorded with justifications, so the veracity of all data points can be assured by end users. Standardized vocabularies and metadata formats are compliant with the United Nations Sustainable Development Goal methodology 14.3.1 methodology are applied in the system or will be implemented in 2020.

This contribution will highlight the challenges and achievements of the global marine biogeochemistry community of how to implement new tools, data from new platforms (e.g. SailDrone) and work flows for the Agenda 2030 and submission number 128: 'Ensuring data reproducibility from sensor to publication for ocean observations' will follow up details on how to use the tool QuinCe for interested parties.

Recent developments on the integrated information system for the support of the implementation of the EU Marine Strategy Framework Directive in Greece

Angelo Lykiardopoulos, Hellenic Centre for Marine Research (Greece), angelo@hcmr.gr
Paraskevi Drakopoulou, Hellenic Centre for Marine Research (Greece), vivi@hcmr.gr
Athanasia-Sissy Iona, Hellenic Centre for Marine Research/Hellenic National Oceanographic Data Centre (Greece), sissy@hnodc.hcmr.gr
Vassiliki Loukaidi, Hellenic Centre for Marine Research (Greece), vloukaidi@hcmr.gr
Christina Damianou, Hellenic Centre for Marine Research/Hellenic National Oceanographic Data Centre (Greece), ch.damianou@hcmr.gr
Vasilis Lakes, Hellenic Centre for Marine Research (Greece), vaslak@hcmr.gr
Konstantinos Kalkavouras, Hellenic Centre for Marine Research (Greece), ckalkav@hcmr.gr

According to the Marine Strategy Framework Directive (MSFD), EU Member States must operate monitoring programmes for the ongoing assessment of the environmental status of their marine waters on the basis of the indicative lists of elements set out in Annex III and the list set out in Annex V, and by reference to the environmental targets established pursuant to Article 10. Monitoring programs aim to evaluate the 11 Descriptors, including 40 Criteria and 56 Environmental Indicators, as defined in the MSFD (2008/56 / EU). Based on the 11 Descriptors, it is estimated that the Good Environmental Status (GES) is achieved or not. This means that monitoring should provide data which support the indicators in order to assess if Good Environmental Status (GES) has been achieved or is maintained, to measure progress towards environmental targets and evaluate the effectiveness of measures to achieve or maintain GES.

In respect to the monitoring requirements, the Hellenic Centre for Marine Research (HCMR), as the responsible organization for carrying out the required actions for the monitoring of the quality of the Greek marine waters, inter alia, is aiming to produce services capable to be used as indication tools of Good Environmental Status (GES) of marine waters across Europe and for Greek territory waters. The services produced are compliant to Inspire Directive and they are also using the Open Geospatial Consortium Web Services technologies.

The services have been divided into three different levels according to their aim and usability as following:

- The first level deals with the acquisition of monitoring data sets produced and gathered under the activities of the MSFD monitoring programs concerning Greek Waters. These data sets were processed and stored into a Data Base System.
- The second level deals with data sets from other Projects dealing as well with water monitoring. The Projects contributed for this aim mainly are: a) Poseidon Project, b) Argo Floats Project and c) WFD and others monitoring Projects.

These data sets from the Projects above have been also processed and either stored into the Data Base System, either are embedded to the platform via a specially developed machine to machine interface.

- The third level of services will stand as the presentation layer of the gathered water monitoring data. In particular, Good Environmental Status (GES) indicators will be produced for every Descriptor of water monitoring among any of water bodies categories which are: a) Sea Region Subdivisions, b) Territorial Water Bodies and c) Coastal Water Bodies.

In the future phases a Decision Making System will be developed and the Data Base will be enhanced with socioeconomic data for every monitoring area. Thus, the Decision Making System will offer an overall picture of the water health status among the different water body categories assisting and offering guidance to the Decision Makers with the decisions they have to be made.

For the visualization needs a Web based Geospatial portal has been developed for the end users giving the capabilities to navigate and search among the various water bodies areas for specific Descriptors and indicators data to be collected and download for further activities.

The architecture of the portal has as following:

- Postgres Relational Data Base System with stored processed data.
- OGC GeoServer for creating the necessary geospatial layers for any of the water bodies areas and stations of measurement.
- Interactive Web page interface for navigation, searching and downloading data sets.

The figures bellow depict some aspects of the portal.

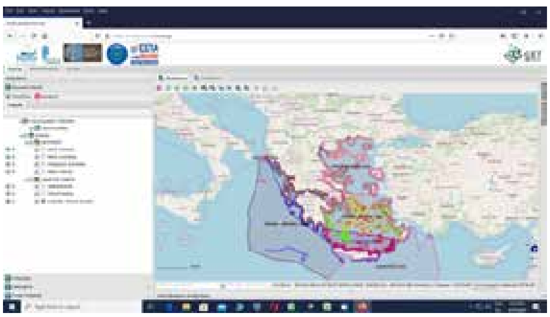


Figure 1: all water bodies on sight

Figure 2: Territorial water bodies and Ship Cruise root

Figure 3: Information of specific territorial water body

In-situ Real-time Underwater Noise Dataflow: from OBSEA to EMODnet

Enoc Martínez, Universitat Politècnica de Catalunya, Electronics Department (Spain), enoc.martinez@upc.edu
Daniel M. Toma, Universitat Politècnica de Catalunya, Electronics Department (Spain), daniel.mihai.toma@upc.edu
Antonio Novellino, ETT (Italy), antonio.novellino@ettsolutions.com
Marco Alba, ETT (Italy), marco.alba@ettsolutions.com
Joaquín del Río, Universitat Politècnica de Catalunya, Electronics Department (Spain), joaquin.del.rio@upc.edu

OBSEA Underwater Noise Monitoring System

Human activities such as shipping, construction, sonar and seismic exploration have been raising the underwater ambient sound level to unprecedented levels in the past decades. In order to reduce its harmful impact on the ecosystem, the European Commission's (EC) Marine Strategy Framework Directive (MSFD) included the long-term monitoring of underwater noise as a relevant indicator to achieve a good environmental status.

OBSEA is a cabled observatory, located off the coast of Vilanova i la Geltrú (Barcelona, Spain) at a depth of 20 meters (Del-Rio et al., 2020) networks of permanent cabled observation systems are being deployed in the ocean. This paper presents a balance of a decade of activity at the OBSEA cabled observatory, as an example of ocean monitoring success and drawbacks. It is not the objective of this article to analyze the scientific and technical aspects already presented by the authors in different publications (Table 4. It is equipped with a wide variety of oceanographic instruments, including CTD, ADCP, an underwater camera and a hydrophone. Since 2017 acoustic data is processed in-situ to provide real-time underwater measurements and can constitute the baseline for D11C2 (best practices guidelines on continuous underwater noise monitoring) according to the MSFD directive and the recommendations of the EC Task Group Noise (Dekeling et al., 2014). These measurements include Sound Pressure Levels (SPL) calculated at different band frequencies: 63, 125, 2000 Hz third-octave bands and full bandwidth.

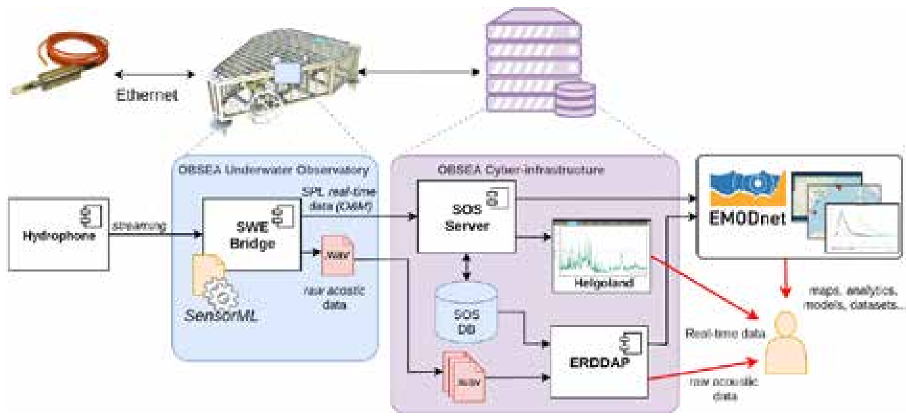


Figure 1: OBSEA's real-time underwater noise monitoring system dataflow

The underwater noise acquisition system, depicted in figure 1, provides an end-to-end real-time in-situ processing, following the FAIR principles (Findable, Accessible, Interoperable and Reusable). The SWE Bridge is a standards-based universal driver configurable through SensorML capable of interfacing almost any oceanographic instrument, including hydrophones (Martinez et al., 2017) Basel, Switzerland. The study of global phenomena requires the combination of a considerable amount of data coming from different sources, acquired by different observation platforms and managed by institutions working in different scientific fields. Merging this data to provide extensive and complete data sets to monitor the long-term, global changes of our oceans is a major challenge. The data acquisition and data archival procedures usually vary significantly depending on the acquisition platf orm. This lack of standardization ultimately leads to information silos, preventing the data to be effectively shared across different scientific communities. In the past years, important steps have been taken in order to improve both standardization and interoperability, such as the Open Geospatial Consortium's SensorWeb Enablement (SWE. It includes embedded real-time SPL algorithm for real-time noise measurements. The processed data is encoded in a Observations and Measurements (O&M) output, compatible with Sensor Observation Services (SOS). Alongside real-time processed data, raw audio files are also generated using the WAV format, allowing further analysis and validation.

Data and Metadata Management

The SensorML file contains a full description of the hydrophone metadata, including data streams, algorithm configuration, description, identification, contacts as well as technical information. To provide unambiguous meaning for each term, SeaDataNet and IOOS controlled vocabularies are used. The metadata contained within the SensorML file is propagated to the SOS service (52north implementation) and stored alongside the processed data, where it is archived and accessible through its standard API.

The WAV format, used for raw audio data, does not provide any standardized way to embed acoustic metadata. Thus, the ID3 container has been adopted. This metadata container, widely used in commercial audio formats such as mp3, embeds standardized metadata tags to a file. This tagging system has been leveraged to include all the hydrophone metadata as user-defined tags, including hydrophone sensitivity, deployment location, timestamp, sensor name, manufacturer serial number, contact information among others.

In order to make the WAV data available in a coherent and standardized manner, an ERDDAP service has been set up. ERDDAP provides a standardized way to retrieve data and metadata, including access to raw files. Using the ERDDAP and SOS interfaces the data is connected with EMODNet Physics data portal, where data is integrated into an underwater noise product. The same data is then used to calibrate underwater noise models and sound maps.

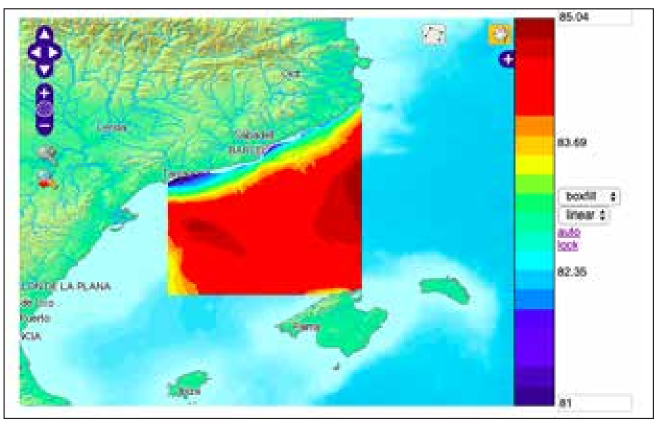


Figure 2: Example of a soundmap from EMODnet Physics in the OBSEA area

Acknowledgments

This work has been supported by the project EMSO-LINK (H2020 grant Agreement No. 731036), by Generalitat de Catalunya "Sistemas de Adquisición Remota de datos y Tratamiento de la Información en el Medio Marino" (SARTI-MAR)" 2017 SGR 371 and by the Spanish Ministry of Education and Science (MEC) with the project "Redes de sensores submarinos autónomos y cableados aplicados a la monitorización remota de indicadores biológicos" TEC2017-8786 and EMODnet Physics EASME/EMFF/2018/1.3.1.8/ Lot3/SI2.810790. This work used the EGI infrastructure with the dedicated support of INFN-CATANIA-STACK. Researchers want to acknowledge the support of the Associated Unit Tecnoterra composed by members of Universidad Politécnica de Catalunya (UPC) and the Consejo Superior de Investigaciones Científicas (CSIC).

References

J. Del-Rio *et al.*, "Obsea: A Decadal Balance for a Cabled Observatory Deployment," *IEEE Access*, vol. 8, pp. 33163–33177, 2020, doi: 10.1109/ACCESS.2020.2973771.

R. P. A. Dekeling et al., Monitoring Guidance for Underwater Noise in European Seas, Part II: Montitoring Guidance Specifications. 2014.

E. Martínez, D. M. Toma, S. Jirka, and J. Del Río, "Middleware for plug and play integration of heterogeneous sensor resources into the sensor web," *Sensors (Switzerland)*, vol. 17, no. 12, pp. 1–28, 2017, doi: 10.3390/s17122923.

Interoperable Provision of Research Vessel Tracking Data via OGC SensorThings API and Sensor Observation Service

Christian Autermann, 52°North GmbH (Germany), c.autermann@52north.org
Simon Jirka, 52°North GmbH (Germany), jirka@52north.org
Dick Schaap, MARIS B.V. (The Netherlands), dick@maris.nl

Within the last year, many applications of Sensor Web technologies to share marine observation data have emerged. This comprises not only data standards for encoding observation data but also the provision of metadata and data access interfaces. Many of these applications have focused on the provision of stationary in-situ data (e.g. data collected by fixed buyos or stations).

Within this paper, we introduce how the Sensor Web Enablement (SWE) standards of the Open Geospatial Consortium (OGC) can be applied to handle near real-time flows of research vessel data. Specific challenges addressed in this paper are the efficient transmission of the data into a Sensor Web infrastructure as well as the subsequent provision via Web applications.

As a starting point of the developments of the project, several requirements were identified that guided the design of the system:

- Integration of different research vessels of different operators
- Achieve an integration beween the EARS (Eurofleets Automatic Reporting System) software system which is running on each of the considered research vessels
- Ensure a timely and lightweight flow of information
- Rely on internaltional interoperability standards to ensure extensibility and sustainable re-use
- Facilitate the visualisation of historic as well as live data

The resulting system architecture is outlined in Figure 1. Each of the research vessels is running an instance of the EARS (Eurofleets Automatic Reporting System) software system. This software collects the generated observation data. From there on, different approaches such as UDP/Iridium but also manual data transfer via USB storage are used to transfer the collected data to the on-shore vessel operator for further storage and archival. Subsequently, the research vessel operators provide access to the available current near-live data via HTTP endpoints.

The HTTP endpoints offering the latest ship data are the basis for enabling the feeding of the data into a Sensor Web infrastructure. For this purpose, a feeder application continuously checks the HTTP endpoints for new ship data. If new data is available, it is downloaded and forwarded via a lightweight MQTT stream to an OGC SensorThings API server.

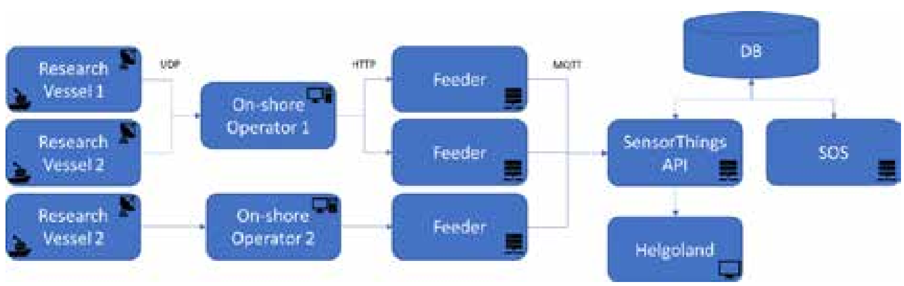


Figure 1: Architecture Overview

The SensorThings API server acts as a sink for the MQTT streams delivering the different types of ship data and metadata. In details the following types of information are handled:

- Navigation data: longitude/latitude, heading, speed, depth, course over ground, speed over ground
- Meteorological data: wind mean velocity, wind gust, wind direction, air temperature, humidity, solar radiation, atomospheric pressure
- Thermosalinity data: salinity, water temperature, raw fluorometry, density (Sigma-t)

After receiving new data via MQTT, the SensorThings API server takes care of ingesting the collected vessel data into a central Sensor Web database (in this case a relational PostgreSQL database). From there on, the data is made available via interoperable interfaces (OGC Sensor Observation Service and OGC SensorThings API).

On top of the SensorThings API, a Web viewer application is deployed. This Web viewer, based on the 52°North Helgoland Sensor Web viewer, allows users to view the current positions and data of the included research vessels. In addition, also historic data of the ships (e.g. trajectories of past journeys) can be discovered and visualised.

Besides establishing the necessary data flows and visualisation tools based on technologies previously developed and enhanced in projects such as SeaDataCloud, the modelling of the research vessel data was a second major task during the design process. In this case, a consistent mapping of the different entity types to the OGC Sensor Observation Service and SensorThings API data models had to be established. Core elements of this model include:

- The collected data is modelled as so called SpatialFilteringProfile measurements as defined by the OGC Sensor Observation Service standard. This means that the latest navigation data is merged with the corresponding thematic observation data into individual observations.
- The tracks of the research vessel are considered as so-called Features of Interest which are dynamically updated with each new message containing ship navigation data.
- The Thing concept of the OGC SensorThings API is mapped to the research vessels (in case of the SOS, the vessels are mapped to the concept of "procedures")

Furthermore, to ensure semantic interoperability, the terms used for referring to the observed properties and to the units of measurements are taken from the NERC Vocabulary Server.

The validation of the developed approach is carried out in close cooperation with Eurofleets. Data delivered by the vessels "Belgica" (Royal Belgian Institute of Natural Sciences) and "Sarmiento de Gamboa" as well as "García del Cid" (both CSIC).

In summary we present an approach how live streams of research vessel data can be collected via interoperable Sensor Web standards. This comprises not only the encoding of the data and metadata, but also lightweight data transmission technologies, semantic as well as syntactic data interoperability as well as Web-based visualisation tools.

Credits

The work presented in this paper was performed as part of the EMODnet Ingestion 2 project in cooperation witht Eurofleets.

Ship of Opportunity Monitoring of the Western Mediterranean Sea using FerryBox

Sana Ben Ismail, INSTM (Tunisia), sana.benismail@instm.rnrt.tn
Sondes Awachri, INSTM (Tunisia)
Mohamed Anis Ben Ismail, INSTM (Tunisia)
Nouha Barraj, INSTM (Tunisia)
Cherif Sammari, INSTM (Tunisia)

Introduction.

The use of "Ships of Opportunity" (SoO) in marine observation networks has been promoted by Global Ocean Observing System to deal with the lack of monitoring systems in ocean that enable continuous observation of coastal seas is a major obstacle when it comes to understand the implications for ecosystem dynamics and functioning. In fact, traditional monitoring of marine environments using oceanographic research vessels is costly and often lacks the spatial coverage and temporal resolution that is required to study variability in physical, chemical and biological conditions on seasonal or interannual time scales. Hitherto, these gaps in the data are a serious problem for accurate assessments of climate- and human-induced changes in marine environments. Unattended autonomous observing systems aboard SoO are cost-effective and reliable alternatives to obtain continuous observations on near-surface parameters with high spatial coverage and temporal resolution. Here after examples are presented that highlight the added value of the recorded data for the study of both long- and short-term variability in water mass stability, plankton communities and surface water productivity in the Western Mediterranean Basin.

By the evaluation of technical and scientific performance, it is evident that FerryBoxes have become a valuable tool in marine research that helps to fill gaps in coastal and open ocean operational observation networks.

Material and methods.

The FerryBox continuously measures oceanographic parameters in a flow-through system (Fig. 1). Depending on the draught of the ship, the water intake is fixed at a depth between 2 and 7 m. A debubbling unit removes air bubbles, which may enter the system during heavy sea. Coupled to the debubbler is an internal water loop in which the water passes different sensors as it circulates. The basic sensors used measure temperature and salinity, turbidity and chlorophyll-a fluorescence. In addition, an oxygen sensor (Clark electrode or oxygen optode) and pH sensor were installed. Housekeeping parameters such as flow rates and pressures inside the water loops were measured to supervise the system which was developed in collaboration with an industrial partner and is commercially available (4HJena engineering GmbH, Germany).

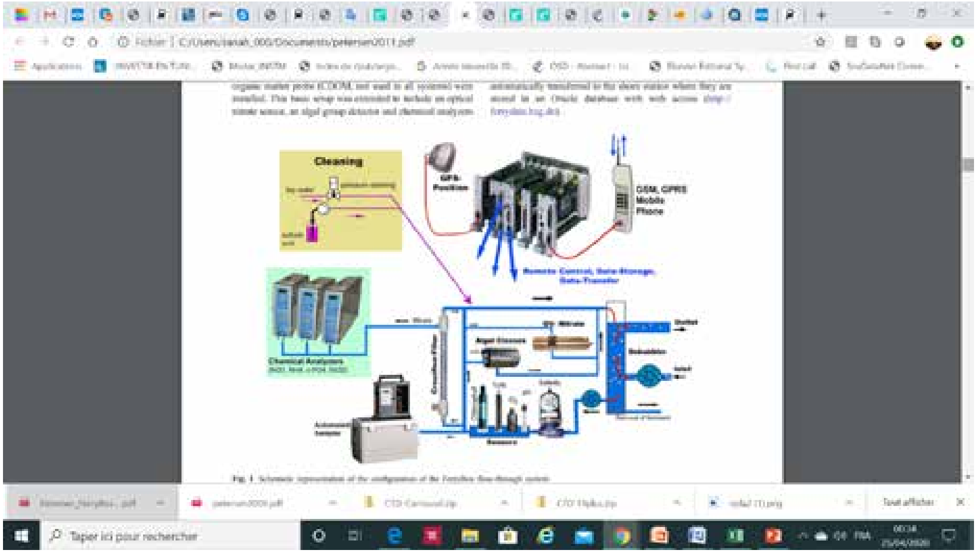


Figure 1: FerryBox system

The FerryBox system is installed on board a Carthage ferry of the Compagnie Tunisienne de Navigation (CTN) (Figure 2). The ferry crosses the western Mediterranean Sea mainly between Tunis and Marseille and between Tunis and Genova on a weekly basis, recently since 2018 the trajectories of the ferry changes in summer season and we had Genova-Zarzis cross.

Figure 2: FerryBox lines operated in the Western Mediterranean Sea

Results.

We outline our experience from 4 years of continuous FerryBox research in the Western Mediterranean and then allude to some of the possible applications. In this study, we give, in the first insight detailed comparisons of inter-annual FerryBox measurements in the Western Mediterranean Sea between 2016 and 2017, which shows a decrease in the Primary Production against an increase in temperature for similar days of each year. In the second place, we give comparisons of FerryBox measurements with satellite data in bloom period in order to demonstrate the value of these data for satellite validation. We give insight in the folowing section about the developement of database and FerryBox data management. The system we are about to create should respect most of the requirements of ISO 9001 quality norms. the data stream from the shore station to the Python MySQL data base is described.

POSTERS

The national research infrastructure NMDC (Norwegian Marine Data Centre) is providing seamless access to marine data

Helge Sagen, Institute of Marine Research (Norway), Helge.Sagen@hi.no

Marine 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.

Figure 1: NMDC portal at www.nmdc.no

Integrated infrastructure

An integrated infrastructure NMDC addressing interoperability, data documentation, archiving, providing data access has been established. 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. An upgrade and further development of the infrastructure is needed to be in the forefront of the technical developments and advancements.

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.

NMDC services

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.

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 centre at IMR holds the certificate Core Trust Seal to ensure good data management practices. 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.

The NMDC portal search interface at www.nmdc.no

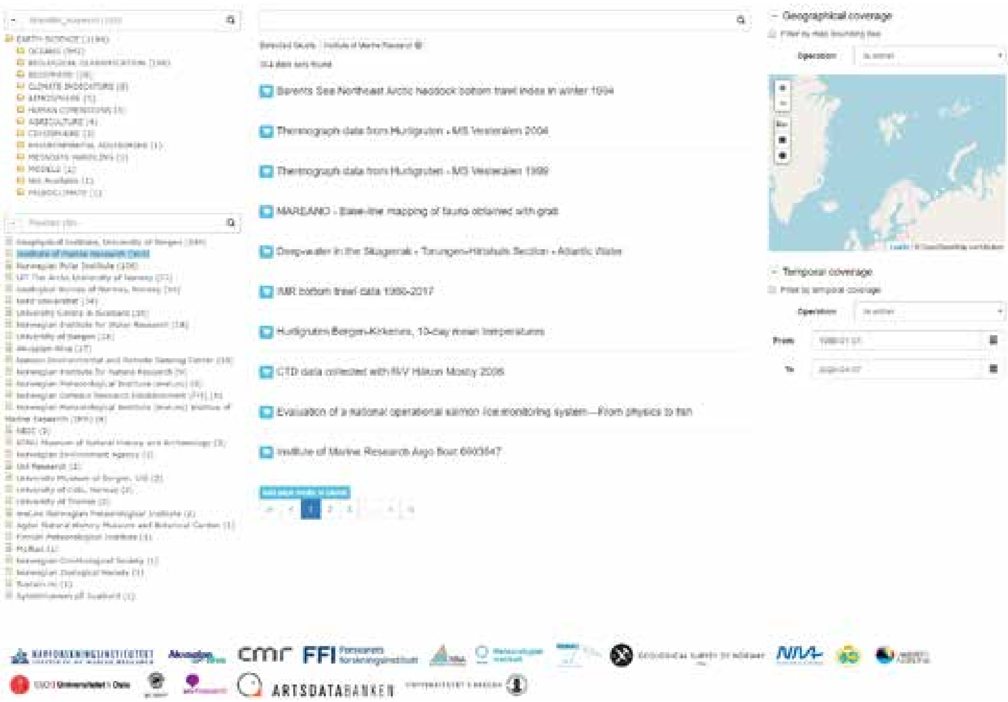


Figure 2: NMDC search interface

The search interface queries several parts of the metadata, free text, scientific keywords, data provider, geographical polygon and time interval. The geographical polygon search supports two cases, either all data set within the polygon or parts of the data set intersects with the polygon. The same functionality is implemented for the temporal coverage also.

ReefTEMPS, FAIRs access to Reef ecosystem environmental measurements

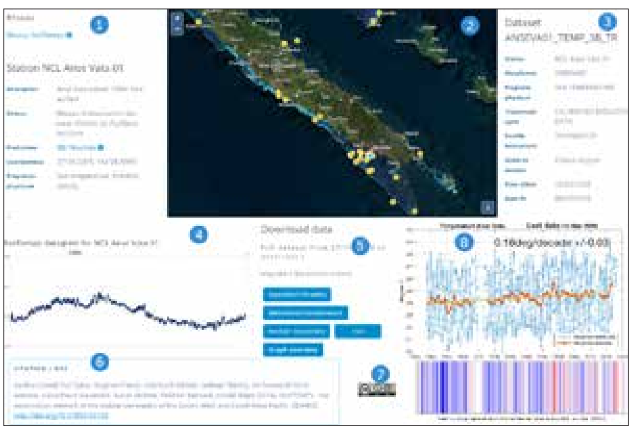
Sylvie Fiat, ENTROPIE, IRD, Univ Réunion, CNRS, Ifremer, Univ Nouvelle-Calédonie (New-Caledonia), Sylvie.fiat@ird.fr
Jérôme Aucan, ENTROPIE, IRD, Univ Réunion, CNRS, Ifremer, Univ Nouvelle-Calédonie (New-Caledonia), Jerome.aucan@ird.fr
Régis Hocdé, MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, (France), regis.hocde@ird.fr

ReefTEMPS is a coastal ocean observatory in the South, South West and West Pacific that provides long-term monitoring of climate change and its effects on the status of coral reefs and their resources. The network has been deploying sensors for temperature, pressure, conductivity, fluorescence and/or turbidity at 98 sites throughout the coastal zone of some 20 island states and territories since 1958.

ReefTEMPS (http://www.reeftemps.science/), labelled *Service National* (for) *Observation* (SNO), is part of the French research infrastructure "coastal ocean and nearshore observations" IR ILICO.

The network's platforms have been strategically positioned in relation to the processes observed and the underlying scientific issues, in locations that are sometimes difficult to access and for which very little data exist. Indeed, satellite remote sensing can not yet provide high resolution and accuracy over coastal waters. Autonomous sensors, with acquisition frequencies ranging from 30 minutes to 1 second (high frequency), are immersed for periods ranging from 6 months to 2 years and are then downloaded. The acquired data are then added to the database by the instrumention engineers. A validation process involving oceanographers qualifies the data and produces long time series or "historical series". The network was deployed throughout the Pacific island region starting in 2010, but has incorporated much older stations, some of which have been in operation since 1958.

ReefTEMPS is used to measure long-term trends, inter-annual changes (i.e. seawater temperature), and rare or extreme events (i.e. tropical storms) and to understand the physical, chemical and biological processes at work in the overall dynamics of the systems (i.e. coral bleaching).



Getting Anse Vata (New Caledonia) temperature data in ReefTEMPS:

1/ station metadata,
2/ station location on interactive map (blue spot),
3/ Dataset metadata,
4/ Interactive datagram to preview data,
5/ Download available,
6/ DOI reference,
7/ Data licence,
8/ Data science products as a result.

FAIR, as Findable, Accessible, Interoperable and Re-usable (FAIR principles, Wilkinson et al. 2016), in its Data principles

(Findable) The ReefTEMPS data and metadata are openly available thanks to a unique identifier provided by Seanoe, a French data warehouse provided by the ODATIS ocean cluster of the Data Terra research infrastructure (see figure 1 case 6)): doi:10.17882/55128.

Several catalogue web services (CSV, SOS) are delivered directly. ReefTEMPS is also referenced in thematic portals (ILICO, Seanoe). Data searches in generic search engines (datasetsearch.research.google.com, search.datacite.org) with a few simple keywords (i.e. oceanographic data pacific) identify ReefTEMPS in the top10.

(Accessible) The ReefTEMPS information system allows access to metadata and data using web services available under the http protocol without access restrictions. Opening up the data was a strong initial choice.

(Interoperable) Numerous standards are used to provide the data to as many user communities as possible, including oceanographers, ecologists, and the public interested in wind or marine based sports. Much of this is based on the Open Geospatial Consortium (OGC): with the Catalog Service for the Web (CSW), Web Map Service (WMS), Web Feature Service (WFS) and Sensor Observation Service (SOS). The other part is based on OpenDAP "Open-source Project for a Network Data Access Protocol" where data are organized using the Cimate and Forecast standard (CF) and the OceanSites data format. The vocabularies used in ReefTEMPS are mainly provided by OceanSites but some more specific ones are the vocabularies defined by Ifremer.

(Reusable) The data are updated six months on the Seanoe archive storage, with incremental versions of the dataset, revised versions always available on request and referenced using unique keys. Exemple of the 2019-09 version of the dataset: doi:10.17882/55128#66815.

The diversity of data delivery modes, whether by direct download or by archiving, allows ReefTEMPS to optimize data reuse. It is published under Creative Commons.

FAIR, as fair, sustainable development (Data ecosystems for sustainable development, UNDP report, 2015), in its service organization

ReefTEMPS provides data management and delivery for 14 Pacific countries divided into 20 territories, most of which has very little data capacity. The architecture of the information system has been designed so that multiple instances can be deployed very simply in various locations while keeping the entire network interoperable and accessible from a single entry point. Indeed, it is preferable, from the point of view of the ethics surrounding the intellectual property of the data, that each country be able to host locally its own productions. Although we have not yet had the opportunity to implement this networking functionality, it is something we have been promoting with our partner.

In our poster, we will present the global architecture of the information system with the open sources tools used to build it according to FAIR concepts. We will specify the standards and norms implemented to publish the metadata and data. A short focus will be devoted to the analysis of the topology and flows of consultations, extractions and use of the data. Finally, we will expose the organizational workflow of data processing with our subnetwork partners from the various countries of the Pacific island region.

Figure 2: ReefTEMPS multi-instance network organisation.

Harmonised ocean monitoring from the UK large research vessels (I/Ocean)

Louise Darroch, National Oceanography Centre - British Oceanographic Data Centre (UK), louise.darroch@bodc.ac.uk
Juan Ward, National Oceanography Centre – National Marine Facilities (UK), juaward@noc.ac.uk
Alexander Tate, British Antarctic Survey – UK Polar Data Centre (UK), ajtate@bas.ac.uk
Justin Buck, National Oceanography Centre - British Oceanographic Data Centre (UK), juck@bodc.ac.uk
Alexandra Kokkinaki, National Oceanography Centre - British Oceanographic Data Centre (UK), alexk@bodc.ac.uk

More than 40% of the human population live within 100 km of the sea. Many of these communities intimately rely on the oceans for their food, climate and economy. However, the oceans are increasingly being adversely affected by human-driven activities such as climate change and pollution. Many targeted, marine monitoring programmes (e.g. GOSHIP, OceanSITES) and pioneering observing technologies (e.g. autonomous underwater vehicles, Argo floats) are being used to assess the impact humans are having on our oceans. Such activities and platforms are deployed, calibrated and serviced by state-of-the-art research ships, multimillion-pound floating laboratories which operate diverse arrays of high-powered, high-resolution sensors around-the-clock (e.g. sea-floor depth, weather, ocean current velocity and hydrography etc.). These sensors, coupled with event and environmental metadata provided by the ships logs and crew, are essential for understanding the wider context of the science they support, as well as directly contributing to crucial scientific understanding of the marine environment and key strategic policies (e.g. United Nation's Sustainable Development Goal 14). However, despite their high scientific value and cost, these data streams are not routinely brought together from UK large research vessels in coordinated, reliable and accessible ways that are fundamental to ensuring user trust in the data and any products generated from the data.

The National Oceanography Centre (NOC) and British Antarctic Survey (BAS) are currently working together to improve the integrity of the data management workflow from sensor systems to end-users across the UK National Environment Research Council (NERC) large research vessel fleet, as part of a UK initiative, I/Ocean. In doing so, we can make cost effective use of vessel time while improving the FAIRness, and in turn, access of data from these sensor arrays. The initial phase of the solution implements an Application Programming Interface (API) framework (Fig. 1) with endpoints tailored towards different end-users such as scientists on-board the vessels as well as the public on land. The framework is interfaced with a web portal, used to register ships and sensors, as well as their associated metadata. An innovative API-based NMEA data logging system, located on the ships (RVDAS), will be used to harmonise on-board access while mediating automated sensor triage in real- time, assuring sensors are working correctly and only the best data are obtained. This system will be supported with a standardised digital event logging API and web interface, allowing data quality issues to be identified and resolved quickly. Novel open-source, data transport formats will be produced that are embedded with well-structured metadata, common standards and provenance information (such as controlled vocabularies), allowing end-users of all abilities to access and understand data across platforms. These will include an extension to SeaDataNet Climate Forecast NetCDF conventions for trajectory data. To allow scientists and data managers to improve the quality of this data, an open- source, community-driven data processing application will be developed that applies quality control to international standards (SAMOS, IOOS Qartod, GoSUD). Data will be democratised through shore- based open data APIs including the National Oceanic and Atmospheric Administration's (NOAA) ERDDAP and the Open Geospatial Consortium's (OGC) Sensor Observation Service (SOS), part of the Sensor Web Enablement (SWE) standard specification. This will allow end-users to discover and access data or layer their own tools to meet their own needs. To preserve provenance throughout the pipeline, each sensor will be unambiguously identified using novel instrument persistent identifiers (PID), part of the latest recommendations from the Research Data Alliance (RDA).

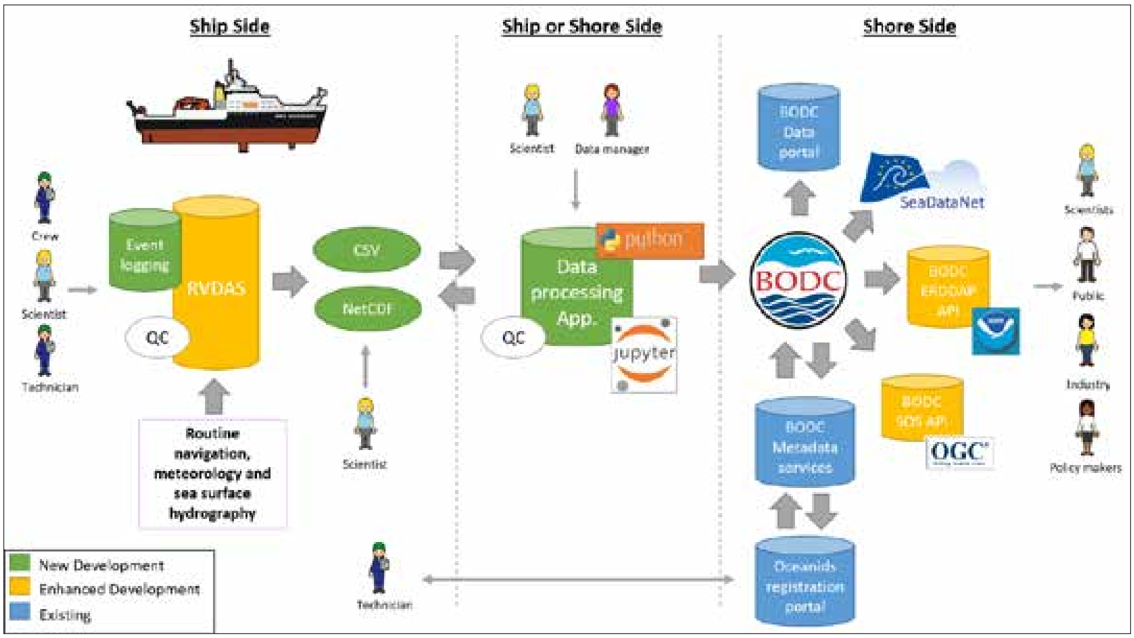


Figure 1: Initial phase of the I/Ocean initiative to improve the integrity of the data management workflow from sensor systems to end-users across the UK National Environment Research Council (NERC) large research vessel fleet.

While the initial phase of I/Ocean is focused on routine navigation, meteorology and sea surface hydrography variables in delayed-mode, our vision is to extend to near real-time data delivery and other strategic data types such as complex swath bathymetry that supports the Seabed 2030 global initiative. Access to universally interoperable oceanic data, with known quality and provenance, will empower a broad range of stakeholder communities (including academia, government and industry), creating opportunities for innovation and impact through data use, re-use and exploitation.

Zooplankton diel vertical migration database.

Vladimir Solovyev, P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences (Russia), sol@ocean.ru
Tamara Shiganova, P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences (Russia), shiganov@ocean.ru
Alexander Ostrovskii, P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences (Russia), osasha@ocean.ru

Zooplankton diel vertical migration, being widespread in a marine and freshwater environment, remains a challenge for researchers: biologists, chemists and acousticians.

The database to support research on vertical migration of zooplankton (ZMD) has been developed in Shirshov Institute of Oceanology, Russian Academy of Sciences. ZMD contains long-term data from the Black Sea observations since 2009 and from the Baltic Sea since 2011 obtained with anchored autonomous Aqualog profilers, ship observations and satellite remote sensing.

The vertical migration of zooplankton depend on many environmental factors such as density, light, oxygen, temperature, salinity, chlorophyll, etc.

The Aqualog moored profiling carrier collects regular time series of vertical profiles of marine environmental data at a fixed geographic position over extended periods of time.

The profiler is equipped with oceanographic sensors and makes repeated round trips up and down a taut mooring wire between the subsurface flotation and the anchor, while maintaining a stable orientation with respect to the direction of current.

When compared to traditional moorings, which use multiple instruments at fixed depths, the profiler acts as a lift for a single set of third-party sensors, thus increasing the vertical resolution of the data while decreasing operational costs and optimizing value.

Aqualog profiler was equipped by high frequency 2MHz ADCP. It provided a higher spatial resolution for the measurements of the migrators velocity and volume backscattering strength than in the previous hydroacoustic studies.

Hydroacoustic methods, enabling sufficiently cheap, neither time consuming nor labor-intensive data collection and permitting the synoptic coverage of a large area, could provide information valuable for biologists and chemists. It could include: a zooplankton migration magnitude (upper and lower boundaries of the depth range covered during the migration), zooplankton migration velocities, biomass and spatial distribution.

The synchronous measurements of temperature, salinity, pressure, dissolved oxygen, Chlorophyll, turbidity were made using a Seabird SBE 52MP or Idronaut 319 CTD sensors interacting with a Seabird SBE 43F or AANDEREA Optode oxygen sensor, Seapoint Chlorophyll Fluorometer and Turbiduty sensor. Simultaneously, the vertical profiles of horizontal current velocity components were measured.

To identify the influence of various parameters on the zooplankton vertical migration, especially temperature, salinity and oxygen concentration, the calibration data of these sensors are stored in the database. The data passed quality control according various criteria.

Additionally, the data collected by the Terra, Aqua, Landsat 8 and Sentinel satellites were used to obtain the information on chlorophyll a (Chl a) surface concentration and diffuse attenuation coefficient (K490), depending on the surface suspended matter concentration.

ZMD also stores data from ship observations near profiler deployment points. Temperature, salinity, oxygen, chlorophyll, turbidity profiles were measured by CTD Rosette System from the ship. Water samples were taken using bathometers from different depths in the water column for chemical analyses. The database also stores data of concentration and species composition of zooplankton was obtained by biological nets.

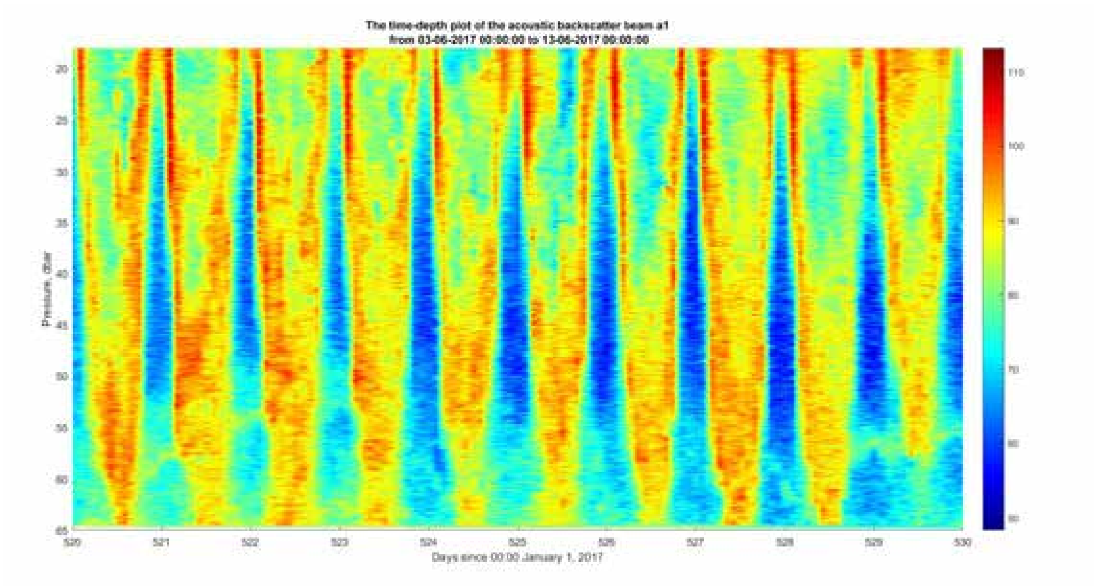


Figure 1: Zooplankton diel migration plot

ZMD allows to request data using observation period, geographical point of observation and sensors set. Output format can be in CVS, ODV, XML, NetCDF. Sensors data are synchronized in time, depth. Figure 1 shows an example of the migration of zooplankton in the Baltic Sea obtained from the ZMD system.

ZMD was developed using Oracle Database Server and Tomcat application server; Matlab is used for data processing.

Fishing gear as a data collection platform: Opportunities to fill gaps in ocean observation networks

Cooper Van Vranken, Berring Data Collective (Denmark), cooper@berringdatacollective.com
Berthe Vastenhoud, Berring Data Collective (Denmark), berthe@berringdatacollective.com
Patrick Gorringe, SMHI (Sweden), Patrick.Gorringe@smhi.se

Fishing gear as a data collection platform

While open oceans are observed by automatic data collection devices, there is a lack of sub-surface physical oceanographic data in coastal and shelf-seas. Commercial fishing gears can act as platform for sensors, which measure physical oceanographic data during fishing operations. While fishermen are catching fish, they can also catch bottom data and then profiles when the net goes down and up.

Quantification and comparison of the existing sub-surface data coverage with fishing activities show that integration with fishing could contribute to filling gaps in existing ocean observation systems in coastal and shelf-seas. Figure 1 maps the mean number of measurements taken by the current sub-surface ocean observation network and the mean of monthly number of fishing events in 2017 and 2018 in Alaska.

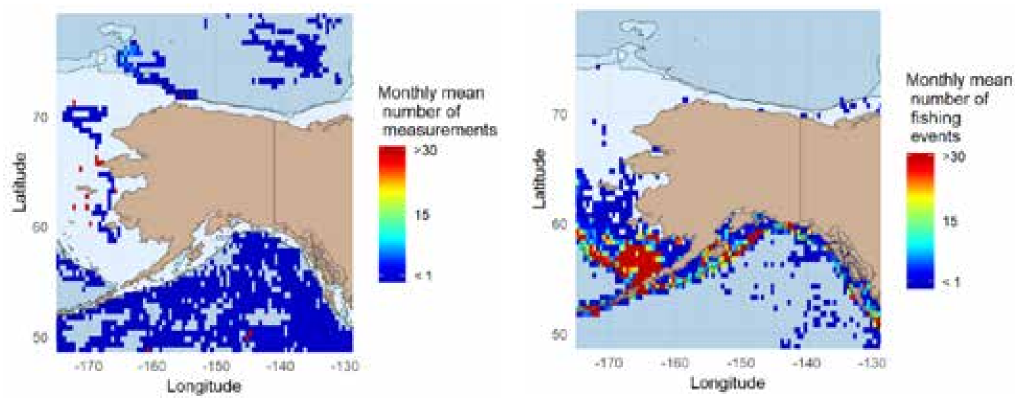


Figure 1: LEFT: Mean monthly number of sub-surface observations. RIGHT: Mean monthly number of fishing events. Data sources: EMODnet Physics, Argo, JCOMMOPS, DBCP, Ocean Sites, IOOS, and Global Fishing Watch.

The pattern shown in Alaska is repeated in most shelf and coastal regions around the world. It is counterintuitive that the holes in our ocean observation networks are close to shore where the majority of maritime activity takes place. These edge zones are subject to more influences and rapid changes compared with the deep ocean such as boundary currents and frontal mixing zones; and therefore, require more in-situ data to accurately monitor and model.

Due to the interdisciplinary use cases for fishing industry operations, fisheries management, as well as operational oceanography there are challenges associated with data standards and management of this type of oceanographic data. Additionally, balancing the needs for industry confidentiality with the FAIR principles for oceanographic is key to the future development and scale-ability of this collaborative approach to ocean observation.

We make the case that fishery data has the potential to complement existing ocean observing systems in areas where oceanographic data is lacking and needed most, in order to meet the needs of ocean data and forecasts users.

The fishing for data network

There is an emerging network of international scientific and industry programs that have used the approach of collecting oceanographic data with fishing gear as a platform (Figure 2). Results from these programs demonstrate that collection of ocean data via fishing gear can be successfully implemented by a diverse range of vessel types, locations, and fisheries. Data is now beginning to be assimilated operationally into models.

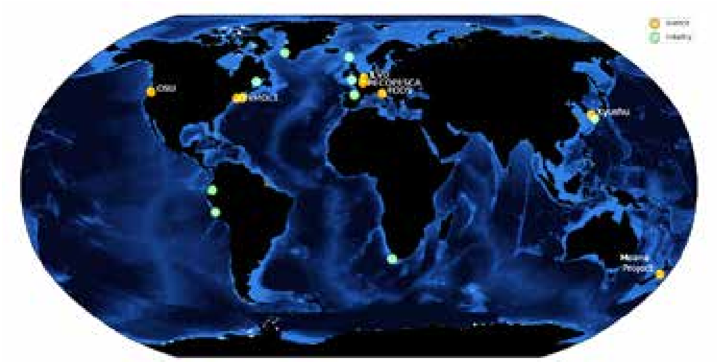


Figure 2: An emerging network of both science and industry lead initiatives in shelf seas around the world.

Berring Data Collective

Berring Data Collective (BDC) is an initiative both outfitting vessels and working to promote communication and data standardization between existing programs. Establishing standards, common data management and centralized distribution pathways is key to getting fishing gear as a widely accepted ocean observation platform (Figure 3).

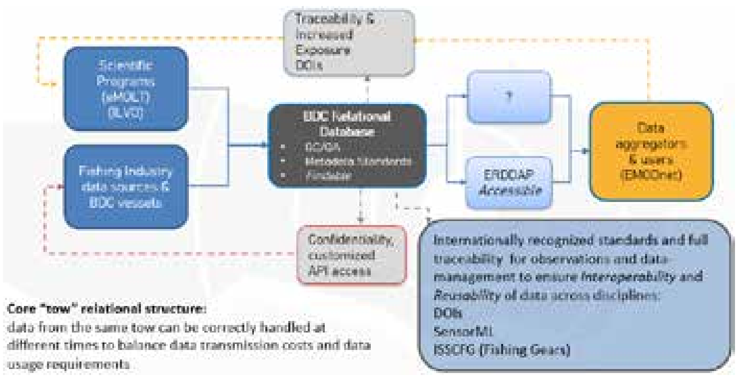


Figure 3: Schematic of the data flow of oceanographic data collected with fishing gear as a platform.

The current centralized data management structures and accompanying data standards are presented, which have been developed in conjunction with EMODnet Physics. Due to the uses for fishing industry operations, there are challenges with data standards and management of this data. Balancing the needs for industry confidentiality with the FAIR principles is key to the future development and scale-ability of this approach to ocean observation. An emerging network of scientific and industry programs is collecting data with fishing gear. We present progress on knowledge sharing, standardization, and data flow for this emerging collaborative approach to ocean observation.

Recent developments in oceanographic data and information system for Polish NODC initiative

Marcin Wichorowski, Institute of Oceanology PAS (Poland), wichor@iopan.pl
Joanna Stoń-Egiert, Institute of Oceanology PAS (Poland), aston@iopan.pl
Krzysztof Rutkowski, Institute of Oceanology PAS (Poland), kr@iopan.pl
Włodzimierz Krzymiński, Institute of Oceanology PAS (Poland), wkrzyminski@iopan.pl
Michał Piotrowski, MEWO S.A. (Poland), mpiotrowski@mewo.eu
Michał Wójcik, ZOO IM UMG (Poland), Michal.Wojcik@im.umg.edu.pl

Polish NODC initiative

Some of the Polish scientific organizations, among others the Institute of Oceanology Polish Academy of Sciences, Institute of Meteorology and Water Management National Research Institute, Polish Geological Institute National Research Institute, National Marine Fisheries Research Institute, University of Gdańsk, Gdańsk University of Technology, Maritime Institute of Maritime University in Gdynia, Pomeranian Academy in Słupsk and University of Szczecin, being involved in marine research for a long time have consolidated their efforts and undertook actions to make Polish oceanographic scientific data resources accessible for public from one national repository recently. These organizations having got well established good international cooperation, being members of international projects, organizations, and committees i.e. SeaDataNet, GeoSeas, EuroGOOS, BOOS etc., provide data separately to ICES, EMODNet and other well recognized international data and information centers and systems so far.

In order to increase Polish input to pan European oceanographic databases infrastructure – SeaDataNet, the idea arose to join efforts and to establish the structure of the Polish National Oceanographic Data Committee some years ago. One of the goals of the first stage of the Polish NODC development plan is to build an infrastructure as a solid basis for future administrative structure of this NODC. Such opportunity appeared thanks to the present Digital Agenda Poland.

Consortium comprising IO PAN, IMUMG, PGI PIB, MIR PIB, UG, APS, US - 7 of mentioned above partners, has successfully submitted project proposal in the frame of the Digital Agenda Poland programme dedicated for scientific organizations and is heading to provide open access to the scientific data resources. The consortium has been awarded for the project eCUDO.pl (Oceanographic Data and Information System). The project aims to harmonize Polish oceanographic data release and open access, make them interoperable through implementation of committed standards for information structure and communication protocols INSPIRE compliant. The project intends to develop and deploy distributed infrastructure for data management, and providing of FAIR and open access to oceanographic data resources.

Present activities encompass harmonization of environmental data collection and its preservation in accordance with INSPIRE requirements and SeaDataNet standards, securing resources for data management and stewardship, as well as digitalization of hardcopy data archives. The most significant results as expected, are better data discovery findability, accessibility, interoperability and finally higher potential for reuse of data as collected during the years of research activity and for the time being accessible only for internal purposes of the data originators.

eCUDO system development

Design of the system according to the programme requirements is user oriented and driven by development of services demanded by perspective users of the system. Except the access for "human" users of the system (including scientists, industry, administration, education, NGOs and general audience) and WCAG2.1 compliant interface of the system, the most expected data traffic with growing throughput in the near future is that generated by automated systems, both effected from metadata and data replication with other data centers, but variety of small but numerous autonomic systems (IoT). For this purpose extended M2M interfaces have to be developed. System is designed to meet 5 Star Open Data Model as well.

The system itself needs to be flexible and scalable. Oceanographic data is acquired and processed by many units and is distributed by their nature. Organizations engaged into system development are used to exploit data management infrastructures they have already deployed for their own purposes. These solutions are based on different technologies and every local instance of the system has to interface to these existing solutions, but provide unique access conditions and information structure on upper layer of data provisioning. Volume of data growth is correlated with sensor technologies development and is expected to growth exponentially in the future, as well as demands for data and number of data requests.

To meet these requirements design of the system base on hyper-convergent infrastructure according to this presented on Fig. 1.

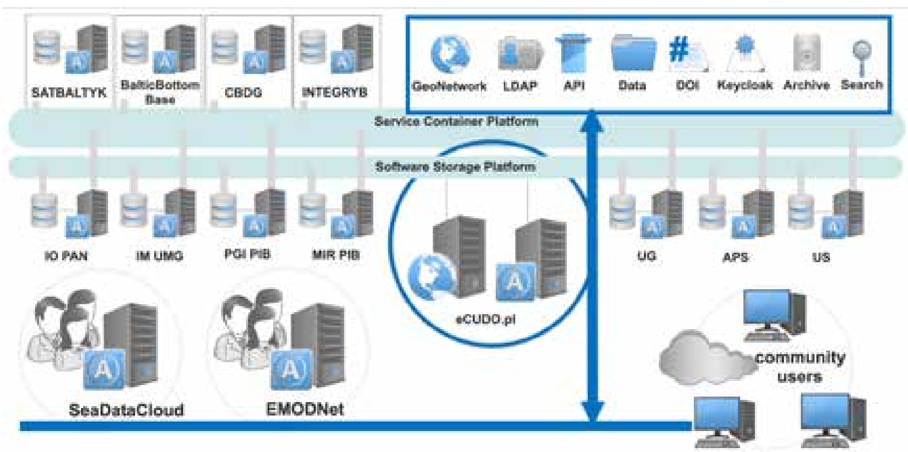


Figure 1: General design of the eCUDO system.

Conclusions

Development and deployment of the unified system providing access to distributed data resources as managed by the key scientific oceanographic institutes in Poland will trigger and provide added value for national economy through increase of data availability for all levels of administration, with simultaneous decrease of the total cost of data acquisition, management and exchange.

Unified data formats and protocols will boost the development of the services based on environmental data. Advanced services provided for clients (including data analysis services) extend availability of oceanographic data both to Polish and European organizations.

System will be open for all stakeholders and ready to accommodate other organizations and their data sources. Up to now cooperation has been agreed with Ministry of Climate, Ministry of Maritime Economy and Inland Navigation – System for Spatial Information of Maritime Authorities, as well as with the Institute of Meteorology and Water Management NRI.

SOCIB Data Infrastructure

Juan Gabriel Fernández, SOCIB (Spain), jfernandez@socib.es
Paz Rotllán, SOCIB (Spain), protllan@socib.es
Miguel Charcos, SOCIB (Spain), mcharcos@socib.es
Biel Frontera, SOCIB (Spain), bfrontera@socib.es
Xisco Notario, SOCIB (Spain), xnotario@socib.es
Inmaculada Ruíz, SOCIB (Spain), iruiz@socib.es
Miquel Àngel Rújula, SOCIB (Spain), mrujula@socib.es
Miquel Gomila, SOCIB (Spain), mgomila@socib.es
Sonia Gómara, SOCIB (Spain), sgomara@socib.es
Joaquín Tintoré, SOCIB and IMEDEA (Spain), jtintore@socib.es

Background

The Balearic Islands Coastal Ocean Observing and Forecasting System (SOCIB), is a Marine Research Infrastructure, a multi-platform and integrated ocean observing and forecasting system that provides streams of data, added value products, and forecasting services from the coast to the open ocean (Tintoré et al., 2013). The SOCIB Data Center is in charge of managing the SOCIB data assets and their data life cycle and supporting the SOCIB products and services strategy. The SOCIB Data Management Program is the framework where the Data Center aims to meet all these requirements.

A new Data Quality Strategy

SOCIB Data Quality Strategy is being re-addressed to better meet the new requirements from the coastal ocean community and to keep up-to-date the IT systems that power SOCIB. This new Data Quality Strategy is driven by the improvement in SOCIB of both (1) data assets FAIRness and (2) data quality processes and procedures. To guarantee that SOCIB benefits from the new Data Quality Strategy, the SOCIB Data Management Program is being updated accordingly. Although not fully implemented, the upgraded Data Management Program is currently offering new tools and processes that improve the quality of the SOCIB data assets.

An improved Data Management Program

Updates on the Data Management Program pivot around two main action lines. The first one includes 3 processes: Quality Assurance, Quality Control and Quality Assessment. The second one requires upgrading the IT architecture to assure data quality. These 2 actions are shown in the figure 1.

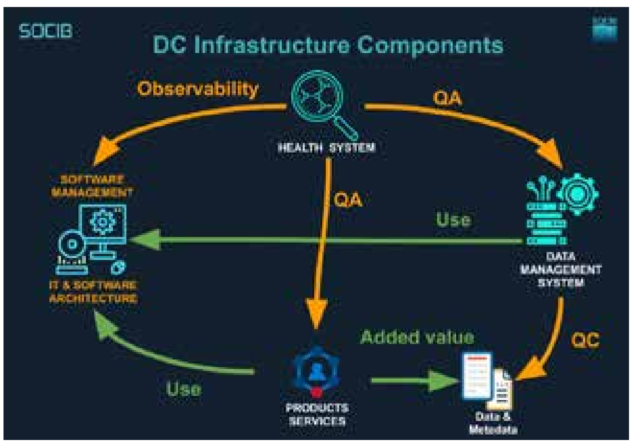


Figure 1: Data Center infrastructure components (orange represents the main improvement actions within the Data Management Program)

Within the SOCIB 2020-2021 Data Management Program update, the following action lines can be highlighted:

Development of a *beta* Data Health Dashboard to facilitate Data Quality Assessment. The data quality assessment system integrates this dashboard, an autonomous Data Quality Checker and a custom Incidences Management System.

Adoption of DevOps methodologies and tools to ensure both software quality and reliability of the SOCIB e-infrastructure. This also includes a migration to a container-based architecture (ie. Docker and Kubernetes).

Improvement of the e-infrastructure observability (logs, metrics, performance, alerts) by adopting state of the art technologies.

Improvement of the data life cycle observability: real time QC metrics, delay of arrival alerts and new data stewardship processes, among others.

In this presentation the above items will be introduced along with their role in the quality assurance of the SOCIB data assets. In addition 2 more action lines will be also introduced, showcasing downstream capabilities directly benefiting from the underlying data infrastructure:

Release of an improved Data Catalog portal (http://apps.socib.es/data-catalog/), which includes embedded metadata in JSON-LD format (compliant with the DataCite metadata scheme), allowing Google Dataset to interoperate and discover the SOCIB data products (ie. the "I" of FAIR).

Implementation of a linked data resource catalog to be integrated in the new SOCIB corporate website. There are also future plans to implement a linked data node.

Supporting processes

The new 2020 SOCIB Data Management Program will benefit from two strategic actions: (1) the implementation of a Quality Management System following the IODE's Quality Management Framework, to become an accredited Associate Data Unit of the World Data System (WDS), and (2) the CoreTrustSeal certification for the SOCIB data repository (with the support of the FAIRsFAIR project). Both certification processes are quite related to each other, since they focus on documenting quality processes of data management interconnected areas. Achieving the requirements of both certifications provides a better understanding of the complexity of the system and the opportunity to improve the processes, documentation and team management.

Conclusions

It is more and more evident today that existing ocean data systems have to meet new quality assurance requirements from the community they support. In addition, the evolution of both (1) existing technologies and methodologies, and (2) the emergence of new ones, create new core or internal requirements. In this scenario, the SOCIB Data Infrastructure, after a process assessment, is being upgraded with new components, processes and systems with the main goal of staying upfront the state of the art of leading international data infrastructures.

Tintore, J. et al. (2013), The Balearic Islands Coastal Ocean Observing and Forecasting System Responding to Science, Technology and Society Needs, Marine Technology Society Journal, 47 (1), doi: 10.4031/MTSJ.47.1.10

The JERICO e-Infrastructure

Miguel Charcos, SOCIB (Spain), mcharcos@socib.es
Jay Pearlman, IEEE (France), jay.pearlman@fourbridges.org
Francoise Pearlman, IEEE (France), jsp@sprintmail.com
Juan Gabriel Fernández, SOCIB (Spain), jfernandez@socib.es
Peter Thijsse, MARIS (The Netherlands), peter@maris.n l
Patrick Gorringe, SMHI (Sweden), patrick.Gorringe@smhi.se
Joaquín Tintoré, SOCIB (Spain) and IMEDEA (Spain), jtintore@socib.es
Miguel Ángel Alcalde, SOCIB (Spain), maalcalde@socib.es

Introduction to e-JERICO

The Joint European Research Infrastructure network for Coastal Observatory (JERICO) integrates a variety of observing platforms and technologies to observe and monitor the coastal areas in Europe. This meta-observing system provides complex and coupled information of the physical, chemical and biological processes through data from fixed buoys, piles, moorings, drifters, ferrybox, gliders, HF radars and coastal cable observatories. Achieving an understanding of the coastal processes requires high-quality data that cannot be obtained without standard and methodical work from the entire JERICO community. JERICO achieved the necessary interconnection between the various partners during the JERICO and JERICO-NEXT projects over the last eight years. This progress resulted in a diversity of resources for users including data, best practices, Sensor Web Enablement, software, manuals and publications. Moreover, JERICO aims to provide high quality data and data products in an optimal way facilitating the outcomes for different data user types including science, society, governments and private sectors.

Providing easy access and disseminating information of this amalgam of multidisciplinary resources requires a virtual infrastructure capable of linking and integrating each of these resources in a single and standard digital platform. The harmonization and connection of the information will have a critical impact on integrating observations of the physical, chemical and biological fields from the various regions that are dispersed around Europe. Virtual infrastructures address the need to integrate resources to support work and research. The current tendency in the international context is to implement these virtual and collaborative environments. For example, the European Plate Observing System (EPOS) succeeded in adopting this strategy to facilitate and integrate resources from distributed research infrastructures in a diversified context that is similar to JERICO. Similarly, for coastal ocean data, e-JERICO will provide access to the distributed resources from JERICO in a seamless and custom way according to the needs of the various users.

e-JERICO conceptial design

Providing organized and custom information to users requires implementing two concepts. On one hand, it is necessary to collect the information of the resources that are available in the community. Then, the information of these resources should be organized, linked and persisted in a metadata database, namely a resources catalog. On the other hand, the system should manage the requests from users, access the information in the resource catalog and provide the user with the custom outputs based on the context of the specific request. These requests may require access to the resources per sé that are distributed in the various infrastructures of the JERICO community and other external systems. e-JERICO is not an aggregator of data, but a data and service provider to the ocean community. It draws on different repositories such as EMODnet, ICOS, the Ocean Best Practices System, Ocean Docs and others. Due to the diverse nature of these resources and the heterogeneity of methods of the infrastructures hosting them, resources can be stored, accessed and interpreted in a myriad ways. The interaction between the main core system with the data centers and infrastructures is realized through an interoperability module that is capable of translating the distributed information into the common standard of e-JERICO. On the other side, the user access or machine-to-machine interaction requires a layer to gather the external petition and returns the information in a customized manner. Figure 1 illustrates the components of the e-JERICO infrastructure and their relationship. Managing the resources of the community and their connections facilitate processes aiming to monitor and evaluate the status of the system, to provide statistics and key performance indicators of the work of the community as well as measurement of FAIRness of the data cycle for each specific flow.

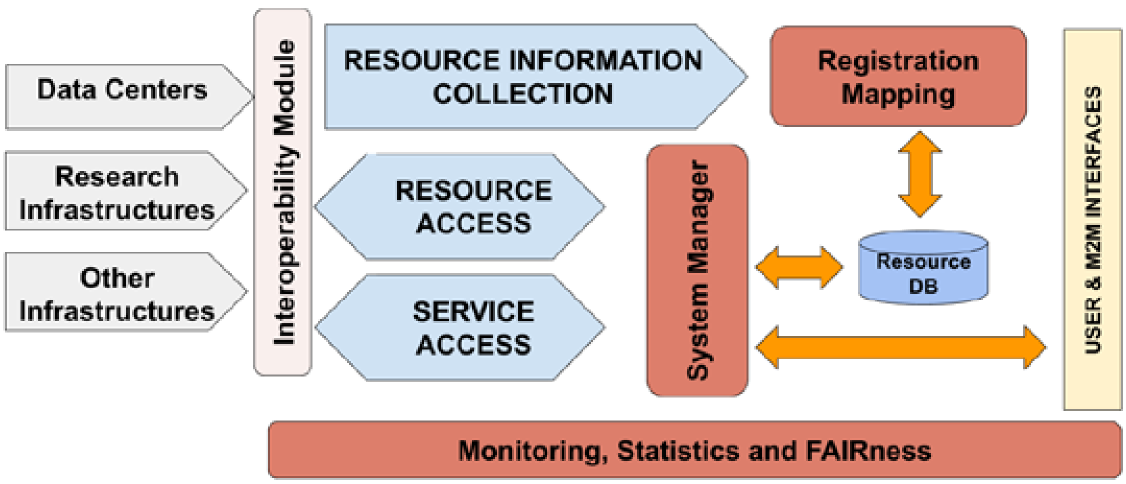


Figure 1: e-JERICO conceptual design

Conclusion

This presentation will review the capabilities and challenges in building the emerging e-JERICO. It is an interesting illustration of working in a very diverse environment supporting users from science to applications and policy, supplying data, data products, services and other resources of the JERICO community in an integrated, free, open, and organized way. Through this e-infrastructure, JERICO will facilitate the access to coastal information and enhance the capacity to manage the interaction between distributed infrastructures worldwide, in particular across Europe.

Authors Index

W

Z

Types of contributions Scientific and technical papers, and review articles are the main contributions accepted. Moreover, the journal will also publish short notes, limited to a length of four pages including references; technical notes dedicated to the presentation of current research activity or significant original data. In the case of articles of strictly regional or local interest, the contribution should be limited to the essentials.

Submission of papers The manuscript should not have been published or submitted for publication elsewhere. Submission of a multi-authored manuscript implies the consent of all participating authors. Manuscripts shoud be submitted online at this web address: www.editorialmanager.com/bgta.

Reviewing process Before publication, each manuscript will undergo a reviewing process. The Editor-in-Chief will assign the manuscript to an Associate Editor who is directly in charge of the reviewing process of the manuscript and will send the manuscript for evaluation to reviewers. The final decision on publication will be made either by the Associate Editor or by the Editor-in-Chief. Editing and styling will be done at the discretion of the publisher.

Manuscript preparation All papers must be written in English and should not normally exceed 20 printed pages, including figures and tables. The manuscript should be written in maximum consistency with the style and layout of the journal.
The title of the paper should unambiguously reflect its contents. A suggestion for an abbreviated running title should be given. The author name(s) and affiliation(s) should follow the title. Where an author has moved permanently or temporarily to a new address, this should be indicated as a footnote on the first page. An author to contact should be specified with full postal address, phone, fax, and e-mail.
Each paper must be provided with an Abstract of about 150-200 words, describing concisely the purpose and results of the paper. If the paper was presented at a meeting, mention should be made thereof in the acknowledgements.
The SI system should be used for all scientific and laboratory data: if it is necessary to quote other units, these should be added in parentheses. Abbreviations for units should follow the suggestions of the American Institute of Physics (AIP) Style Manual. Where abbreviations are likely to cause ambiguity or to not be readily understood by an international readership, units should be written in full. The full stop should not be included in abbreviations, e.g., m (not m.), ppm (not p. p. m.), "%" and "/" should be used in preference of "per cent" and "per".
Footnotes should be avoided, especially if they contain information which could equally well be included in the text.
The use of proprietary names should be avoided. Papers essentially of an advertising nature will not be accepted.

References should be cited at the appropriate point in the text with the author's name followed by the year of publication; thus, "according to Finetti (1982)" or "as shown in an earlier paper (Marussi, 1983)". When there are two or more papers by the same author published in the same year, the distinguishing letters, a, b, etc. should be added to the year. The name of two coauthors must be linked by "and"; for three or more, the first author's name must be followed by "et al.". References should be listed together at the end of the paper and arranged alphabetically without numbering. Only cited references should be included in the list. Examples of layout of references are given below.

Slejko D., Caporali A., Stirling M. and Barba S.; 2010: *Occurrence probability of moderate to large earthquakes in Italy based on new geophysical methods*. J. Seismol., **14**, 27-51, doi: 10.1007/s10950-009-9175-x.

Richter C.F.; 1978: *Elementary Seismology*. W. H. Freeman and Company, San Francisco, 768 pp.

Rikitake T.; 1975: *Dilatancy model and empirical formulas for an earthquake area*. In: Wyss M. (ed), Earthquake prediction and rock mechanics, Birkhuser Verlag, Basel und Stuttgart, pp. 141-147.

Garcia J. and Slejko D.; 2009: *Scenario earthquakes for tsunami generation in the Ionian Sea*. In: Slejko D. and Rebez A. (eds), Gruppo Nazionale di Geofisica della Terra Solida, 28° Convegno Nazionale, Riassunti estesi delle comunicazioni. Stella Arti Grafiche, Trieste, pp. 375-380.

Garcia J., Slejko D., Alvarez L., Peruzza L. and Rebez A.; 2003: *Probabilistic seismic hazard assessment for Cuba*. In: Slejko D. (ed), Gruppo Nazionale di Geofisica della Terra Solida, Atti del 21° Convegno nazionale, CD-Rom, Prospero, Trieste, file 04.20.

DISS Working Group; 2010: *Database of Individual Seismogenic Sources (DISS), Version 3.1.1: A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas*. <http://diss.rm.ingv.it/diss/>, doi: 10.6092/INGV.IT-DISS3.1.1, last access March 2015.

Montaldo V.; 2005: *Seismic hazard and uncertainties assessment in north-eastern Italy: comparing approaches with varied geological and seismological background*. PH.D. Thesis in Geological Sciences and Geo-Tecnologies for Environment and Territory, XVIII cycle, Università degli Studi di Milano - Bicocca, 146 pp.

Carulli G.B.; 2006: *Carta geologica del Friuli Venezia Giulia*. SELCA, Firenze.

Koulakov I.; 2011: *High-frequency P and S velocity anomalies in the upper mantle beneath Asia from inversion of worldwide traveltime data*. J. Geophys. Res., **116**, B04301, 22 pp., doi: 10.1029/2010JB007938.

Figures .jpg or .tiff files of each figure, with a 300 dpi resolution, should be provided upon request. Colour illustrations are accepted, and will be not charged to the authors.
Size of characters should be such that after reduction the smallest character will not be less than 1.5 mm.

Numbering and captioning Each illustration should be given a figure number in Arabic numerals. Each table should be given a table number in Arabic numerals.
Figures and tables must have a caption or title which is referred to in a list at the end of the manuscript. Captions should never be incorporated in the drawings. Scales should be indicated graphically and not numerically.
Equations should all be numbered sequentially with Arabic numbers, and refered to within the text by Eq. or Eqs. followed by the pertinent number in brackets.

Proofs The (contact) author will receive a set of proofs to be checked by typesetting editing. No new material may be inserted in the text at the time of proof reading unless accepted by the Editor-in-Chief. Proofs should be returned within a week.

Page charges and reprints There will be no page charges, including colour pictures. The .pdf file of the printed paper will be freely available at the BGTA web site. Paper reprints can be ordered on a reprint order form that will be sent with the page proofs and will be charged at current printing prices.

Copyright Publication of an article implies the transfer of the copyright from the author to the publisher, by signing the "Transfer of Copyright" agreement after manuscript is received. Permission to quote from this journal with acknowledgement of the source is granted.

Bollettino
di Geofisica
teorica ed applicata

Vol. 62
SUPPLEMENT 1

IMDIS 2021
International Conference on Marine Data and Information Systems

12-14 April, 2021
Online

Book of Abstracts

Guest Editors: Michèle Fichaut, Vanessa Tosello, Alessandra Giorgetti

imdis

Ifremer

IODE

NIOZ
Sea Research

SeaDataNet

OGS

OGS National Institute of Oceanography and Applied Geophysics

ISSN 0006-6729