Project Information	
Project full title	EuroSea: Improving and Integrating European Ocean Observing and Forecasting Systems for Sustainable use of the Oceans
Project acronym	**EuroSea**
Grant agreement number	862626
Project start date and duration	1 November 2019, 50 months
Project website	https://www.eurosea.eu

Deliverable information	
Deliverable number	D3.1
Deliverable title	**DMP**
Description	Data Management Plan
Work Package number	WP3
Work Package title	Network Integration and Improvements
Lead beneficiary	IFREMER
Lead authors	Sylvie Pouliquen
Contributors	Antonio Novellino, George Petihakis, Joaquin Tintoré
Due date	30/04/2020
Submission date	31/05/2020
Resubmission date	07/12/2021 (Executive Summary and Conclusions added)
Comments	Delayed due to COVID-19 event

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862626.

https://doi.org/10.3289/eurosea_d3.1

Table of contents

Executive Summary

The EuroSea project Data Management Plan (DMP) is building upon the existing European marine data Infrastructures (i.e. European Global Ocean Observing System – EuroGOOS, Copernicus Marine Environmental Monitoring Service - CMEMS, European Marine Observation and Data network - EMODnet) that have been developed and enhanced in the past decades. Since their establishment, the EuroGOOS and the SeaDataNet network of National Oceanographic Data Centers have widely contributed to International Oceanographic Data Exchange and more in general to GOOS policies in sharing ocean data and co-production of oceanographic services. Moreover, they have widely contributed to design and set up the operational data infrastructures of the CMEMS and EMODnet data systems, which, mimicking the international organization and management of the observations network, are often organised by observing platform (i.e.: Argo, gliders, Research vessels, fixed point observatories …).

Capitalizing on national and European projects and programmes, in situ data FAIRness (Findable, Accessible, Interoperable and RE-Usable) has significantly improved in some domains. This includes better harmonization, interoperability, and integration of data and metadata services at regional or observing networks level. However, many gaps and barriers still exist and must be addressed. To this end, the purpose of EuroSea DMP is to propose a plan to enhance the FAIRness of these network data systems to facilitate their use either directly or throughout existing operational integrated services such as CMEMS and EMODnet. These data systems have been developed and maintained for the past two decades and EuroSea proposed and deployed empowerment or service enhancement will be up-taken and continued by these networks and service providers after the end of the EuroSea project.

1 Introduction to EuroSea

Although the Ocean is a fundamental part of the global system providing a wealth of resources, there are fundamental gaps in ocean observing and forecasting systems, limiting our capacity in Europe to sustainably manage the ocean and its resources. Ocean observing is “big science” and cannot be solved by individual nations; it is necessary to ensure high-level integration for coordinated observations of the ocean that can be sustained in the long term. For Europe, EuroSea will point the way for the current and future cooperation between science and industry, politics and the public with the common goal of a sustainable blue economy and the responsible handling of the sensitive marine ecosystems. The project will make a significant contribution to not only generating, processing and linking information about our oceans, but also to make long-term and extensive use of this and the resulting knowledge in a wide variety of areas. As a link between sectors and disciplines, EuroSea faces a very big challenge.

2 Data Summary

This Data Management Plan (DMP) is produced as part of the Open Research Data Pilot that EuroSea complies with. The DMP sets the framework for the handling of data produced in EuroSea from acquisition over curation to dissemination, and shall thereby assure the implementation of best practice procedures for lifecycle management of EuroSea data during and beyond the lifetime of the project. This data management plan (DMP) describes the data that will be authored and how the data will be managed and made openly

accessible throughout the lifetime of EuroSea. In line with the TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP) v2.0 the content of the data management plan includes:

- the types of data to be managed;
- the standards that would be applied, for example format and metadata content;
- provisions for archiving and long-term preservation;
- access policies and provisions; and
- quality assurance

2.1 What is the purpose of the data collection/generation and its relation to the objectives of the project? What types and formats of data will the project generate/collect?

EuroSea has a specific approach in regard to data management as the project builds up on existing observing networks, e.g. for coastal platforms it will consider outcome from and contribute to the JERICO S3 DMP, as well as on the achievements realised within the H2020 AtlantOS project that issued a set of best practices that will be further enhanced within the EuroSea project. The consequence thereof is that WP3 task 3.10 mainly deals with harmonizing data management procedures and implementing FAIR principles with a first target to serve the Copernicus Monitoring Environment Marine Service (CMEMS) and EMODnet, as well as SeaDataNet and historical National Oceanographic Data Centres at a later stage. Therefore, we foresee that more specific information will be provided with regards to the provenance, discoverability and accessibility to data coming from individual networks that are involved in EuroSea during the course of the project in the deliverables provided by task 3.10.

2.2 Will you re-use any existing data and how? What is the origin of the data?

EuroSea builds on an integrated data system of systems developed within the EuroGOOS context in close link with Copernicus, EMODnet and the European Regional Observing System (ROOS) on one hand and the OCG (Observation Coordination Group) international ocean observing networks on another hand. Based on EuroGOOS and OCG data recommendations, this system of systems harmonizes work-flows, data processing, quality assurance procedures, and distribution across in-situ observing network systems in EuroSea WP3, WP5, WP6 and WP7 and integrates in-situ observations into existing European and international data infrastructures, in EuroSea termed data Integrators.

The primary data providers in EuroSea are existing observing Networks as described in WP3 that have been established before the start of EuroSea mostly under EuroGOOS Task Team activities:

Observation Networks

- Argo
- Underway gliders
- Research vessels and Ferrybox
- Fixed point observatories (moorings and repeat vessels)
- GLOSS (Tide gages)

- HF Radars
- Autonomous Surface Vehicles

Thematic Networks

- Augmented observatories
- Across observing scales

These data streams will be complemented by other type of data acquired within the Demonstrators WP5, WP6 and WP7.

To facilitate the harmonization across the observing networks, EuroSea relies on existing data integrators, i.e. no separate data portal or data archiving system will be built. The integrators in EuroSea are:

- For marine environmental data: The In-Situ Thematic Assembling Centre (INS TAC) component of Copernicus Marine Environment Monitoring Service (CMEMS[1]) and EMODnet (in particular Physics[2] and Chemistry[3] lots)
- for marine biodiversity data: the EMODnet biology[4]

2.3 What is the expected size of the data? To whom might it be useful ('data utility')?

The size of the data will vary from a few gigabytes for some networks to a few terabytes for others and are managed within data system already set up by the Networks. At the Integrators level it counts in Terabytes but the existing systems have been set up to manage such amount of data in an operational way.

One goal is to ensure that data from different and diverse in-situ observing networks are readily accessible and useable by the wider community, including the international ocean science community and other stakeholders in this field. To achieve this, the strategy is to move towards an integrated data system within EuroSea that harmonizes work flows, data processing and distribution across the in-situ observing network systems, and integrates in-situ observations into existing European and international data infrastructures such as Copernicus INS TAC, SeaDataNet[5] NODCs, EMODnet, etc.

The targeted integrated system deals with data management challenges that must be met to provide efficient and reliable data service to users. These include:

- Common quality control for heterogeneous and near real time data
- Standardization of mandatory metadata for efficient data exchange
- Interoperability of Network and Integrator data management systems

1 CMEMS – www.marine.copernicus.eu
2 EMODnet Physics – https://www.emodnet-physics.eu/
3 EMODnet Chemistry https://www.emodnet-chemistry.eu/
4 EMODnet Biology https://www.emodnet-biology.eu/
5 SeaDataNet - https://www.seadatanet.org/

EuroSea will foster integration of old and new observations into the existing Network data systems and make these observations free and open accessible to maximise impact and demonstrate the value of the ocean observing system to those users targeted by EuroSea in various domains such as Climate Change, Ocean Health and Operational Services. We are cooperating closely with our partner EU action "Blue Cloud" in facilitating the access and interoperability of ocean data.

3 FAIR data

During the past two decades, a series of standards for data and metadata formats as well as exchange protocols have been established within the marine community where projects and organizations like JCOMM, RDA (Research Data Alliance)[6], EuroGOOS, EMODnet and Copernicus played a significant role.

During the H2020 BG7 AtlantOS project the individual data processing and distribution steps were analysed for the networks involved in AtlantOS, to identify possible gaps and impediments in the use of standards, to build on these specifications and to facilitate the implementation within the involved data networks. This analysis was summarised in the deliverable Data Harmonization Report [1] that contained recommendations on data harmonization and data processing to facilitate the interoperability of the systems. A Data Management Handbook [2] was also produced with the elements of standardization across the Networks relying on existing European and international standards and protocols, and to describe the data exchange backbone of the AtlantOS system, providing guidelines to Networks on how to set them up. This handbook also describes ways to facilitate data discovery at the Network level and enhancements at the Integrator level for better fit-for-purpose services to users.

Within the EuroSea project, the Data Management Handbook will be revised to take into account the progress made in the past 5 years. In particular, EuroSea is concerned with improving the interoperability among Networks and Integrators, through standardization of basic data features within the EuroSea community.

Within EuroSea we will work towards ***enabling FAIR data (***Figure 1***), supporting integration of ocean data into Copernicus Marine Service, EMODnet and SeaDataNet portfolios***

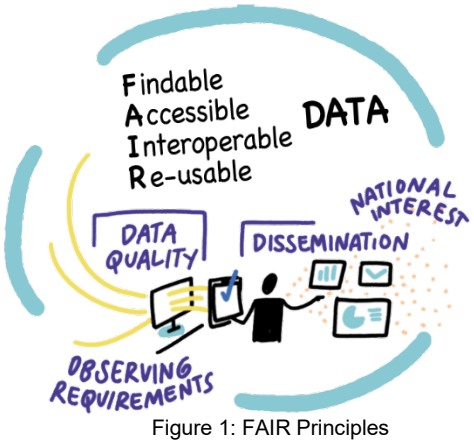

Figure 1: FAIR Principles

We will follow recommendations provided in the OceanObs19 white paper [3] and work together with Networks and Integrators in fostering the implementation of the following principles:

6 Research Data Alliance : https://www.rd-alliance.org/

- **Findable:** Each dataset should be identified by a unique persistent identifier and described by rich, standardized metadata that clearly include the persistent identifier. The metadata record should be indexed in a catalogue and carried with the data.

- **Accessible:** The dataset and its metadata record should be retrievable by using the persistent identifier and a standardized communications protocol. In turn, that protocol should allow for authentication and authorization, where necessary. All metadata records should remain accessible even when the datasets they describe are not easily accessible.

- **Interoperable:** Both metadata and datasets use formal, accessible, shared, and broadly applicable vocabularies and/or ontologies to describe themselves. They should also use vocabularies that follow FAIR principles and provide qualified references to other relevant metadata and data. Importantly, the data and metadata should be machine accessible and parsable.

- **Reusable:** To meet this principle, data must already be findable, accessible, and interoperable. Additionally, the data and metadata should be sufficiently richly described that it can be readily integrated with other data sources. Published data objects should contain enough information on their provenance to enable them to be properly cited and should meet domain-relevant community standards.

The EuroSea project consists of many mature data providers and data integrators with well-established workflows and data policies, which administer a heterogeneous pool of data resources. Consequently, a prescriptive one-model-fits-all rulebook is neither feasible nor desirable. Hence, EuroSea shall take a pragmatic approach to improve standardization among the involved Networks and Integrators, within the European and International landscape (see Figure 2) by defining an essential set of minimum requirements, and set guidelines for best practice by enhancing and further developing those provided by the AtlantOS project [1 - 2].

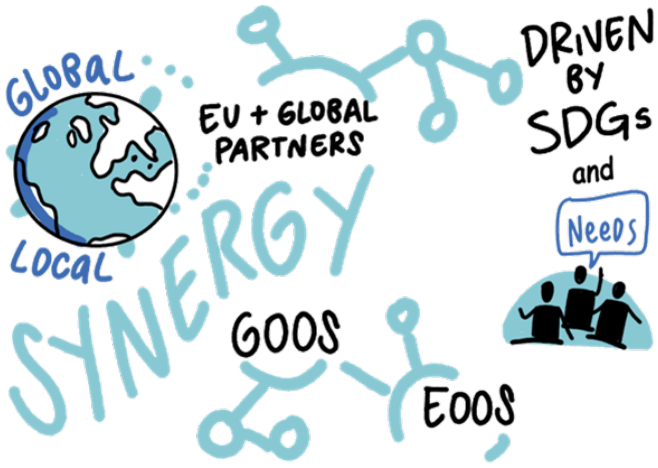

Figure 2: *International and European integration*

In the coastal areas, EuroSea data providers will coordinate their activities with ongoing EU coastal ocean initiatives such as Jerico-S3.

3.1 Making data findable, including provisions for metadata

The data produced and/or used in the project will be discoverable with metadata and identifiable.

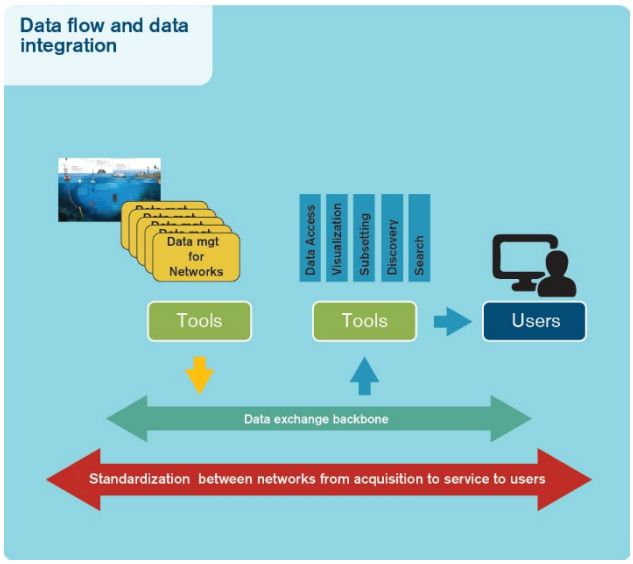

Figure 3: *Schematic representation of data flow from observing units*

Data management will be developed in coordination and collaboration with already existing infrastructures and integrators (e.g., JCOMMOPS, CMEMS, EMODnet) to avoid the duplication of effort and to facilitate a fast adoption and availability of the produced data. Once connected to these systems, each new dataset can be automatically and immediately exposed by means of all the available interoperability interfaces (the TOOLS in Figure 3) and data portals. Within WP3, there is a focused action to support an end-to-end data and information management, search, discovery and access system from Networks to targeted Integrators (CMEMS, EMODnet and SeaDataNet). One key action under the EuroSea project will be the enhancement of the data provenance to provide the user with the end-to-end data management pipeline information. Exhibiting infrastructures are already collaborating and sharing data, by providing the user with the provenience metadata he will be able to understand who and how data has been managed.

Each Network has its naming convention that have been agreed often at International level with JCOMM. EuroSea will not change the Network data system but will provide recommendations, based on the work done internationally (RDA[7] , CF [8]) as well as at European scale by EuroGOOS DATAMEQ[9] working group and previous AtlantOS project. However, mandatory information to be provided with the data at network level, data exchange and access will be enhanced with Integrators and for users.

Existing FAIR vocabularies will be enhanced on the basis of existing metadata and naming conventions through JCOMMOPS, CMEMS, EMODnet and SeaDataNet systems (among others).

In term of metadata, EuroSea will follow the recommendations that have been set up by the EuroGOOS DATAMEQ, developed under CMEMS, EMODnet Physics and SeaDataNet/SeaDataCloud and further extended at international level under the AtlantOS project.

The applicable recommendations relying on existing international standards that will ensure cross platform coherence and also facilitate data discovery for users and data integration:

- **Platforms** should have a unique identifier that will be either WMO for most platforms or ICES code for ships.

7 Research Data Alliance https://www.rd-alliance.org/
8 CF : http://cfconventions.org/
9 EuroGOOS DATAMEQ : http://eurogoos.eu/data-management-exchange-quality-working-group-data-meq/

- **Metadata** used by the networks for parameters should be “mappable” on standard vocabularies existing and EU (SeaDataNet vocabularies) or international (CF or WoRMS for Taxa). More specifically, metadata are based on P01-P09 (parameter), P07 (CF variable), P06 (units) L06 (platform types) from SeaDataNet controlled vocabularies managed by NERC/BODC (Vocabulary Server version 2.0). Within the AtlantOS project An AtlantOS Essential Variables list of terms (aggregated level), related to ECV –EOV or other, has been defined and was published in June 2016 on the NERC/BODC Vocabulary Server (version 2.0) as A05 vocabulary (https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/A05/). Moreover, dedicated mandatory fields will be proposed: data provenience to describe the different centres and integrators.
- **Institutions** used in a data file shall be identified by a unique code from the EDMO (European Directory of Marine Organizations) existing catalogue. EDMO shall be enhanced with the Network’ s needs.
- **QC information** will be attached to the data; both Quality flags that can be mapped to SeaDataNet flag scale (available in the SeaDataNet Common Vocabularies[10] as list L201), and, whenever known, processing level information (“qualified in RT using automated procedures” or “processed in delayed DM by Scientist”). Extension of QC information to introduce a “sensors setting” field will be studied with Networks and Jerico-S3. Such metadata field would allow to provide link to information about the sensor (type, accuracy, calibration procedures), last calibration, calibration best practices using SensorML.

3.2 Making data openly accessible

Observing nodes under the umbrella of EuroSea will follow the principle of free and open access to data produced by their facilities and feel committed to work towards the realization of this principle. For the Networks under OCG the data distribution is performed through well identified data centres most of the time certified within IODE [11]and JCOMM[12]. EuroSea aims to identify unnecessary or obsolete barriers towards open access to their data and will continuously work towards the stepwise elimination of these obstacles. The use of CC-BY[13] licence will be encouraged 

Access to metadata of each observing networks contributing to EuroSea should be free and open access without any restrictions. Observing networks data centre holding facilities shall provide and enable ease data discovery for fitness-for-use evaluation of the host data. Data published by observing networks contributing to EuroSea shall be made available free of charge though the data access tools developed by the Networks. Machine to Machine interface will be encouraged. Specific regulations deviating from this general rule may apply when data is used for e.g. non-scientific or commercial purposes. Separate fees may apply for the reproduction and delivery of data when web-based transfer of data is not possible to cover reproduction costs.

10 Seadatanet Vocab http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp
11 IODE : https://www.iode.org/
12 JCOMM: https://www.jcomm.info/
13 Creative Commons "CC BY ": https://creativecommons.org/

In general, data shall be made available by observing networks contributing to EuroSea as soon as possible and without undue delay. Each observing networks contributing to EuroSea shall nominate ‘data stewards’ to facilitate such data requests. This 'data stewards' are participating in WP3 work package.

Observing networks contributing to EuroSea acknowledge national and international intellectual property rights regulations. Each data infrastructure is responsible for the warranty of copyrights and intellectual property rights which may apply for its data holdings. Observing networks contributing to EuroSea will clearly indicate licenses and terms of use for each dataset in the corresponding metadata.

If data or information produced by observing networks contributing to EuroSea is used in published or unpublished work, attribution for the used resources is required. Data citations shall exclusively use the information provided within the metadata of each data set. To this aim, EuroSea DMP will explore the adoption and adaption of available controlled vocabularies, e.g. SDN:EDMERP[14] for proper program citation and acknowledgement.

Observing Networks and Integrators are in charge to provide the necessary documentation and tools to facilitate data access to users. Using open source code to facilitate user uptake is already an on-going activity at Network and Integrator level that will be further encourage within EuroSea.

14 https://www.seadatanet.org/Standards/Metadata-formats/EDMERP

3.3 Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

See previous sections §3.1and §3.2

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

See section §3.2

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

See section §3.2

3.4 Increase data re-use (through clarifying licences)

Data and products from observing networks contributing to EuroSea will be integrated in existing data integrators facilitating their re-use and re-distribution. As the data will be distributed with a minimum set of metadata that will trace their originators, feedback on their use through the integrators will be feasible. Re-use of non-open and free data will not be facilitated by EuroSea as these data won't be integrated in the EuroSea identified Integrators (CMEMS, EMODnet, SeaDataNet), as well as their linked network of initiatives and sources (e.g. ICES[15], PANGAEA[16], etc). For those restricted access data, only discovery will be facilitated by integration in European and international catalogues.

15 https://www.ices.dk/

16 https://pangaea.de

4 Allocation of resources

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

The cost of making data FAIR is managed and supported within the Networks and the Integrators involved in the project and EuroSea only covers the networking cost that allows to define common best practices shared and agreed among all the partners.

Who will be responsible for data management in your project?

The data management responsibility stays at the level of the Networks and Data Systems but within task3.10, data managers from networks and integrators will work together towards the implementation of the FAIR principles. This DMP is the continuity of similar activities carried on within the AtlantOS project and within EuroGOOS DATAMEQ WG in partnership with JCOMM networks and Demonstration WPs.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

As EuroSea doesn't build any data system but relies on Data Systems that are managed either by the Networks or the Integrators as defined in section §3.4, the resources for long-term preservation of the data are covered by those entities.

5 Data security

As EuroSea doesn't build any data system but relies on Data Systems that are managed either by the Networks or the Integrators as defined in section §3.4 all the issues related to Data Security are managed at this level. These Networks and Integrators are managed since decade(s) by professional data centres that apply standards in term of data security.

6 Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

No

Projects participating to the ORDP might present information relevant to the ethical aspects (data protection) in the DMP. In such a case, the ethics chapter of the DoA may simply refer to the DMP for more information on the details of the ethics aspects related to data.

Nothing to report

Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

Whenever the EuroSea project will implement surveys, questionnaires, or collect personal data for any reason (e.g., attendance to organized events), European GRDP law will be used as reference and the user will be informed about the use of personal data. In general, EuroSea will not transfer personal data (e.g., email addresses) to other entities and the only use will be setting up a distribution list to inform users about project progress. User will be always able to change his consensus and ask for being removed from the distribution channel.

7 Other issues

Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

Nothing to report

8 Further support in developing your DMP

Nothing to report

9 Conclusion

It is essential to improve the quantity, quality, accessibility and interoperability of marine information for decision making and enable new opportunities in the maritime sectors, for the benefit of citizens. These drivers and goals set the requirements for improved ocean data management.

Within the European framework, these needs are addressed by the inter-European-projects, inter-EuroGOOS-institutes working group titled the Data Management, Exchange and Quality (DATAMEQ). This team works to: (i) conceptually guide EuroGOOS in situ data management considering the existing systems; (ii) propose solutions to unlock data for operational and research purposes; and (iii) deliver standards for data matching principles of Findability, Accessibility, Interoperability and Reusability (FAIR).

One key goal of the EuroSea DMP is to enhance the FAIRness of the data managed within the EuroSea project that can be achieved by a tight cooperation with the European actors (in particular the EUROGOOS partners) in collaboration with the DATAMEQ team. Notably, these European actors operate a significant part of the European observing systems and serve both CMEMS and EMODnet.

10 References

[1]. Koop-Jakobsen K, Waldmann C, Huber R , Harscoat Valerie, Pouliquen Sylvie (2016). Data Harmonization Report. AtlantOS project. D7.1. https://doi.org/10.13155/47090

[2]. Harscoat Valerie, Pouliquen Sylvie, AtlantOS Wp7 Partners (representatives Of Networks And Integrators) (2016). Data Management Handbook AtlantOS. D7.4. https://doi.org/10.13155/48139

[3]. Tanhua T, Pouliquen S, Hausman J, O'Brien K, Bricher P, de Bruin T, Buck JJH, Burger EF, Carval T, Casey KS, Diggs S, Giorgetti A, Glaves H, Harscoat V, Kinkade D, Muelbert JH, Novellino A, Pfeil B, Pulsifer PL, Van de Putte A, Robinson E, Schaap D, Smirnov A, Smith N, Snowden D, Spears T, Stall S, Tacoma M, Thijsse P, Tronstad S, Vandenberghe T, Wengren M, Wyborn L and Zhao Z (2019) Ocean FAIR Data Services. Front. Mar. Sci. 6:440. doi: https://doi.org/10.3389/fmars.2019.00440.