Topic	H2020 – INFRAIA-2018-2020
Short Title	EurofleetsPlus
Title	An alliance of European marine research infrastructures to meet the evolving requirements of the research and industrial communities
Project Number	824077
Delivery Date	01/06/2019
Deliverable No	4.10
Lead Beneficiary	RBINS
Dissemination Level	Public

Eurofleets+ Data Management guidelines for cruise and programme applicants and grantees

Document information	
Document Name	EF+_data_management_guidelines_PI_1.2.docx
Document ID	EF+ data management guidelines PI 1.2
Revision	1.2
Revision Date	2019-07-30
Author	Thomas Vandenberghe
Security	Public

Approvals	Name	Organisation	Date
Coordinator			
Activity Coordinator			
WP Leader			

History			
Revision	Date	Modification	Author
V1.0	2019-07-12	First doc	Thomas Vandenberghe
V1.1	2019-07-26	Add guidelines for DMP Roadmap	Thomas Vandenberghe
V1.2	2019-07-30	Add annexes	Thomas Vandenberghe

any means to any third party, in whole or in parts, except with prior written consent of the EUROFLEETS+ Coordinator.

The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

<div align="center">TABLE OF CONTENTS</div>

Contents

1. Applicability

This document describes creating a data management plan and the general data management, and is intended for applicants and grantees of all the Eurofleets+ Access Programmes.

2. 'FAIR', Open Data and licenses

The Horizon 2020 Programme asks that the research output of its beneficiaries adheres to the FAIR principle and furthermore, that open access to research data is ensured to a maximal extent. Open access refers to the practice of providing online access to scientific information (publications and research data) that is free of charge to the end-user and reusable. FAIR data is findable, accessible, interoperable and readable. FAIR data may still be (partially) private, eg. when it is only shared under certain conditions. Since 2017, open research data is the default option for H2020, while it is still possible to opt-out and select a more closed data model. For general H2020 projects, this may be because you want to protect your data before any publication has taken place, keep intellectual property rights for commercial purposes, or have privacy or security concerns. In summary, any Eurofleets+ research data must be Findable, Accessible, Interoperable and Reusable for the group of users you as a research cruise principal investigator have defined; this may be anyone, with unrestricted access, in which case your data is considered open data.

Within open data there are gradations, going from fully open (public domain), to legally obliging attribution, to barring commercial re-use etc. A good overview of open licenses can be found on CreativeCommons, https://creativecommons.org/licenses. A further benefit is that these licenses are ready-for-use, unambiguous (as they already have a legal description), and machine-readable.

The default Eurofleets+ license for cruise data is the public domain CC0-license, https://creativecommons.org/publicdomain/zero/1.0**, as re-using data should not be made unnecessarily complex.**

3. Data Management Plan

Eurofleets+ applies the FAIR and Open principles of the H2020 framework on three levels: the EF+ project itself, the JRA research output and the TNA funded cruises. For this reason there will be two Data Management Plan (DMP) types in use: a general one for Eurofleets+, and one for each funded programme and cruise. Your Data Management Plan should detail the steps you will take to get the data ready before ingestion into the Eurofleets infrastructure (which is detailed in the general DMP). **The DMP for cruises comes in two forms, a preliminary DMP at submission time and a final DMP after the cruise funding has been awarded.**

We have based ourselves on the Guidelines on FAIR Data Management in Horizon 2020 (http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf) to define the template for the Cruise DMP. Some of the specifics, for instance those on the timing of any opt-out (see p. 3), make less sense in the Eurofleets+ context, and are not applicable.

For Eurofleets+, opting out of open data is possible at the time the preliminary DMP is created or after the project has been granted, at the time the final DMP is submitted, but only for data with a distinct commercialibility or security concerns, or by extending the embargo period. The embargo period may last up to two years after the cruise. However, the data must be provided to the Eurofleets+ data centres in a timely fashion (it will be curated but not be disseminated further).

Crucial elements like embargoes, opting-out data and licensing must be included as soon as possible in the DMP for the different types of datasets.

Project proposals will be judged on the completeness of the Data Management Plan. The final DMP helps proper data management. It is a living document that follows the research lifecycle and should be updated during the project lifetime, especially after the cruise ended, to reflect any changes to your procedures. A data management plan prerecords the procedures you plan to follow, to ensure that they are actually followed.

Applicants to the three programmes should complete their DMP as follows:

- Ship-time access of SEA programme:
  - Preliminary DMP + Full DMP mandatory
- Marine Equipment Application of SEA programme:
  - Preliminary DMP + Full DMP mandatory
- Co-PI programme:
  - Preliminary DMP + Full DMP mandatory
- Remote Transnational Access (RTA) programme
  - Preliminary DMP optional unless embargo or data opt-out

The data management practices (i.e. (meta)data reporting using EMODnet and SeaDataNet) need to be followed for all programmes.

4. DMP Roadmap

Because it is a dynamic process and as versioning is difficult to manage using text documents, we use an online DMP tool, called DMP Roadmap. This also facilitates managing a multitude of submissions.

The Data Management Plan (both the preliminary and full DMP) of a cruise proposal MUST be created with DMP Roadmap on the website http://dmp.ef-ears.eu and kept up-to-date after funding and cruise end.

A preliminary DMP is created at the time of the project proposal. This preliminary DMP contains a more limited set of questions. You can leave out details at this stage. After the proposal is granted, the Data Management Plan will have to be completed in full. For this, you need to answer additional questions and extend your previous answers more in-depth. Three reference data management centres, assigned to a funded project, will review your DMP and provide feedback at that stage. The Data Management partners are the Hellenic Centre for Marine Research (HCMR) in Greece, the Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS) in Italy and the Belgian Marine

Data Centre (BMDC). The data centres are also responsible for receiving and processing your datasets, so that they are interoperable for systems like EMODnet and SeaDataNet. They might return to you with questions.

You should not repeat the information that is present in these guidelines, nor repeat the data treatment the Eurofleets+ reference data managers will perform. **The DMP exists so that you feel more comfortable with your own data management because it becomes a forethought, not an afterthought.** You should keep in mind not only the period before, but also beyond the submission of the data in the Eurofleets+ portal. E.g. how will you continue to store all these large datasets after the project? On your own hard disks or on a external safe repository? Short and to the point answers are ok. If (part of) a question is not applicable to your proposal, please indicate this with 'n/a'. Please feel free to share your DMP (or future DMPs) with collaborators so that they are rapidly up-to-date with how you (plan to) work.

5. DMP Roadmap step-by-step guide

Templates

The web application can host multiple data management plan templates, but this instance hosts only one, the Eurofleets+ funded cruise DMP one. This DMP template consists of two phases, one for the preliminary DMP (before the application) and one for the full DMP (Post-Award DMP). One plan per cruise or programme is created and includes both phases.

User accounts

Users have to create their own account and can log in afterwards with this. The organisation of the user is provided at time of registration as a text string but can be changed in the 'Edit profile' screen, upon clicking the username in the top right button.

Pages

Dashboard

The main page after logging in is the Dashboard, which lists all the plans of a user. Form this page, an existing plan can be edited, shared, downloaded, copied or deleted. For sharing, see below. Users can also create a new plan from here.

Create a new plan

A new plan can be created by clicking 'Create plan' in the Dashboard. On the next screen, the Cruise project title, the primary research institute and the primary funding organisation must be completed. The primary research institute comes pre-filled with 'Your Research Institute'. The tick box 'No research organisation associated with this plan or my research organisation is not listed' should be enabled. The primary funding organisation should be completed with 'Eurofleets+' (type an 'e' to see this dropdown menu option). Click 'Create plan' to continue.

Edit plan page

This brings you to the main screen of the application, where you can complete the DMP.

Please note that for each step (tab) and answer, you need to click 'Save', else the information will be lost.

This page has 6 tabs: 'Project details', 'Plan overview', 'Phase 1', 'Phase 2', 'Share' and 'Download'.

Project Details

Complete or adapt here the Title, Grant Number, Abstract and ID. The Grant number should only be completed when your cruise is granted. The ID is the project acronym as you provided it in the online proposal submission website, as is the abstract. If you will not be the data contact person, please specifiy this person's contact details. Please be consistent with the answer in '*Who will be responsible for data management in your project?*'.

Plan overview

This tab provides guidelines and resumes the questions of phase 1 and 2.

Phase 1: Preliminary DMP

The 14 questions are organised per theme, according to the Horizon 2020 DMP template. All questions are mandatory, but the level of detail can be limited. You can format your answer if needed (bold, bulleted lists etc.).

Per question, it is possible to provide comments that collaborators can review. Collaborators are persons that can be invited to review/edit your DMP (see 'Share').

Please click 'Save' for each answer and comment.

Phase 2: Full DMP

The 26 questions of this phase needs to be completed more in full. Please copy-paste your 14 answers from phase 1 and adapt them as needed. All questions are mandatory.

Please click 'Save' for each answer and comment.
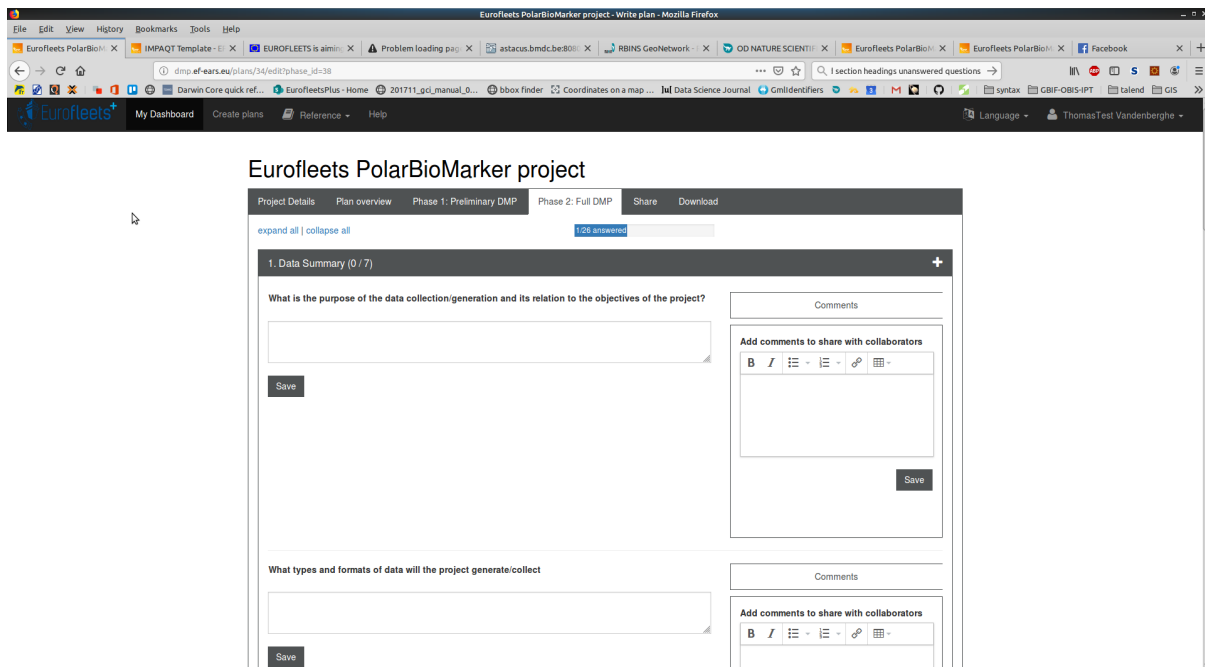
Share

While it is possible to make your plan publically visible on the whole website, this is not advised. You can also share your plan with other colleagues (collaborators) who are then invited to create an account on the site (if they have none yet). These persons cannot further share the DMP.

Download

This tab enables you to download a specific phase of your DMP as a docx or pdf file. Use this functionality to provide your preliminary DMP (phase 1) as a pdf in the proposal submission form.

Please mark all three tick boxes ('project details coversheet', 'question text and section headings' and 'unanswered questions') upon download of the file.

6. Further support in creating the DMP

The following organisations are supporting FAIR data management for European marine research:

- SeaDataNet extended the ISO 19115 geographic metadata standard, and, by using standard vocabularies, provides data access services covering many different fields.

- The European Marine Observation and Data Network (EMODNet) provides access to data and data products and strengthens and makes extensive use of the SeaDataNet infrastructure.

The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

The UK Data Service has a costing tool to help you formulate data management costs and avoid overtime near the end of the submission deadline: *UK Data Service (2013). Data management costing tool. UK Data Archive, University of Essex.*

7. When to provide data to the reference data centres

There are two periods considered: a) the period during which the scientist should entrust the data centres with his/her data; b) the embargo period, which may extend beyond period a). Data delivered under a) will be processed by the reference data centres, imported, QCed and normalised against vocabularies, but if the embargo still applies the data will not be made available for download anywhere. In any case the data must be provided well before the end of the whole Eurofleets+ project (before M45, 01/11/2022) otherwise there is no time for the data centres to make it interoperable.

The raw data and metadata produced during your cruise must be provided withing the following timelines:

- Metadata of the cruise (SeaDataNet Cruise Summary Report): within two weeks after the cruise (cfr. the 'Guidelines for applicants', 'Reporting'.)

- Metadata of the datasets of the cruise: within two months after the cruise

- en-route data: via EARS, automatically

- CTD data and data of deployed devices: within two months after the results are obtained

- "manual" data, e.g. observations and/or measurements on samples: within two months after the result is obtained (to allow lab analyses, does not include QC), together with operations metadata generated by the EARS software.

- embargo: optional, up to two years after the cruise, justified in the DMP

8. How this is done

The Eurofleets+ project will provide the necessary infrastructure to make the data that is gathered during the funded cruises FAIR.

The Eurofleets+ infrastructure takes care of manual data (data originating from specific manual operations like sampling and deployments) and en-route data acquisition (salinity, temperature, depth etc. that are continuously measured). The Eurofleets+ infrastructure should be installed on the vessel before the cruise begins. It:

- takes care of en-route data acquisition (salinity, temperature, depth etc. that are continuously measured). The EARS acquisition software is responsible for this.

- ensures that "manual" data (data originating from specific manual operations like sampling and deployments) can be enriched with the necessary meta-information (e.g. sampling devices, time of sampling, person responsible for sampling). The EARS services and front end-software is responsible for this.

Eurofleets+ makes use of an existing portal for you to submit your metadata and data, the EMODnet Data Ingestion Portal (DIP). It also bridges with and from SeaDataNet to other systems. Your datasets will show up in the SeaDataNet Common Data Index data repository, your cruise will be findable on the Cruise Summary Report Inventory, on the relevant EMODnet portal and the European Geological Data Infrastructure (EGDI, if applicable). With all these steps complete, FAIRness is achieved.

The data submitter in the DIP should be the responsible of the specific granted Eurofleets programme dataset (the PI or co-PI in case of a cruise or the responsible for the sampling programme), and should be in possession of all relevant data files, as there can only be one dataset submission per granted programme, with one zip archive of data files. You may also store the data in any other repository, but this will not be accepted as the official submission.

The different data types are listed in table 1. The data should be in the formats listed in column 8.

The complete data workflow is provided below:
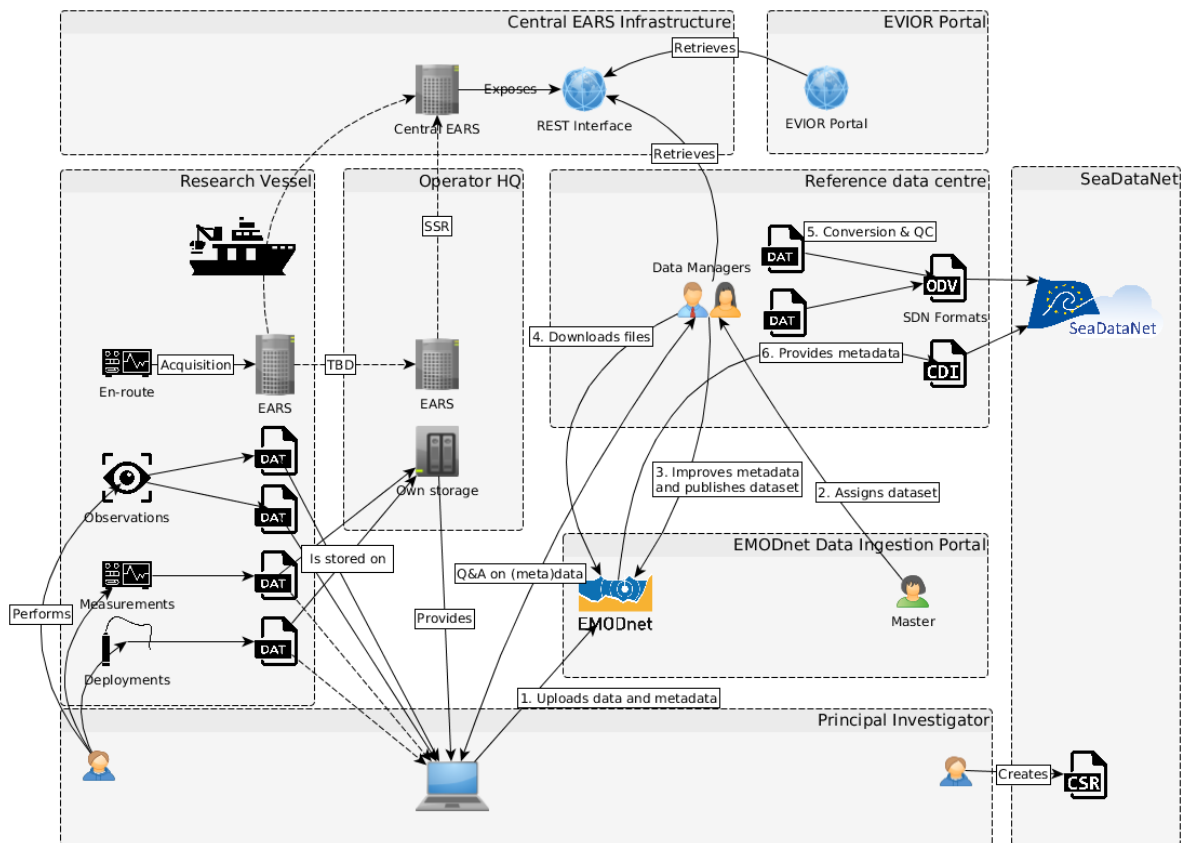
In order to be able to submit the dataset in EMODnet DIP, the data submitter should create a Marine ID for him/herself first. The ingestion portal is hosted on https://submission.emodnet-ingestion.eu/insite.php?module=submitter. The submitter starts phase 1, completes the metadata of the dataset and uploads the data files as a zip. He should report Eurofleets+ in the 'Project / Programme' field of the 'Data types' tab.

Category	Platform	Science	Matrix	Measurement type	spatiotemporal geometry	format at acquisition time	format at submission time by PI	final format by DM
En-route data acquisition	R/V, ROV, AUV, Float	Physics		Location	trajectory	Proprietary output files/network message	csv/ODV/NetCDF EARS for R/V	ODV/NetCDF
En-route data acquisition	R/V, ROV, AUV, Float	Physics	air	Meteorology	trajectory	Proprietary output files/network message	csv/ODV/NetCDF EARS for R/V	ODV/NetCDF
En-route data acquisition	R/V, ROV, AUV, Float	Physics, chemistry	water	Thermosalinometry	trajectory	Proprietary output files/network message	csv/ODV/NetCDF EARS for R/V	ODV/NetCDF
En-route data acquisition	R/V, ROV, AUV, Float	Physics, chemistry	water	AUMS/FerryBox	trajectory	Proprietary output files/network message	csv/ODV/NetCDF EARS for R/V	ODV/NetCDF
Long-term timeseries by deployed sensors	compound sensor	Physics, chemistry	water	Frame with multiple sensors	Point timeseries	Proprietary output files	csv/ODV/NetCDF	ODV/NetCDF
Long-term timeseries by deployed sensors	Single sensor	Physics, chemistry	water	Any single sensor (current meter, sea level,…)	Point timeseries	Proprietary output files	csv/ODV/NetCDF	n/a
Human operations	R/V	Physics	air, water, sediment	CTD profiles	Profile	Proprietary output files	csv/ODV/NetCDF	ODV/NetCDF
Human operations	R/V	Physics	air, water, sediment	Ad hoc physical observations	Single point in time and space	Proprietary output files/spreadsheet	csv/ODV/NetCDF	ODV
Human operations	R/V	Physics, Chemistry, Geology, Biology	water, sediment, biota	Water, sediment or biota samples	Single point in time and space	spreadsheet	csv/ODV	ODV
Human operations	R/V	Geology, Biology	Sediment, ecology	Human observations	Single point in time and space	spreadsheet	csv	ODV
Imagery data	R/V, ROV, AUV, Float	Geology, Biology	Sediment, ecology	Images	Single point in time and space	Image	n/a	n/a

Table 1. Overview of data types, purposes and formats. Not all data types are relevant for Eurofleets+, for instance deployed sensors fixed-point timeseries are not expected. Note that there is no pathway for imagery or sound data; data derived from these data types should be stored as csv

6. Summary

1. Follow the webinar on FAIR, DMPs and Data management via https://www.eurofleets.eu/2019/07/26/eurofleets-data-management-plan-dmp-webinar-how-to-prepare-a-dmp/

2. Prepare your proposal

3. Create an account on the EF+ DMP Roadmap (http://dmp.ef-ears.eu) and complete the preliminary DMP

4. Submit your proposal on the official Eurofleets+ submission website, together with a pdf of the DMP generated from DMP Roadmap

5. Await confirmation that you are awarded the project

6. Complete the final DMP on http://dmp.ef-ears.eu

7. Receive additional feedback from the reference data centres on improving the DMP

8. Participate in the cruise

9. Within 2 weeks after the cruise, provide the SeaDataNet Cruise Summary Report following the 'Guidelines for applicants', chapter 'Reporting'.

10. Within 2 months after the cruise, provide the Cruise Report to the EUROFLEETS+ Evaluation Office at AWI (cfr. the 'Guidelines for applicants', chapter 'Reporting'.)

11. Perform the necessary QC on your raw data so that it is ready for publication in a data repository.

12. The data submitter (the person responsible for the data) gathers all the raw data collected during the cruise:

- CTD data, ROV/AUV data: provide during the data submission

- Multibeam data and imagery should not be submitted. Data derived from them should.

- "Manual" data: provide during the data submission, together with

- Metadata on all ship operations as you have collected it by using the EARS Front end software.

- En route data will be transferred through the EARS acquisition software to the reference data centres, or separately provided by the vessel operator. Not to be included in the data submission.

13. Within 2 months after the cruise, the data submitter provides the data fils and metadata of the dataset in the EMODnet Data ingestion Portal, following the guidelines on https://www.emodnet-ingestion.eu/media/emodnet_ingestion/org/documents/helpguide_ds_22sept2017.pdf. Note that there can only be one submission per cruise/programme and that all data files must be zipped into one archive.

14. Depending on the sea area the cruise takes place in, a data centre will be assigned to your dataset(s). A data manager will contact you if there are any questions about making your data interoperable.

7. Annexes

Metadata of your datasets

The following metadata elements should be completed on the EMODnet Data Ingestion Portal:

- Title: clearly mention 'Eurofleets+', the EF+ cruise identifier and the name of the cruise or programme

- Abstract

- Keywords of the observation types

- File format types

- Provenance and lineage:

    - Data quality processing: handling of outliers, missing and suspect values and null observations) and data harmonisation (code, label and QC flag explanations, consistent use of headers,data formatting, conforming to standards).

    - Summary of processing methodology: Different basic steps how the data was derived from the device and how it was combined, transformed and processed (including software used) to the stage presented now.

- Dataset documentation: any relevant files. Please attach the DMP.

- Creation date of the dataset

- Author of the dataset: the submitter is automatically recorded as the responsible of the dataset, and is considered the author.

    - Full name of the organization

    - Name

    - Email address

    - Country

    - Organisation

- Data holding organisation: country, name, email

- Project or programme: please select Eurofleets+ from the dropdown list

- Sea areas

- Coordinate Reference System (CRS): WGS 84, ETRS89,...

- Spatial Extent (Bounding Box): The spatial coordinates of the widest geographical extent (westBoundLongitude, eastBoundLongitude, southBoundLatitude, northBoundLatitude)

- Temporal Extent: start and stop date of the collection points

- Taxonomic coverage: Report the taxa (use appropriate levels like families, orders or classes) in the abstract.

- Conditions for data sharing: license.

- Embargo: please upload the DMP pdf, so that information on embargoes is known.

- Files: upload one zipped archive

Operational information of your data

The following is a maximal set of meta-information elements that are relevant for one individual observation or measurement. Don't worry about structuring this information, a tabular format, repeating shared meta-information elements for values, is in most cases just fine. Elements that are required by SeaDataNet are marked in bold. This section is still under construction and will be finished near december 2019. Elements in bold are required if applicable. Please note that only the elements in italic are considered the 'value', but for proper interoperability and reusability, all elements are needed.

- **Provide the vessel name, and any platform on that vessel (i.e. ROV, AUVs, RIBs,…) used for sampling**

- **Campaign code as used by Eurofleets+**

- A description of the different features of interest (e.g. sea water column at depth range x, sediment layer at depth range y, benthos communities of gravel beds,…) of each individual measurement.

- Sampling in the case when an ex-situ measurement is performed on the feature of interest:

  - **The sampling location coordinates and their Coordinate Reference System**

  - The name of the location

  - **In case of a station, the station code**

  - **The time, sampling depths and locations at the beginning of the sampling, and if applicable also at the end**

  - Sea bottom depth at sampling location if relevant

  - Vertical datum: depth reference system (mean sea level,...)

  - The time, length, swath and width over which sampling took place

  - The sampling device: **type**, **model**, characteristics, calibration information,...

  - **Subsamples** for chemistry, biology or geology: the extent of the subsample, from which part/depth/organ it was taken,…

- The observed property you have observed or measured:

  - The scientific purpose of the measurement or observation operation

  - The *parameter*

  - **Statistical modifiers** (time-averaged, percentile, standard error,...)

  - The *time at which the measurement was taken*

  - The **time reference** at which the measurement was taken (UTC, time offset to UTC)

- The procedure you applied to take the measurement:

  - Any **preparatory steps**: sieving, filtration, mixing,,…

- An indication on the **fraction** (matrix) or combined sample these steps resulted in (e.g. dissolved, particles (180-300um), wet weight/dry weight,…)

- For chemistry, a description of all **analytical steps** taken

- The algorithm you have applied to the raw device output if applicable

- The measurement device: **type**, model, characteristics, calibration information,...

- The **time** (plus time reference) at which the result of the procedure was known (if it is different from the time at which the measurement was taken)

- The result of the measurement:

  - *The **value***

  - *The **unit** of the value, expressed in units that are considered a community standard*

  - Any information on accuracy and **quality control flags** (plus explanation of their meaning)