



**DELIVERABLE TITLE:**

**JERICO-S3 D6.3 - Data Management Best practices report for physical and BGC platforms**

**DELIVERABLE NUMBER: D6.3**

**WORK PACKAGE N° and NAME: WP 6 Data management for multidisciplinary coastal data**

**Authors: Yoana Voynova, Lorenzo Corgnati, Nikolaos Zarokanellos, Miguel Charcos, Gerasimi Anastasopoulou, Leonidas Perivoliotis**

**Involved Institution: Hereon, CNR, SOCIB, HCMR**

**Due date // Submission date: 28.02.2022 // 30.05.2022**

**Nature:R**

(R = Report, P = Prototype, D = Demonstrator, O = Other)

**Dissemination level : PU**

PU = Public, PP = Restricted to other programme participants (including the Commission Services), RE = Restricted to a group specified by the consortium (including the Commission Services), CO = Confidential, only for members of the consortium (including the Commission Services)

**GRANT N°:** 871153

**PROJECT ACRONYM:** JERICO-S3

**PROJECT NAME:** Joint European Research Infrastructure for Coastal Observatories - Science, services, sustainability

**COORDINATOR:** Laurent DELAUNEY - Ifremer, France - jerico@ifremer.fr

*According to the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) and the 78-17 modified law of 6 January 1978, you have a right of access, rectification, erasure of your personal data and a right of restriction to the data processing. You can exercise your rights before the Ifremer data protection officer by mail at the following address: IFREMER – Délégué à la protection des données- Centre Bretagne – ZI de la Pointe du Diable – CS 10070 – 29280 Plouzané - FRANCE or by email: [dpo@ifremer.fr](mailto:dpo@ifremer.fr) + [jerico-s3@ifremer.fr](mailto:jerico-s3@ifremer.fr)  
Ifremer shall not hold your personal data for longer than necessary with regard to the purpose of the data processing and shall destroy it thereafter.*

## DOCUMENT TECHNICAL DESCRIPTION

<b>Document ID</b>	JERICO-S3-WP6-D6.3-02.05.2022-V1.0
--------------------	------------------------------------

<b>JERICO-S3 DELIVERABLE</b>	
Joint European Research Infrastructure network for Coastal Observatory <b>Science, Services, Sustainability</b>	
<b>DELIVERABLE n° WP and full title</b>	JERICO-S3 D6.3 - Data Management Best practices report for physical and BGC platforms
<b>Description</b>	This document summarizes the data management best practices applied to specific JERICO-S3 platforms such as FerryBoxes, HF Radars and Gliders.
<b>Lead beneficiary</b>	HCMR
<b>Lead Authors</b>	Yoana G. Voynova (Hereon), Lorenzo Corgnati (CNR), Nikolaos Zarokanellos (SOCIB), Miguel Charcos Llorens (SOCIB), Gerasimi Anastasopoulou (HCMR), Leonidas Perivoliotis (HCMR)
<b>Contributors</b>	Emma Reyes (SOCIB), Anna Rubio (AZTI), Gisbert Breitbach (Hereon)

<b>REVISION HISTORY</b>			
Revision	Date	Modification	Author
V1.0	13.05.2022	Complete version for review	
V2.0	27.05.2022	Final version	

<b>APPROVALS</b>				
	Name	Organisation	Date	Visa
<b>Coordinator</b>	Delauney Laurent	Ifremer	18/03/2023	DL
<b>External</b>	Andrew King	NIVA	10/03/2023	AK
<b>WP Leaders</b>	Peter Thijsse	MARIS	25.05.2022	PT
	Veronique Creach	CEFAS	30.05.2022	VC
	Leonidas Perivoliotis	HCMR	13.05.2022	LP

<b>Diffusion list</b>			
Consortium beneficiaries	Third parties	Associated Partners	other
<b>X</b>			



**PROPRIETARY RIGHTS STATEMENT**

THIS DOCUMENT CONTAINS INFORMATION, WHICH IS PROPRIETARY TO THE **JERICO-S3** CONSORTIUM. NEITHER THIS DOCUMENT NOR THE INFORMATION CONTAINED HEREIN SHALL BE USED, DUPLICATED OR COMMUNICATED EXCEPT WITH THE PRIOR WRITTEN CONSENT OF THE **JERICO-S3** COORDINATOR.

## TABLE OF CONTENTS

<b>1. EXECUTIVE SUMMARY</b>	<b>6</b>
<b>2. INTRODUCTION</b>	<b>7</b>
<b>3. MAIN REPORT</b>	<b>8</b>
<b>3.1 High Frequency Radar</b>	<b>8</b>
3.1.1 Description of Data	8
3.1.2 Data value chain: from acquisition to delivery	9
Brief description of sensors	9
Methodologies used for data collection	10
Metadata information and data format	11
Data policy	17
Data dissemination – Link to European/International Data Banks	17
3.1.3 Standards used /Contributing Best Practices	17
3.1.4 Quality Control/Quality Assurance	18
Near-Real Time QC	18
Delayed Mode QC	23
3.1.5 Issues	24
3.1.6 Training materials and Contacts	26
3.1.7 Examples of HFR Data Management Plan	28
<b>3.2 Glider</b>	<b>29</b>
3.2.1 Description of data	29
3.2.2 Data value chain: from acquisition to delivery	32
Brief description of sensors	32
Methodologies used for data collection	33
Metadata information and data format	35
Data policy	36
Data dissemination – Link to European/International Data Banks	36
3.2.3 Standards Used / Contributing Best Practices	37
3.2.4 Quality control/Quality Assurance	40
Near-Real Time QC	40
Overview of Near-Real Time and recovery data QC/QA	40
Valid Range Check for sensors and the global ocean	44
Spike Test	46
Gradient Test	46
Near-Real Time visualization	47



Delayed Mode QC	48
Removing bad profiles	48
Data Interpolation	48
Outliers (IQR and STD)	48
Spikes Correction	48
Thermal lag Correction	49
Salinity Correction	49
Dark count Correction	49
Quenching correction	49
3.2.5 Issues	49
3.2.6 Training Materials and Contacts	50
<b>3.3 FerryBox</b>	<b>51</b>
3.3.1 Description of Data	51
3.3.2 Data value chain: from acquisition to delivery	53
Metadata Information	56
3.3.3 Standards used / Contributing Best Practices	57
3.3.4 Quality Control/Quality Assurance Delayed Mode QC	59
Near-Real Time QC	59
Delayed Mode QC	59
3.3.5 Issues	64
3.3.6 Training Materials and Contacts	64
<b>4. ANNEXES AND REFERENCES</b>	<b>65</b>



## 1. EXECUTIVE SUMMARY

This document provides information on the data management best practices for three widely used platforms for the data collection in the coastal zone: HF-Radars, ocean gliders and FerryBoxes. The existing standards and best practices from marine community efforts were reviewed and compiled by specific JERICO-S3 partners (CNR for HF-Radars, SOCIB for Gliders and Hereon for FerryBoxes) all selected due to their long experience, and expertise.

The provided best practices and standards in this report regard the processing steps from the acquisition to the data delivery for each platform, the best practices concerning the data processing, the quality control and quality assurance (both in NRT and DM), the upcoming issues and vulnerabilities of the data management encountered in each platform, training materials and contacts, as well as examples of data management plans. The implementation of the best practices described in this document is expected to increase the data FAIRness, facilitating the data integration into the relevant European Data aggregators, such as the Copernicus Marine, the EMODnet and the SeaDataNet.

## 2. INTRODUCTION

JERICO-RI's observatories collect physical and biogeochemical data along the European coastal zone, making a significant contribution to global ocean monitoring efforts. The variety of the platforms currently operating in the coastal zone makes necessary the release of tools and documentation that will facilitate the access and use of these data by the stakeholders. WP6 aims to provide comprehensive data management guidelines for JERICO-RI coastal platforms through the application of best practices from a multi-platform perspective, covering the whole data lifecycle from data acquisition, processing, storage and preservation to publishing in European data aggregators. These guidelines will be used to process and release the data collected through the combined PSS (Pilot Super Sites) and IRS (Integrated Regional Sites) activities scheduled in WP3 and WP4.

This document contains data management best practices for the HF radars, the gliders, and the FerryBoxes, three platforms that are widely used for a long time in the coastal zone, and consequently, the data acquisition and handling procedures have been thoroughly discussed in the marine scientific community. According to the project's approved DoW, specific partners were assigned the responsibility to identify the existing standards and data management best practices for each one of these platforms due to their long experience, involvement in the platform community and specific expertise: Hereon for the description of BPs regarding FerryBoxes, CNR for HF-Radars and SOCIB for gliders. Furthermore, this document was compiled in conjunction with the WP5-D5.2 deliverable regarding the best practices for deployment and acquisition to form a complete JERICO-RI handbook for platform deployment, data processing and management.

A common template has been used for all the three chapters related to each platform to achieve the BPs homogenization for the data management between the platforms, but any supplementary information considered as valuable could be also added. The provided template includes:

(a) Short description of the data the platform collects as well as the instruments, the data types and more,

(b) Data value chain description, i.e. from data acquisition to data Delivery, as in a short description of the sensors, the methodologies used for the data collection, the metadata information and the data format, the data policies, the data dissemination and the provision of a link to European/ International Banks,

(c) Description of the standards used and/or the best practices regarding the data processing for each platform,

(d) Procedures regarding the quality control (near-real time and delayed mode) and the data quality assurance for each platform,

(e) Any occurring issues each platform is confronted with such as its drawbacks and vulnerabilities,

(f) Training materials and contacts, and finally

(g) Disclosure of examples of Data Management Plans.

### 3. MAIN REPORT

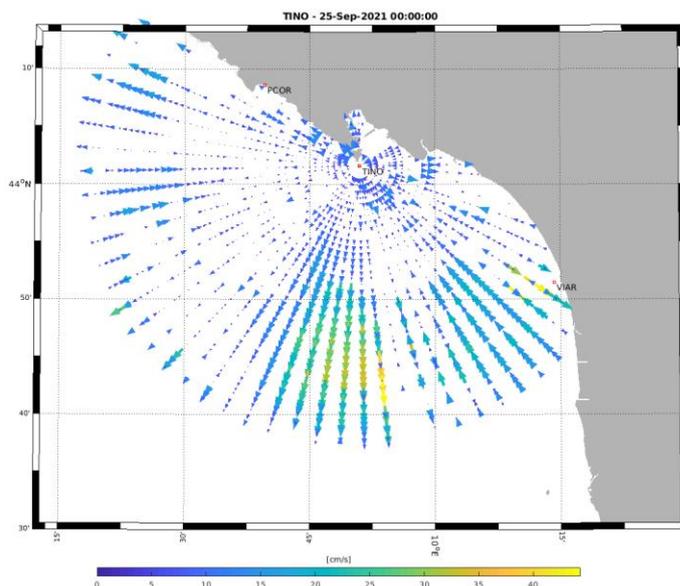
#### 3.1 High Frequency Radar

##### 3.1.1 Description of Data

This section briefly describes the types of data to be managed. The main geophysical variable measured by High Frequency Radars (HFR) is the near-surface sea water current velocity. Data generated from HFR measurements are two-dimensional gridded data and they are of two types: radial data and total data. The term 'radial' refers to current velocities lying on radial lines centered in each of the measurement sites. The term 'total' refers to the effective near-surface current velocity, which is obtained from the combination of the radial data from at least two radar sites. Radial velocity vectors can be considered the radial components of the total velocity vectors.

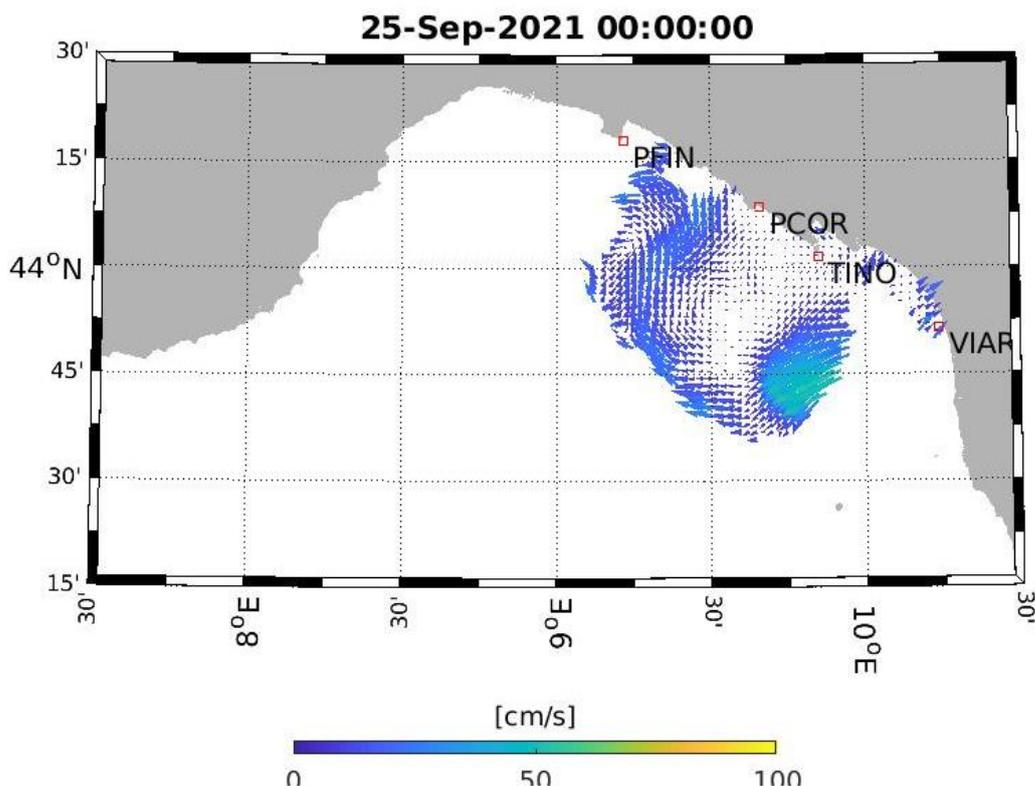
Depending on the instrument manufacturer and on the acquisition method, raw data acquired by HFR systems present different variable sets and different names for common variables. The geophysical content of data from all HFR systems can be considered the same.

Radial data contain magnitude and direction of near-surface sea water radial current velocities, near-surface zonal and meridional components of sea water radial current velocities, standard deviation of near-surface zonal and meridional components of sea water radial current velocities, quality flags and metadata. **Figure 3.1.1.1** shows an example of radial velocity data measured by HFR systems.



**Figure 3.1.1.1.** Radial velocity field measured by the TINO HFR station (located at Isola del Tino, in the Ligurian Sea), belonging to HFR-TirLig network.

Total data contain near-surface zonal and meridional components of sea water current velocities, standard deviation of near-surface zonal and meridional components of sea water current velocities, Geometrical Dilution of Precision (GDOP) (Chapman et al., 1997; Kim et al., 2007), quality flags and metadata. **Figure 3.1.1.2** shows an example of total velocity data measured by HFR systems.



**Figure 3.1.1.2.** Total velocity field measured by the HFR-TirLig network in the Ligurian Sea.

In addition to surface coastal ocean currents, HFR can measure directional wave spectrum and derived parameters (i.e. significant wave height, wave period, mean wave direction) as well as wind intensity and direction, but these data types are out of the scope of the present document, since standardization is not yet existing. Thus, only HFR radial and total sea water current velocities will be considered in this document.

### 3.1.2 Data value chain: from acquisition to delivery

#### *Brief description of sensors*

HFR systems are based on the principle of Bragg scattering of the electromagnetic radiation over the rough conductive sea surface (Crombie, 1955). They infer the radial current component from the Doppler shift of radio waves backscattered by surface gravity waves of half their electromagnetic wavelength. Each single radar site is configured to estimate radial currents moving toward or away from the receive antenna.

The speed of the wave is derived from linear wave theory and then the velocity of the underlying ocean surface currents is retrieved by subtraction. The distance to the backscattered signal is determined by range-gating the returns. Depending on the hardware settings and the methodology used to determine the incoming direction of the scattered signal, commercial HFR systems can be differentiated into two major types: Beam Forming (BF) and Direction Finding (DF). BF radars use linear phased arrays of receive antennas (8 to 16 antennas in a linear array) to electronically steer the sensing beam (Gurgel et al, 1999). DF radars (Barrick et al., 1985) measure the return signal continuously over all angles, exploiting the directional properties of a three-element antenna system (two directionally dependent orthogonal crossed loops and a single omnidirectional monopole) and use the Multiple

Signal Characterization (MUSIC) DF algorithm (Schmidt, 1986) for determining the direction of the incoming signals.

HFR systems operate at specific frequencies within the 3–50-MHz band and provide radial measurements which are representative of current velocities in the upper 0.5–2 meters of the water column. In regions of overlapping coverage from two or more sites, radial current estimations are geometrically combined to estimate total current vectors on a predefined Cartesian regular grid.

The specific geometry of the HFR domain and the intersection angles of radial vectors influence on the accuracy of the total current vectors resolved at each grid point is quantified via a dimensionless parameter named Geometrical Dilution of Precision (GDOP) (Chapman, 1997).

Best Practices on deployment, operation and maintenance of HFR systems are detailed in (Mantovani et al., 2020).

### *Methodologies used for data collection*

High Frequency Radars are land-based systems for remote sensing of ocean surface currents and waves, thus a continuous access to their measurements is possible via different communication technologies. This allows the establishment of Near Real Time (NRT) dataflows. Depending on the HFR manufacturer and operational configurations, the native data formats may be different, as well as the variable names and the metadata. In order to foster the coordinated development of HFR technology and its products for ensuring the full exploitation of its potential within the development of the European operational oceanography, a common European QC, data and metadata model has been defined (Corgnati et al., 2018) and operational tools and services have been developed to automatically ingest and harmonise data coming from different HFR data sources.

In this framework, the European High Frequency Radar Node (EU HFR Node) was established in 2018 by AZTI, CNR-ISMAR and SOCIB, under the coordination of the EuroGOOS HFR Task Team (<http://eurogoos.eu/high-frequency-radar-task-team/>), as the focal point and operational asset in Europe for HFR data management and dissemination, also promoting networking between EU infrastructures and the Global HFR network. The EU HFR Node is fully operational since December 2018 in distributing tools and support for standardization to the HFR providers. At present, the EU HFR Node manages data from 16 European HFR networks (built by 53 radar sites) and integrates US HFR network data (173 radial stations, grouped in 5 networks).

In particular, the EU HFR Node implements the following functions:

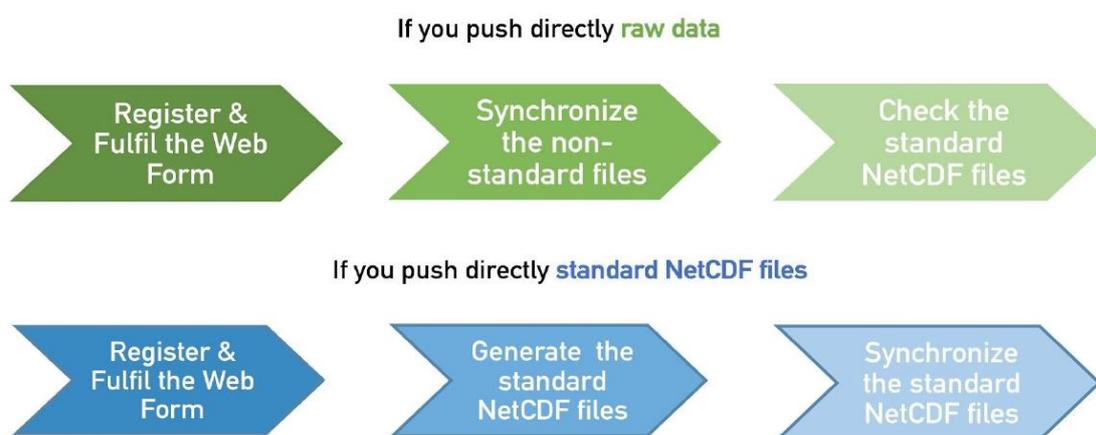
- data acquisition and harvesting
- Quality Control (QC)
- conversion to the European standard data format for HFR current data
- validation/assessment
- delivery of NRT and historical HFR current data with different reprocessing levels.

The EU HFR Node service is founded on a simple and very effective rule: if the data provider can set up the data flow according to the defined standards, the node only collects, checks and distributes the datasets. If the data center cannot set up the data flow, the EU HFR Node directly harvests the raw data from the provider, harmonizes, quality-controls and formats these data and makes them available to the marine portals. The strength and flexibility of this solution reside in the architecture of the European HFR node, which is based on a centralized database, fed and updated by the operators via a webform (<http://150.145.136.36>). The database contains updated metadata of the HFR networks and the needed information for processing/archiving the data.

The collection of HFR data is operated either in real time directly from the radial stations (for providers relying on the node processing service) or from the storage systems (for providers sending standardized data) via rsync protocol. The data are synchronized towards the EU HFR Node servers, where they are processed and/or validated.

The guidelines on how to set the data flow from HFR providers to the EU HFR Node are described in (Reyes et al., 2019).

**Figure 3.1.2.1** shows the possible dataflows implemented by the EU HFR Node for data harmonization and distribution.



**Figure 3.1.2.1.** Workflow for the integration of HFR data in the EU HFR node (picture from (Reyes et al., 2019)).

#### *Metadata information and data format*

This section details the standards applied to data and metadata format and syntax using community agreed vocabularies. The EU HFR Node generates datasets of HFR-derived current measurements compliant to the European common QC, data and metadata model for real-time HFR current data (Corgnati et al., 2018), that is the official European standard for HFR real-time current data, allowing to ensure efficient and automated HFR data discovery and interoperability with tools and services across distributed and heterogeneous earth science data systems.

The model was implemented according to the standards of Open Geospatial Consortium (OGC) for access and delivery of geospatial data, and compliant with the Climate and Forecast Metadata Convention CF-1.6, the Copernicus Marine Environment Monitoring Service In Situ Thematic Assembly Center (CMEMS-INSTAC) conventions, the Unidata NetCDF Attribute Convention for Data Discovery (ACDD), the OceanSITES convention and the INSPIRE directive. The model also follows the guidelines of the Data Management, Exchange and Quality Working Group (DATAMEQ) and fulfills the recommendations given by the Radiowave Operators Working Group (US ROWG).

The European common QC, data and metadata model for real-time HFR current data integrates the SeaDataNet (SDN) requirements about the SeaDataNet metadata services (<https://www.seadatanet.org/Metadata>) for enforcing discovery and access of HFR data and in order to gain visibility and valorization for the projects and the institutions producing HFR data.

The HFR related metadata directories constituting the SeaDataNet metadata services are:

- European Directory of Marine Organizations (**EDMO**): it contains up-to-date addresses and activity profiles of research institutes, data holding centers, monitoring agencies,

governmental and private organizations, that are in one way or another engaged in oceanographic and marine research activities, data and information management and/or data acquisition activities (<https://www.seadatanet.org/Metadata/EDMO-Organisations>).

- European Directory of the Initial Ocean-Observing Systems (**EDIOS**): it gives an overview of the ocean measuring and monitoring systems operated by European countries. This directory includes discovery information on location, measured parameters, data availability, responsible institutes and links to data-holding agencies plus some more technical information on instruments (<https://www.seadatanet.org/Metadata/EDIOS-Observing-systems>).
- Common Data Index (**CDI**): it gives users a highly detailed insight in the availability and geographical spreading of marine data sets and it provides a unique interface for requesting access, and if granted, for downloading datasets from the distributed data centers across Europe (<https://www.seadatanet.org/Metadata/CDI-Common-Data-Index>).

EDMO, EDIOS and CDI entries are xml files to be prepared using Mikado software (<https://www.seadatanet.org/Software/MIKADO>). Entries have to be mailed to [sdn-userdesk@seadatanet.org](mailto:sdn-userdesk@seadatanet.org) for ingestion.

Each HFR data provider is mandatorily asked to have EDMO, EDIOS and CDI entries.

SDN is also managing the following catalogs:

- European Directory of Marine Environmental Research Projects (**EDMERP**): it covers marine research projects for a wide range of disciplines. Research projects are described as metadata factsheets with their most relevant aspects. The primary objective is to support users in identifying interesting research activities and in connecting them to involved research managers and organizations across Europe (<https://www.seadatanet.org/Metadata/EDMERP-Projects>).
- European Directory of Marine Environmental Datasets (**EDMED**): it is a comprehensive reference to the marine data sets and collections held within European research laboratories, so as to provide marine scientists, engineers and policy makers with a simple mechanism for their identification. It covers a wide range of disciplines (<https://www.seadatanet.org/Metadata/EDMED-Datasets>).

EDMERP and EDMED entries are xml files to be prepared using Mikado software (<https://www.seadatanet.org/Software/MIKADO>). Entries have to be mailed to [sdn-userdesk@seadatanet.org](mailto:sdn-userdesk@seadatanet.org) for ingestion.

HFR data providers are invited to provide EDMERP and EDMED entries.

The European common QC, data and metadata model for real-time HFR data uses **NetCDF** (Network Common Data Form), a set of software libraries and machine-independent data formats that is an international standard for sharing scientific data.

The recommended implementation of NetCDF is based on the community-supported Climate and Forecast Metadata Convention (CF), which provides a definitive description of the data in each variable, and the spatial and temporal properties of the data. The used version is **CF-1.6** and it must be identified in the *Conventions* attribute.

The European common data and metadata model for real-time HFR data adds some requirements to the CF-1.6 standard, to fulfill the requirements of CMEMS-INSTAC and SDC CF extension.

In particular:

- where time is specified as a string, the **ISO8601 standard "YYYY-MM-DDThh:mm:ssZ"** is used; this applies to attributes and to the base date in the *units* attribute for time. There is no default time zone; **UTC** must be used, and specified.
- Global attributes from **Unidata's NetCDF Attribute Convention for Data Discovery (ACDD)** are implemented.
- **INSPIRE directive compliance** is recommended.
- Variable names from **SeaDataNet (SDN) P09 controlled vocabulary** are used.

The QC, data and metadata model applies to both real-time radial velocity data and real-time total velocity data.

The European common format for HFR real-time data is **netCDF-4 classic model format**.

NetCDF-4 is the state of the art version of the netCDF library and it has been launched in 2008 to support per-variable compression, multiple unlimited dimensions, more complex data types, and better performance, by layering an enhanced netCDF access interface on top of the HDF5 format.

At the same time, a format variant, netCDF-4 classic model format, was added for users who needed the performance benefits of the new format (such as compression) while keeping backward compatibility with previous versions.

The components (dimensions, variables and attributes) of NetCDF data set are described in the following.

The **global attribute** section of a netCDF file describes the contents of the file overall, and allows for data discovery. All fields should be human-readable and use units that are easy to understand. Global attribute names are case sensitive.

The European common QC, data and metadata model for real-time HFR data divides global attributes to be adopted for HFR data in three categories: Mandatory Attributes, Recommended Attributes and Suggested Attributes.

The Mandatory Attributes include attributes necessary to comply with CF-1.6, OceanSITES and CMEMS-INSTAC conventions (Copernicus-InSituTAC-FormatManual-1.41. Copernicus-InSituTAC-SRD-1.5, Copernicus-InSituTAC-ParametersList-3.2.0). The global attributes required for the SDC Common Data Index (CDI) scheme and the SDC CF extension are mandatory as well.

The Recommended Attributes include attributes necessary to comply with INSPIRE and Unidata Dataset Discovery conventions.

The Suggested Attributes include attributes that can be relevant in describing the data, whether it is part of the standard or not.

Attributes are organized by function: Discovery and Identification, Geo-spatial-temporal, Conventions used, Publication information and Provenance.

The complete list of global attributes, their description and the required syntax are reported in Section 5.2 of (Corngati et al., 2018), available in the Ocean Best Practices Repository at <http://hdl.handle.net/11329/1441>.

NetCDF **dimensions** provide information on the size of the data variables, and additionally tie coordinate variables to data. CF recommends that if any or all of the dimensions of a variable have the interpretations of "date or time" (T), "height or depth" (Z), "latitude" (Y), or "longitude" (X) then those dimensions should appear in the relative order T, Z, Y, X in the variable's definition.

In the specific case of HFR radial data files, if the radial measurements are taken by the instruments based on a polar geometry (e.g. Codar.ruv files), the X and Y axis dimension shall be "bearing" (Y) and "range" (X). In this case, anyway, latitude and longitude shall be present in the netCDF file as data variable.

Since HFR data have only one depth layer of measurement, i.e. the surface layer, the depth dimension must have size equal to 1 and value equal to 0 meters.

If non-physical variables are present in the data file, e.g. the processing parameters of the HFR device generating the data or the codes of the sites contributing to a total velocity data, related non-physical dimensions may be defined to expose the variables in the model.

The complete list of dimensions, their description and the required syntax are reported in Section 5.3 of (Corgnati et al., 2018), available in the Ocean Best Practices Repository at <http://hdl.handle.net/11329/1441>.

NetCDF **coordinates** are a special subset of variables. **Coordinate variables** orient the data in time and space; they may be dimension variables or auxiliary coordinate variables (identified by the *coordinates* attribute on a data variable).

Coordinate variables have an *axis* attribute defining that they represent the X, Y, Z, or T axis. The only exception is the *crs* variable, that is an ancillary coordinate variable required by the SDC CF extension. Missing values are not allowed in coordinate variables.

The latitude and longitude datum is **WGS84**, i.e. the default output of GPS systems.

Bearing and range are the coordinate variables for radial velocity data. For radial data measured on a polar geometry (e.g. Codar .ruv files), latitude and longitude are data variables since they are evaluated starting from bearing and range.

The complete list of coordinate variables, their description and the required syntax are reported in Section 5.4 of (Corgnati et al., 2018), available in the Ocean Best Practices Repository at <http://hdl.handle.net/11329/1441>.

The SDN extensions to CF were concerned with providing storage for standardized semantics and metadata included in the SDN profiles format. In addition to extending coordinate variables and attributes within variables, there are a number of **SDN namespace variables** that form part of the SeaDataCloud extension.

The complete list of SDN namespace variables, their description and the required syntax are reported in Section 5.5 of (Corgnati et al., 2018), available in the Ocean Best Practices Repository at <http://hdl.handle.net/11329/1441>.

**Data variables** contain the actual measurements and information about their quality, uncertainty, and mode by which they were obtained.

The European common QC, data and metadata model for real-time HFR data divides data variables to be adopted for HFR data in two categories: Mandatory Variables and Recommended Variables.

When an appropriate CF standard name is available, it is required to be used; if no such name exists in the CF standard, the *standard\_name* attribute should not be used. In those cases, the *long\_name* attribute has to be used. Please refer to the CF Standard Names table online for authoritative information (definitions, canonical units) on standard names.

It is recommended that variable names be a 4-character-capitalized-letters name. They are not strictly standardized, however; one should use the CF *standard\_name* attribute to query data files. Note that a single standard name may be used more than once in a file, but variable names are unique.

Data variables required in the SDC CF extension are mandatory as well.

Each data variable is equipped with specific attributes, named **variable attributes**. Variable attributes can be mandatory or recommended, however the European QC, data and metadata model for real-time HFR data recommends that all other attributes be used and contain meaningful information, unless technical reasons make this impossible. Variable attributes required in the SDC CF extension are mandatory.

Even if CF conventions prefer the use of coordinate variables as dimensions, because it conforms to COARDS (Cooperative Ocean-Atmosphere Research Data Service) convention and because it simplifies the use of the data by standard software, in order to comply with SDC CF extension data model, the European common QC, data and metadata model for NRT HFR data mandates to declare the variables with all their dimensions and also to have the *coordinates* attribute filled with the list of dimensions. The complete list of data variables and related attributes, their description and the required syntax are reported in Section 5.6 of (Corgnati et al., 2018), available in the Ocean Best Practices Repository at <http://hdl.handle.net/11329/1441>.

**Quality Control variables** are variables storing the results of specific Quality Control (QC) tests to be applied to data, as mandated by the European common QC, data and metadata model for NRT HFR data (see Section 3.1.4).

Since in HFR data the quality control values vary along one or more axes of the data variables, they are provided as separate numeric flag variables, with at least one dimension that matches the 'target' variable.

When QC information is provided as a separate flag variable, CF-1.6 requires that these variables carry the *flag\_values* and *flag\_meanings* attributes. These provide a list of possible values and their meanings.

QC variables can also exist not linked to a target physical variable (e.g. GDOP threshold QC variable linked to GDOP variable), but also as standalone variables reporting the results of a specific QC test, e.g. Over-water test (see Section 3.1.4).

No CF-1.6 standard names exist for QC variables, thus long names have to be used. QC variables must be of type short.

Each QC variable is equipped with specific variable attributes, which can be mandatory or recommended. However, the European QC, data and metadata model for real-time HFR data recommends that all other attributes be used and contain meaningful information, unless technical reasons make this impossible.

The complete list of QC variables and related attributes, their description and the required syntax are reported in Section 5.7 of (Corgnati et al., 2018), available in the Ocean Best Practices Repository at <http://hdl.handle.net/11329/1441>.

The European QC, data and metadata model for real-time HFR data strongly recommends the application of **data packing**, i.e. a method for reducing the data volume by reducing the precision of the stored numbers. It is implemented using the variable attributes *add\_offset* and *scale\_factor*. After the data values of a variable have been read, they are to be multiplied by the *scale\_factor* and have *add\_offset* added to them. If both attributes are present, the data are scaled before the offset is added. When scaled data are written, the application should first subtract the offset and then divide by the scale factor. The units of a variable should be representative of the unpacked data. If the *scale\_factor* and *add\_offset* attributes are of the same data type as the associated variable, the unpacked data is assumed to be of the same data type as the packed data. However, if the *scale\_factor* and *add\_offset* attributes are of a different data type from the variable (containing the packed data) then the unpacked data should match the type of these attributes, which must both be of type float or both be of type double. An additional restriction in this case is that the variable containing the packed data must be of type byte, short or int. It is not advised to unpack an int into a float as there is a potential precision loss. When data to be packed contains missing values, the attributes that indicate missing values (*\_FillValue*, *valid\_min*, *valid\_max*, *valid\_range*) must be of the same data type as the packed data. Please refer to the NetCDF Climate and Forecast Metadata Conventions

(<https://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html>) for further details about data packing.

The **data type** is a bigram used in file names for a quick identification of the file content, in the framework of the adopted **naming convention**. According to CMEMS-INSTAC requirements, data filenames must contain the two bigrams '**XX\_YY**', where:

- the bigram 'XX' indicates the type of measurement;
- the bigram 'YY' indicates the data type.

For HFR data the two bigrams 'XX' and 'YY' are defined as:

- XX=TV (Total Velocity) for total current data files;
- XX=RV (Radial Velocity) for radial current data files;
- YY=HF

Thus, the two bigrams 'XX\_YY' inside the filenames are:

- 'TV\_HF' for total current velocity data files;
- 'RV\_HF' for radial current velocity data files.

Please refer to CMEMS-INSTAC System Requirement Document (<http://archimer.ifremer.fr/doc/00297/40846/50211.pdf>) for further details about the adopted naming convention.

In order to fulfill the specific requirements of CMEMS-INSTAC, EMODnet Physics and SDC Data Access, that are operationally distributing NRT and historical HFR data since 2019, the European common QC, data and metadata model for NRT HFR data was declined for those specific applications: the manual for the standard QC, data and metadata model adopted in CMEMS-INSTAC and EMODnet Physics is described in (Copernicus Marine In Situ Tac Data Management Team, 2020), the one for SDC Data Access is described in (Cognati et al., 2019).

The following table reports the links to the documents that build the QC, data and metadata standard model for NRT HFR current data.

**Table 3.1.2.1.** Links to the documents that build the QC, data and metadata standard model for NRT HFR current data.

Manual	Link
European common QC, data and metadata model	<a href="https://repository.oceanbestpractices.org/handle/11329/1441">https://repository.oceanbestpractices.org/handle/11329/1441</a>
CMEMS-INSTAC Product User Manual	<a href="https://archimer.ifremer.fr/doc/00620/73192/">https://archimer.ifremer.fr/doc/00620/73192/</a>
CMEMS-INSTAC Quality Information Document	<a href="https://archimer.ifremer.fr/doc/00301/41256/">https://archimer.ifremer.fr/doc/00301/41256/</a>
CMEMS-INSTAC NetCDF format manual	<a href="https://archimer.ifremer.fr/doc/00488/59938/">https://archimer.ifremer.fr/doc/00488/59938/</a>
SeaDataNet HFR model	<a href="https://repository.oceanbestpractices.org/handle/11329/1511">https://repository.oceanbestpractices.org/handle/11329/1511</a>

### Data policy

A common and standardized data policy for HFR data does not exist: the access to HFR data is regulated by each data provider. Anyway, the EU HFR Node recommends the free distribution of HFR data by the adoption of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

At present, data from 15 out of the 16 European HFR networks and from the whole US network (i.e. data from 223 out of the 226 HFR radial sites, counting for 98.7% of the managed radial sites) are freely distributed using the Creative Commons Attribution 4.0 International License. These data are accessible for discovery and download via the EU HFR Node THREDDS catalog at:

[http://150.145.136.27:8080/thredds/HF\\_RADAR/HFradar\\_CMEMS\\_INSTAC\\_catalog.html](http://150.145.136.27:8080/thredds/HF_RADAR/HFradar_CMEMS_INSTAC_catalog.html)

### Data dissemination – Link to European/International Data Banks

Since December 2018 the EU HFR Node operationally distributes standardized Near Real Time HFR derived current data (both radial and total velocity data) from European and US networks towards CMEMS-INSTAC (<http://www.marineinsitu.eu/>) and EMODnet Physics (<https://portal.emodnet-physics.eu/>) marine data portals.

NRT HFR current data are available and freely accessible in the following CMEMS-INSTAC data products:

- [INSITU\\_GLO\\_UV\\_NRT\\_OBSERVATIONS\\_013\\_048](#)
- [INSITU\\_GLO\\_NRT\\_OBSERVATIONS\\_013\\_030](#)
- [INSITU\\_ARC\\_NRT\\_OBSERVATIONS\\_013\\_031](#)
- [INSITU\\_BAL\\_NRT\\_OBSERVATIONS\\_013\\_032](#)
- [INSITU\\_IBI\\_NRT\\_OBSERVATIONS\\_013\\_033](#)
- [INSITU\\_MED\\_NRT\\_OBSERVATIONS\\_013\\_035](#)
- [INSITU\\_NWS\\_NRT\\_OBSERVATIONS\\_013\\_036](#)

NRT HFR data are available and freely accessible within the EMODnet Physics data portal by applying the filter 'Radar' in the 'Platform Type' menu.

The EU HFR Node operationally distributes standardized Delayed Mode HFR derived current data (both radial and total velocity data) from European and US networks towards CMEMS-INSTAC (<http://www.marineinsitu.eu/>) and SDC Data Access (<https://www.seadatanet.org>) marine data portals.

Delayed Mode HFR current data are available and freely accessible in the following CMEMS-INSTAC data products:

- [INSITU\\_GLO\\_UV\\_L2\\_REP\\_OBSERVATIONS\\_013\\_044](#)

Delayed Mode HFR current data are freely accessible within SDC Data Access portal by using the 'HFR' keyword in the search field of the CDI discovery page at <https://cdi.seadatanet.org/search>.

### 3.1.3 Standards used /Contributing Best Practices

The international standards and conventions used for standardized HFR data processing are the following:

- Open Geospatial Consortium (OGC) for access and delivery of geospatial data: <https://www.ogc.org/docs/is>
- Climate and Forecast Metadata Convention CF-1.6:

<https://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html>

- Unidata NetCDF Attribute Convention for Data Discovery (ACDD):  
<https://www.unidata.ucar.edu/software/netcdf-java/v4.6/metadata/DataDiscoveryAttConvention.html>
- ISO8601 standard date and time format:  
<https://www.iso.org/iso-8601-date-and-time-format.html>
- OceanSITES:  
[http://www.oceansites.org/docs/oceansites\\_data\\_format\\_reference\\_manual.pdf](http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf)
- INSPIRE Directive:  
<https://inspire.ec.europa.eu/>
- SeaDataNet Metadata Formats:  
<https://www.seadatanet.org/Standards/Metadata-formats>
- SeaDataNet Data Transport Formats:  
<https://www.seadatanet.org/Standards/Data-Transport-Formats>
- CMEMS-INSTAC System Requirements Document (SRD):  
<https://archimer.ifremer.fr/doc/00297/40846/>
- Integrated Ocean Observing System (IOOS) Quality Assurance / Quality Control of Real Time Oceanographic Data (QARTOD) Manual for Real-Time Quality Control of High Frequency Radar Surface Current Data:  
<https://repository.oceanbestpractices.org/handle/11329/288>

The following controlled vocabularies are used as well in the standardized HFR data processing:

- National Environment Research Council (NERC) Vocabulary P01:  
[https://vocab.nerc.ac.uk/search\\_nvs/P01/](https://vocab.nerc.ac.uk/search_nvs/P01/)
- National Environment Research Council (NERC) Vocabulary P06:  
[https://vocab.nerc.ac.uk/search\\_nvs/P06/](https://vocab.nerc.ac.uk/search_nvs/P06/)
- National Environment Research Council (NERC) Vocabulary P09:  
<http://vocab.nerc.ac.uk/collection/P09/current/>

### 3.1.4 Quality Control/Quality Assurance

#### *Near-Real Time QC*

The European common QC, data and metadata model for NRT HFR current data (Corgnati et al., 2018) requires both NRT radial and total data to be mandatorily processed by a battery of specific Quality Control (QC) tests. These tests were selected (and modified when needed) by the dedicated working group (composed by the HFR operators and by the EuroGOOS HFR Task Team members) among the ones defined in the IOOS QARTOD Manual for Real-Time Quality Control of High Frequency Radar Surface Current Data, according to the defined hierarchy.

These mandatory QC tests are manufacturer-independent, i.e. they do not rely on particular variables or information provided only by a specific device.

These standard sets of tests were defined both for radial and total velocity data and they are the required ones for labelling the data as Level 2B (for radial velocity) and Level 3B (for total velocity) data. **Table 3.1.4.1** reports the processing level definitions.

Processing Level	Definition	Products
LEVEL 0	Reconstructed, unprocessed instrument/payload data at full resolution; any and all communications artifacts, e.g. synchronization frames, communications headers, duplicate data removed.	Signal received by the antenna before the processing stage. (No access to these data in Codar systems)
LEVEL 1A	Reconstructed, unprocessed instrument data at full resolution, time-referenced and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing.	Spectra by antenna channel
LEVEL 1B	Level 1A data that have been processed to sensor units for next processing steps. Not all instruments will have data equivalent to Level 1B.	Spectra by beam direction
LEVEL 2A	Derived geophysical variables at the same resolution and locations as the Level 1 source data.	HFR radial velocity data
LEVEL 2B	Level 2A data that have been processed with a minimum set of QC.	HFR radial velocity data
LEVEL 2C	Level 2A data that have been reprocessed for advanced QC.	Reprocessed HFR radial velocity data
LEVEL 3A	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency	HFR total velocity data
LEVEL 3B	Level 3A data that have been processed with a minimum set of QC.	HFR total velocity data
LEVEL 3C	Level 3A data that have been reprocessed for advanced QC.	Reprocessed HFR total velocity data
LEVEL 4	Model output or results from analyses of lower level data, e.g. variables derived from multiple measurements	Energy density maps, residence times, etc.

**Table 3.1.4.1.** Processing Levels for HFR data.

The mandatory QC tests for radial velocity data are listed and described in **Table 3.1.4.2**.

QC test	Meaning	QC variable type
Syntax	<p>This test will ensure the proper formatting and the existence of all the necessary fields within the radial netCDF file.</p> <p>This test is performed on the netCDF files and it assesses the presence and correctness of all data and attribute fields and the correct syntax throughout the file.</p>	N/A, it is a test on the netCDF file structure, not on data content.
Over-water	<p>This test labels radial vectors that lie on land with a "bad_data" flag and radial vectors that lie on water with a "good_data" flag.</p>	gridded
Velocity Threshold	<p>This test labels radial velocity vectors whose module is bigger than a maximum velocity threshold with a "bad_data" flag and radial vectors whose module is smaller than the threshold with a "good_data" flag.</p>	gridded
Variance Threshold	<p>This test labels radial vectors whose temporal variance is bigger than a maximum threshold with a "bad_data" flag and radial vectors whose temporal variance is smaller than the threshold with a "good_data" flag.</p> <p>The Codar manufacturer suggests not to use variance data for real-time QC, as documented in the fall 2013 CODAR Currents Newsletter. The indication is due to the fact that the CODAR parameter defining the variance is computed at each time step, and therefore considered not statistically solid.</p> <p>Thus, this test is applicable only to Beam Forming (BF) systems. Data files from Direction Finding (DF) systems will apply instead the "Temporal Derivative" test reporting the explanation "Test not applicable to Direction Finding systems. The Temporal Derivative test is applied." in the comment attribute.</p>	gridded
Temporal Derivative	<p>For each radial bin, the current hour velocity vector is compared with the previous and next hour ones. If the differences are bigger than a threshold (specific for each radial bin and evaluated on the basis of the analysis of a one-year-long time series), the present vector is flagged as "bad_data", otherwise it is labelled with a "good_data" flag.</p> <p>Since this method implies a one-hour delay in the data provision, the current hour file should have the related QC flag set to 0 (no QC performed) until it is updated to the proper values when the next hour file is generated.</p>	gridded

Median Filter	For each source vector, the median of all velocities within a radius of <RCLim> and whose vector bearing (angle of arrival at site) is also within an angular distance of <AngLim> degrees from the source vector's bearing is evaluated. If the difference between the vector's velocity and the median velocity is greater than a threshold, then the vector is labelled with a "bad_data" flag, otherwise it is labelled with a "good_data" flag.	gridded
Average Radial Bearing	This test labels the entire datafile with a 'good_data' flag if the average radial bearing of all the vectors contained in the data file lies within a specified margin around the expected value of normal operation. Otherwise, the data file is labeled with a "bad_data" flag. The value of normal operation has to be defined within a time interval when the proper functioning of the device is assessed. The margin has to be set according to site-specific properties. This test is applicable only to DF systems. Data files from BF systems will have this variable filled with "good_data" flags (1) and the explanation "Test not applicable to Beam Forming systems" in the comment attribute.	scalar
Radial Count	Test labelling radial data having a number of velocity vectors bigger than the threshold with a "good_data" flag and radial data having a number of velocity vectors smaller than the threshold with a "bad_data" flag.	scalar

**Table 3.1.4.2.** Mandatory QC tests for HFR radial data.

The mandatory QC tests for total velocity data are listed and described in **Table 3.1.4.3.**

QC test	Meaning	QC variable type
Syntax	This test will ensure the proper formatting and the existence of all the necessary fields within the total netCDF file. This test is performed on the netCDF files and it assesses the presence and correctness of all data and attribute fields and the correct syntax throughout the file.	N/A, it is a test on the netCDF file structure, not on data content.
Data Density Threshold	This test labels total velocity vectors with a number of contributing radials bigger than the threshold with a "good_data" flag and total velocity vectors with a number of contributing radials smaller than the threshold with a "bad_data" flag.	gridded
Velocity Threshold	This test labels total velocity vectors whose module is bigger than a maximum velocity threshold with a "bad_data" flag and total	gridded

	vectors whose module is smaller than the threshold with a "good_data" flag.	
Variance Threshold	<p>This test labels total vectors whose temporal variance is bigger than a maximum threshold with a "bad_data" flag and total vectors whose temporal variance is smaller than the threshold with a "good_data" flag.</p> <p>The Codar manufacturer suggests not to use variance data for real-time QC, as documented in the fall 2013 CODAR Currents Newsletter. The indication is due to the fact that the CODAR parameter defining the variance is computed at each time step, and therefore considered not statistically solid.</p> <p>Thus, this test is applicable only to Beam Forming (BF) systems. Data files from Direction Finding (DF) systems will apply instead the "Temporal Derivative" test reporting the explanation "Test not applicable to Direction Finding systems. The Temporal Derivative test is applied." in the comment attribute.</p>	gridded
Temporal Derivative	<p>For each grid cell, the current hour velocity vector is compared with the previous and next ones. If the differences are bigger than a threshold (specific for each grid cell and evaluated on the basis of the analysis of a one-year-long time series), the present vector is flagged as "bad_data", otherwise it is labelled with a "good_data" flag.</p> <p>Since this method implies a one-hour delay in the data provision, the current hour file should have the related QC flag set to 0 (no QC performed) until it is updated to the proper values when the next hour file is generated.</p>	gridded
GDOP Threshold	This test labels total velocity vectors whose GDOP is bigger than a maximum threshold with a "bad_data" flag and the vectors whose GDOP is smaller than the threshold with a "good_data" flag.	gridded

**Table 3.1.4.3.** Mandatory QC tests for HFR total data.

Each QC test results in a flag related to each data vector, which is inserted in the specific test variable. These variables can be matrices with the same dimensions of the target data variable, containing, for each cell, the flag related to the vector lying in that cell, in case the QC test evaluates each cell of the gridded data, or a scalar, in case the QC test assesses an overall property of the data.

An overall QC variable reports the quality flags related to the results of all the QC tests: it is a "good data" flag if and only if all QC tests are passed by the data.

The flagging policy is not to modify the data, but only to label them with flags. Thus, each geophysical variable in the standard output files contains exactly the measured data, and the QC variables containing flags can be used as masks to the geophysical variables for having information about data quality.

The adopted QC flagging scheme is the ARGO QC flag scale (which extends the UNESCO scale), as reported in **Table 3.1.4.4**.

**Table 3.1.4.4.** ARGO Quality Control flag scale.

Code	Meaning	Comment
0	unknown	No QC was performed
1	good data	All QC tests passed
2	probably good data	These data should be used with caution
3	potentially correctable bad data	These data are not to be used without scientific correction or re-calibration
4	bad data	Data have failed one or more QC tests
5	value changed	Data may be recovered after transmission error
6	value below detection	Data value is below detection limit
7	nominal value	The provided value is not measured but comes from a priori knowledge (instrument design or deployment), e.g. instrument target depth
8	interpolated value	Missing data may be interpolated from neighboring data in space or time
9	missing value	

The flags used for NRT HFR data QC are 0: no QC performed; 1: good data; 4 bad data.

For some of these tests, HFR operators need to select the best thresholds. Since a successful QC effort is highly dependent upon selection of the proper thresholds, this choice is not straightforward, and may require trial and error before final selections are made. These thresholds should not be determined arbitrarily, but based on historical knowledge or statistics derived from historical data.

The threshold values are reported in the *comment* variable attribute of each QC variable. The flagging scheme is reported as well in the *flag\_values* and *flag\_meanings* variable attributes of each QC variable.

The EU HFR Node produces and distributes NRT HFR current data compliant with the European common QC, data and metadata format for NRT HFR data, thus all the NRT HFR data files generated by the EU HFR Node are processed with the aforementioned QC tests.

#### Delayed Mode QC

Delayed Mode (DM) HFR current data must be mandatorily processed with the NRT QC tests. Furthermore, DM HFR data can be processed with additional Advanced Quality Control (AQC) procedures. Following the HFR Node procedures for the production of DM HFR current data, it is

recommended that the Historical HFR data time series are screened by means of yearly plots in order to allow the inspection of human experts for assessing the effective quality of HFR data. The following plots by year and system shall be produced:

- temporal series of the spatial average of the current velocity module, its standard deviation and the total spatial coverage
- temporal series of the QC flags for all the grid nodes with data
- maps of the mean value of QC flags for the target year
- maps of mean velocity module and its standard deviation for the target year
- spatial (x-axis) vs. temporal (y-axis) coverage 80/80 annual metric (this allows to check if the system has reached the goal of providing surface currents over the 80% of the area during 80% of the time)
- maps of the mean velocity field in the area of 80% temporal coverage

Based on the data screening and the analysis of the plots listed above, the HFR Node generates reports on the HFR data time series quality per network/sites, where the performance is analyzed year by year and periods for reflagging (expert but subjective analysis) are proposed. Reflagging introduces the use of flags 2: probably good data; 3: potentially correctable bad data, until the data can be corrected or is finally discarded.

In addition, possible changes in the processing of the data (namely in the thresholding strategy) are proposed.

The EU HFR Node produces and distributes DM HFR current data compliant with the European common QC, data and metadata format for HFR data, thus all the DM HFR data files generated by the EU HFR Node are processed with the aforementioned AQC procedures. In particular, the EU HFR Node performs the data screening, produces the plots and the reports. Then the proposition for reflagging/reprocessing is divided in two stages, in close collaboration with the data providers:

- in a first step, the report is sent to the provider for its validation and agreement or feedback on the comments and the reflagging/reprocessing proposed.
- Then, after the provider's feedback, changes in the original data series (reflagging or reprocessing) are performed. A final version of the report is produced.

Reflagging is only performed if the provider validates the reflagging proposition. If no reflagging is performed, the DM data contains the same information of the NRT data.

The DM AQC procedures are described in the CMEMS-INSTAC Quality Information Document (Copernicus Marine In Situ Tac, 2020).

### 3.1.5 Issues

The main issue in HFR data management is the **lack of a unique standard internationally adopted**. US IOOS HFR network and Australian IMOS HFR network, and, in general, each national HFR network outside Europe, apply their own strategy for QC and data models. In Europe three main QC, data and metadata standards exist, i.e. the European common QC, data and metadata model for NRT HFR data, its declinations for distribution towards CMEMS-INSTAC and SDC Data Access data portals, even if they only slightly differ in minor elements. Anyway, this could generate confusion in the data providers and the data users.

- Under the leadership of OceanOPS, the contribution of the European HFR community is crucial towards the achievement of the metadata standardization and integration across the global ocean observing networks, as one of OceanOPS's five goals.

- Currently, representatives of the European HFR Node are participating in the revisiting of the QARTOD manual (U.S. Integrated Ocean Observing System, 2016) with a plan to update it, taking part also of a working group to review, organize and recommend HFR best practice documents for a GOOS endorsed best practice list, together with US colleagues.

Another vulnerability in HFR data management is the **absence of a common data policy** for HFR data. This should be achieved with priority, in order to enhance the HFR data stream and foster the use of HFR data in societal and scientific applications.

- Giving due credit to all those contributing in the development of new datasets is imperative to incentivize the data publication. Different options listed in Tanhua et al., 2019 are: i) the use of data citation tools, such as DOIs (Digital Object Identifiers) or PID (Persistent Identifiers for Data and/or products) attached to the platform (e.g. WMO number for Argo, etc.); ii) the inclusion of metadata records identifying the data provider; iii) the automation of data citation (e.g. ORCID); iv) the definition of consistent ways to count data usage (e.g. MakeDataCount project).
- In the case of HFR network, different metadata (e.g. EDMO, EDIOS and CDI entries) are included referring to the data provider, as mentioned above. However, the definition of the DOI minting strategy for European HFR data (like the one established for Argo profilers) seeking the convergence in the granularity level, versioning, metadata scheme, citation tracking method, information to be included in the landing page of the DOI, etc. as well as the selection of a FAIR-aligned open repository and the implementation of tools to automate data citations, are still outstanding issues in the HFR community.

**Double data distribution** from the data provider in non-standard format and from the main European data portals in standard format can also create confusion to the data users, since they often ask which data should they finally use.

- Although metadata and identifiers may benefit double-distributed data traceability, harmonization of data processing and distribution in standard format should be prioritized in a coordinated way (i.e. adopting the European HFR node as focal point for a common data delivery approach) to avoid duplication and heterogeneity, seeking for interoperability to serve both, the data exchanges between networks and user-friendly tools (Tanhua et al., 2019).

**Slow but continuous progress of the adoption of the standards**, since only 40% of the 69 ongoing HFR sites in Europe are synchronized with the European HFR node in near real-time (Rubio et al., 2021). As mentioned by Tanhua et al. (2019), the implementation of the standards can often be beyond the technical capabilities of many scientific communities and, if not, these time-demanding tasks are usually carried out on a voluntary basis from the data provider, being typically funded by science activities. Fortunately, funding and allocation of costs for data management are now encouraged, being no longer an afterthought.

- Leveraging existing software tools, as those listed in section 3.1.6, together with offering support (i.e. by setting up a helpdesk in the European HFR Node), training and outreach (i.e. by webinars and the development of a website explaining the European HFR node role, the responsibilities, the benefits of a centralized standardization, the return to the data providers, etc) to the teams, can help HFR community significantly increase their level of data interoperability with a minimum of resources.
- In addition, the direct provision of the data in the accepted standards could be an added-value option on the side of the manufacturers (Tanhua et al., 2019).

Mainly due to its recent creation in December 2018 (Rubio et al., 2021), the European HFR node still **lacks a well-founded data management plan** and **regular long-term financial** support.

- The development and the publication of a well-established and standardized European HFR node data management plan, evolving throughout the life cycle of the HFR data would benefit the FAIRness of the data. Furthermore, this DMP will be available for all data providers of the European HFR node to be provided for European projects, when requested.
- Effective coordination of the European HFR network and the further establishment and implementation of a robust and sustained governance structure and framework, as the one recently proposed by (Rubio et al., 2021), will contribute to support the human and infrastructure resources required to deliver the HFR long-term strategy.

Finally, a minor vulnerability consists in the **absence of a standard procedure for determining the thresholds** of the QC tests. The definition of such a procedure, based on strong scientific foundations, would greatly enhance the effectiveness of the standardized HFR data management.

- Variability of the ocean circulation and, particularly extreme events, often lead to data anomalies. By using novel methods based on deep learning the short-term trend of ocean observation data can be predicted and the error threshold can be better set.

### 3.1.6 Training materials and Contacts

The core of the services provided by the EU HFR Node consists in the continuous development of the data model and the processing standards through discussion with operators, providers, distributors and international experts. Based on this, the EU HFR Node maintains and updates manuals, procedure guidelines and software tools, and pushes them towards the HFR operators, providers and managers via repositories and training workshops.

The documentation and the training material related to the implementation of the workflow for HFR current data standardization is reported in **Table 3.1.6.1**.

**Table 3.1.6.1.** Documentation and the training material related to the implementation of the workflow for HFR current data standardization.

Documentation / Training material	Link
Matlab processing software for generation of standardized HFR data	<a href="https://cnrsc-my.sharepoint.com/:f/g/personal/lorenzo_corgnati_cnr_it/EnOTpz7b8H1Am9RGmlwWx8oBtF8uLWkh1ACy9ho7sM3Mkg?e=DC4HtU">https://cnrsc-my.sharepoint.com/:f/g/personal/lorenzo_corgnati_cnr_it/EnOTpz7b8H1Am9RGmlwWx8oBtF8uLWkh1ACy9ho7sM3Mkg?e=DC4HtU</a>
EU HFR Node software solutions	<a href="https://cnrsc-my.sharepoint.com/:f/g/personal/lorenzo_corgnati_cnr_it/EtlyL9_9XRxAAtBC2j1o99csBrL-hFekUWn41oZ42F_Vgaw?e=uwM5dw">https://cnrsc-my.sharepoint.com/:f/g/personal/lorenzo_corgnati_cnr_it/EtlyL9_9XRxAAtBC2j1o99csBrL-hFekUWn41oZ42F_Vgaw?e=uwM5dw</a>
EU HFR Node software dependencies	<a href="https://cnrsc-my.sharepoint.com/:f/g/personal/lorenzo_corgnati_cnr_it/ErCNy-LX6L9Pv-">https://cnrsc-my.sharepoint.com/:f/g/personal/lorenzo_corgnati_cnr_it/ErCNy-LX6L9Pv-</a>

	<a href="https://doi.org/10.25704/9XPF-76G7">L25vT8b8QBAxqN5XCI2e1OF7sIKE0MqQ?e=DA8zpl</a>
EU HFR Node webform for insertion of information and metadata related to HFR networks and stations	<a href="https://cnrsc-my.sharepoint.com/:f:/g/personal/lorenzo_corgnati_cnr_it/EotpWERJVCRGryPEkjddf7kBmJ0fqXJ1LMhZyKYEZQgug?e=bG4dTB">https://cnrsc-my.sharepoint.com/:f:/g/personal/lorenzo_corgnati_cnr_it/EotpWERJVCRGryPEkjddf7kBmJ0fqXJ1LMhZyKYEZQgug?e=bG4dTB</a>
SeaDataNet metadata for discovery	<a href="https://cnrsc-my.sharepoint.com/:f:/g/personal/lorenzo_corgnati_cnr_it/Ev4JXQHI85IEimSRUrFj7YBZnX50jklZR-rsPGcPL--g?e=tnQlkO">https://cnrsc-my.sharepoint.com/:f:/g/personal/lorenzo_corgnati_cnr_it/Ev4JXQHI85IEimSRUrFj7YBZnX50jklZR-rsPGcPL--g?e=tnQlkO</a>
Guidelines on how to sync HFR radar data from data providers to the EU HFR Node	<a href="https://doi.org/10.25704/9XPF-76G7">doi:10.25704/9XPF-76G7</a>

The software tools for processing native HFR data for QC and converting them to the standard format for distribution are continuously made available to HFR operators via public GitHub repositories and releases with DOI assigned, for guaranteeing the complete traceability of the processing chain. **Table 3.1.6.2** reports the software tools available for data providers who want to locally perform the HFR data standardization.

**Table 3.1.6.2.** Software tools available for data providers who want to locally perform the HFR data standardization.

Software package	Link / DOI
Matlab package for HFR data standardization to be locally run by data providers	<a href="https://doi.org/10.5281/zenodo.2639555">DOI:10.5281/zenodo.2639555</a>
Add-on containing all dependencies needed by the Matlab package for HFR data standardization	<a href="https://cnrsc-my.sharepoint.com/:u:/g/personal/lorenzo_corgnati_cnr_it/ETqXtGwvyOpAq2a1a_mtiisBpQgGtwuuaO4mAfJIQprfQA?e=RL6QnK">https://cnrsc-my.sharepoint.com/:u:/g/personal/lorenzo_corgnati_cnr_it/ETqXtGwvyOpAq2a1a_mtiisBpQgGtwuuaO4mAfJIQprfQA?e=RL6QnK</a>
JRadar Java Tool for an easy conversion of HFR radial and total CODAR files into compliant HFR standard netcdf files	<a href="https://doi.org/10.5281/zenodo.5081995">DOI:10.5281/zenodo.5081995</a> <a href="https://github.com/Fundacion-AZTI/JRadar">https://github.com/Fundacion-AZTI/JRadar</a>

The Jupyter Notebooks for HFR data analysis developed in the context of CMEMS training actions and other tools developed for obtaining HFR added-value data (i.e. gap-filled data) are also available, as reported in **Table 3.1.6.3**.

**Table 3.1.6.3.** Software tools available for HFR data analysis.

Software package	Link / DOI
Copernicus Marine In Site Arctic Training	<a href="https://mybinder.org/v2/gh/CopernicusMarineInsitu/2020-ARC-TrainingWorkshop/dde7e08a502cc4ef92bd3ee0eda10e1c570cc4f2?filepath=13-05-NearRealTime-product-managing-files-hfradars.ipynb">https://mybinder.org/v2/gh/CopernicusMarineInsitu/2020-ARC-TrainingWorkshop/dde7e08a502cc4ef92bd3ee0eda10e1c570cc4f2?filepath=13-05-NearRealTime-product-managing-files-hfradars.ipynb</a>
Video Tutorial -Arctic Ocean – In Situ data: Managing In Situ data from HFR radars	<a href="https://www.youtube.com/watch?v=blqGOa_99QE">https://www.youtube.com/watch?v=blqGOa_99QE</a>
DIVAnd FREE TOOL developed by GHER (University of Liège) for the interpolation of HFR data available on github	<a href="https://github.com/gher-ulg/DIVAnd_HFRadar.jl">https://github.com/gher-ulg/DIVAnd_HFRadar.jl</a>
HFR data visualization routines in python	<a href="https://www.seanoe.org/data/00697/80874/">https://www.seanoe.org/data/00697/80874/</a> <a href="https://github.com/CopernicusMarineInsitu/2021-sicomar-summer-school">https://github.com/CopernicusMarineInsitu/2021-sicomar-summer-school</a>

For any further information, other requests or needs, please contact the EU HFR Node by mailing to [eufrnode@azti.es](mailto:eufrnode@azti.es).

### 3.1.7 Examples of HFR Data Management Plan

The examples of data management plans (DMPs) reported in **Table 3.1.7.1** are from diverse HFR providers. The plans are integrated documents that describe how the data and operations are handled, also providing an overview of the responsibilities and roles from the different actors involved through the value chain.

**Table 3.1.7.1.** Examples of Data Management Plans for HFR data.

HFR Data Management Plan	Link/DOI
SOCIB - Coastal High Frequency Radar Data Management Plan (Version 1.0) (Marasco et al., 2021)	<a href="https://doi.org/10.25704/ydas-qz53">https://doi.org/10.25704/ydas-qz53</a>
NANOOS - Northwest Association of Networked Ocean Observing Systems	<a href="http://www.nanoos.org/documents/certification/DMP/1.DMP.HFRadar.pdf">http://www.nanoos.org/documents/certification/DMP/1.DMP.HFRadar.pdf</a>
IMOS - Integrated Marine Observing System	<a href="https://s3-ap-southeast-2.amazonaws.com/content.aodn.org.au/Docume">https://s3-ap-southeast-2.amazonaws.com/content.aodn.org.au/Docume</a>

	<a href="https://www.imos.fr/Workflows/OceanRadar_CODAR_Workflow.pdf">nts/IMOS/Workflows/OceanRadar_CODAR_Workflow.pdf</a>
Harlan, J., Terrill, E., & Crout, R. (2008). NOAA IOOS national HF radar network data management: Status and Plans. <i>2008 IEEE/OES 9th Working Conference on Current Measurement Technology</i> , 156–159.	<a href="https://doi.org/10.1109/CCM.2008.4480860">https://doi.org/10.1109/CCM.2008.4480860</a>

## 3.2 Glider

### 3.2.1 Description of data

Ocean Gliders (hereafter gliders) are autonomous vehicles that can operate for long periods of time underwater. They can collect a large variety of physical and biogeochemical data at high temporal and spatial resolutions. The main Essential Ocean Variables (EOV) from the gliders are temperature, conductivity (converted to salinity), pressure (converted to depth), depth average current, oxygen, chlorophyll fluoresce and backscatter. Depending on the platform and the needs, other low-energy sensors can be added depending on the scientific needs.

Gliders navigate autonomously by using their buoyancy to ascend and descend the water columns between the surface and the maximum depth depending on how they have been programmed to perform. Their wings allow them to move forward along a trajectory. Therefore, data generated by these platforms consist of sawtooth-like vertical profiles at various locations in a specific region. The measured variables depend on the payload sensors installed in the glider before each deployment. Moreover, these measurements may vary along with the mission because their sensor configuration can be changed remotely using Iridium satellite communication (**Figure 3.2.1.1**). This periodic remote communication also allows obtaining quasi-real-time data each time the glider surfaces. A subset of the measured data can be sent, considering the transmission of the satellite cost, surface time, and energy consumption (**Figure 3.2.1.2**). Additionally, the complete data set can be downloaded and collected after the glider is recovered in the laboratory. The remote configuration provides a way to appropriately navigate the glider and optimize energy consumption by modifying the sampling frequency and switching sensors where necessary. These changes could have an impact on the output data.

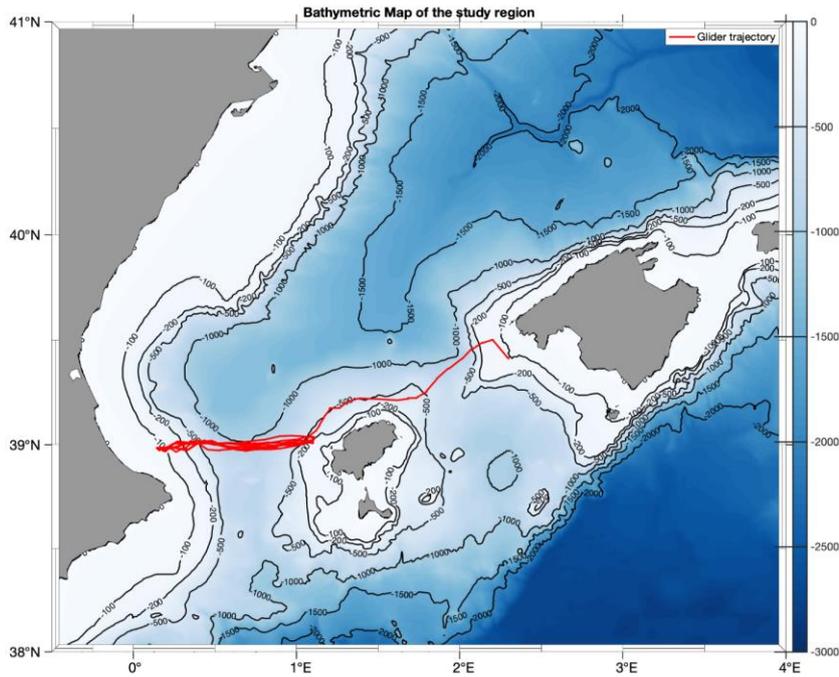


Figure 3.2.1.1. A glider trajectory is plotted with red over the bathymetry.

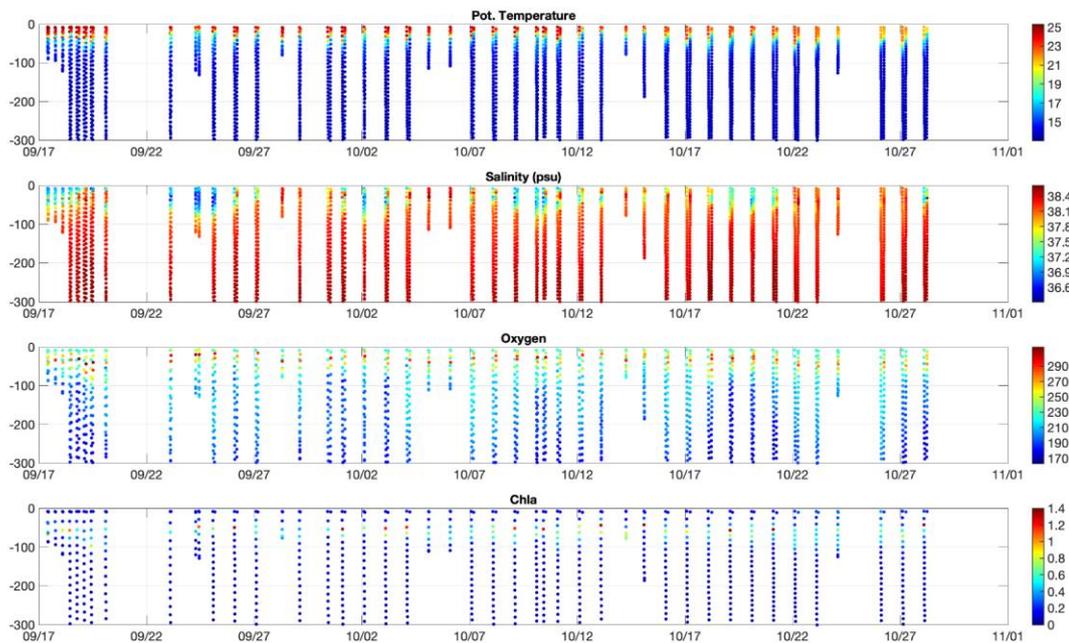
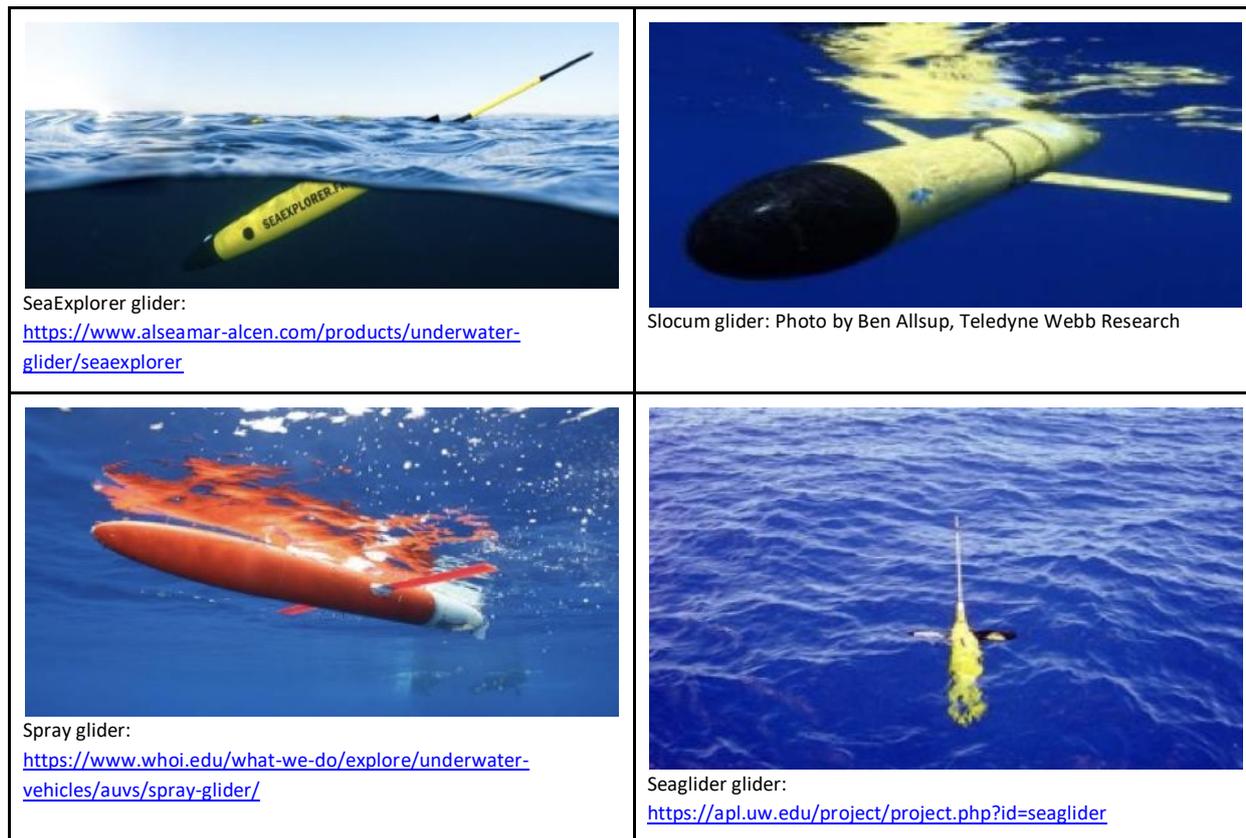


Figure 3.2.1.2. Real time plot of temperature, salinity, oxygen and chlorophyll fluorescence during the mission

The four most commonly used glider platforms presently available are: Slocum, SeaExplorer, Spray, and Seaglider (Figure 3.2.1.3), although there are other types. Gliders are flexible to host a variety of physical and biogeochemical sensors that may provide a diversity of variables. Each of them produces different format types of raw data. Moreover, the raw data also contains engineering variables that are used to operate, control and navigate the glider and to support delayed mode quality control.

Engineering data depends too on the type of platform, adding complexity to the description of raw glider data.



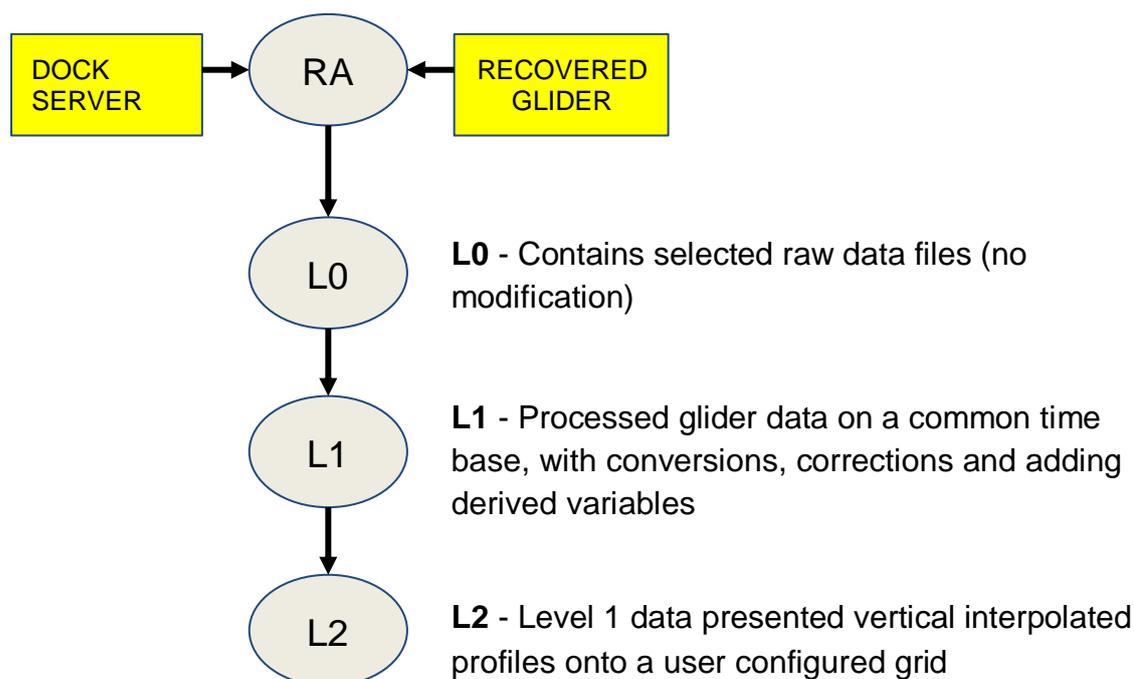
**Figure 3.2.1.3.** Major types of gliders

In summary, the complexity of the glider data originates from:

- Diversity of physical and biogeochemical sensors
- Diversity of data for glider formats
- Remote glider and sensor configuration during mission
- Diversity of data: real time, recovery and delayed mode
- Coexistence of the different science and engineering parameters

There are various approaches leading to processed data from glider observations. One good practice that considers the heterogeneous nature of the glider data is to parse the raw data and create three levels (L0, L1, and L2) of standard netCDF files (see **Figure 3.2.1.4**). The L0 level contains both scientific and engineering parameters sent from the glider. The L1 level files contain observations, calibrations, unit corrections, and derived variables such as salinity. It also includes delayed mode corrections like the thermal lag (Garau et al., 2011) or salinity cross-calibration (Allen et al., 2020). The L2 level files are gridded observations from the glider profiles. This profile-like data can be helpful for modeling applications and comparing the information with Argo profile data. Historically, the format of these files is homogeneous within each institution. However, substantial efforts have been made to harmonize the format at the international level. The Everyone's Gliding Observatories (EGO) defined the EGO glider NetCDF standard (Thierry et al., 2021), aiming to harmonize the L1 data. Later, the OceanGliders (OG) program created the OG1.0 format to support interoperability within international

standards in the USA, Australia, and Europe. This standard aims to support FAIR principles and to strengthen the network community.



**Figure 3.2.1.4.** Processing levels of glider data

In general, glider data management must align with the OG principles and the best practices and adopt their recommendations and formats to ensure interoperability with the international community and the long-term durability of JERICO glider data.

### 3.2.2 Data value chain: from acquisition to delivery

#### *Brief description of sensors*

Ocean gliders represent a technological revolution because they require very little human assistance and therefore are suited to safely collecting data in local and remote locations at a relatively low cost. Additionally, they allow sampling of the ocean where it is impractical for human access, such as in the middle of a hurricane or under sea ice. Due to these attractive properties, a large diversity of scientific sensors has been developed during the past decades (Testor et al., 2019). Gliders provide engineering information that informs of the status of the glider and the trajectory, as well as some basic physical information such as pressure, temperature, and currents. It is not easy to make a comprehensive list of the available sensors that have been used and installed in gliders. Different glider platforms (Slocum, Seaglid, Spray, and SeaExplorer) provide a high diversity on the available sensors carried out with their specific features for each sensor.

From a general perspective, glider manufacturers focused initially on measuring essential physical variables such as pressure, temperature, and salinity. These sensors have been improved over the years and are very reliable. Other physical variables were incorporated and elaborated over time as currents and turbulence (Cauchy et al., 2018). Due to the price and ability of gliders to reach remote locations, manufacturers developed a full range of sensors to perform biogeochemical measurements

relative to the marine ecosystem's biogeochemical processes. In particular, sensors have been developed to measure chlorophyll fluorescence, turbidity, light backscatter at different wavelengths (most common 460, 532, 650, and 880 nm), dissolved oxygen, CDOM, irradiance, nitrate, and recently acoustic sensors (acoustic backscatter or ambient sound) that have become popular for fisheries and marine diversity research. There is also the possibility of mounting and integrating new commercial and custom sensors that enhance their potential and promising future. There is also the possibility of mounting and integrating new commercial and custom sensors that enhance their potential and promising future.

The glider community has developed a set of Standard Operating Procedures (SOP) for the most mature sensors in the framework of the OceanGliders program. These SOPs contain a more detailed list of the commonly used sensors based on the measured variable (see section 3.2.3).

### *Methodologies used for data collection*

Gliders are long-endurance autonomous vehicles that move through the water column along a predefined path to respond to a specific scientific goal. Data collection procedures start at the deployment preparation stage when making the observational and operational plan. The data collection flow includes the following steps:

- Planning
- Deployment
- Data transmission
- Data processing
- Data archiving
- Data dissemination and visualization
- Data distribution

The Eurogoos Glider Task Team (<https://eurogoos.eu/gliders-task-team/>) defined the standard data flow of glider data, ensuring in particular their dissemination and long-term persistence. Usually, the glider operators define the planning of a glider mission, the deployment of the glider, and the different phases of the transmission of the glider observations. Data processing, archiving, dissemination and visualization is performed by the Data Assembly Centers (DAC), which plays an essential role in the dissemination process through the Global Data Assembly Center (GDAC). Some organizations have the capabilities to operate gliders and assemble the data as a DAC. Otherwise, a network of DAC is available to allow the data flow to the GDAC. Coriolis is the European GDAC, and it ensures that the glider data flow in the required format to the international infrastructures, such as GTS for use in Numerical Weather Prediction and ocean forecasting. DAC must guarantee that the data is provided in the standard format: EGO or OG for international interoperability with other GDAC. Coriolis guarantees that the data flows to the European data aggregators such as SeaDataNet, CMEMS Instac, and EMODnet. **Figure 3.2.2.1** is the specific case of the Data Management Plan (DMP) of SOCIB (Marasco et al., 2021) as an example of the possible flow of glider data. SOCIB is a specific case of an organization that operates gliders and acts as a DAC for glider data, including external organizations internationally. Glider users should comply with the OG Data management plan for a more general data flow.

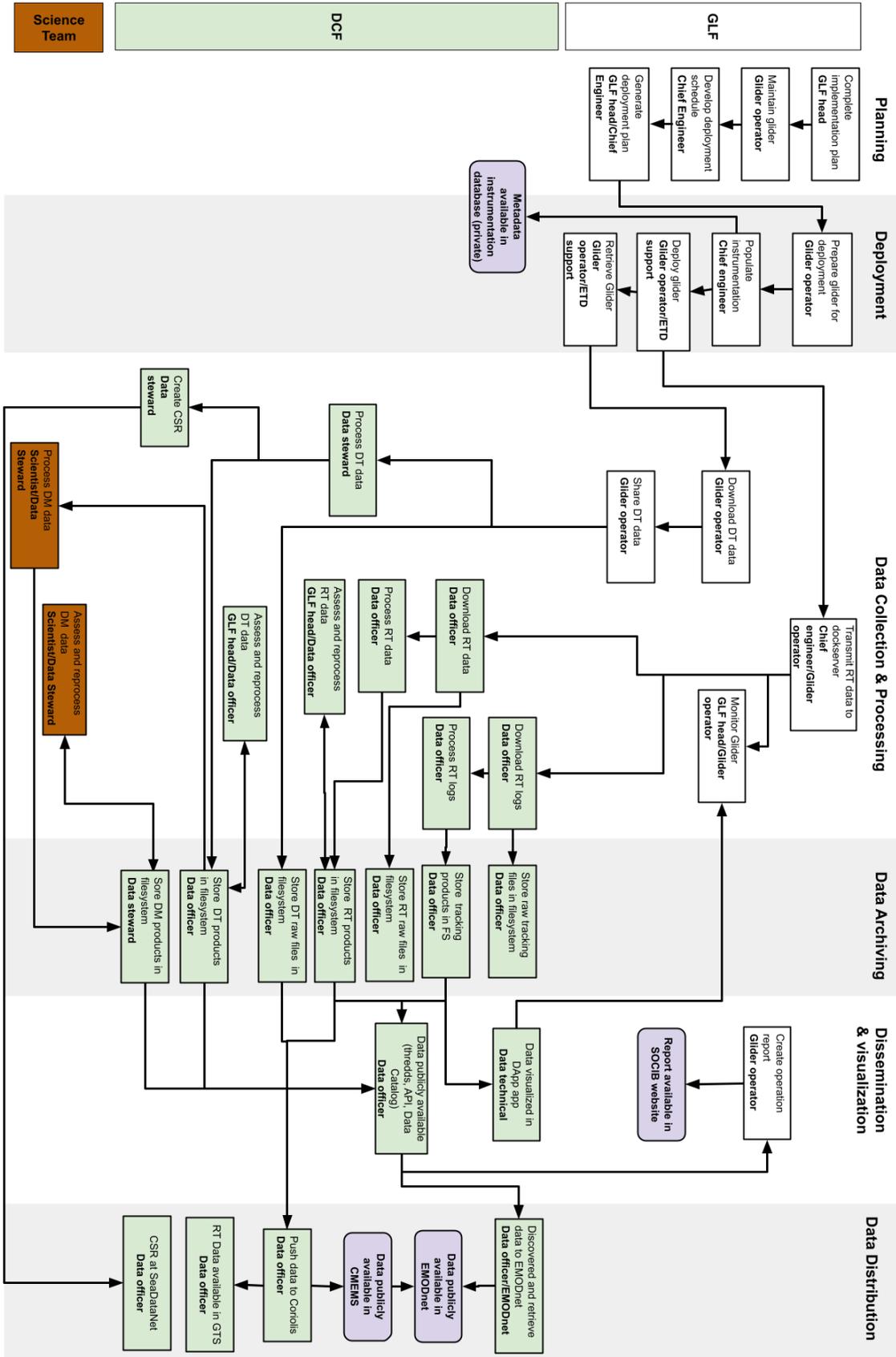
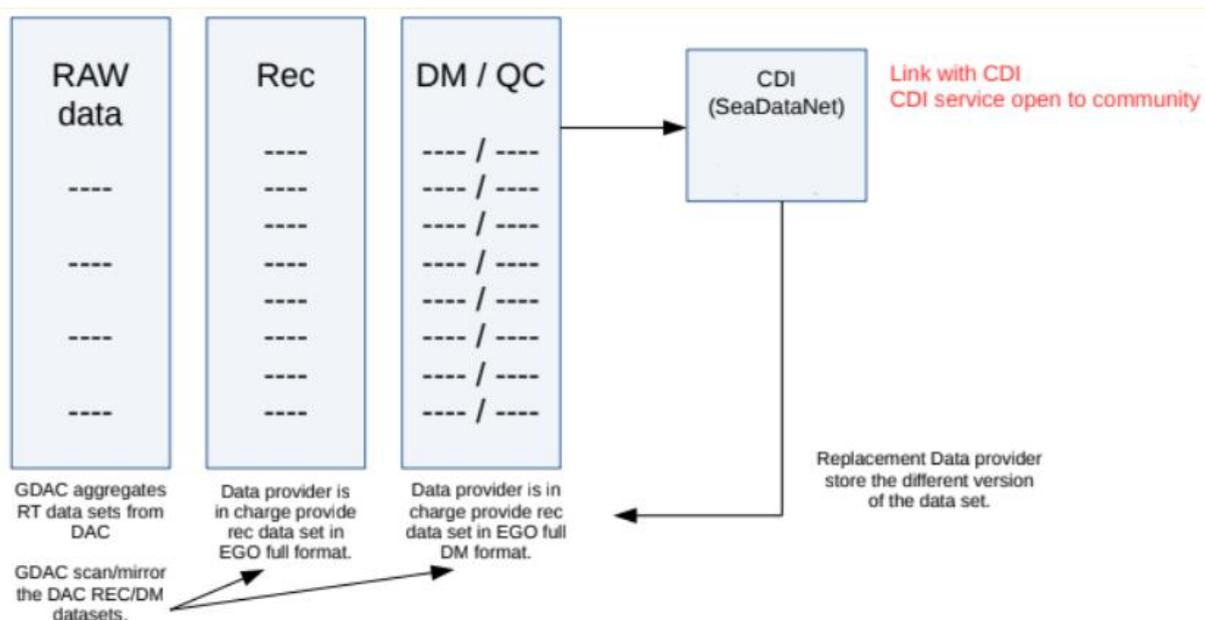


Figure 3.2.2.1. SOCIB Data Management Plan (Marasco et al., 2021) as an example of data flow from operation to data collection and distribution from DAC to GDAC and Aggregators.

We have to differentiate between near-real-time (NRT), recovery, and delay mode data in the data flow, as **Figure 3.2.2.2** indicates. Real-time data is transmitted via satellite to the glider operator server. Basic QC is applied to this data by GDAC or possibly (but not mandatorily) by DAC. DAC needs to promptly provide real-time data to GDAC to be used by GTS in the forecast models. Real-time data is only a subset of the total amount of measured data. The complete set of data is collected after the recovery of the glider. Similar to real-time, recovery data may contain QC, and they are provided to DAC/GDAC soon after collection. Delayed mode data is usually produced with human intervention. It contains corrected data such as thermal lag, salinity correction, or data interpolation. Additionally, it must contain the QC of each of the variables. Various versions of the dataset can be provided to GDAC at different times as the correction process evolves.

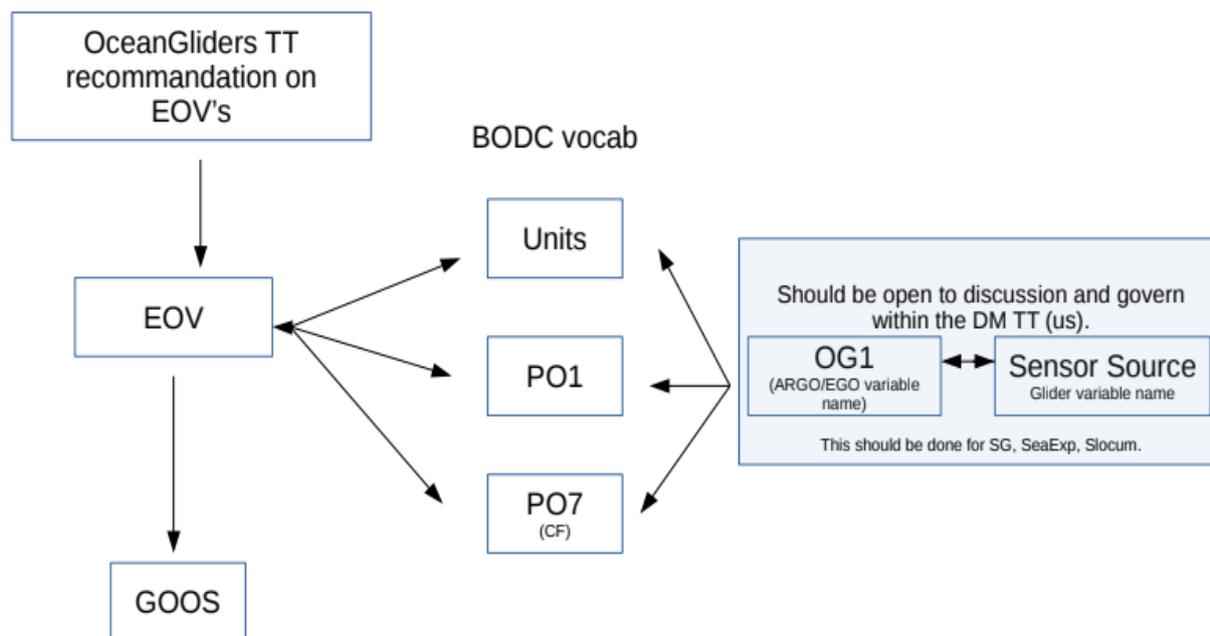


**Figure 3.2.2.2.** Data distribution strategies for real time, recovery data and delayed mode glider data.

### Metadata information and data format

As mentioned earlier, due to the diversity of sensors and platforms, metadata and formats of raw data from gliders are heterogeneous. Furthermore, the metadata for each glider is complex because it includes scientific, engineering and sometimes custom variables of the manufacturer. As an example, John Kerfoot at Rutgers University published the [list of Slocum metadata](#) [20].

On the contrary, OceanGliders aims to harmonize the L1 level products for real-time, recovery, and delayed mode data. It works toward global interoperability between standards of different continents in order to reach FAIR principles adopted by the Global Ocean Observing System (GOOS). The data model is compliant with the Climate and Forecast Metadata Convention CF-1.8 specifications. Data is recorded as a trajectory discrete geometry where each data file contains a series of profiles that represent the entire mission of the glider. The format follows the Attribute Conventions for Data Discovery (ACDD) 1.3 convention. The vocabulary collections are hosted in different places (i.e., NERC Vocabulary Server -NVS, OceanOPS, and ICES) as shown in **Figure 3.2.2.3**.



**Figure 3.2.2.3.** Vocabulary conventions used for glider data as decided by the EuroGOOS Glider Task Team in 2019.

The conventions adopted by OG are described in detail at the [OG format user manual](#) [30]. This document includes the description of the global attributes, the variable names and conventions, vocabularies, coordinate and time standards, variable attributes, and dimension definitions. This convention uses SDN/BODC vocabularies as well as SeaDataNet metadata services, including EDMO, EDIOS, EDMERP, CDI. It also promotes the use of standard identifiers such as the Research Organization Registry (ROR) and The International Council for the Exploration of the Sea (ICES). Despite the use of controlled vocabularies and global identifiers, an effort has to be made related to identifying other resources in the JERICO catalog. For example, datasets should ideally include the JERICO identifier of the software or service-producing them as well as the identifiers of the Best Practices used along with the data flow. It is also recommended that ORCID identifiers are used to identify related actors.

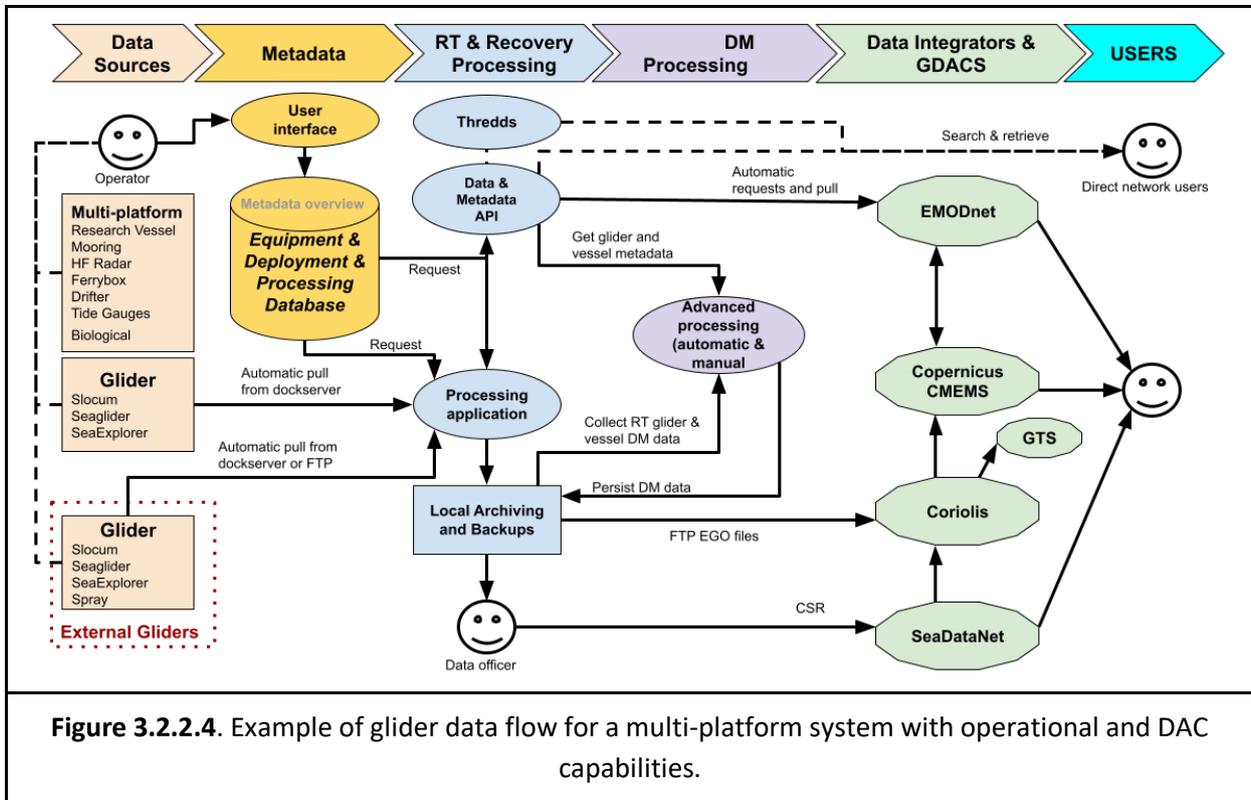
### Data policy

Glider data flows to the European aggregators via Coriolis. The access to the data in this portal follows the data policy of these infrastructures. They comply with the open-access distribution of data. At the data provider level, it is also recommended to adopt a free distribution approach using Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

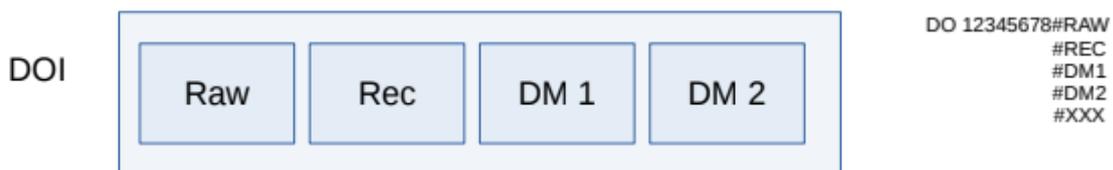
### Data dissemination – Link to European/International Data Banks

Several data providers may distribute glider data, but the European marine portals play an essential role, as stated by OG. The distribution in Europe of glider data to the EMODnet and CMEMS-INSTAC is performed via Coriolis as agreed by the community with the support of the EuroGOOS Glider Task Team during the international meeting in Genoa, Italy in 2018 (see **Figure 3.2.2.4**). Additionally, it is

recommended that the data is also registered in SeaDataNet using CSR. Thus, physical data from glider will be available in the main European marine data portals CMEMS-INSTAC (<http://www.marineinsitu.eu/>), EMODnet Physics (<https://map.emodnet-physics.eu/>) and SeaDataNet (<https://www.seadatanet.org/>). It will also be available globally for ocean forecasting and weather predictions in the WMO information system because Coriolis provides the data to GTS in the required format.



In order to improve the access to and persistence of the data, DAC should provide a DOI for the different levels of the dataset. Ideally, a DOI should be included in the metadata of the dataset. Optionally, the same DOI may be used for all levels of the dataset. Each level may be identified by adding a suffix that identifies the data level and version (see **Figure 3.2.2.5**)



### 3.2.3 Standards Used / Contributing Best Practices

The standards and conventions are described in the OG data format. These standards and conventions are summarized below:

**Table 3.2.3.1.** Data format standards and conventions

Standard/Convention	URL
OG1.0 Format	<a href="https://github.com/OceanGlidersCommunity/OG-format-user-manual/blob/main/OG_Format.adoc">https://github.com/OceanGlidersCommunity/OG-format-user-manual/blob/main/OG_Format.adoc</a>
EGO1.3 Format	<a href="https://archimer.ifremer.fr/doc/00239/34980/">https://archimer.ifremer.fr/doc/00239/34980/</a> <a href="https://www.ego-network.org/dokuwiki/lib/exe/fetch.php?media=public:datamanagement:v1.3:guidelines_to_fill_a_json_file_v1.3.1.docx">https://www.ego-network.org/dokuwiki/lib/exe/fetch.php?media=public:datamanagement:v1.3:guidelines_to_fill_a_json_file_v1.3.1.docx</a>
Climate and Forecast Metadata Convention CF-1.8	<a href="http://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html#trajectory-data">http://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html#trajectory-data</a>
Unidata NetCDF Attribute Convention for Data Discovery (ACDD)	<a href="https://www.unidata.ucar.edu/software/netcdf-java/v4.6/metadata/DataDiscoveryAttConvention.html">https://www.unidata.ucar.edu/software/netcdf-java/v4.6/metadata/DataDiscoveryAttConvention.html</a>
SeaDataNet Metadata Formats	<a href="https://www.seadatanet.org/Standards/Metadata-formats">https://www.seadatanet.org/Standards/Metadata-formats</a>

The following National Environment Research Council (NERC) vocabularies are recommended:

**Table 3.2.3.2.** Vocabularies

Vocabulary	URL
P01	<a href="https://vocab.nerc.ac.uk/collection/P01/current/">https://vocab.nerc.ac.uk/collection/P01/current/</a>
P07	<a href="https://vocab.nerc.ac.uk/collection/P07/current/">https://vocab.nerc.ac.uk/collection/P07/current/</a>
L06	<a href="https://vocab.nerc.ac.uk/collection/L06/current/">https://vocab.nerc.ac.uk/collection/L06/current/</a>
L06 platforms	<a href="http://vocab.nerc.ac.uk/collection/L06/current/27/">http://vocab.nerc.ac.uk/collection/L06/current/27/</a>
C86	<a href="https://vocab.nerc.ac.uk/collection/C86/current/">https://vocab.nerc.ac.uk/collection/C86/current/</a>
OG1	<a href="https://vocab.nerc.ac.uk/collection/OG1/current/">https://vocab.nerc.ac.uk/collection/OG1/current/</a>

**Table 3.2.3.3.** Glider Data Management Best Practices

Best Practice	URL
OG Standard Operating Procedures	<a href="https://github.com/OceanGlidersCommunity/DataAssemblyCenter_SOP">https://github.com/OceanGlidersCommunity/DataAssemblyCenter_SOP</a>
IMOS/ANFOG Glider Data Management	<a href="https://catalogue-imos.aodn.org.au/geonetwork/srv/api/records/a681fdb-a-c6d9-">https://catalogue-imos.aodn.org.au/geonetwork/srv/api/records/a681fdb-a-c6d9-</a>

	<a href="https://www.oceansites.org/docs/oceansites_user_manual_ver1.1.pdf">44ab-90b9-113b0ed03536/attachments/ANFOG_data_management3_1.pdf</a>
OceanSITES User's Manual	<a href="http://www.oceansites.org/docs/oceansites_user_manual_ver1.1.pdf">http://www.oceansites.org/docs/oceansites_user_manual_ver1.1.pdf</a>
Handbook for Data Management activities regarding data flow and data integration	<a href="https://archimer.ifremer.fr/doc/00370/48139/48242.pdf">https://archimer.ifremer.fr/doc/00370/48139/48242.pdf</a>
Report harmonization in data and data processing to facilitate the interoperability of the systems	<a href="https://www.atlantos-h2020.eu/download/deliverables/7.1%20Data%20Harmonization%20Report.pdf">https://www.atlantos-h2020.eu/download/deliverables/7.1%20Data%20Harmonization%20Report.pdf</a>
SeaDataNet data management protocols for glider data	<a href="https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c853ef45&amp;appId=PPGMS">https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5c853ef45&amp;appId=PPGMS</a>
OceanGliders 1.0 Harmonizing format across OceanGliders Terms of References	<a href="https://drive.google.com/file/d/1lcck2Uxm3Yzhu4-k7szj7kkkoSTYVG3W/view">https://drive.google.com/file/d/1lcck2Uxm3Yzhu4-k7szj7kkkoSTYVG3W/view</a>
The Ocean Gliders Data Management Task Team (OGDMTT) proposes a new BUFR sequence to report the full suite of glider observations along a trajectory profile, including biogeochemical parameters, on the GTS.	<a href="https://github.com/wmo-im/BUFR4/issues/16">https://github.com/wmo-im/BUFR4/issues/16</a>

**Table 3.2.3.4. Glider QA/QC Best Practices**

Best Practice	URL
Guidelines for the delayed mode scientific correction of glider data	<a href="https://repository.oceanbestpractices.org/handle/11329/884">https://repository.oceanbestpractices.org/handle/11329/884</a>
IMOS Delayed Mode QA/QC 3.0	<a href="http://content.aodn.org.au/Documents/IMOS/Facilities/Ocean_glider/Delayed_Mode_QAQC_Best_Practice_Manual_OceanGliders_v3.0.pdf">http://content.aodn.org.au/Documents/IMOS/Facilities/Ocean_glider/Delayed_Mode_QAQC_Best_Practice_Manual_OceanGliders_v3.0.pdf</a>
Seaglider Quality Control Manual	<a href="https://gliderfs2.coas.oregonstate.edu/sgliderweb/Seaglider_Quality_Control_Manual.html">https://gliderfs2.coas.oregonstate.edu/sgliderweb/Seaglider_Quality_Control_Manual.html</a>
Argo quality control manual 2.9	<a href="http://www.argodatamgt.org/content/download/20685/142877/file/argo-quality-control-manual_version2.9.pdf">http://www.argodatamgt.org/content/download/20685/142877/file/argo-quality-control-manual_version2.9.pdf</a>

IMOS/ANFOG Management	Glider	Data	<a href="https://catalogue-imos.aodn.org.au/geonetwork/srv/api/records/a681fdbac6d9-44ab-90b9-113b0ed03536/attachments/ANFOG_data_management3_1.pdf">https://catalogue-imos.aodn.org.au/geonetwork/srv/api/records/a681fdbac6d9-44ab-90b9-113b0ed03536/attachments/ANFOG_data_management3_1.pdf</a>
IOOS Glider DAC - Manual for Quality Control of Temperature and Salinity Data Observations from Gliders			<a href="https://cdn.ioos.noaa.gov/media/2017/12/Manual-for-QC-of-Glider-Data_05_09_16.pdf">https://cdn.ioos.noaa.gov/media/2017/12/Manual-for-QC-of-Glider-Data_05_09_16.pdf</a>
OOI - Data Product Specification for Global Range Test (1.01)			<a href="https://oceanobservatories.org/wp-content/uploads/2015/09/1341-10004_Data_Product_SPEC_GLBLRNG_OOI.pdf">https://oceanobservatories.org/wp-content/uploads/2015/09/1341-10004_Data_Product_SPEC_GLBLRNG_OOI.pdf</a>
Guidelines for the delayed mode scientific correction of glider data			<a href="https://www.jerico-ri.eu/download/jerico-next-deliverables/JERICO-NEXT-Deliverable_5.15_Final.pdf">https://www.jerico-ri.eu/download/jerico-next-deliverables/JERICO-NEXT-Deliverable_5.15_Final.pdf</a>
Manual for Real-Time Oceanographic Data Quality Control Flags (version 1.2)			<a href="https://repository.library.noaa.gov/view/noaa/24982">https://repository.library.noaa.gov/view/noaa/24982</a>

**Table 3.2.3.4.** Glider QA/QC Best Practices

### 3.2.4 Quality control/Quality Assurance

Quality control and assurance (QC/QA) has an essential role in the life cycle of glider observations. In recent years the glider community has started to develop the necessary protocols and standards to homogenize and define the different levels of data quality. QC in the glider observations, we refer to the application of methods and procedures that determine whether the observations meet the overall quality international standards and defined quality criteria for individual parameters. The QA, on the other hand, defines the various processes that research institutes seek to ensure that the glider observations maintain the international standard or have improved.

DAC or GDAC applies quality control procedures in an automated mode to facilitate real-time data QC in the latest phases by identifying some of the problems. These automated flags also support the ability to filter data that have been sent to GTS. Problems that are not detected by these automated procedures should be caught in the delayed mode QC with the interaction of a human expert. This section provides details about the QC/QA processes in real-time and delayed mode.

#### *Near-Real Time QC*

Overview of Near-Real Time and recovery data QC/QA

In contrast to delayed mode QC, what defines a QC technique as 'real time' or 'near real time' for gliders is that both scientists and engineers may have a subset of all the information on land during the glider deployment due to satellite bandwidth. The data availability depends on the reflecting cost of the satellite communication, surface time, and operation risks. In real-time, we should always expect limited and incomplete data. However, the limited observations should offer us enough information regarding data quality on a horizontal and vertical scale. The particular issue of real-time

QC is to evaluate the most recent data point. Based on the data nature, we have defined the role of real-time QC data:

- 1) Real time data is acquired typically seconds to hours and can be instantly used for model ingestion. This is often the case for operational modeling communities that require data within 24 hours, requiring automated procedures, and are aware that the data has not been subjected to climate grade QC. They require data that has undergone a rough QC to ensure that data assimilations are not harmed. However, the outcomes of real time QC are used to inform later stages of QC, such as when data tagged 3 (possibly poor) is evaluated, it is changed to 4 (bad data) based on the flagging scheme that the user applied.
- 2) The real time data can be a useful source for manager decisions and policy makers.
- 3) To ensure that high quality glider observations gather and detect failure with the on board sensors.

Although the majority of the automatic checks are often connected with real-time QC, and while they do better meet the needs and requirements for the real-time data stream, they should not be confined to them. The close monitoring of the real-time observations by pilots and scientists at least once a day or every couple of days is one of the main features of the real-time QC. Based on the nature of the real-time QC, we could distinguish the following levels:

- a) Checks for quick response are performed automatically.
- b) Automatic update once more information becomes available to better support an evaluation, referred to as near real-time (like more vertical profiles available, deeper profiles to stable waters, comparison with other platforms available in the study area, changes on the instrument configurations, etc.)
- c) Manual assessment is highly recommended. Pilots and PIs keep a tight eye on the real time observations in the hopes of spotting something suspicious on the data.

Basic automatic QC tests can be applied to any geophysical parameter. Similar automatic QC tests have also been developed for biogeochemical observation. The **Table 3.2.4.1** summarizes a series of QC tests that are automatically run to flag geophysical variables.

**Table 3.2.4.1.** Basic QC tests applied automatically to data

QC method name	Transfer parameters	Description
NaN Check	Any variable	Flag NaN values as 9
Impossible date check	Time	Tests that time values are within the timeframe associated with the deployment and mark bad entries as 4
Impossible location check	Longitude, Latitude	Tests for impossible values ( $-90 < x <= 90$ and $-180 < y <= 180$ ), and that the values are sensible for this deployment and mark bad entries as 4
Valid range check	Scientific data	Tests if values are within the expected ranges. Four types of range tests should be applied according to these aspects: Sensor ranges

		Global ranges Regional and local ranges Mark entries that fail the test with the specific flag.
Spike test	Scientific data (primarily for T and S)	Identifies spikes based on Argo Data Management (Argo, 2013) test. Mark bad entries as 6.
Gradient test	Scientific data	Test the difference between vertically adjacent measurements. Values that are too steep are flagged as 4.
Surface data test	Optical sensor data	Data from optical sensors are usually noisy at the surface and are flagged as 4.

In both cases, data are not modified, but instead, they are labeled in a separate QC variable with the same parameter name as the suffix `_QC`. The aim is to identify issues in the data and flag them in a separate variable used to mask the data in the original geophysical variables. Several quality flag schemes can be used and adopted in the glider observations. Depending on the flag scheme that the user applies, each flag can have different meanings. However, most of the available flag schemes are focused on the same approach and aim to characterize the observed values in the most precise and accurate way. **Table 3.2.4.2** shows an example of the mapping between the SEADATANET flag scheme with the other available flag schemes. A [mapping between the available schemes<sup>1</sup>](#) has been created to homogenize the different quality flag schemes between the organizations [37].

In 2013 UNESCO/IOC recommended a [flagging standard<sup>2</sup>](#) [31] where the QARTOD adopted it in 2014. In recent years many institutions have followed and used the UNESCO/IOC standard as it has been recommended by the glider experts. The adopted QC flagging scheme for glider data is the [ARGO QC flag scale<sup>3</sup>](#) (which extends the UNESCO scale, [19]).

<sup>1</sup> [https://odv.awi.de/fileadmin/user\\_upload/odv/misc/ODV4\\_QualityFlagSets.pdf](https://odv.awi.de/fileadmin/user_upload/odv/misc/ODV4_QualityFlagSets.pdf)

<sup>2</sup> [https://www.nodc.noaa.gov/oads/support/MG54\\_3.pdf](https://www.nodc.noaa.gov/oads/support/MG54_3.pdf)

<sup>3</sup> <https://argo.ucsd.edu/data/how-to-use-argo-files/>

**Table 3.2.4.2.** UNESCO/IOC flag scheme.

Flag*	Description
Pass=1	Data have passed critical real-time quality control tests and are deemed adequate for use as preliminary data.
Not evaluated=2	Data have not been QC-tested, or the information on quality is not available
Suspect or Of High Interest=3	Data is considered to be either suspect or of high interest to data providers and users. They are flagged suspect to draw further attention to them by operators.
Fail=4	Data are considered to have failed one or more critical real-time QC checks. If they are disseminated at all, it should be readily apparent that they are not of acceptable quality.
Missing data=9	Data is missing; used as a placeholder.

\* UNESCO scale

Even though the most common flag scheme is the UNESCO/IOC flag scheme, high heterogeneity in the flag schemes exists among the community. The **Table 3.2.4.3** shows an example of the mapping between the different flag schemes of the SeaDataNet with other available flag schemes. ([http://seadatanet.maris2.nl/v\\_bodc\\_vocab/browse\\_export.asp?l=L201&all=yes&description=1&x=45&y=15](http://seadatanet.maris2.nl/v_bodc_vocab/browse_export.asp?l=L201&all=yes&description=1&x=45&y=15))

**Table 3.2.4.3.** Mapping of the SeaDataNet flag scheme.

Flag description	ODV	GTSPP	ARGO	SEADATANET	ESEAS	WOD	QAETOD	BODC	SMHI	Ocean SITES	UNESCO /IOC
no quality control (QC) was performed	1	0	0	0	0	0	0	Q	*	0	1
QC was performed; good data	0	1	1	1	1	0	3	*	*	1	3
QC was performed; probably good data	0	2	2	2	1	0	3	*	*	2	4
QC was performed; probably bad data	4	3	3	3	3	4	2	K	?	3	3

QC was performed; bad data	8	4	4	4	4	4	1	K	B	4	4
QC was performed; the value was changed as a result of QC	1	5	5	5	2	0	0	R		5	2
QC was performed; the value below detection	1	0	0	6	0	0	0	<	<	0	2
QC was performed; the value in excess	1	0	0	7	2	0	0	>	>	0	2
QC was performed; the value is missing	1	9	9	9	9	0	9	N	B	9	9
QC was performed; the value phenomenon uncertain	1	0	0	A	0	0	0	Q	B	0	2

#### Valid Range Check for sensors and the global ocean

To define the ranges for each parameter that QC will be applied to, it is necessary to understand the limitations of each instrument on the glider and the study area. There are various ranges based on the type of the sensors and the user's specific requirements. The first ranges to consider are the ones given by the instrument specifications (example of different types of sensors; **Table 3.2.4.4**). Values of the measurement must be within the range specified by the manufacturer of the sensor. The calibration datasheet of each of the sensors used in the glider provides the required information to perform the sensor range test.

**Table 3.2.4.4.** Shows the ranges for some sensors manufacturers

Variable	Sensor range	Sensor
Temperature [° C]	-5 to 42	Slocum Glider Payload CTD
Conductivity [S/m]	0 to 9	Slocum Glider Payload CTD
Turbidity [NTU]	0 to 25	Wetlabs
Chla-Flu [mg/m <sup>3</sup> ]	0 to 50	Wetlabs
Oxygen concentration [µmol/l]	0 to 500	Aanderaa oxygen optode
Oxygen sat [%]	0 to 120	Aanderaa oxygen optode

In general, some rules can be applied to any sensor manufacturer. For example, chlorophyll, turbidity, oxygen concentration, and oxygen saturation should be positive and below a very high threshold. Concerning global ranges of the various physical and biogeochemical variables, depth specific thresholds can be adopted from the User Manual of the World Ocean Database 2013 (Johnson et al., 2013). However, QC is usually applied independently of the depth. The global values are specified by different sources, as summarized in **Table 3.2.4.5**.

**Table 3.2.4.5.** Global ranges for various physical and biogeochemical variables

Variable	Global range	Source
Temperature [° C]	-2.5 to 40	DATAMEQ, 2010
Salinity [PSU]	2 to 41	DATAMEQ, 2010
Conductivity [S/m]	0 to 8.5	ARGO User manual. V3.2 (2015)
Turbidity [NTU]	0 to 50	GROOM / ARGO qc V2.8, 2013
Chla-Flu [mg/m <sup>3</sup> ]	0 to 50	Argo QC Manual For Biogeochemical Data
Oxygen conc [mmol/l]	0 to 600	ARGO
Oxygen sat [%]	-	-

As we aim to observe in specific regions, values can fit to specific observational areas. These values should be clearly within the global range, and they allow a more restrictive way to identify bad measurements. These regional values depend on the location of the observation. Each operator and DAC should consider the values defined for their specific area. For example, ARGO/DATAMEQ defined the values for the Mediterranean Sea of the temperature (10 to 40 °C) and salinity (20 to 40 PSU). SOCIB has been able to reduce these global ranges based on 10 years of monitoring Ibiza and Mallorca channels, as **Table 3.2.4.6** indicated.

**Table 3.2.4.6.** Local range values around the Balearic Islands in Spain. These values were defined by SOCIB based on 10 years of sustained observations.

Variable	Local ranges in Balearic Islands
Temperature [°C]	10 to 30
Salinity [PSU]	35 to 40
Conductivity [S/m]	4 to 6.5
Density [kg/m <sup>3</sup> ]	990 - 1035
Turbidity [NTU]	0 to 10
Chla-Flu [mg/m <sup>3</sup> ]	0 to 5
Oxygen conc [mmol/l]	140 to 350
Oxygen sat [%]	50 to 120

#### Spike Test

A spike in both size and gradient is the difference between sequential measurements when one measurement is notably different from nearby ones. The objective is to run the data via a rolling filter (along the time dimension), which can be used as a starting point. Spikes distinguish the data from the original.

$$\text{Test value} = | V2 - (V3 + V1)/2 | - | (V3 - V1) / 2 |$$

where V2 is the measurement being tested, and V1 and V3 are the values above and below.

For the physical parameters like temperature and salinity, this test ignores the pressure differences. Instead, it assumes a sampling scheme that accurately replicates temperature and salinity changes as a function of pressure. This algorithm applied to vertical temperature and salinity profiles.

- **Temperature:** The V2 value is flagged when the test value exceeds 6.0°C for pressures less than 500 dbar, or the test value exceeds 2.0°C for pressures greater than or equal to 500 dbar.
- **Salinity:** The V2 value is flagged when the test value exceeds 0.9 PSU for pressures less than 500 dbar, or the test value exceeds 0.3 PSU for pressures greater than or equal to 500 dbar.

#### Gradient Test

This test fails when the difference between vertically adjacent measurements is too significant. The test overlooks pressure variations because it assumes that the changes of variables implicitly contain these variations.

Test value =  $| V2 - (V3 + V1)/2 |$  where V2 is the measurement being tested, and V1 and V3 are the values above and below.

For example, temperature and salinity variables are flagged as 4 in the following scenarios:

- **Temperature:** The V2 value is flagged when the test value exceeds 9.0°C for pressures less than 500 dbar, or the test value exceeds 3.0°C for pressures greater than or equal to 500 dbar.
- **Salinity:** The V2 value is flagged when the test value exceeds 1.5 PSU for pressures less than 500 dbar, or the test value exceeds 0.5 PSU for pressures greater than or equal to 500 dbar.

### Near-Real Time visualization

The **Table 3.2.4.7** summarizes a list of plots that are highly recommended to perform to ensure QA on the main observed physical and biogeochemical parameters during the glider deployment. These plots help to detect spikes, outliers, and sensor issues.

**Table 3.2.4.7.** List of plots during the glider deployment to ensure QA in the observations

Scientific parameters	Frequency
Map of the glider trajectory with the last profile indicated	Daily
Time series plot per sensor <ul style="list-style-type: none"> <li>• CTD (Temperature, Conductivity and Salinity)</li> <li>• Optical sensor/s (CHL fluorescence, Turbidity, Backscatter and CDOM)</li> <li>• Oxygen sensor (Oxygen concentration, Oxygen Saturation and Temperature)</li> </ul>	Daily
Time series vs depth plot per sensor color code <ul style="list-style-type: none"> <li>• CTD (Temperature, Conductivity and Salinity)</li> <li>• Optical sensor/s (CHL fluorescence, Turbidity, Backscatter and CDOM)</li> <li>• Oxygen sensor (Oxygen concentration, Oxygen Saturation and Temperature)</li> </ul>	Daily
Vertical profiles for the whole depth range and for the first 200m <ul style="list-style-type: none"> <li>• Depth vs Temperature, Conductivity and Salinity</li> <li>• Depth vs CHL fluorescence, Turbidity, Backscatter and CDOM</li> <li>• Depth vs Oxygen concentration, Oxygen Saturation and Temperature</li> </ul>	Daily
All profiles superimposed the last profile for the whole depth range <ul style="list-style-type: none"> <li>• Depth vs Temperature, Conductivity and Salinity</li> <li>• Depth vs CHL fluorescence, Turbidity, Backscatter and CDOM</li> <li>• Depth vs Oxygen concentration, Oxygen Saturation and Temperature</li> </ul>	Daily
TS plots <ul style="list-style-type: none"> <li>• All profiles superimposed the last profile</li> <li>• All profiles superimposed the last profile for the deep waters</li> </ul>	Every few days
TS color code with CHL fluorescence, Turbidity, Backscatter, CDOM and Oxygen	Daily
Temperature (from oxygen sensor) vs Oxygen color code with depth	Daily
Comparison between sensors <ul style="list-style-type: none"> <li>• Pressure of CTD vs navigation pressure of the glider</li> <li>• Oxygen temperature vs temperature of the CTD vs fluorometer</li> </ul>	Daily

Intercomparison with other available platforms in the region (Ship-based observation and Argo floats).	Daily
TS diagram between downcasting and upcasting for detecting issues with the CTD pump	Every few days

### Delayed Mode QC

As for real time data, further delayed mode QC can flag bad values when identified by an experienced user. In addition, delayed mode QC may also modify values by using advanced algorithms. These changes do not occur in the original variable to avoid losing the original information. An additional variable with the `_corrected` suffix should be created that contains a copy of the variable values. These corrected variables may be modified during the delayed mode QC process. The most common corrections applied to the data are:

- Removing bad profiles
- Data interpolation
- Outliers
- Spike correction
- Thermal lag correction
- Salinity correction
- Dark counts
- Quenching correction

A higher level of QC is carried out on the recovery data by the DAC. In the following sections, we explain more details about some of these corrections.

#### Removing bad profiles

In this step you can remove vertical profiles from bad glider dives taking into account the first STD or the median with the first STD at the reference depth.

#### Data Interpolation

Data interpolation consists of applying interpolation algorithms to fill NaN values that were identified during the real time process. Linear interpolation is commonly used for time, latitude and longitude.

#### Outliers (IQR and STD)

We can remove outliers by using the interquartile range and standard deviations of the whole dataset. Filters can be made more or less severe by adjusting the multipliers. By following the value ranges for each parameter based on past observations, we can define the degree of the adjusting. Most of the time, erroneous measurements are frequently made vertically.

#### Spikes Correction

By removing spikes in each parameter, we are smoothing the vertical profiles. The data can be filtered using two different rolling filters. A rolling median is the counterpart of the median approach. Data is first subjected to a rolling minimum and then a rolling maximum in the min and max technique. This is especially valuable in optics data because spikes are particles in the water column that are not evenly distributed. The median approach is probably ideal in the case of salinity because "spikes" could be both positive and negative (Gaussian distribution).

### Thermal lag Correction

Pinot et al. (1997) established based on in situ data in the Mediterranean that the effect of thermal lag in Slocum G1 gliders is considerable in areas with a sharp thermocline. The salinity and the density are strongly biased due to the thermal lag effects on conductivity measurements. In fact, the salinity errors result from the mismatch between temperature (measured outside the conductivity cell) and conductivity (measured inside the conductivity cell). More modern gliders use pumps and minimize the effect of the thermal lag. However, it is critical to correct thermal lag when the pump is not available in the CTD. Garau et al. (2011) propose two algorithms to correct this effect. One is based on estimating the conductivity that would be measured outside the cell. The second one infers the temperature inside the conductivity cell.

### Salinity Correction

The calibration salinity from the glider can be performed using the vessel data from missions in the same transects as the glider close to the dates of the same missions. A semi-automatic delay mode correction may be used based on white maximization image analysis (Allen et al. 2020) to calibrate the salinity against vessel data.

### Dark count Correction

The calculation of a dark count in situ is required to compensate for sensor drift from factory calibration. The 95<sup>th</sup> percentile of bio-optical data between 200 and 400 meters is used to determine the dark count values for the optical parameters.

### Quenching correction

Gliders give us the ability to measure chlorophyll fluorescence and help us to better understand the phytoplankton distributions. However, the validity of these datasets can be jeopardized by underestimating the daytime fluorescence derived from different regional and temporal scales. Existing approaches in the literature have adjusted for quenching. However, these methods rely on assumptions that aren't valid in all places or seasons. Thomalla et al. (2018) shows that by multiplying a mean nighttime fluorescence to backscattering ratio by daytime backscattering profiles from the surface to the depth of quenching (defined as the depth at which the day fluorescence profile diverges from the mean night profile), we are able to correct daytime quenched fluorescence [40]. The approach described here overcomes some of these assumptions, resulting in adjusted surface fluorescence throughout the day that nearly matched profiles from the prior (or subsequent) night, with a difference of less than 10% in the observed chlorophyll fluorescence.

### 3.2.5 Issues

There is a technical debt regarding the present workflow, which includes the link to Coriolis, which ideally should be readdressed in order to make use of machine to machine tools from data providers, such as Data API and ERDAPP servers.

In addition, the availability of a global database would simplify the process of collecting glider and missions' metadata across borders. This database should of course provide the opportunity to link the metadata to datasets in a flexible way so the metadata can be updated afterwards by data operators. This metadata catalog should account for the fact that the data lives by nature in a distributed landscape. It should also address the fact that the data is duplicated in various formats in the servers of data providers and main European data portals. It will also be beneficial to have an API supporting the access and visibility of the metadata. A global metadata catalog would also support the processing

at a global scale by GDACs by providing a common metadata repository that assimilates the information of the operations comprehensively.

Finally, QC testing at the data provider level is lacking due to the absence of tools to support the process.

### 3.2.6 Training Materials and Contacts

A series of workshops and glider schools were organized by the community along the years. These workshops were intended to strengthen the glider network and standardize procedures and formats.

**Table 3.2.6.1** reports the list of courses and workshops.

**Table 3.2.6.1.** List of workshops and materials related to glider data management

Title	Date	Details and Materials
1st EGO Workshop	2-3 October, 2006	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2006">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2006</a>
2nd EGO Workshop and Glider School	25-31 October, 2007	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2007">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2007</a>
3rd EGO Workshop and Glider School	27-31 October, 2008	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2008">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2008</a>
4th EGO Workshop and Glider School	17-21 November, 2009	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2009">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2009</a>
5th EGO Workshop and Glider School	14-18 March, 2011	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2011">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2011</a>
6th EGO meeting & Final Symposium of the COST Action ES0904	16-17 June, 2014	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2014">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2014</a>
7th EGO Conference	26-29 September, 2016	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2016">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2016</a>
8th EGO Meeting	20-24 May 2019	<a href="https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2019">https://www.ego-network.org/dokuwiki/doku.php?id=public:ego_workshops:ego2019</a>

Glider School Plocan Website	Since 2011-ongoing/every year	<a href="https://gliderschool.eu/previous-edition/">https://gliderschool.eu/previous-edition/</a>
------------------------------	-------------------------------	---

There are a few software packages that are available to support processing, management and quality control of glider observations (see **Table 3.2.6.2**).

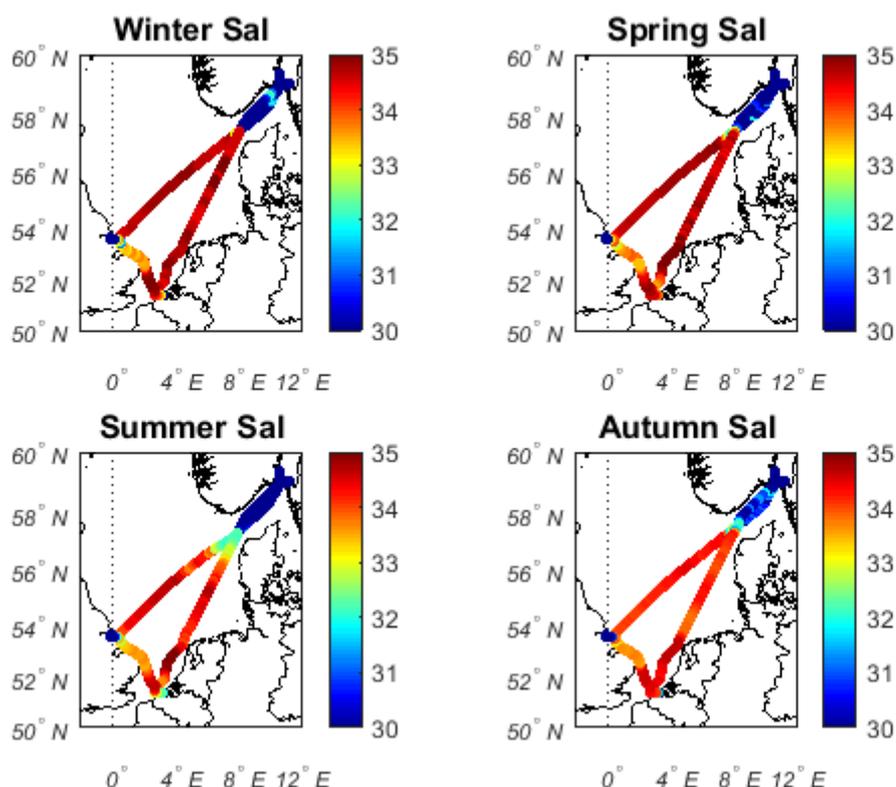
**Table 3.2.6.2.** List of software packages for support processing, management and quality control of glider observations

Software Package	URL
SOCIB Glider Toolbox [5]	<a href="https://github.com/socib/glider_toolbox">https://github.com/socib/glider_toolbox</a>
SOCIB salinity correction toolbox [21]	<a href="https://github.com/socib/salinity-correction-toolbox">https://github.com/socib/salinity-correction-toolbox</a>
EGO data checker [13]	<a href="https://www.seanoe.org/data/00344/45538/">https://www.seanoe.org/data/00344/45538/</a>
CoTeDe is an Open-Source Python package to quality control (QC) oceanographic data such as temperature and salinity [8]	<a href="https://github.com/castelao/CoTeDe">https://github.com/castelao/CoTeDe</a>
Glider tools is a Python 3.6+ package designed to process data from the first level of processing to a science-ready dataset (delayed mode quality control) [17]	<a href="https://zenodo.org/record/4815417#.YebNrVjMLzc">https://zenodo.org/record/4815417#.YebNrVjMLzc</a>

### 3.3 FerryBox

#### 3.3.1 Description of Data

FerryBox systems provide data of a number (about 20) essential ocean variables (EOVs), along with housekeeping parameters in surface coastal, estuarine and open ocean waters. The standard EOVs include temperature, salinity, dissolved oxygen (concentration and % saturation), pH, chlorophyll fluorescence, and turbidity. Additional parameters can be added, depending on scientific interest and applications. These include partial pressure of carbon dioxide ( $pCO_2$ ), colored dissolved organic matter (CDOM), nutrient measurements (UV, lab-on-chip, wet-chemistry methods), total alkalinity, micro-plastics (filtering setup), and algal classes (Petersen et al. 2008; Lips and Lips, 2008; Petersen and Colijn, 2017; Petersen et al. 2014; Voynova et al. 2019; Assmann et al. 2011). Housekeeping parameters include status of the FerryBox flow-rate, speed of vessel, timestamp, coordinates, quality flags, and statistical information. The statistical information includes estimates of minimum, maximum, variance and counts (Petersen and Colijn, 2017).



**Figure 3.3.1.1.** Seasonal plots of salinity in the North Sea in 2014 along the SOO Lysbris Seaways route between Immingham (UK) – Moss/Halden (NO) – Zeebrugge/Ghent (NL).

Depending on where FerryBoxes are deployed, data acquisition frequency can vary. In an underway system, where a research vessel or a ship of opportunity (SOO) crosses different regions along its route, more frequent sampling is required, sometimes down to a few seconds to properly capture regional variability (Lips and Lips, 2008; Macovei et al. 2020). This allows for capturing environmental variability, such as seasonal, regional and water mass changes (**Fig. 3.3.1.1**). At fixed stations, sampling rate may be lowered, but the environmental variability should be properly resolved. At a tidal station for example, where tidal, daily, current variabilities exist, sampling should be done at least every 2-3 hours, preferably every 20 minutes to 1 hour (Voynova et al. 2017).

Since the beginning of European FerryBox activities in the Baltic Sea in the 1980s, and later on, on a European level with the FP5 project “FerryBox” (2003-2005), there has been an increasing number of FerryBox lines as well as ongoing activities on harmonization of operation. However, progress in visibility and accessibility of the FB data has been slow, even if long-term FerryBox data sets exist, especially in the Baltic Sea and North Sea. These data sets are often available upon request to the data originator and mostly based on files (ASCII or NetCDF) (Petersen and Colijn, 2017). An effort has been made to make ferrydata.hereon.de, a Hereon COSYNA-supported data portal available for the European FerryBox community. Even though many FerryBox operators do provide data there, and this data portal is considered as the EuroGOOS FerryBox TaskTeam data portal, this is still under discussion with the community. Many FerryBox operators also directly report data to international and European data aggregators like EMODnet.

### 3.3.2 Data value chain: from acquisition to delivery

Sensors used in FerryBoxes vary, depending on operator preference, size requirements, energy consumption and sometimes age of each FerryBox. One large advantage of the FerryBox is that, depending on how it is made and programmed, it can accommodate many different types of sensors and in a modular manner - technology updates are possible. Instrument companies like 4H Jena have a typical set of parameters (<https://www.4h-jena.de/en/maritime-technologies/flow-systems/ferrybox/>), but most likely, they can provide a setup required by the FerryBox operator. Some of the sensors used in FerryBox operations are listed in **Table 3.3.2.1**, but this list is not exhaustive.

**Table 3.3.2.1.** A list of some sensors, used to measure EOVs in FerryBoxes, and information on accuracy, precision and potential sources of bias and uncertainty. Table was modified from Table 1 in Petersen et al. (2018).

Parameter	Range	Unit	Accuracy	Precision	Uncertainty /Bias	Instrument	Manufacturer
Water Temperature (1)	0 to 50	°C	0.1	0.01	Potential warming inside tubing (max 0.5 °C)	Excell ETSG2	Falmouth Scientific Inc., USA
Water Temperature (1)	-5 to 35	°C	0.001-0.002		Potential warming inside tubing (max 0.5 °C)	Sea-Bird SBE45 and SBE38	Sea-Bird Scientific, USA
Salinity (2)	0 to 50		0.02	0.001	Potential drift with fouling	Excell ETSG2	Falmouth Scientific Inc., USA
Salinity (2)	-5 to 35		0.005		Potential drift with fouling	Sea-Bird SBE45	Sea-Bird Scientific, USA
Dissolved Oxygen (3)	0 to 500	µmol L <sup>-1</sup>	5-8%	1	Potential Drift and initial offset	Optode (3830,..)	Aanderaa/Xylem, Norway, USA
Chlorophyll Fluorescence (4)	0 to 200	µg L <sup>-1</sup>	10%	0.5	Changing fluorescence yield, biofouling	SCUFA/C3/ ECOFLNTU	Turner Designs, TriOS, WetLabs
Turbidity (5)	0 to 50	NTU		0.05	Offset due to small bubbles	SCUFA/C3	Turner Designs, WetLABS/SeaBird

pH (6)		no unit		0.01	Drift in electrodes, requires frequent calibration or pH check	Glass electrode, ISFET	Sensortechnik Meinsberg, Endress + Hauser
CDOM Fluorescence (7)	0-1500	ppb	0.1		biofouling	C3, microFlu, NanoFlu	Turner Designs, TriOS

An example of some routes operated in European Seas are shown in **Table 3.3.2.2**. Data are typically transferred once a ship reaches port and a stable connection can be established, but can also be transmitted by satellite communications. At Hereon, a subset of data of the main parameters and log files are transferred to the dataserver via an http transfer. The larger working files, which store all information and housekeeping parameters are downloaded during regular maintenance.

**Table 3.3.2.2.** Some FerryBox platforms operated in European waters. Table is a subset from table in Petersen and Colijn (2017), pp.12-13.

Institution	Ports	Platform	Parameters (Table 3.3.2.1)	Website	Start of Operation	Route Frequency
HCMR	Souda Bay – Piraeus	Olympic Champion	1 - 6	<a href="http://www.poseidon.hcmr.gr">www.poseidon.hcmr.gr</a>	2002	daily
Hereon	North Sea	Lysbris Seaways	1-7, pCO <sub>2</sub>	<a href="http://www.cosyna.de">www.cosyna.de</a>	2007	2 weeks - 1 month
Ifremer	Portsmouth-Santander-Plymouth-Roscoff-Cork (St Malo)	MV Pont-Aven	1-5,7			
IMR	Bergen-Kirkenes	MS Vesterålen	1,2,4		2006	11-day roundtrip
INSTM	Tunis-Marseilles, Tunis-Genoa	Carthage	1-6, pCO <sub>2</sub> , sound velocity		2016	

MSI/TUT	Tallinn – Helsinki	MS Europa	Silja	1,2,4,5, pCO <sub>2</sub> ; nutrients	<a href="http://ferrybox.msi.ttu.ee">http://ferrybox.msi.ttu.ee</a>	1997	daily
NIVA	Oslo – Kiel	MS Fantasy	Color	1,2,4,5,7, cyanobacteria, nutrients, irradiance, radiance	<a href="http://www.ferrybox.no">www.ferrybox.no</a>	2008	daily
SYKE	Helsinki – Travemünde	Finnpartner, Finnmaid		1,2,4,5,7, nutrients, Phycocyanin	<a href="http://www.itameriportal.fi/en/italineriintiedotus/en_GB/levatiedotus">http://www.itameriportal.fi/en/italineriintiedotus/en_GB/levatiedotus</a>	1998	daily

Data are stored on a FerryBox computer, using LabView or Python system controls. The commercially available FerryBox from 4H Jena Engineering has a LabView system, similar to the one used in Hereon. SMHI operates using a python-based system (Petersen and Colijn, 2017). All parameters are filtered/flagged by housekeeping parameters. These include status of the FerryBox, flow rate, speed of the vessel (for moving platforms).

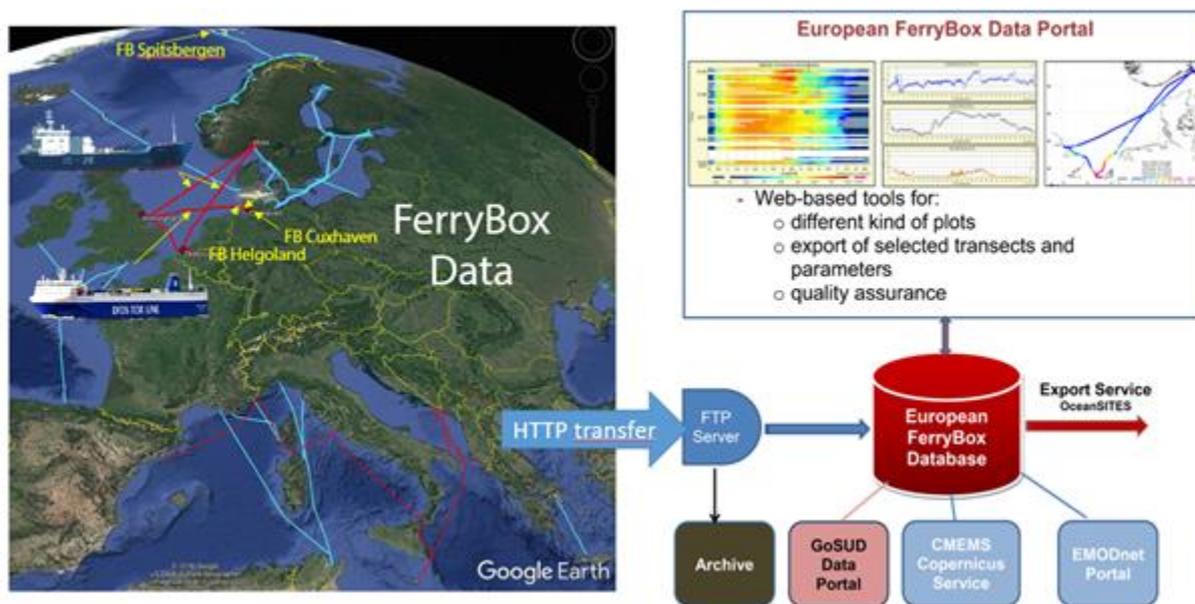


Figure 3.3.2.1. FerryBox Data Portal data transfer and dissemination.

Automated transfer of data and quality checks from processing allows for import into the Hereon FerryBox Database ([ferrybox.org](http://ferrybox.org), <http://ferrydata.hereon.de> for fixed routes and [tsdata.hereon.de](http://tsdata.hereon.de) for fixed stations). Working files contain metadata including sensor type and identification, data quality flags, range, minimum, maximum, variance, location data, raw values. An example is shown in Fig. 3.3.2.2.

```

$METADATA
$Project; FerryBox
$Hostname; PC1/R07036-PC1
$DateTime; 2021.12.05 00:00:04
$Filename; CO2FT_A_PCO2_Corr_20211205.txt
$Type; HydroC CO2FT (CO2T-1019-004*4b)
$Formula; Meas=a0+a1*Raw+a2*Raw^2+a3*Raw^3
$Ranges; 1
$Range1.a0; 0.0000000000E+0
$Range1.a1; 10.0000000000E-3
$Range1.a2; 0.0000000000E+0
$Range1.a3; 0.0000000000E+0
$QUALITYBITS
$Bit; Value, Meaning
$0; 1, Unknown measuring value.
$5; 32, Processstatus = unknown.
$6; 64, Processstatus 1 = clean.
$7; 128, Processstatus 2 = standby.
$8; 256, Processstatus 3 = empty.
$9; 512, Processstatus 9 = error.
$10; 1024, Processstatus 8 = undefined state.
$11; 2048, Simulated data.
$12; 4096, Measuring delay.
$13; 8192, Value from zero-cycle.
$14; 16384, Value from flush-cycle.
$INFOBITS
$Bit; Value, Meaning
$0; 1, Unknown measuring value.
$1; 2, Data overflow.
$2; 4, Data underflow.
$3; 8, Device timeout.
$4; 16, Channel mismatch.
$5; 32, Receiverwarning (GPS only).
$6; 64, Lower than 2 satellites (GPS only).
$7; 128, No GPS signal (GPS only).
$8; 256, CRC error (GPS only).
$9; 512, Latitude and longitude are 0 (GPS only).
$10; 1024, Formaterror.
$FORMATS
$1; Timestamp, Date Time; YYYY.MM.DD hh:mm:ss
$2; PCO2_Corr, µAtm; Float
$3; Quality, Flags; Int
$4; MeasCount, Cnt; Int
$5; MeanTime, Sec; Int
$6; Range, MR; Int
$7; Minimum, µAtm; Float
$8; Maximum, µAtm; Float
$9; Variance, Units; Float
$10; Longitude, Deg; Float
$11; Latitude, Deg; Float
$12; Rawvalue, Units; Float
$13; Info, Flags; Int
$DATASETS
$Timestamp PCO2_Corr Quality MeasCount MeanTime Range Minimum Maximum Variance Longitude Latitude
$Date Time µAtm Flags Cnt Sec MR µAtm µAtm Units Deg Deg Units Flags
2021.12.05 00:00:00 469.480000 0 35 55 1 468.860000 470.010000 0.114900 NaN NaN 46948.000000 0
2021.12.05 00:01:00 469.394722 0 36 54 1 468.930000 469.940000 0.073317 NaN NaN 46939.472222 0
2021.12.05 00:02:01 469.399091 0 33 57 1 468.730000 470.090000 0.183421 NaN NaN 46939.909091 0
2021.12.05 00:02:00 469.399091 0 36 57 1 468.930000 469.940000 0.142258 NaN NaN 46939.487222 0

```

**Figure 3.3.2.2.** An example of a working file for pCO<sub>2</sub> data in December 2021 with metadata and data descriptions in file header.

### Metadata Information

Different types of metadata exist for FerryBox data:

1. Metadata for the sensors stored as SensorML. These metadata from Hereon are stored inside the sensor database of the AWI (<https://sensor.awi.de/>) which is generally used for German marine data. For example the FerryBox on Lysbris which operates on the route covering large parts of the North Sea is <https://hdl.handle.net/10013/sensor.51009c8a-9d2c-4c53-ac7a-615ce4ac3974>.
2. Metadata for transects and data stored as ISO19115. These metadata are relevant for the FAIR use of data. The Hereon FerryBox type 2 metadata are stored in ISO 19115-3:2018 format available at <https://hcdc.hereon.de/geonetwork/srv/eng/catalog.search>.

3. An additional metadata system for type 2 metadata provide OGC Web Feature Services used by the COSYNA data portal CODM. From there you could use the FerryBox data in findable, accessible and interoperable way to reuse the data.

### 3.3.3 Standards used / Contributing Best Practices

Access to Hereon FerryBox data could be achieved with the help of the COSYNA Data Portal CODM (<https://codm.hereon.de/codm>) described in Breitbach et al., 2016.

COSYNA data management is based on different standards like:

- ISO 19115: <https://www.iso.org/standard/53798.html>
- INSPIRE Directive:  
<https://inspire.ec.europa.eu/>
- ISO8601 standard date and time format:  
<https://www.iso.org/iso-8601-date-and-time-format.html>
- Climate and Forecast Metadata Convention CF-1.6:  
<https://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html>
- OPeNDAP:  
<https://www.opendap.org>
- Open Geospatial Consortium (OGC) for access and delivery of geospatial data:  
<https://www.ogc.org/docs/is>
- OGC Sensor Observation Service (SOS) as webservice to access data from the database used for in-situ data, OGC WMS to visualize geodata
- OceanSITES:  
[http://www.oceansites.org/docs/oceansites\\_data\\_format\\_reference\\_manual.pdf](http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf)
- CMEMS-INSTAC System Requirements Document (SRD):  
<https://archimer.ifremer.fr/doc/00297/40846/>
- EuroGOOS recommendations for in-situ data Near Real Time Quality Control:  
<https://archimer.ifremer.fr/doc/00251/36230/>

The following controlled vocabularies are used as well in the FerryBox data processing:

- National Environment Research Council (NERC) Vocabulary P02:  
[https://vocab.nerc.ac.uk/search\\_nvs/P02/](https://vocab.nerc.ac.uk/search_nvs/P02/)
- National Environment Research Council (NERC) Vocabulary P07:  
[https://vocab.nerc.ac.uk/search\\_nvs/P07/](https://vocab.nerc.ac.uk/search_nvs/P07/)

Hereon FerryBox data from Ferrydata or TS data could be downloaded using SOS. This web service requires the fixed part:

<http://sos.hereon.de/sos.py?request=GetObservation&service=SOS>

In addition, the dynamic URL parameters:

- offering – the name of the offering (for example, Gothenburg-Immingham for the FerryBox operating on the SOO Magnolia on the route Gothenburg-Immingham)
- observedProperty – in most cases the P02 parameter name. The name has 4 capital letters like PSAL for salinity
- eventTime – begin and end of the data as ISO 8601 format

Putting this together the example URL:

<http://sos.hereon.de/sos.py?request=GetObservation&service=SOS&eventTime=2022-01-17T20:39:54.579Z/2022-01-18T13:07:54.579Z&offering=Gothenburg-Immingham&observedProperty=PSAL>

will show the salinity data of the Magnolia transect starting at the 17.01.2022 20:39.

The file responded to the URL above will be:

```
<om:Observation xsi:schemaLocation="http://www.opengis.net/om
http://amb25.stccmop.org/schemas/sos/current/sosGetObservation.xsd" gml:id="PSAL">
<gml:description>None</gml:description>
<gml:name>Gothenburg-Immingham</gml:name>
<gml:location>
<!-- geometry containing all tuples in this observation -->
<gml:Envelope>
<gml:lowerCorner srsName="urn:ogc:def:crs:EPSG:6.5:4326">53.0 -5.0</gml:lowerCorner>
<gml:upperCorner srsName="urn:ogc:def:crs:EPSG:6.5:4326">58.2 12.2</gml:upperCorner>
</gml:Envelope> </gml:location>
<!-- Time of response- use TimePeriod for a series of data
-->
<!-- or TimeInstant for a single measurement -->
<gml:TimePeriod gml:id="DATA_TIME">
<gml:beginPosition>2022-01-17T20:39:54.579Z</gml:beginPosition>
<gml:endPosition>2022-01-18T13:07:54.579Z</gml:endPosition>
</gml:TimePeriod>
<!-- Procedure --><om:procedure/> <!-- the property measured -->
<om:observedProperty xlink:href="PSAL"/>
<!-- Feature Of Interest --> <om:featureOfInterest xlink:href="None"/>
<!-- Result Structure and Encoding --> <om:resultDefinition>
<swe:DataBlockDefinition> <swe:components name="PSAL">
<swe:DataRecord>
<swe:field name="time"><swe:Time definition="urn:ogc:def:phenomenon:time:iso8601"/></swe:field>
<swe:field name="latitude">
<swe:Quantity definition="urn:ogc:def:phenomenon:latitude:wgs84">
<swe:uom code="deg"/></swe:Quantity></swe:field>
<swe:field name="longitude">
<swe:Quantity definition="urn:ogc:def:phenomenon:longitude:wgs84">
<swe:uom code="deg"/></swe:Quantity></swe:field>
<swe:field name="depth">
<swe:Quantity definition="cf:depth">
<swe:uom code="urn:ogc:unit:meter"/></swe:Quantity></swe:field>
<swe:field name="PSAL">
<swe:Quantity definition="PSAL">
<swe:uom xlink:href="urn:mm.def:units#PSU"/></swe:Quantity></swe:field>
<swe:field name="quality flag">
<swe:Quantity definition="SeaDataNet Quality Flag Definition">
<swe:uom xlink:href="http://vocab.ndg.nerc.ac.uk/list/L201/current"/></swe:Quantity></swe:field>
</swe:DataRecord></swe:components>
<swe:encoding><swe:AsciiBlock tokenSeparator="," blockSeparator="|" decimalSeparator="."/>
</swe:encoding></swe:DataBlockDefinition></om:resultDefinition>
<om:result>
2022-01-17T20:45:00Z,53.549791,0.137263,2,30.7159,2|2022-01-
17T20:45:20Z,53.549136,0.139254,2,30.2796,2|2022-01-
17T20:45:40Z,53.548523,0.141172,2,30.0625,2|2022-01-
17T20:46:00Z,53.547919,0.143132,2,30.0368,2|
...
```

```
18T13:07:20Z,55.474905,8.346988,2,28.3699,2
</om:result>
</om:Observation>
```

The response to the SOS URL contains the whole information needed to interpret the data as well as the data itself in <om:result>.

Using the SOS web service, BSH as the data manager of the Northwest Shelf Regional Ocean Observing System (NOOS) transfers the Hereon FerryBox data to CMEMS and EMODnet.

### 3.3.4 Quality Control/Quality Assurance Delayed Mode QC

#### *Near-Real Time QC*

Near Real Time Quality Check is performed after each cruise. File transfer, in general of 500K Data points/week via FTP, is done when a secure connection is established. In general, file transfer can cover about 20 parameters, and 3-6 housekeeping parameters: timestamp, coordinates, quality flags, statistical information (minimum and maximum values, variance, counts).

Near Real Time Quality Check is performed inside the database within one hour after data transfer following the procedures recommended by the EuroGOOS DATA-MEQ working group.

Near Real Time Quality Controlled data ([ferrybox.org](http://ferrybox.org), <http://ferrydata.hereon.de> for fixed routes) are communicated to CMEMS Copernicus Service and EMODnet Portal, and exported in NetCDF format to OceanSITES

#### *Delayed Mode QC*

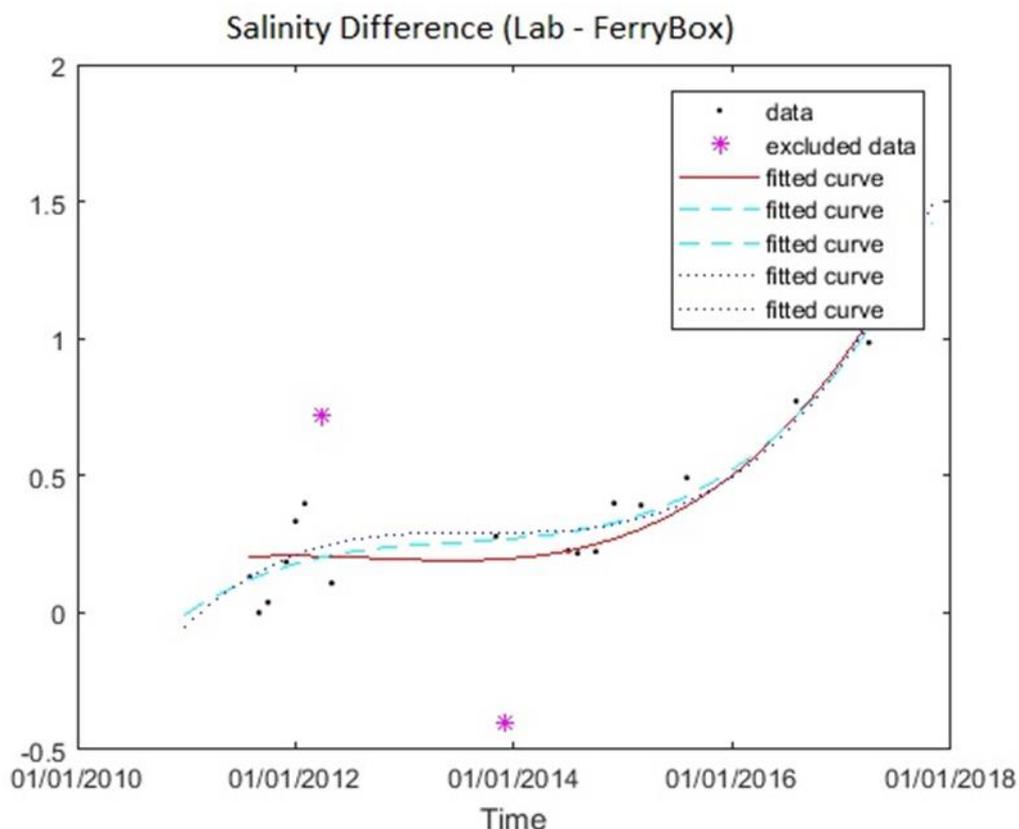
Data for a number of parameters need to be additionally verified using delayed quality controls. These protocols are still under development, and will be further described with some examples below, or (concerning the carbonate system parameters for example) in WP2 deliverables. The methods include: Assessing measurements against multiple sensors in the field. Examples include:

- Temperature
- Chlorophyll fluorescence
- Turbidity

Assessing measurements against laboratory samples for salinity, turbidity, chlorophyll a fluorescence, dissolved oxygen, nutrients. The QC-ed results are then published on the Pangaea or similar PID databases. An example is shown here: Petersen et al. 2017, <https://doi.pangaea.de/10.1594/PANGAEA.883824>.

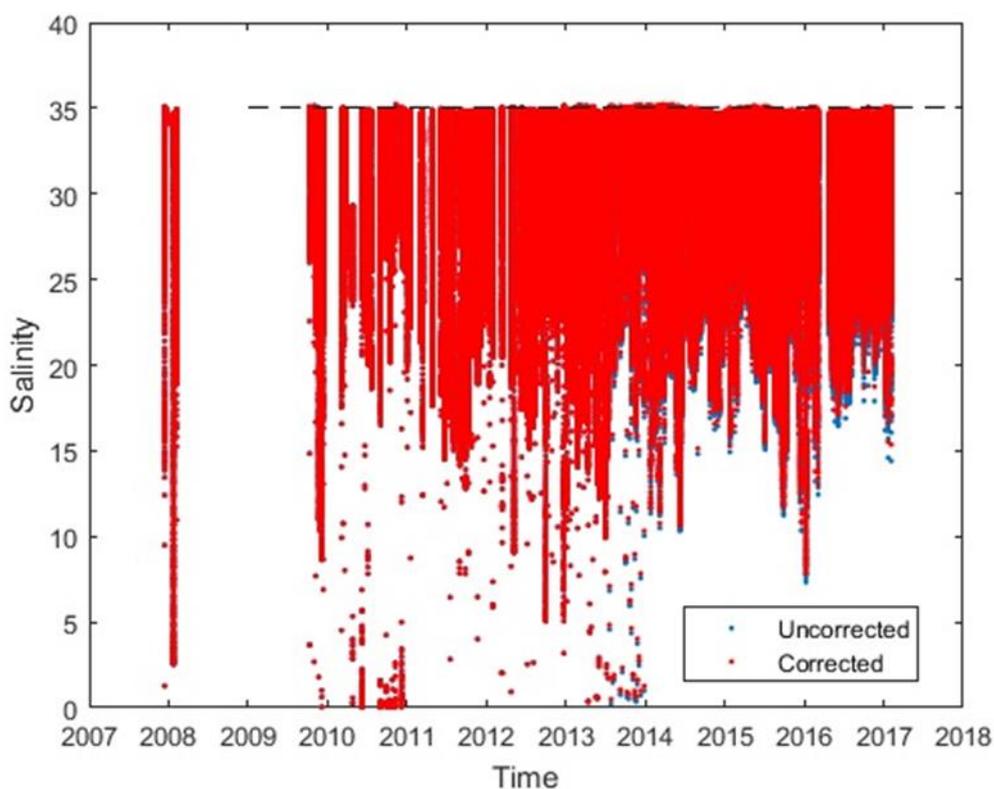
In another Pangaea dataset, Macovei et al. (2021) published example data for pCO<sub>2</sub>, salinity and temperature: <https://doi.pangaea.de/10.1594/PANGAEA.930383>. In this dataset is an example of a long term deployments on FerryBox platforms, including multiple sensors, and respective drifts. Salinity, measured by the FSI sensor (Teledyne, USA) on the CV Lysbris Seaways for example, started drifting around the middle of July, 2013 and until April, 2017, when the sensor was finally changed. The drift was not linear, and it was difficult to apply an offset to the salinity measurements, since the ferry transect varied in time and space. Instead, to correct for the drift, salinity samples were used, collected along the Halden-Bruges/Gent FerryBox route, and measured in the lab by a high-precision salinometer (OPTIMARE, Bremerhaven, Germany). It is recommended that this type of high-precision laboratory measurements are available to check salinity measurements in the field, if the instruments measuring salinity in the FerryBox are not calibrated frequently (for example every year).

The difference between the lab measurements and the FerryBox measurements was used to calculate a drift in the FSI measurements. There was a large spread (from -11.0 to +9.4) in the individual sample differences for all salinities (out of 291 samples) for salinities from 0 to 35. Most likely the large spread was related to the timing of sample collection. The FerryBox measurements were collected in random regions along the ship route, and sometimes in areas with large salinity gradients. The route between Halden and Bruges-Gent passes through the ports of Gent, Bruges and Halden, where salinity reaches 0. At the same time in the nearshore regions of the Skagerrak and Rhine River outflow, salinity fluctuations frequently vary. When sample water is collected along a rapidly changing gradient, it is possible that the comparison to the lab measurement will not match. Thus, only the samples with salinity more than 28 were considered in this comparison, so the average lab-FerryBox difference for each route was used, and this difference was plotted over time (see **Fig. 3.3.4.1**).



**Figure 3.3.4.1.** Difference between lab measured and FerryBox measured salinity of select samples with salinity > 28 (black points). Also shown are polynomial fits, to determine a change in the salinity difference with time. Outliers are shown (stars).

Outliers, determined via a median average difference, were not considered in the fit. A polynomial fit was used to this difference in measurements with time, and the correction was applied to the salinity measurements after the middle of July, 2013 and up to April 2017, to obtain the corrected values in **Fig. 3.3.4.2**.



**Figure 3.3.4.2.** Corrected and uncorrected salinity for Immingham-Moss route on Lysbris Seaways FerryBox platform, operated by Hereon. Difference between lab-based and FerryBox salinity measurements from an FSI instrument were applied to correct the salinity records, which are now published by Macovei et al. 2021, at <https://doi.pangaea.de/10.1594/PANGAEA.930383>.

Assessing other parameters according to instrument calibration data, with necessary pre- and post-calibration applications. Examples include:

- $p\text{CO}_2$ , which includes comparison to standard SOCAT database measurements (Macovei et al. 2020)
- total alkalinity and pH, which includes comparison to collected samples measured in the laboratory (Voynova et al. 2019). Measurements are standardized using reference materials.
- automated nutrient measurements, which include assessment to additional nutrient samples measured in the laboratory, as well as examination of individual periodic calibrations and check standards. Measurements are standardized using reference materials.

Web-based tools available at the European FerryBox Database allow for data visualization and help with Delayed Quality Control These are tools developed specifically for the use of underway observations (**Fig. 3.3.4.3**).



Figure 3.3.4.3. Visualization tools available at the European FerryBox Database.

- Transect Plot I: one or more parameter along one transect (one cruise). Flagged data from realtime quality are marked in red (Fig. 3.3.4.4):

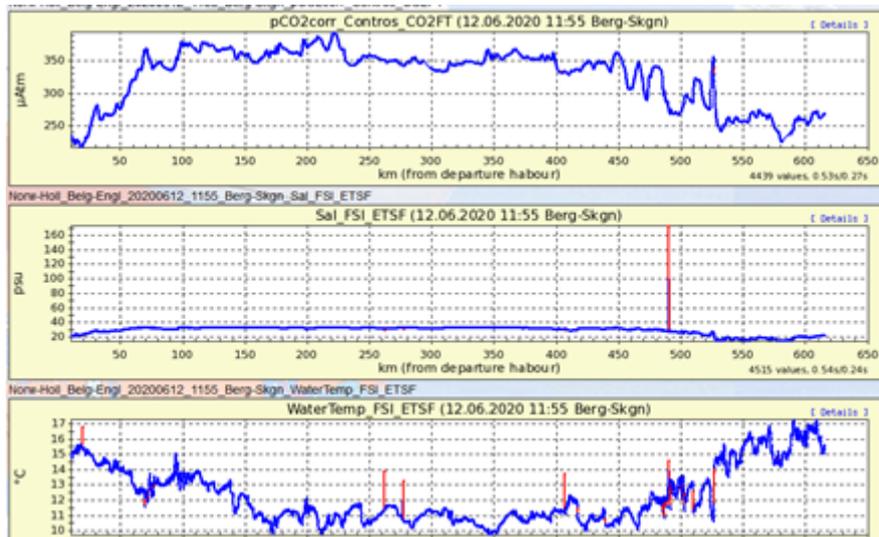
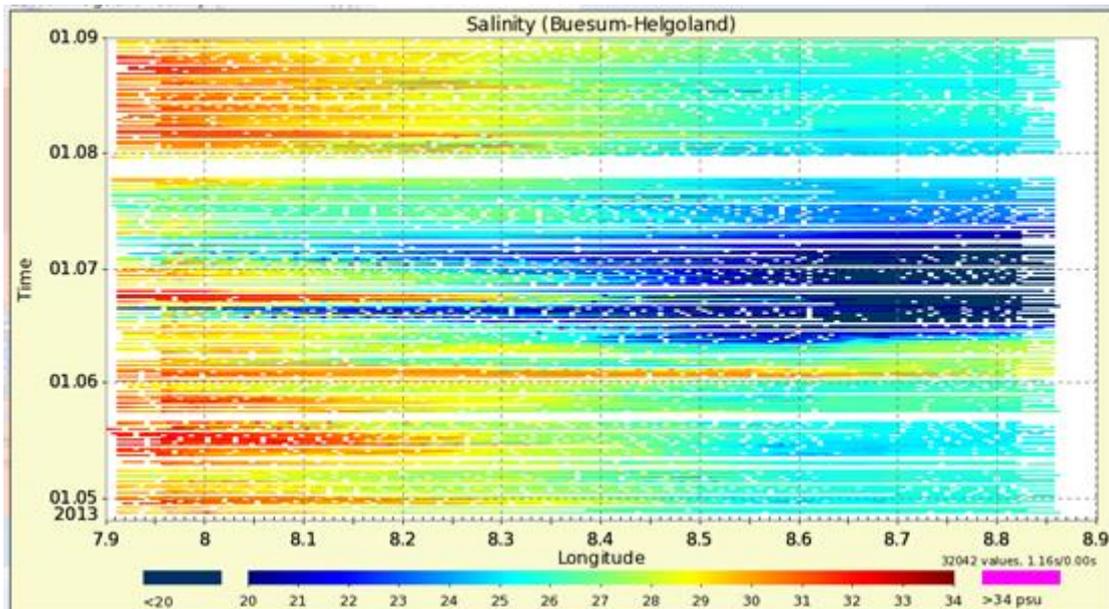


Figure 3.3.4.4. pCO<sub>2</sub>, salinity, temperature along Bergen-Skogn route in June, 2020

- Transect Plot II: one parameter along multiple transects (comparison of repeated measurement)
- Scatter Plot: Hovmöller plot: Plotting data (color coded) vs. position (e.g. distance, longitude or latitude) and time. Flood event in 2013 was visualized in the North Sea coast (Fig. 3.3.4.5).



**Figure 3.3.4.5.** Hovmöller plot of salinity in the German Bight. Figure was published in Callies et al. 2021.

- Map Plot: Plotting data as color-coded information in a map
- Time Series Plot: Time series of one or more parameter at a certain position

From transect plot I it is possible to visualize the data as SOS V2 (<https://52north.org/software/software-projects/sos/>) using the menu details of the plot. This will lead to a visualization like **Figure 3.3.4.6**. Please note that only http is available.



**Figure 3.3.4.6.** SOS V2 plot of the transect Immingham-Esbjerg (22/01/2022). Shown are the measured data for oxygen saturation in a spatial as well as temporal representation.



### 3.3.5 Issues

One limitation of FerryBox systems is that they provide only data on surface water properties and, if necessary, have to be complemented with depth profiles obtained by conductivity, temperature and depth (CTD) measurements from research vessels or buoys. Another disadvantage is that a SOO cannot stop along its route to sample additional stations and depths, as a research vessel is able to. Also, often SOO change routes and may not be available in a certain region of scientific interest (Petersen et al. 2011).

Moving platforms can change routes, therefore if a port is not identified, data are not transferred to the database. This requires manual verification, and subsequent addition of ports. One way to solve this would be to use AIS data, with many ports identified in advance.

Data in the database is sorted according to routes, therefore with some limitations, like the necessity to define ports, which are used in the working files. One way is to avoid this, is to not predefine ports, which is already in place for FerryBox data from research vessels.

### 3.3.6 Training Materials and Contacts

A practical exercise how to use the quality assessment implemented inside the European FerryBox Database [ferrydata.hereon.de](http://ferrydata.hereon.de) during the 7<sup>th</sup> FerryBox Workshop ([https://ferrybox.org/imperia/md/images/hzg/institut\\_fuer\\_kuestenforschung/koi/ferrybox/programm.pdf](https://ferrybox.org/imperia/md/images/hzg/institut_fuer_kuestenforschung/koi/ferrybox/programm.pdf)) was given by M. Gehrung and G. Breitbach.

## 4. ANNEXES AND REFERENCES

- [1] Aßmann, S., C. Frank, A. Kötzinger, Spectrophotometric high-precision seawater pH determination for use in underway measuring systems Ocean Sci. Discuss., 8 (3) (2011), pp. 1339-1367
- [2] Barrick, D.E., Lipa, B.J. and Crissman, R.D. (1985), Mapping surface currents with CODAR, Sea Technol., 43–48.
- [3] Bartolomé Garau, Simón Ruiz, Weifeng G. Zhang, Ananda Pascual, Emma Heslop, John Kerfoot, and Joaquín Tintoré (2011), Thermal Lag Correction on Slocum CTD Glider Data. <https://doi.org/10.1175/JTECH-D-10-05030.1>
- [4] Breitbach, G., Krasemann, H., Behr, D., Beringer, S., Lange, U., Vo, N., and Schroeder, F.: Accessing diverse data comprehensively – CODM, the COSYNA data portal , Ocean Sci., 12 (2016), 909–923, <https://doi.org/10.5194/os-12-909-2016>
- [5] C. Troupin, J.P. Beltran, E. Heslop, M. Tornera, B. Garau, J. Allen, S. Ruiz, J.Tintoré (2015). <https://doi.org/10.1016/j.mio.2016.01.001>
- [6] Callies U, M Kreus, W Petersen, YG Voynova (2021). On Using Lagrangian Drift Simulations to Aid Interpretation of in situ Monitoring Data. Frontiers in Marine Science, 8:666653. DOI: 10.3389/fmars.2021.666653.
- [7] Carval Thierry, Gourcuff Claire, Rannou Jean-Philippe, Buck Justin J.H., Garau Bartolome (2021), EGO gliders NetCDF format reference manual. <http://dx.doi.org/10.13155/34980>
- [8] Castelão, G. P., (2020). A Framework to Quality Control Oceanographic Data. Journal of Open Source Software, 5(48), 2063, <https://doi.org/10.21105/joss.02063>
- [9] Cauchy, P., Heywood, K. J., Merchant, N. D., Queste, B. Y., and Testor, P. (2018). Wind speed measured from underwater gliders using passive acoustics. J. Atmos. Oceanic Technol. 35, 2305–2321. <https://doi.org/10.1175/JTECH-D-17-0209.1>
- [10] Chapman, R. D., Shay, L. K., Graber, H. C., Edson, J. B., Karachintsev, A., Trump, C. L., and Ross, D. B. (1997), On the accuracy of HF radar surface current measurements: Intercomparisons with ship-based sensors, J. Geophys. Res., 102(C8), 18737– 18748, [doi:10.1029/97JC00049](https://doi.org/10.1029/97JC00049).
- [11] Copernicus Marine In Situ Tac (2020). 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. <https://doi.org/10.13155/41256>
- [12] Copernicus Marine In Situ Tac Data Management Team (2021). Copernicus Marine In Situ NetCDF format manual. <https://doi.org/10.13155/59938>
- [13] Copernicus, EGO, OceanSITES, Argo Data Management (2019) NetCDF file format checker for Argo floats, Copernicus In Situ TAC, EGO gliders, OceanSITES. Version 1.15 with new rule files. Plouzane, France, SEANOE. DOI: <https://doi.org/10.17882/45538>
- [14] Corgnati, L., Mantovani, C., Novellino, A., Jousset, S., Cramer, R. N., and Thijsse, P (2019), SeaDataNet data management protocols for HF Radar data, WP9 - Deliverable D9.12. Version 1.6. SeaDataNet, 83pp. DOI: <http://dx.doi.org/10.25607/OBP-1011>.
- [15] Corgnati, L., Mantovani, C., Novellino, A., Rubio, A. and Mader, J. (2018), Recommendation Report 2 on improved common procedures for HFR QC analysis. JERICO-NEXT WP5-Data Management, Deliverable 5.14, Version 1.0. Brest, France, IFREMER, 82pp, (JERICO-NEXT-WP5-D5.14-V1.). DOI: <http://dx.doi.org/10.25607/OBP-944>
- [16] Crombie, D. (1955), Doppler Spectrum of Sea Echo at 13.56 Mc./s.. Nature 175, 681–682, [doi:10.1038/175681a0](https://doi.org/10.1038/175681a0)

- [17]Gregor L, Ryan-Keogh TJ, Nicholson S-A, du Plessis M, Giddy I and Swart S (2019) GliderTools: A Python Toolbox for Processing Underwater Glider Data. *Front. Mar. Sci.* 6:738. [doi: https://doi.org/10.3389/fmars.2019.00738](https://doi.org/10.3389/fmars.2019.00738)
- [18]Gurgel, K.W., Antonischki, G., Essen, H.H., and Schlick, T. (1999), Wellen radar (WERA): A new ground wave radar for remote sensing. *Coastal Engineering*, 37, 219–234, [doi:10.1016/S0378-3839\(99\)00027-7](https://doi.org/10.1016/S0378-3839(99)00027-7)
- [19]How to use Argo profile files. <https://argo.ucsd.edu/data/how-to-use-argo-files/>
- [20]John Kerfoot. <https://marine.rutgers.edu/~kerfoot/slocum/masterdata/source>
- [21]John T. Allen, Cristian Munoz, Jim Gardiner, Krissy A. Reeve, Eva Alou-Font and Nikolaos Zarokanellos (2020), Near-Automatic Routine Field Calibration/Correction of Glider Salinity Data Using Whitespace Maximization Image Analysis of Theta/S Data. <https://doi.org/10.3389/fmars.2020.00398>
- [22]Johnson, D.R., Boyer, T.P., Garcia, H.E., Locarnini, R.A., Baranova, O.K. and Zweng, M.M. (2013) World Ocean Database 2013 User's Manual. Silver Spring, MD, National Oceanographic Data Center, 172pp. (National Oceanographic Data Center Internal Report 22). DOI:10.7289/V5DF6P53.
- [23]Kim, S. Y., Terrill, E., and Cornuelle, B. (2007), Objectively mapping HF radar-derived surface current data using measured and idealized data covariance matrices, *J. Geophys. Res.*, 112, C06021, [doi:10.1029/2006JC003756](https://doi.org/10.1029/2006JC003756).
- [24]Lips, I., U. Lips, Abiotic factors influencing cyanobacterial bloom development in the Gulf of Finland (Baltic Sea) *Hydrobiologia*, 614 (2008), pp. 133-140
- [25]Macovei, VA, YG Voynova, M Becker, J Triest, W Petersen. Long-term intercomparison of two pCO<sub>2</sub> instruments based on ship-of-opportunity measurements in the Skagerrak, *Limnology and Oceanography: Methods*, 19(1): 37-50, DOI: 10.1002/lom3.10403.
- [26]Macovei, VA, YG Voynova, M Gehrung, W Petersen. 2021. Ship-of-Opportunity, FerryBox-integrated, membrane-based sensor pCO<sub>2</sub>, temperature and salinity measurements in the surface North Sea since 2013. PANGAEA, DOI: 10.1594/PANGAEA.930383
- [27]Mantovani, C., Corgnati, L., Horstmann, J., Rubio, A., Reyes, E., Quentin, C., Cosoli, S., Asensio, J.L., Mader, J. and Griffa, A. (2020), Best Practices on High Frequency Radar Deployment and Operation for Ocean Current Measurement, *Front. Mar. Sci.* 7:210, [doi: 10.3389/fmars.2020.00210](https://doi.org/10.3389/fmars.2020.00210)
- [28]Marasco, M., Reyes, E., Fernández, J.G., Charcos, Notario, X., M., Rotllán, P., Rújula, M.A., Tintoré, J. (2021). SOCIB - Coastal High Frequency Radar Data Management Plan (Version 1.0). Balearic Islands Coastal Observing and Forecasting System, SOCIB. DOI: <https://doi.org/10.25704/ydas-qz53>
- [29]Marasco, M., Zarokanellos, N.D., Charcos, M., Fernández, J.G., Miralles, A., Rotllán, P., Tintoré, J. (2021), SOCIB Glider - Canales Endurance Line Data Management Plan (Version 1.0). Balearic Islands Coastal Observing and Forecasting System, SOCIB. <https://doi.org/10.25704/vj13-eh49>
- [30]OceanGliders 1.0 Harmonizing format across OceanGliders Terms of References. [https://github.com/OceanGlidersCommunity/OG-format-user-manual/blob/main/OG\\_Format.adoc](https://github.com/OceanGlidersCommunity/OG-format-user-manual/blob/main/OG_Format.adoc)
- [31]Paris. Intergovernmental Oceanographic Commission of UNESCO. 2013. Ocean Data Standards, Vol.3: Recommendation for a Quality Flag Scheme for the Exchange of

- Oceanographic and Marine Meteorological Data. (IOC Manuals and Guides, 54, Vol. 3.) 12 pp.  
[https://www.nodc.noaa.gov/oads/support/MG54\\_3.pdf](https://www.nodc.noaa.gov/oads/support/MG54_3.pdf)
- [32] Petersen, W. and F. Colijn, eds. FerryBox White Paper. EuroGOOS Publication 2017
- [33] Petersen, W., F. Schroeder, F.-D. Bockelmann, FerryBox — application of continuous water quality observations along transects in the North Sea Ocean Dyn., 61 (2011), pp. 1541-1554
- [34] Petersen, W., H. Wehde, H. Krasemann, F. Colijn, F. Schroeder, FerryBox and MERIS — assessment of coastal and shelf sea ecosystems by combining in situ and remotely sensed data, Estuar. Coast. Shelf Sci., 77 (2) (2008), pp. 296-307
- [35] Pierre Testor, Brad de Young, Daniel L. Rudnick, Scott Glenn, Daniel Hayes, Craig M. Lee, Charitha Pattiaratchi, Katherine Hill, Emma Heslop, Victor Turpin, Pekka Alenius, Carlos Barrera, John A. Barth, Nicholas Beaird, Guislain Bécu, Anthony Bosse, François Bourrin, J. Alexander Brearley, Yi Chao, Sue Chen, Jacopo Chiggiato, Laurent Coppola, Richard Crout, James Cummings, Beth Curry, Ruth Curry, Richard Davis, Kruti Desai, Steve DiMarco, Catherine Edwards, Sophie Fielding, Ilker Fer, Eleanor Frajka-Williams, Hezi Gildor, Gustavo Goni, Dimitri Gutierrez, Peter Haugan, David Hebert, Joleen Heiderich, Stephanie Henson, Karen Heywood, Patrick Hogan, Loïc Houpert, Sik Huh, Mark E. Inall, Masso Ishii, Shin-ichi Ito, Sachihiko Itoh, Sen Jan, Jan Kaiser, Johannes Karstensen, Barbara Kirkpatrick, Jody Klymak, Josh Kohut, Gerd Krahnmann, Marjolaine Krug, Sam McClatchie, Frédéric Marin, Elena Mauri, Avichal Mehra, Michael P. Meredith, Thomas Meunier, Travis Miles, Julio M. Morell, Laurent Mortier, Sarah Nicholson, Joanne O'Callaghan, Diarmuid O'Conchubhair, Peter Oke, Enric Pallàs-Sanz, Matthew Palmer, JongJin Park, Leonidas Perivoliotis, Pierre-Marie Poulain, Ruth Perry, Bastien Queste, Luc Rainville, Eric Rehm, Moninya Roughan, Nicholas Rome, Tetjana Ross, Simon Ruiz, Grace Saba, Amandine Schaeffer, Martha Schönau, Katrin Schroeder, Yugo Shimizu, Bernadette M. Sloyan, David Smeed, Derrick Snowden, Yumi Song, Sebastian Swart, Miguel Tenreiro, Andrew Thompson, Joaquin Tintore, Robert E. Todd, Cesar Toro, Hugh Venables, Taku Wagawa, Stephanie Waterman, Roy A. Watlington and Doug Wilson (2019), OceanGliders: A Component of the Integrated GOOS.  
<https://doi.org/10.3389/fmars.2019.00422>
- [36] Pinot, Jean-Michel; Vélez-Belchí, Pedro; Tintoré, Joaquín; López-Jurado, José Luis (1997), The thermal-lag effect in SBE-25 CTDs: Importance of correcting data collected in the Mediterranean summer thermocline. <http://hdl.handle.net/10261/28196>
- [37] Reiner Schlitz (2013), Oceanographic quality flag schemes and mappings between them. [https://odv.awi.de/fileadmin/user\\_upload/odv/misc/ODV4\\_QualityFlagSets.pdf](https://odv.awi.de/fileadmin/user_upload/odv/misc/ODV4_QualityFlagSets.pdf)
- [38] Reyes, E., Rotllán-García, P., Rubio, A., Corgnati, L., Mader, J., and Mantovani, C. (2019), Guidelines on how to sync your High Frequency (HF) radar data with the European HF Radar node (Version 1.2), Balearic Islands Coastal Observing and Forecasting System, SOCIB, [doi:10.25704/9XPF-76G7](https://doi.org/10.25704/9XPF-76G7).
- [39] Rubio, A., Reyes, E., Mantovani, C., Corgnati, L., Lorente, P., Solabarrieta, L., Mader, J., Fernandez, V., Pouliquen, S., Novellino, A., Karstensen, J., & Petihakis, G. (2021), European High Frequency Radar network governance: Vol. D3.4 (EuroSea Deliverable, Issue doi:10.3289/eurosea\_d3.4). DOI: [https://doi.org/doi:10.3289/eurosea\\_d3.4](https://doi.org/doi:10.3289/eurosea_d3.4)
- [40] Sandy J. Thomalla, William Moutier, Thomas J. Ryan-Keogz, Luke Gregor, Julia Schutt (2018), An optimized method for correcting fluorescence quenching using optical backscattering on autonomous platforms. [doi: 10.1002/lom3.10234](https://doi.org/10.1002/lom3.10234)



- [41]Schmidt, R. (1986), Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag., AP-34, no. 3, 276–280.
- [42]Tanhua, T., Pouliquen, S., Hausman, J., O'Brien, K., Bricher, P., de Bruin, T., Buck, J. J. H., Burger, E. F., Carval, T., Casey, K. S., Diggs, S., Giorgetti, A., Glaves, H., Harscoat, V., Kinkade, D., Muelbert, J. H., Novellino, A., Pfeil, B., Pulsifer, P. L., ... Zhao, Z. (2019), Ocean FAIR Data Services. *Frontiers in Marine Science*, 6. DOI: <https://doi.org/10.3389/fmars.2019.00440>
- [43]U.S. Integrated Ocean Observing System (2016), Manual for Real-Time Quality Control of High Frequency Radar Surface Current Data: a Guide to Quality Control and Quality Assurance for High Frequency Radar Surface Current Observations. Version 1.0. Silver Spring, MD, U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Integrated Ocean Observing System, 58pp. DOI: <https://doi.org/10.25607/OBP-1464>
- [44]Voynova, YG, H Brix, W Petersen, S Wiegelt-Krenz, M Scharfe (2017). Extreme flood impact on estuarine and coastal biogeochemistry: the 2013 Elbe flood, *Biogeosciences*, 14: 541-557, DOI: 10.5194/bg-14-541-2017.
- [45]Voynova, YG, W Petersen, M Gehrung, S Aßmann, AL King (2019). Intertidal regions changing coastal alkalinity: the Wadden Sea-North Sea tidally-coupled bioreactor, *Limnology and Oceanography*, 64(3): 1135-1149, DOI: 10.1002/lno.11103.