



<b>Topic</b>	H2020 – INFRAIA-2018-2020
<b>Short Title</b>	EurofleetsPlus
<b>Title</b>	An alliance of European marine research infrastructures to meet the evolving requirements of the research and industrial communities
<b>Project Number</b>	824077
<b>Delivery Date</b>	30/09/2019
<b>Deliverable No</b>	4.7
<b>Lead Beneficiary</b>	RBINS
<b>Dissemination Level</b>	Public

## Data handling chain setup



Document information	
<b>Document Name</b>	D4.7 Data handling chain setup v1.1 RBINS
<b>Document ID</b>	D4.7 Data handling chain setup v1.1 RBINS
<b>Revision</b>	1.1
<b>Revision Date</b>	2019-09-17
<b>Author</b>	Thomas Vandenberghe
<b>Security</b>	Public

Approvals				
	Name		Organisation	Date
<b>Coordinator</b>				
<b>Activity Coordinator</b>	Serge Scory		RBINS	2019-09-20
<b>WP Leader</b>	Anneli Strobel		AWI	2019-09-24

History			
Revision	Date	Modification	Author
V0.5	2019-09-11	First version	Thomas Vandenberghe
V0.8	2019-09-16	Revision; more info on en-route data	Thomas Vandenberghe
V1.0	2019-09-20	Colleague and WP3 feedback	Thomas Vandenberghe
V1.1	2019-09-24	WP4 and MI feedback	Thomas Vandenberghe

This document contains information, which is proprietary to the EUROFLEETS+ consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by 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.

## TABLE OF CONTENTS

### Contents

Introduction .....	3
Data handling chain .....	3
General data division and data treatment.....	4
En-route data by fixed sensors .....	6
Timelines.....	6
Data types.....	7

### Introduction

Eurofleets+ has three Access Programmes and four calls that will each generate data which should be published in open data repositories and should receive a FAIR data curation: SEA-REGIONAL, SEA-OCEANS, Co-PI and RTA.

The data management planning phase is covered by previous deliverables D4.4, D4.5 and especially D1.3. This deliverable extends the handling chain already detailed in D1.3, where needed. The actual data management work is covered under Tasks 4.4 and 9.4. Data management steps will not be specifically assigned to either tasks, since task 9.4 is quite broad. It includes a) both the actual data management (i.e. up to the curation level needed for SeaDataNet); and b) the data integration that the EVIOR portal will demonstrate in order to provide a unique Eurofleets+ outlook on the data, by integrating the tools used in European Marine Data Management. There are data management follow-up reports due for D4.9 (July 2020) and D4.12 (January 2022); they will report the whole scope of a).

Three data types are expected:

- En-route (underway) data acquisition by fixed sensors on the platform: location, meteorology, thermosalinometry, FerryBox
- Human operations: physical measurements such as a CTD profile or water transparency; measurements and observations derived from water, sediment or biota samples; occurrence observations; multibeam data
- Long-term time series by sensors deployed on frames, ROVs, AUVs or floats

The data types and handling steps for human operations and long-term time series have already been described in D1.3, as they are handled by the PI, with or without interaction with the vessel operator. For en-route data, we define extra data handling steps below.

### Data handling chain

The below image is taken from Deliverable 9.4 (MARIS) and details the complete data handling chain for all data types. Please note that the "Central EARS infrastructure" is a wished feature that is not yet currently available (i.e. within EARS2).

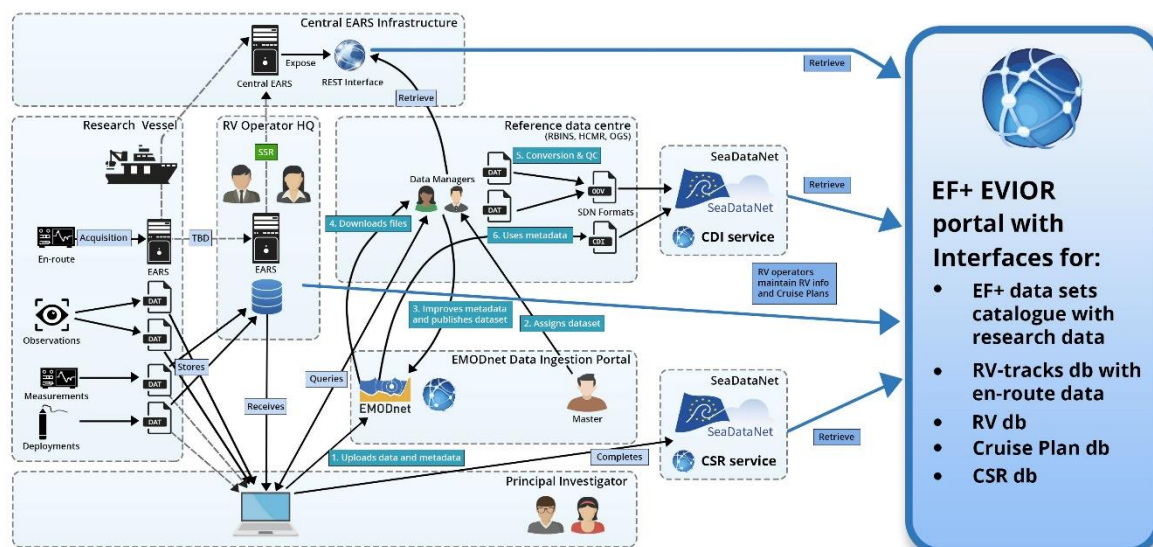


Image 1: complete data flow of the Eurofleets+ project. Abbreviations: EARS: Eurofleets Automatic Reporting System; RV: Research Vessel; CDI: Common Data Index, CSR: Cruise Summary Report.

## General data division and data treatment

The PI or at least someone who has all the data performs one submission in the EMODnet Data Ingestion Portal (DIP) within 2 months after obtaining the results (taking into account eventual lab tests), and includes all data files (including en-route and multibeam data, see below). The metadata record is created at submission time within the DIP web application and is annotated with the EDMERP (EUROPEAN DIRECTORY OF MARINE ENVIRONMENTAL RESEARCH PROJECTS) code for Eurofleets+ and the CSR identifier. The DIP Master disseminates the files to the right data centre. Within the four different call types, the data will be spread evenly over the RDCs (Reference Data Centres). A priori, the datasets are assigned alternately between the three reference data centres (both in alphabetic order), regardless of their scientific content. A correction based on the number of cruise days can be made. Each data centre should be able to handle all data types, however, multibeam data is treated by OGS (Istituto Nazionale di Oceanografia e di Geofisica sperimentale) and HCMR (Hellenic Centre for Marine Research). In the case the data package of one cruise is to be handled by two RDCs, the cruise submission should get assigned still to one main RDC. The main RDC communicates with a supportive data centre in case experience with multibeam data is needed.

The RDCs follow the DIP guidelines on [https://www.emodnet-ingestion.eu/media/emodnet\\_ingestion/org/documents/helpguide\\_dc\\_22sept2017.pdf](https://www.emodnet-ingestion.eu/media/emodnet_ingestion/org/documents/helpguide_dc_22sept2017.pdf).

The RDCs make available to SeaDataNet all gathered data from the whole cruise.

The RDC will communicate with the PI on the following topics:

- Request missing data based on the CSR submitted to SeaDataNet, the Cruise Report submitted to AWI, the Cruise Plan submitted to EVIOR and the DMP; this is done in the DIP by rejecting the form, and letting the submitter resubmit. Please note the DMP must be included in the data submission. The data is kept as-is, only appended to if a data file is lacking.

- Improvements of the metadata: improve the existing fields if needed and add the platforms, instruments, parameters/variables, marine organisms, vertical extents, spatial resolution and frequency of observations.
- QC issues of specific parameters. These should be documented and disseminated to the other RDCs.

Phase I is concluded by making the data available in the Summary Records service.

For EMODnet DIP Phase II, the RDCs will incorporate the data into the SeaDataNet CDI service.

It is not required for Eurofleets+ that the records are included in the main marine data management system (i.e. typically a database system) of the RDC, as these might fall of scope of the projects normally assigned to the institute. The same data system or a clone may be used, or data files may be treated as csv files (modus 3). The RDCs may apply modus 1, 2 or 3 workflows as illustrated in image 2, concerning the workflows commonly used in data centres. The datasets should be quality controlled, converted to standard formats, and made available for uptake by using the NEMO, Octopus and Mikado applications.

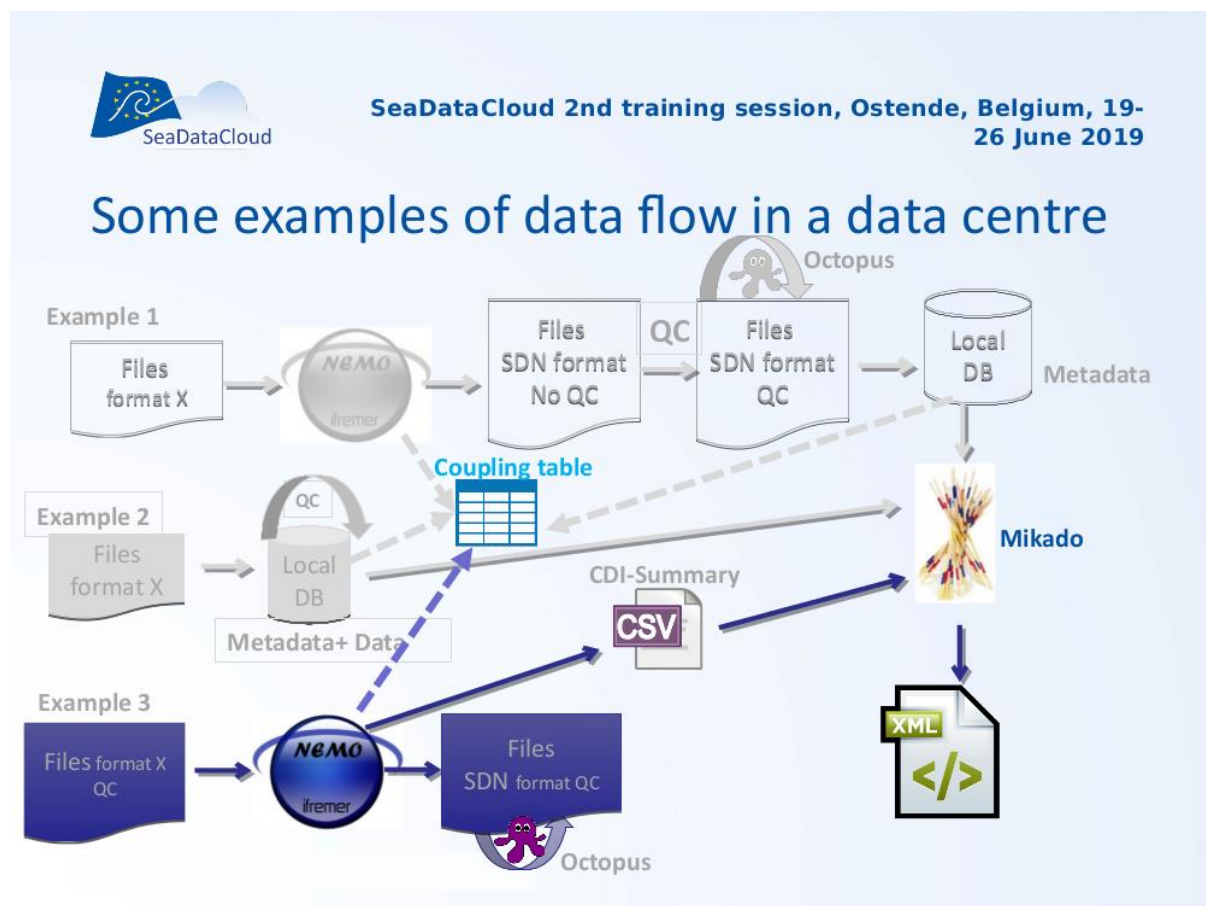


Image 2: Data centre workflows in the SeaDatNet context.

The RDCs should follow the same solutions for commonly encountered QC problems, and should describe these in the EMODnet DIP portal Phase II log (detailing the dataset provenance information). QC and other choices should be shared collaboratively. A first template document with commonly applied QC steps should already be described. RDCs should refer to it as first reference procedure and updates should be added when new cases are encountered. A first version of this

document can be found here:

[https://marineinst.sharepoint.com/:w:/r/sites/EurofleetsPlus/Shared%20Documents/Work%20Packages/WP4%20Call%20Management%20%26%20Proposal%20Evaluation/Data%20Management/QC\\_DM\\_procedures.docx?d=w5122e7a99a924c2094b1ea786235386c&csf=1&e=dYekuY](https://marineinst.sharepoint.com/:w:/r/sites/EurofleetsPlus/Shared%20Documents/Work%20Packages/WP4%20Call%20Management%20%26%20Proposal%20Evaluation/Data%20Management/QC_DM_procedures.docx?d=w5122e7a99a924c2094b1ea786235386c&csf=1&e=dYekuY)

The RDCs use EMODnet DIP as a tracking tool. For the follow-up reports (D4.13 and D4.14) the RDCs should export the "Data centre key indicators" as a csv and send it to RBINS. Both the DM and QC reference procedures and a tracking summary (export from the DIP) should be hosted on the Marine Institute's SharePoint (<https://marineinst.sharepoint.com/sites/EurofleetsPlus>).

## En-route data by fixed sensors

### Timelines

This data originates on the R/V and is disseminated in a semi(-automatic) way through the EARS system. RBINS, CSIC and IFREMER validate the Eurofleets2 setup on their vessels during a trial period (September-October 2019). There are two periods, each using a different approach:

- During 2020 the R/V operators need to follow the Eurofleets 2 procedure. As a result of ongoing discussions (as of 16 September) among the WP3 EARS partners, this procedure might slightly change or be optimized. Therefore EARS should be installed in November-December 2019, and should also include the vessel-to shore solutions.
- Remainder of the project: EARS v3 should be ready by December 2020. From January 2021 (D3.12) new data pathways will be used. This will be specified in D3.2. Rollout to R/V operators should conclude not long after that period.

#### 2020:

The R/V operator installs EARS. A simple docker container installation is being created, together with a separate install of the EARS Front-end application. The docker container will:

- Store the oceanographic events as needed, by interacting with the EARS Front-end application
- Store acquisition data in the EARS database and as NetCDF files.
- Create the Ship Summary Report (SSR) in a continuous loop

The SSR is used to power the visualisation of the ship's track on the EVIOR portal and is, as of EARS2 not optimal for creation of SeaDataNet CDIs (format is in XML, lacks sensor information). To make EARS2 fully functional, the R/V operator should also create a procedure to send the SSR files from vessel to shore and make them available online, so that they can be read by MARIS and made available on the EVIOR portal. They will receive the Eurofleets2 documentation for an example implementation.

For the publication of the en-route data by the RDCs to SeaDataNet, the data centres will make use of NetCDF files generated continuously by the EARS acquisition module. These should be extracted by the R/V operator from the EARS server and provided to the PI so that he can submit them in the complete data package (which includes all manual data and operational metadata and is associated with a dataset metadata record).

In order to build the CDIs, the RDCs make use of the NetCDF files. The disadvantage of the SSR files is that they are encoded in XML which makes conversion to the SeaDataNet Standards more difficult.

### Remainder of the project:

The RDCs download the data either directly from a centralized EARS infrastructure or from EVIOR's Dynamic Vessel Tracking & Events System, depending on the outcome of D3.2.

### Data types

The data will be prioritised as described in T3.1.6:

- Meteorological data (wind speed and direction, air temperature, humidity, solar radiation, air pressure)
- Sea temperature, salinity, currents, oxygen, nutrients, carbon, plankton biomass
- 3.5 kHz or Chirp light seismic
- Multi-beam bathymetry

The data types the Eurofleets2 EARS acquisition module is equipped to deal with out-of-the-box are underlined. D3.2 will implement the additional data types. EARS is not equipped to store light seismics and multibeam bathymetry. For this, additional operational procedures need to be agreed with the RV operators, both for 2020 and in the remainder of the project. Their submission should also follow the general guidelines, namely the use of EMODnet DIP, which means that the RV operator provides the PI with the necessary files.