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	D3.7
Deliverable title	Network harmonization recommendations
Description	Harmonisation data management procedures and implementing FAIR principles with the target to serve the data infrastructures: Copernicus Marine Service and EMODnet (first stage) as well as SeaDataNet and historical National Oceanographic Data Centres (later stage)
Work Package number	WP3
Work Package title	Network integration and Improvement
Lead beneficiary	IFREMER
Lead authors	Sylvie Pouliquen, Dominique Obaton
Contributors	Antonio Novellino, Alessandra Giorgetti, Abed El Rahman Hassoun and the network experts (displayed within the text)
Due date	31.08.2022
Submission date	02.12.2022
Comments	

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_d3.7

Table of contents

Executive summary 1

1. Glossary 1

2. Introduction 2

3. Context 3

3.1. Data management landscape 3

3.2. Data management actors 4

3.3. Aim of the document 4

4. Proposition for data harmonisation 5

4.1. Harmonisation of data information 6

Identification 6

Data common vocabularies 6

Citation and traceability 7

4.2. Harmonisation of actor information 8

Identification 8

Citation 8

5. Proposition for access harmonisation priorities 9

5.1. Data format 9

5.2. Technical access 9

5.3. Licence to use data 9

6. Networks current status 10

6.1. Argo 10

Recommended basic information 10

Additional information on network 11

6.2. Gliders 12

Recommended basic information 12

Additional information on network 13

6.3. Vessels 14

Recommended basic information 14

Additional information on network 15

6.4. Eulerian observations (fixed platforms) 15

Recommended basic information 15

6.5. Tide gauges 17

Recommended basic information 17
6.6. HR radar platforms 18
Recommended basic information 18
Additional information on network 20
6.7. Autonomous surface vehicles 20
Recommended basic information 20
6.8. Augmented observatories 21
Recommended basic information 21
Additional information on network 23
Conclusion 23
References 24

Executive summary

This document proposes recommendations on metadata and information to be associated with marine data from ocean observation networks. The objective is to reach a common basis of metadata and information for any in situ networks that will ease the interoperability and their integration in the various European data integrators such as Copernicus marine, EMODnet or SeaDataNet while being in line with what has been done at international level. Proposed recommendations are built from previous work through projects, different initiatives and thanks to EuroGOOS and the European Ocean Observing System (EOOS). They are, and will need to be, complemented by additional metadata and information specific to the network considered. Each recommendation is associated with a criterion based on the FAIR principles as proposed by the international collective FORCE11. The output table obtained from these proposed basic recommendations is then filled by the different EuroSea in situ networks which highlight similarities and differences and the maturity of the networks. It gives a good overview of the existing metadata and information used by the observation networks for further discussions and improvements.

1. Glossary

Acronym	Meaning
ASV	Autonomous Surface Vehicles
CC	Creative Commons (data license)
CF	Climate and Forecast (metadata convention)
DAC	Data Assembly Centre
DATAMEQ	Data Management, Exchange and Quality working group (of EuroGOOS)
DOI	Digital Object Identifier
EBI	European Bioinformatics Institute
EDIOS	European Directory of the initial Ocean-observing Systems
EDMERP	European Directory of Marine Environmental Research Project
EDMO	European Directory of Marine Organisation
EGO	European Gliders Observatories
ENA	European Nucleotide Archive
EOOS	European Ocean Observing System
EOSC	European Ocean Science Cloud
EOV	Essential Ocean Variables
FAIR	FAIR principles. FAIR for Findable, Accessible, Interoperable, Reusable
GDAC	Global Data Assembly Centre
GLOSS	Global Sea Level Observing System
GOOS	Global Ocean Observing System
GO-SHIP	Global Ocean -SHIP
GTS	Global Telecommunication System
GUI	Graphical User Interface
HFR, HF Radar	High Frequency Radar
ICES	International Council for the Exploitation of the Sea
ID	identifier
IEEE	Institute of Electrical and Electronics Engineers

IMO	International Maritime Organisation
IOC	Intergovernmental Oceanographic Commission (of Unesco)
IODE	International Oceanographic Data Exchange (of Unesco)
ISO	International Organisation for Standardisation
IT system/development	Information Technology system/development
JCOMM	Joint WMO/IOC Technical Commission for Oceanography and Marine Meteorology
JRC	Joint Research Center (EU Commission's science and knowledge service)
NVS	Nerc Vocabulary Server
OBPS	Ocean Best Practices System
ORCID	Open Researcher and Contributor ID
PI	Principal Investigator
PSMSL	Permanent Service for Mean Sea Level
QC	Quality Control
RDA	Research data Alliance
SDN	SeaDataNet
SHIPC	SeaDataNet Ship and Platform Code
SOOP	Ship Of Opportunity Programme
SOT	Ship Observations Team
URN	Uniform Resource Name
UTC	Coordinated Universal Time
VOS	Voluntary Observing Ship
WGS	World Geodetic System
WIGOS	WMO Integrated Global Observing System
WMO	World Meteorological Organisation
WoRMS	World Register of Marine Species

2. Introduction

The ocean observing system needs to be ensured by high-level integration and coordination to guarantee its long-term sustainability, efficient accessibility and usability by a wide range of users. Enormous advancements and efforts toward these objectives have been already reached in Europe, partly through the activities of the IOC-UNESCO's International Oceanographic Data and Information Exchange (IODE) and EuroGOOS DATAMEQ working group, although there is still room for additional progress and gaps to be addressed. 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, organisations and data integrators like JCOMM[1], RDA (Research Data Alliance)[2], EuroGOOS[3], EMODnet[4], SeaDataNet[5] and Copernicus[6] played a significant role.

[1] JCOMM: https://www.goosocean.org/index.php?option=com_oe&task=viewGroupRecord&groupID=78
[2] Research Data Alliance: https://www.rd-alliance.org/
[3] EuroGOOS: https://eurogoos.eu/
[4] EMODnet: https://emodnet.ec.europa.eu/en/data-portals-overview
[5] SeaDataNet: https://www.seadatanet.org/
[6] Copernicus: https://www.copernicus.eu/en/copernicus-services

EuroSea project is paving the way towards improving current and future co-operations between science, industry, politics and the public, aiming to enforce and widespread a sustainable blue economy vision and the responsible management and protection of the vulnerable marine ecosystems. The project is making a significant contribution in generating, processing and linking information about our ocean through providing long-term and extensive use of the resulting knowledge in a wide variety of areas.

Taking into consideration that harmonised data are a key element in maintaining a usable and interoperable ocean observing system, this deliverable, built on previous assessment, aims to provide recommendations for the harmonisation of the marine in situ networks involved in EuroSea. It will be a useful product for the European data integrators, particularly EMODnet, SeaDataNet and Copernicus Marine service. Moreover, based on it, a technical paper has been accepted to the IEEE MetroSea 2022 conference and its associated publication. This larger diffusion of this deliverable will allow hopefully feedback from the different data integrators and improving the proposed harmonised recommendations.

3. Context

3.1. Data management landscape

Harmonisation of marine in situ data is necessary for scientists of different disciplines and experts in various fields to share information and tackle together a specific marine phenomenon or threat. Tanhua et al. (2019) gave an exhaustive review of in situ data management highlighting the FAIR principles that Wilkinson et al. (2016) proposed as a guideline to data management, to conclude with some recommendations. Recently, Révelard et al. (2022) showed the importance of trans-disciplinarity in marine sciences to, among others, work and collaborate more efficiently and to be more aligned with societal needs. These different important and relevant studies provide an overview of the marine data landscape and some useful guidance.

The marine data landscape has been well reviewed in several studies that provided recommendations for in situ networks and/or research infrastructures in the context of different projects and initiatives with the same objective to improve data management for producers, operators and the IT systems. For example, the AtlantOS project has already made several recommendations to move towards an integrated EU data system (Koop-Jakobsen et al., 2016; Harscoat and Pouliquen, 2016). It dealt with data management challenges and included propositions of common quality control procedures for heterogeneous and near-real-time data, standardisation of mandatory metadata for efficient data exchange and interoperability of networks and integrators data management systems. Afterwards, the ENVRI-FAIR project has enhanced the recommendations related to data harmonisation for and between networks and integrators, introducing the FAIR principles and the connection to the EOSC- European Ocean Science Cloud, for a large set of environmental data including the marine ones (Thijsse and Schaap, 2019). The outcomes will allow establishing technical preconditions for the implementation of an ENVRI-hub by 2023. In addition, the COPiLOtE[7] project already applied the FAIR principles, detailed in several characteristics and indicators, to the French initiative ODATIS (Ocean Data Information and Services)[8] ocean cluster (Quimbert et al., 2022), giving a good illustration on how to translate the FAIR principles to recommendations and *vice versa*.

[7] https://www.odatis-ocean.fr/en/activites/projets-en-cours/projets-nationaux/copilote-certification-pole-ocean-certification-ocean-data-cluster#COPiLOte

[8] https://www.odatis-ocean.fr/en/

3.2. Data management actors

Institutions are generally in charge of the management of the observing system they operate and the first level of data processing for their own applications. Therefore, each institution collects, controls and distributes data according to its own rules.

Operational oceanography involves major investments in infrastructures, including observing systems and high-performance computing hardware, as well as human resources with appropriate training. Such investments are difficult to be made by a single country, thus active national and international cooperation is key for the development of a synchronised Global Ocean Observation System (GOOS)[9] supported by its European part EOOS. This is particularly relevant for ocean systems at regional and global scale. Since 1994, in order to tackle this critical issue, EuroGOOS is coordinating the development and operation of five regional European operational systems: the Arctic (Arctic ROOS), the Baltic (BOOS), the North West Shelf (NOOS), the Ireland-Biscay-Iberia area (IBI-ROOS) and the Mediterranean (MONGOOS). Later, in 2014, EuroGOOS established Task Teams for coordinating networks of ocean observing systems. Task Team members collaborate in the areas of shared priorities, exchange best practices and feed data to the integrators EMODnet and Copernicus Marine Service. EuroGOOS is also part of the GOOS Regional Alliance and a partner of OceanOPS[10]. These assemblies are key structures to discuss and promote active collaborations at different levels, in order to maximise the efficiency of national resources and investments in operational oceanography.

Ocean data management and exchange processes supported by EuroGOOS and OceanOPS are indeed intended to reduce duplication of efforts among agencies, to improve data quality and reduce costs related to geographic information, thus making oceanographic data more accessible and helping to establish key partnerships to increase data availability.

To avoid duplication and heterogeneity, a common data management approach must be adopted by all actors and organisations involved in data acquisition and management. Recommendations and best practices should evolve in an 'agile' manner to follow the progress of research and to handle new platforms /sensors/variables adopted by the marine community.

3.3. Aim of the document

This document takes into account the progress made in the past 5 years and delivers reviewed and updated recommendations for an efficient harmonised data management. It takes a pragmatic approach to improve standardisation among the involved networks and integrators to reach some common and basic data features. In particular, EuroSea is concerned with enhancing the interoperability among the networks involved in the project to support integration of marine data, particularly into Copernicus Marine Service, EMODnet and SeaDataNet data portals.

In the following sections, recommendations are provided in the framework of the FAIR principles, which have as a main objective to facilitate the data management for both the operators and the IT systems. In section 4 we focus on metadata and data while in section 5 we concentrate on data accessibility. Proposed recommendations are made according to the FAIR principles broken down into the 15 characteristics laid

[9] https://www.goosocean.org/
[10] https://www.ocean-ops.org/board

down by the FORCE11[11] collective and clearly described in Quimbert et al. (2022). This transdisciplinary bottom-up approach follows Révelard et al. (2022) top-down data management recommendations.

All recommendations proposed in the following sections will be modified and updated following the networks and integrators' feedback in order to reach an agreed-upon document that will be shared through Github or equivalent. In addition to this common basis, each network may have some specific and relevant information according to the data they manage.

4. Proposition for data harmonisation

Keeping in mind that EuroSea deals with various and different in situ networks (ARGO, Gliders, Vessels, Eulerian Observation network, Tide gauge, HF Radar, Autonomous surface vehicles, Augmented observatories) from which the output data can and should be available through the Copernicus Marine Service and EMODnet data integrators, we have defined basic common elements (metadata, information) to associate to the data and to the actors in order to ease the data management. This step, in line with OceanOPS work, is an enhancement from what has been proposed by Harscoat & Pouliquen (2016) for in situ EOVs (temperature, salinity, current, sea level, oxygen, chlorophyll, nitrate and carbon) and ideally for any in situ marine data. This proposition will need to be readjusted based on the feedback from the in situ networks and should continuously be improved with time (this will be reported in the next deliverable due D3.13 data handbook).

The following proposed recommendations should define and characterise the measured data to ensure consistency and understanding between the different in situ oceanographic data networks and support data management. These recommendations are provided at the European level taking into account the various networks and relying on existing international standards. They are divided in 2 parts namely the harmonisation of data information and the harmonisation of the actors' information which themselves are divided into several sub-categories. The most relevant characteristics of the FAIR principles (proposed by the FORCE11 community) are associated to each recommendation with their meanings, as briefly recalled in the table below.

Table 1. FAIR principles proposed by the FORCE11 community

Findable	
F1	(meta)data are assigned a globally unique and eternally persistent identifier.
F2	data are described with rich metadata.
F3	(meta)data are registered or indexed in a searchable resource.
F4	metadata specify the data identifier.
Accessible	
A1	(meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1	the protocol is open, free, and universally implementable.
A1.2	the protocol allows for an authentication and authorization procedure, where necessary.
A2	metadata are accessible, even when the data are no longer available.

[11] https://force11.org/info/the-fair-data-principles/
FORCE11 is a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing.

Interoperable	
I1	(meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2	(meta)data use vocabularies that follow FAIR principles.
I3	(meta)data include qualified references to other (meta)data.
Re-usable	
R1	(meta)data have a plurality of accurate and relevant attributes.
R1.1	(meta)data are released with a clear and accessible data usage license.
R1.2	(meta)data are associated with their provenance.
R1.3	(meta)data meet domain-relevant community standards.

4.1. Harmonisation of data information

Identification

Recommendation: Platform identification

Each platform/station should be identified by a unique ID, which is:

A WMO code for:

- Drifting buoys
- Fixed ocean observing platforms
- Autonomous vehicles (e.g. gliders, saildrones)
- Profiling floats (e.g. Argo)
- Marine mammals

Refer to the OceanOPS website https://www.ocean-ops.org/ for identification or request of a WMO code

An ICES code for:

- Vessels (SHIPC)

Refer to https://vocab.ices.dk/?ref=315 for the list of ships and associated codes; it is also from the ICES site that you can ask for additional code.

Coastal platforms such as HR radar or tide gauges do not have a reference database yet. A platform unique ID could be proposed by the platform network to an international organisation (such as WMO / OceanOPS) to become sustainably identifiable. EDIOS series ID has been proposed for HR radar and could be a good alternative.

Link to FAIR: Interoperability (platform identification)

I1: (meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2: (meta) data use vocabularies that follow FAIR principles

Data common vocabularies

The key metadata that should be associated with the data allow defining the measure, the variable and providing information on the time and on the geographical position.

Recommendation: Variable.

Basic metadata to be associated with the variables are: the measuring device (instrument type) used, the precise definition of the variable, its standard name, the unit used and the quality flag associated. This information is provided by the Nerc Vocabulary Server (NVS) via the SeaDataNet (SDN) web interface[12] except for the standard name that is on the CF metadata convention website[13]

- Instrument type, refer to SDN L22 (i.e. NETTZZZZ or TOOLZZZZ) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L22
- Definition of variable, refer to SDN P01 & subset (i.e. SDN:P01::VVVVZZXX) https://vocab.seadatanet.org/bandit/browse_step.php
- Standard name, follow the CF convention https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html
- Unit of the variable, refer to SDN P06 (i.e. ZZZZ, 4 uppercase letters) https://vocab.nerc.ac.uk/collection/P06/current/
- Quality flag, refer to SDN L20 (i.e. number between 0 and 9 or letter : A, B, Q) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L20

Recommendation: Time

The time associated to the data should be written following the ISO 8601 format i.e.:

- Date is expressed as YYYY-MM-DD
- Time is in 24-hour mode and UTC, e.g. T13:05:15Z meaning 13 hours 5 minutes 15 seconds UTC (representing by Z)

Recommendation: Geographical position (latitude and longitude coordinates)

The reference coordinate system to be used to characterise the data is the WGS84, the standard for GPS.

The above proposition for common data vocabularies apply to physical and chemical variables. For the biological variables, taxonomy used refers to the World Register of Marine Species (WoRMS[14]). Classification and other common vocabularies to describe the data should be added.

Link to FAIR: Interoperability and findability (information –metadata- attached to data)

I1: (meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2: (meta) data use vocabularies that follow FAIR principles
F2: data are described with rich metadata

Citation and traceability

Recommendation: Dataset

The datasets should be identified by a DOI, persistent identifier for object and ISO standard. The two main reference DOI publishers in Europe are ZENODO[15] for any research fields (and including data, papers, software …) and SEANOE[16] for marine research data.

[12] https://vocab.seadatanet.org/search
[13] https://cfconventions.org/Data/cf-standard-names/current/build/cf-standard-name-table.html
[14] https://www.marinespecies.org/aphia.php?p=search
[15] https://help.zenodo.org/
[16] https://www.seanoe.org/html/doi-complementarity-with-databases.htm

The granularity of the dataset, to which a DOI should be assigned, is not homogeneous and not yet consolidated. For example, a DOI can be attached to a platform, a project or a network. An important point is that the different DOIs that refer to a same dataset, either this only dataset or this one among others, should be linked together (e.g. a DOI assigned to a platform and a DOI assigned to the network of all the platforms) to allow traceability.

Link to FAIR: Findability (dataset citation)

F1: (meta) data are assigned a globally unique and eternally persistent identifier.

4.2. Harmonisation of actor information

Identification

Recommendation: Institution identification

The institution responsible (operating) for the marine in situ data should be displayed. This should be done through an EDMO code that references marine institutions all over the world. The information and any organisation code can be found on SeaDataNet website[17].

Link to FAIR: Reusability (institution identification)

R1: (meta) data have a plurality of accurate and relevant attributes.
R1.2: (meta) data are associated with their provenance

Citation

Recommendation: Person

Actors associated with the data should be referenced by a persistent digital identifier, as for example an ORCID code[18].

Recommendation: Project

When data is acquired during a project, this last one needs to be associated with the data and identified by its EDMERP code. The code (5 digits) of a project can be found or obtained for a new project on SeaDataNet website[19].

Link to FAIR: Reusability (person and project citation)

R1: (meta) data have a plurality of accurate and relevant attributes.
R1.2: (meta) data are associated with their provenance

[17] https://edmo.seadatanet.org/results
[18] Information and registration: https://orcid.org/
[19] https://www.seadatanet.org/Metadata/EDMERP-Projects

5. Proposition for access harmonisation priorities

This section follows the section 4 format and deals with data accessibility.

5.1. Data format

Recommendation: NetCDF CF

The netCDF CF (v1.6 or greater) file format should be preferred as it is commonly used by the marine community and by the data integrators for in situ data as well as for satellite and modelling ones. It is a self-describing format, which eases the understanding of the file content.

Link to FAIR: Interoperability and reusability (data format)

I1: (meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation
R1.3: (meta) data meet domain-relevant community standards

5.2. Technical access

Recommendation: Data warehouse

The network dataset should be located and available in a sustainable data warehouse such as Data Assembly Centres (national, regional or global) or equivalent. The data warehouse should provide standard communication protocols (that allows communication between machines).

Recommendation: Access protocol

The metadata and data must be retrievable via standardised communication protocols

- ftp → for direct download of data
- https → by implementing an ERDDAP server that allows access to discrete data (as in situ ones)

ERDDAP[20] is an Apache based data server that offers an easy and consistent way to download subsets of gridded and discrete datasets in common file formats and make graphs and maps. The peculiarity of ERDDAP is that it unifies the different types of data servers so you have a consistent way to get the data you want, in the format you want. In particular, ERDDAP reformats the request into the format required (.html table, ESRI .asc and .csv, Google Earth .kml, OPeNDAP binary, .mat, .nc, ODV .txt, .csv, .tsv, .json, and .xhtml) by the remote server, sends the request to the remote server, gets the data, reformats the data into the format that you requested and sends the data to the requester. ERDDAP is a useful tool, easy to use, understand and maintain with some limitations in terms of responsiveness, platform metadata, discovery, sustainability and aggregation of thousands of large files.

Link to FAIR: Findability and accessibility (technical access)

F3: (meta) data are registered or indexed in a searchable resource
A1: (meta) data are retrievable by their identifier using a standardised communications protocol

5.3. Licence to use data

Recommendation: Licence.

The data should always reference the creator and be widely (as much as possible) open to users. In addition, the licence format should be understandable for a person and a machine. Therefore, the characteristics of the licence should be:

[20] https://coastwatch.pfeg.noaa.gov/erddap/index.html

- When possible, to give open and free access to the data. Note that this access can be done through authorisation or authentication if needed.
- To provide to the data actor (creator) a standardised way to grant permission to use his/her work done under copyright law.
- To be clear and accessible to the user or data actor and readable by a machine

The licence "Creative Commons" (CC[21]) gathers these characteristics. It lists 6 different licence types from most to least permissive with the common point that credits must be given to the creator. The most permissive: CC-BY (with the only limitation that credit must be given to creator) should be preferred.

Link to FAIR: Accessibility and reusability (licence)

A1.1: the protocol is open, free, and universally implementable
A1.2: the protocol allows for an authentication and authorization procedure, where necessary
R1.1: (meta) data are released with a clear and accessible data usage license

6. Networks current status

The EuroSea networks are listed below to provide their status in regards to the harmonisation recommendations proposed, as well as remarks and additional information.

6.1. Argo

Author: Thierry Carval. Thierry.carval@ifremer.fr

Recommended basic information

Variables: Temperature and salinity

Table 2. ARGO network status

ARGO NETWORK **recommendations**	**YES** **value/code** **comment**	**NO** **what is done? Expected? Comment** **OTHER INFORMATION**
Platform identification	WMO platform code for each float	
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: no SDN P01: no CF standard name: Yes SDN P06: no SDN L20: no	NVS R25 (Argo sensor type) NVS R03 (Argo parameter code) linked to SDN P01 - NVS R03 (Argo parameter code) linked to P06 NVS RD2 (delayed mode measurement flag scale) and NVS RR2 (real time measurement flag scale)

[21] Refer to https://creativecommons.org/about/cclicenses/ for all detailed information on CC licences.

ARGO NETWORK **recommendations**	**YES** **value/code** **comment**	**NO** **what is done? Expected? Comment** **OTHER INFORMATION**
Metadata associated to variable (time)	ISO 8601: no	days since 1950-01-01 00:00:00 UTC
Metadata associated to variable (geographical position)	WGS84: yes	
Dataset citation (doi)	Argo GDAC dataset https://doi.org/10.17882/42182	
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: yes	
Project citation	EDMERP: no	work underway to manage a list of projects linked to EDMERP
Data format (netCDF CF)	yes	
Data warehouse	Argo GDAC	One global dataset mirrored on two physical GDACs
Access protocol (https, ftp)	Data by profiles https://data-argo.ifremer.fr Whole dataset (one click download) https://doi.org/10.17882/42182 Data subsetting GUI https://data selection.euro-argo.eu Metadata GUI https://fleetmonitoring.euro-argo.eu ERDDAP API https://www. ifremer.fr/erddap/index.html	
Licence	Creative Commons: CC-BY (open and free access)	

Additional information on network

Near-real time data. Every effort made to deliver Argo data with the shortest delays possible after acquisition and with extensive quality control is in line with the "near-real time data" concept. Hence, there are real-time (less than 24 hours) data and delayed mode delivery systems. The quality control procedures are highest and most stringent for the delayed-mode data stream, which is designed to deliver data for climate quality.

Additional FAIR data services for open science cloud. These services (complementary to the data subsetting GUI and metadata GUI listed in the above table) are based on Argo bigdata (a continuum of NetCDF – Elasticsearch – parquet - spark – Cassandra resources).

- Search engine: OpenSearch API https://opensearch.ifremer.fr. This search engine on Argo data is used among others by EU EOSC Blue-Cloud data portal, Copernicus Eumetsat for Argo-satellite matchup.

- Argo vocabulary server: http://www.argodatamgt.org/Documentation/Argo-vocabulary-server
- Argo ontology**:** http://www.argodatamgt.org/Documentation/Argo-vocabulary-server/ Argo-linked-data-and-SPARQL-endpoint
- Semantic request: Argo SPARQL endpoint https://co.ifremer.fr/co//argo-linked-data/html/ Argo-HTML-SPARQL/ (makes use of Argo ontology)
- Library: https://github.com/euroargodev/argopy. This is a python library dedicated to Argo data access, visualisation and manipulation for regular users as well as Argo experts and operators.

6.2. Gliders

Authors: Victor Turpin (vturpin@ocean-ops.org), Pierre Testor (pierre.testor@locean.ipsl.fr), Soeren Thomsen (soeren.thomsen@locean.ipsl.fr)

Recommended basic information

Variables: Temperature, Salinity, Chla, O_2

Table 3. GLIDER network status

Gliders NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment OTHER INFORMATION**
Platform identification	WMO for each glider	
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: yes	
	SDN P01: yes	
	CF standard name: yes	
	SDN P06: yes	
	SDN L20: yes but*	* QC is not mandatory for gliders' data at the European level. However, it is well documented: http://doi.org/10.13155/51485
Metadata associated to variable (time)	ISO 8601: no	Expected
Metadata associated to variable (geographical position)	WGS84: yes	
Dataset citation (doi)	No	DOI allocation procedures exist. But it is not automatically allocated to each dataset (glider deployment).

Gliders NETWORK **recommendations**	**YES** **value/code** **comment**	**NO** **what is done? Expected? Comment** **OTHER INFORMATION**
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: yes	
Project citation	EDMERP: no	Will be considered
Data format (netCDF CF)	yes	
Data warehouse	Coriolis GDAC	
Access protocol (https, ftp)	Data by profiles https://co.ifremer.fr/co//ego/ego/v2/ Whole dataset https://erddap.ifremer.fr/erddap/tabledap/OceanGlidersGDACTrajectories.html Data subsetting GUI https://dataselection.coriolis.eu.org/ ERDDAP API https://erddap.ifremer.fr/erddap/tabledap/OceanGlidersGDACTrajectories.html	
Licence	Creative Commons: CC-BY	

Additional information on network

Format. The current European glider format is called "EGO format" https://archimer.ifremer.fr/doc/00239/34980/ (where EGO stands for European Gliders Observatories).

Global harmonisation for glider data format (OG1.0) is under progress: the OG1.0 format, vocabularies as well as more information can be found here: https://github.com/OceanGlidersCommunity/OG- format-user-manual

EGO format and OG1.0 are both compliant with OceanOPS (GOOS-OCG) requirements (described in https://www.ocean-ops.org/metadata/)

6.3. Vessels

Authors: Martin Kramp (mkramp@ocean-ops.org), Anthonin Lizé (alize@ocean-ops.org)

Recommended basic information

Variables :

- SOT (Ship Observations Team):
  - VOS (Voluntary Observing Ship), for atmosphere: pressure, temperature, wind speed and direction, humidity; for marine: sea surface temperature
  - SOOP (Ship of Opportunity Program): temperature, salinity, carbon & micro plastics
- GO-SHIP (Global Ocean – Ship): pressure, temperature, salinity, currents, nutrients, oxygen, carbonate system, bathymetry, meteorology

Table 4. VESSELS network status

Vessels NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment OTHER INFORMATION**
Platform identification	IMO otherwise ICES (SHIPC) codes: for the ship WMO: for the data flux (platform)	For GO-SHIP (research) and SOT (opportunity) vessels. Vessel considered as host with various measurements depending on instruments on board
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: yes SDN P01: yes CF standard name: yes SDN P06: no SDN L20: no	 P02 also available (variable thematic) - WIGOS reference Not yet this metadata
Metadata associated to variable (time)	ISO 8601: yes	
Metadata associated to variable (geographical position)	WGS84: yes	
Dataset citation (doi)		To be implemented
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: yes	

Vessels NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment OTHER INFORMATION**
Project citation	EDMERP: no	Yes when relevant, e.g. Eurofleets+
Data format (netCDF CF)	Yes for SOOP & GO-SHIP (prototype) No for others	On going
Data warehouse	yes	
Access protocol (https, ftp)	https, ftp, thredds	
Licence		Under discussion GO-SHIP

Additional information on network

Vessels are the backbone of many platforms and measurements sensors. They are needed for the deployment of the Argo floats, the gliders or the fixed platforms at sea as well as for the measurement of biology (augmented observatory network) that is made from a ship. Considering the EuroSea networks, only the HF radar and tide gauges networks can exist without vessels.

6.4. Eulerian observations (fixed platforms)

Authors: Laurent Coppola (coppola@obs-vlfr.fr), Ivan Rodero (ivan.rodero@emso-eu.org), Long Jiang (ljiang@ocean-ops.org)

Recommended basic information

Variables: Temperature, Salinity, O_2, currents, particles fluxes (sediment traps), geophysics

Table 5. EULERIAN OBSERVATIONS network status

Eulerian observations NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment OTHER INFORMATION**
Platform identification	WMO: yes	

Eulerian observations NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment OTHER INFORMATION**
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: yes SDN P01: yes CF standard name: yes SDN P06: yes SDN L20: yes	
Metadata associated to variable (time)	ISO 8601: yes	
Metadata associated to variable (geographical position)	WGS84: yes	
Dataset citation (doi)	Yes, one DOI per site. Here an example 10.17882/43749	
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: yes	
Project citation	EDMERP: yes	
Data format (netCDF CF)	yes	
Data warehouse	GDAC CORIOLIS for OceanSites/EMSO (for most of the sites)	
Access protocol (https, ftp)	Data by site (platform) and whole dataset: ftp://ftp.ifremer.fr/ifremer/oceansites/ ERDDAP API for geophysics data	
Licence	Creative Commons: CC-BY	

6.5. Tide gauges

Authors: Begoña Pérez Gomez (bego@puertos.es), Elizabeth Bradshaw (elizb@noc.ac.uk)

Recommended basic information

Variables: sea level (water surface height above a datum)

Table 6. TIDE GAUGE network status

Tide gauge NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment** **OTHER INFORMATION**
Platform identification	WMO: no	Expected. Lack of a unique ID in the tide gauge network. Dialogue between OceanOPS, IOC Tsunami and TT-TG (Task Team -Tide Gauges) on going. Expected to allow information according to recommended definitions of tide gauge/tide gauge station/tide gauge site (related to vertical land movement).
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: no SDN P01: no CF standard name: yes SDN P06: yes SDN L20: yes	Expected Expected
Metadata associated to variable (time)	ISO 8601: yes	
Metadata associated to variable (geographical position)	WGS84: yes	
Dataset citation (doi)		Not yet, but recommended. Several initiatives on-going, e.g.: new DOI for the PSMSL dataset.
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: yes	
Project citation	EDMERP: no	If EDMERP could also include programmes as GLOSS, citation could be included in the future

Tide gauge NETWORK **recommendations**	**YES** **value/code** **comment**	**NO** **what is done? Expected? Comment** **OTHER INFORMATION**
Data format (netCDF CF)	yes	
Data warehouse		Data available at different repositories: GLOSS data centres, Copernicus marine, EMODnet Physics, JRC. Ongoing harmonisation work at European (EuroGOOS) and global scale (GLOSS).
Access protocol (https, ftp)	ERDDAP API for some GLOSS data centres	
Licence	Creative Commons: CC-BY	

6.6. HR radar platforms

Authors: Julien Mader (jmader@azti.es), Lohitzune Solabarrieta (lsolabarrieta@azti.es), Anna Rubio (arubio@azti.es), Lorenzo Corgnati (lorenzo.corgnati@sp.ismar.cnr.it), Pablo Lorente (plorente@puertos.es), Emma Reyes (ereyes@socib.es)

Recommended basic information

Variables: surface currents and waves

Table 7. HR RADAR network status

HF radar NETWORK **recommendations**	**YES** **value/code** **comment**	**NO** **what is done? Expected? Comment** **OTHER INFORMATION**
Platform identification	WMO: no	WMO attribute is included in the netcdf files as a global attribute but it is still blank. EDIOS Series ID of the network: • "site_code" (total and radial current data files) • "platform_code" (platform identification)
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: no SDN P01: yes ("sdn_parameter_urn" variable attributes). CF standard name: yes	SDN P06 codes and labels are present to identify the method of production of the original data.

HF radar NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment** **OTHER INFORMATION**
	SDN P06: yes ("sdn_uom_urn" variable attributes). SDN L20: yes. ("sdn_conventions_urn" QC variable attributes)	
Metadata associated to variable (time)	ISO 8601: no	days since 1950-01-01 00:00:00 UTC
Metadata associated to variable (geographical position)	WGS84: yes	and EPSG:4326
Dataset citation (doi)	no	European HF Radar Node is working on the DOI allocation procedure for future versions
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: no	Will be considered
Project citation	EDMERP: partially	The project is cited in the "citation" and "project" global attributes.
Data format (netCDF CF)	yes	
Data warehouse	http://150.145.136.27:8080/thredds/HF_RADAR/HFradar_catalog.html	Soon to be migrated to https://thredds.hfrnode.eu/thredds
Access protocol (https, ftp)	Data by platform and whole dataset: http://150.145.136.27:8080/thredds/HF_RADAR/HFradar_catalog.html (Soon to be migrated to https://thredds.hfrnode.eu/thredds) Data subsetting GUI available for each sea patch HFR systems (via THREDDS NetCDFSubset feature) ERDDAP API through EMODnet data integrator ERDDAP service from the European HFR Node will be implemented in the future	
Licence	Creative Commons: CC-BY 4.0	

Additional information on network

Formats. The current European HF Radar format is netcdf-4 classic model. The standard vocabulary follows the NetCDF CF Metadata Convention Standard Name, Table Version 1.6. HF Radar files are compliant with CF-1.6 Copernicus-InSituTAC-FormatManual-1.42, Copernicus-InSituTAC-SRD-1.5, Copernicus-InSituTAC-Parameters List-3.2.1.

History. Rose HFR datasets follow the recommendations of the European HFR community published in OBPS (Corgnati *et al.*, 2018, 2019).

The European HFR Node was established in 2018 under the coordination of the EuroGOOS HFR task team (Rubio *et al.*, 2017). It is the focal point and the operational asset in Europe for HFR data management and dissemination by promoting networking between EU infrastructures, marine data portals and the global HFR network. The European HFR Node has been fully operational since December 2018 to distribute tools and support for standardisation to HFR providers, to provide standardised near-real-time delayed-mode HFR radial and total current data to the Copernicus Marine Service, EMODnet Physics and SeaDataNet. The results of the integrated efforts in the European HFR community has allowed the harmonisation of data management and standardisation of HFR data access and tools (Mantovani *et al.*, 2020; Lorente *et al.*, 2022).

6.7. Autonomous surface vehicles

Authors: Christoph Waldmann (waldmann@marum.de), Carlos Barrera (carlos.barrera@plocan.eu)

Recommended basic information

Variables: Temperature, Salinity, Turbidity, Chlorophyll-A, Oxygen, Currents …

Table 8. AUTONOMOUS SURFACE VEHICLES (ASV) network status

ASV NETWORK **recommendations**	**YES** **value/code** **comment**	**NO** **what is done? Expected? Comment** **OTHER INFORMATION**
Platform identification	WMO: no	Platform identification by URN is realised for MARUM Wave Glider and would be recommended for other ASVs
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: yes SDN P01: yes CF standard name: yes SDN P06: - SDN L20: no	QC is based on NOAA recommendations (QARTOD) https://ioos.noaa.gov/project/qartod/
Metadata associated to variable (time)	ISO 8601: yes	

ASV NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment** **OTHER INFORMATION**
Metadata associated to variable (geographical position)	WGS84: yes	
Dataset citation (doi)	https://doi.pangaea.de/ 10.1594/PANGAEA.922825	
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: yes	
Project citation	EDMERP: no	
Data format (netCDF CF)	yes	
Data warehouse		Data available at www.panagaea.de
Access protocol (https, ftp)	Via https	
Licence	Creative Commons: CC-BY 4.0	

6.8. Augmented observatories

Authors: Daniele Iudicone (iudicone@szn.it)

Recommended basic information

Variables: oceanic microbiomes (viruses, bacteria, phyto and zooplankton)

Table 9. AUGMENTED OBSERVATORIES network status

AUGMENTED OBSERVATORIES NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment** **OTHER INFORMATION**
Platform identification	WMO: not applicable WoRMS: yes	Oceanic microbiomes: - Taxonomy (from genetics / metabarcoding)

AUGMENTED OBSERVATORIES NETWORK recommendations	**YES value/code comment**	**NO what is done? Expected? Comment** **OTHER INFORMATION**
		- Gene mapping and expression (thanks to nucleic acids, DNA/RNA)
Metadata associated to variable (L22=instrument type, P01=definition of variable, standard name, P06=unit, L20=quality flag)	SDN L22: not applicable SDN P01: not applicable CF standard name: not applicable SDN P06: not applicable SDN L20:	Quality flags to be established, to be used together with DNA/RNA extraction methods.
Metadata associated to variable (time)	ISO 8601: no	To be done
Metadata associated to variable (geographical position)	WGS84: no	To be done
Dataset citation (doi)	no	To be created when the database will be released (also using ZENODO)
Institution identification (operating institution)	EDMO: yes	
Person citation	ORCID: yes	
Project citation	EDMERP: yes when relevant	To have programme as well would be useful
Data format (netCDF CF)	Not relevant	Specific to genomics/genetics
Data warehouse		ENA/EBI: https://www.ebi.ac.uk/ena/browser/home Others, to be established.
Access protocol (https, ftp)	To be defined	
Licence		No data released yet

Additional information on network

Given the recent advancements in fundamental biology due to the introduction of genomic approaches, several research institutions are moving into introducing new -omics protocols to the regular sampling of marine ecosystems and specifically of the oceanic microbiomes (viruses, bacteria, phyto and zooplankton). In practice this corresponds to filter the seawater using different tools (e.g. nets) and filter sizes and, subsequently, to extract sequences of parts of or full genomes (DNA for taxonomy and to identify potential functions) as well as the RNA (to monitor the genes activity in various abiotic and biotic conditions). The raw sequences are then processed using bioinformatics tools to produce datasets for direct scientific use. There are still several limitations. For instance, the scarcity of known genomes strongly limits the taxonomic and functional assignations. Moreover, the standardisation of protocols and processing methods is crucial for data comparability but it is far from being achieved. Finally, the ancillary (oceanographic) data collection is not always produced in a way easy to be adopted by the biologist researchers.

The aim of the network is to support the dissemination of best practices by, first, implementing them in an augmented observatory. Then, the coordination of the activities of marine stations to move into this same direction will be promoting, thanks to a EuroGOOS Biology WG.

However, it remains a challenge to match the standardised oceanographic data categories (as other EuroSea networks) with this one (augmented observatory) since, among others, the biological community is not tightly linked to the physical oceanographic one.

Conclusion

The main objective of this document is, through the aforementioned clear recommendations, to support the networks and their actors to have shareable data information for their main targets, which are the operators and the IT systems. Propositions made (homogeneous data information by platform type or by network) will be useful for data integrators such as EMODnet, Copernicus Marine Service or SeaDataNet to ease their ingestion procedure and ensure the proper visibility to the original source.

Harmonised information and metadata on marine observation data could give rise to important IT developments that will provide access to in situ data by variable, region, period, depth … (in addition to the current display by platforms) while keeping the traceability of the original data. This step forward will be very interesting for many communities of users and among them the scientific users (non-network experts) who look for marine variables available over a region – and multi networks- without knowing about the networks themselves.

Users can have access to the same in situ data through different networks and integrators. They should know and have clear information on, depending on the network or integrator access, the QC assigned to the data, the transformation undergone on the data (e.g. interpolated gridded data) and any specificity “added” to the data. This information shall ensure they choose the best access to the observation data depending on their needs.

The aim is to have a value chain from producer to user and from user to producer: producers get information about who uses their data and when and users get to know what data they use and why.

References

Copernicus Marine In Situ Tac Data Management Team (2021). Copernicus Marine In Situ NetCDF format manual. doi.org/10.13155/59938

Copernicus Marine in situ TAC (2020a). Copernicus in situ NRT current product user manual (PUM). CMEMS-INS-PUM-013-048. doi.org/10.13155/73192

Copernicus Marine In Situ Tac (2020b). For Global Ocean-Delayed Mode in-situ Observations of surface (drifters and HFR) and sub-surface (vessel-mounted ADCPs) water velocity. Quality Information Document (QUID). CMEMS-INS-QUID-013-044. doi.org/10.13155/41256

Corgnati L. *et al.* (2018). Recommendation Report 2 on improved common procedures for HFR QC analysis.

Corgnati L. *et al.* (2019). SeaDataNet management protocols for HF Radar data. https://repository.Oceanbestpractices.org/handle/11329/1511

Harscoat V., Pouliquen S. (2016). Data management handbook. AtlantOS project. EU Horizon 2020 research and innovation programme under grant agreement No 633211.

Koop-Jakopsen *et al.* (2016). Data harmonisation report. AtlantOS project. EU Horizon 2020 research and innovation programme under grant agreement No 633211.

Lorente P. *et al.* (2022). Coastal high-frequency radars in the Mediterranean -Part 1: Status of operations and a framework for future development. 761-795. doi.org/10.5194/os-18-761-2022

Mantovani C. *et al.* (2020). Best Practices on High Frequency Radar Deployment and Operation for Ocean Current Measurement. Frontiers in Marine Science 7:210. doi.org/10.3389/fmars.2020.00210

Quimbert E. *et al.* (2022) Guide principes FAIR. Principes FAIR dans le contexte du pôle ODATIS doi.org/10.13155/87107

Révelard A. *et al.* (2022). Ocean integration : the needs and challenges of effective coordination within the ocean observing system. Frontiers in Marine Science 8:737671. doi:10.3389/fmars.2021.737671

Rubio A. *et al.* (2017). HF radar activity in European coastal seas: next steps towards a pan-European HF radar network,Frontiers in Marine Science, 20, 8. doi.org/10.3389/fmars.2017.00008

Tanhua T. *et al.* (2019) ocean FAIR data services. Frontiers in Marine Science, 6:440. doi:10.3389/fmars.2019.00440

Thijsse P., Schaap D. (2019). Marine subdomain FAIRness roadmap. ENVRI-FAIR project. EU Horizon 2020 research and innovation programme under grant agreement No 824068

Wilkinson, M.D. *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. doi.org/10.1038/sdata.2016.18