Joint European Research Infrastructure network for Coastal Observatory – Novel European eXpertise for coastal observaTories - **JERICO-NEXT**

Deliverable title	Specifications for a European FerryBox data management system
Work Package Title	WP 5 – Data Management
Deliverable number	D5.3
Description	Document describing the roadmap for FerryBox data at a European level, from data provider to end user.
Lead beneficiary	SHMI
Lead Authors	Johanna Linders
Contributors	
Submitted by	J.Linders
Revision number	V1.1
Revision Date	04/09/17
Security	Public

The JERICO-NEXT project is funded by the European Commission's H2020 Framework Programme under grant agreement No. 654410
Project coordinator: Ifremer, France.

History			
Revision	**Date**	**Modification**	**Author**
1.0	24/08/2017	First version	J.Linders and Dick M. A. Schaap
1.1	04/09/2017	Including SeaDataNet in the roadmap	J.Linders and Dick M. A. Schaap

Approvals				
	Name	**Organisation**	**Date**	**Visa**
Coordinator	Patrick Farcy	IFREMER	12.09.17	PF
WP Leaders	Leonidas Perivoliotis	HCMR	08.09.17	LP
	Patrick Gorringe	EuroGOOS	11.09.17	PG

Table of contents

1. Acknowledgement

This document is partly based on work done by the Ferrybox Task Team.

2. Executive Summary

2.1. Present availability of FerryBox data in near real time and delayed mode

Today observational data from FerryBoxes in both near real time and delayed mode is collected and stored by different data providers. The stored data may differ in both format and in quality controls performed. Some of the observations are made public and shared with open access, others are not. It is time-consuming for the users to visit different sites/data bases from different data providers and also complicated if the data offered is in different formats. This is applicable for most marine observational data and the reason for international databases/data like EMODnet, SeaDataNet (delayed mode) and CMEMS (near real time and delayed mode) with standardized formats and quality controls.

The Copernicus Marine Environmental Monitoring Service (CMEMS) provides open access to marine data and information, with both a regional European and a global perspective. CMEMS focuses on monitoring and forecasting. The online service includes satellite data and products, marine observational data and model data. The different data types are offered in a standard netCDF format to ensure compatibility and continuity. The observational data is managed by CMEMS's In Situ TAC (Thematic Assembly Centre), a cooperation by the Regional Ocean Observing Systems (ROOS's) in Europe. The ROOS's are represented by one national institute respectively, the dissemination unit, in the In Situ TAC. The ROOS's, via the dissemination units, have the responsibility to collect, encourage to apply open access and disseminate to CMEMS, any marine observational data, including FerryBox data, that can be of interest in their regions and to make sure that standardized automatic quality controls have been performed. The strength of CMEMS is that the observational data is updated daily at the CMEMS site in netCDF format and has undergone specified quality controls, data also relay on a set parameter list stating standardized parameter names and units. By using the CMEMS's track FerryBox data follow how other open marine observational data in Europe is handled and made available. Ferrybox data benefit from the automatic quality controls and standard formats and is straightforward to handle. It is easy to use with all other types of data that is made available online on the CMEMS site.

While CMEMS is the route, in particular, for near real time data, SeaDataNet offers the possibility of delayed mode quality control and long term storage, discovery and access of validated data sets. SeaDataNet has a close cooperation with CMEMS and they work together on climatologies which are used for calibrating numerical models. SeaDataNet, CMEMS and EuroGOOS are the 3 main pillars under the EMODnet Physics portal and bring together past, present, and future data for a major part of all European physical oceanographic observations. Data collected by FerryBoxes is not much different than data collected by other operational systems and benefits from use of the existing European infrastructure and capacities of both the operational data nodes and the delayed mode data centres.

The HZG FerryBox Database, established at the Institute of Coastal Research at HZG, can be an alternative, a Common European FerryBox Database, for those data providers that don't' provide their data to CMEMS or SeaDataNet. The reasons can be that the data providers for some reason don't want their data to be disseminated via the ROOS's nor giving open access to the data. A Common European Ferrybox Database may then step in and store the data and even perform quality controls or offer the online tools to the data provider before making the data available to CMEMS, if data is marked as open access. Data is stored as transects and is easily viewed in the online software and compared with other transects.

2.2. Standardized automatic quality control

In CMEMS the dissemination units perform automatic quality controls in near real time (and on historical data) to assure e.g. presence of valid time and position stamps, but also to search for e.g. values out of range, missing values and if the flow through at the sea water intake is to small. All observations having undergone the quality controls get a quality control flag to state that the quality controls have been performed and the quality status of the observation. The quality controls are further described in the CMEMS Quality Information Document (http://cmems-resources.cls.fr/documents/QUID/CMEMS-INS-QUID-013-030-036.pdf). Observations not available through CMEMS need quality control documentation to be linked to the data set explaining the routines being performed.

2.3. Possibility to update data: more quality controls and more parameters in delayed mode

The need of observational data in near real time makes the minimizing of the time step between the time of the taken observation and the time when the observations is made available via open access is very important. The first quality controls have to been performed before publishing the data in near real time, but to be able to ensure that FerryBox data is of high quality, more quality controls in delayed mode are needed to be able to see long-term trend and to validate with observations from other platforms, also more parameters may be added in delayed mode depending on performed analysis. Hence it is of importance that the systems/databases providing the FerryBox data to the users are flexible and allows for frequently updates of the data sets to ensure that the data with the most recent quality controls and analysed parameters are available. CMEMS In Situ TAC is looking for the most recent quality controlled data to disseminate and SeaDataNet is provided with validated dataset from the National Institutes. In Jerico-NEXT, SMHI will deliver open source software to investigate and perform the manual analysis for e.g. FerryBox data in deliverable 5.12 (*Software for QC of biochemical data from FerryBox and fixed platforms*) that can be used by data providers and others to investigate FerryBox data and compare with other platforms.

2.4. Acceptance and availability of new data types

The present development of sensors applicable for use with Ferryboxes is an ongoing procedure and the need of acceptance for new parameters and types of data is important, e.g. photos taken with cytocams to identify phytoplankton automatically/manually is a data type not really made way for yet in some of the international databases. It also enhances the need of comparing biological, chemical and physical observations in a smooth way.

3. Introduction

3.1. Ferryboxes provides observations in a wider data coverage both spatially and temporally

Monitoring programs with research vessel and extensive observations, water sampling and analysis are very important to get a qualitative view of the environmental situation but they are also expensive and time consuming. It is common to wish for a more highly frequent monitoring program to be able to resolve features seen in the basic monitoring program, the FerryBox soloution is then a very good complement. The FerryBoxes, underway sampling by sensors attached to a sea water intake below the hull, provides observations in a wider data coverage both spatially and temporally, than is possible with the regular monitoring programs. While many ferries have regular routes covering the same area serval of times each week or month, a more high resolute picture of the covered marine system appears. This is of importance to be able to more in detail study e.g. patterns and life cycles of algae blooms that may occur in between the regular monitoring services. By installing FerryBoxes on commercial shipping vessels and, upcoming, on private ships and vessels, areas not covered by the monitoring programs may also be reached. And by new technics more analysis can be done underway even though the water samples taken underway serves as important validation material and as samples for analysis to be performed in laboratory.

3.2. Near Real Time FerryBox data

There are six EuroGOOS Regional Ocean Observing Systems (ROOS's) each covering a specific region see Fig.1. The ROOS's have contact with the national data centres, institutes and regional data providers to have an ongoing dialogue about the monitoring and modelling situation in the region and identify/collect observational data available for downloading both in near real time and in delayed mode.

Figure 1 Geographically divided regions of the six Regional Ocean Observing Systems (ROOS's) in Europe, with associated institute as they are represented in INSTAC CMEMS, including the global region manage by the coordinator.

The ROOS's provide near real time observational data to the Copernicus Marine Environmental Monitoring Service. Copernicus is a European Union Program to monitoring the Earth, the Copernicus Marine Environmental Monitoring Service (CMEMS) provides open access to marine data and information, with both a regional European and a global perspective. CMEMS focuses on monitoring and forecasting. The online service includes satellite data and products, marine observational data and model data. The different data types are offered in netCDF format to ensure compatibility and continuity. The observational data is managed by the CMEMS In Situ TAC (Thematic Assembly Center), a cooperation by the Regional Ocean Observing Systems (ROOS's). Each ROOS's is represented by a national institute from each region respectively also known as the dissemination units in the CMEMS In Situ TAC. The ROOS's, via the dissemination units, have the responsibility to collect, encourage to apply open access and disseminate to CMEMS any marine observational data that can be of interest in their regions and to make sure that standardized quality controls have been performed. The observational data is updated daily at the CMEMS site in netCDF format and relay on a parameter list stating standardized parameter names and units. The ROOS's are the link between the regional data providers and CMEMS to achieve a closer and active communication with the data providers and within the region.

3.3. Delayed mode FerryBox data

Both CMEMS and SeaDataNet includes delayed mode data. Especially SeaDataNet is focused on long term storage and is provided with validated data set of FerryBox data from the national institutes and has an obvious role when it comes to delayed mode validation of FerryBox data. SeaDataNet has a close cooperation with CMEMS and they work together on climatologies which are used for calibrating numerical models.

3.4. Acceptance of new data types e.g. automated analysis of plankton data

New data types e.g. photos taken with cytocams to identify phytoplankton automatically/manually is constantly developed and need to be handle and made available in a similar way as the more standard parameters to increase the compatibility and enhance the possibility to easy include different data in analysis. This is why acceptance for new parameters and types of data is so important and cannot be handled by individual institutes, instead international resolutions has to be approved.

4. Main report

4.1. Main road map for data management system

The national institutes that collect the FerryBox data are in charge of the master data. The most recent quality controlled data are to be stored there and they should have the responsibility to provide their respective ROOS with access to the open data, but also to communicate larger findings of errors and enhancements. A free and open source software to investigate and quality control FerryBox data will be provided by SMHI within the Jerico-NEXT in deliverable 5.12 (*Software for QC of biochemical data from FerryBox and fixed platforms*) to simplify the quality work in delayed mode for e.g. national institutes and data providers. The dissemination unit, the national institute in respective ROOS working as region leader in CMEMS In Situ TAC, collect and update their data files daily and are responsible to provide CMEMS with automatic quality controlled FerryBox data and make it available in near real time for the end users. CMEMS is updated daily, and hence the available data is updated and may include new quality flags if provided by the data providers in delayed mode.

SeaDataNet get deliverables of validated dataset from the national institutes including FerryBox data.

A Common European FerryBox Database can be an alternative for those data providers that don't' provide their data to CMEMS. For some reason maybe the data provider doesn't want their data to be disseminated via the

ROOS's or to give open access to the data. The HZG FerryBox Database as a Common European Database may then step in and store the data and even perform quality controls before making the data available to CMEMS, if data is marked as open access. Data can also be worked with and compared within the software offered together with the database.

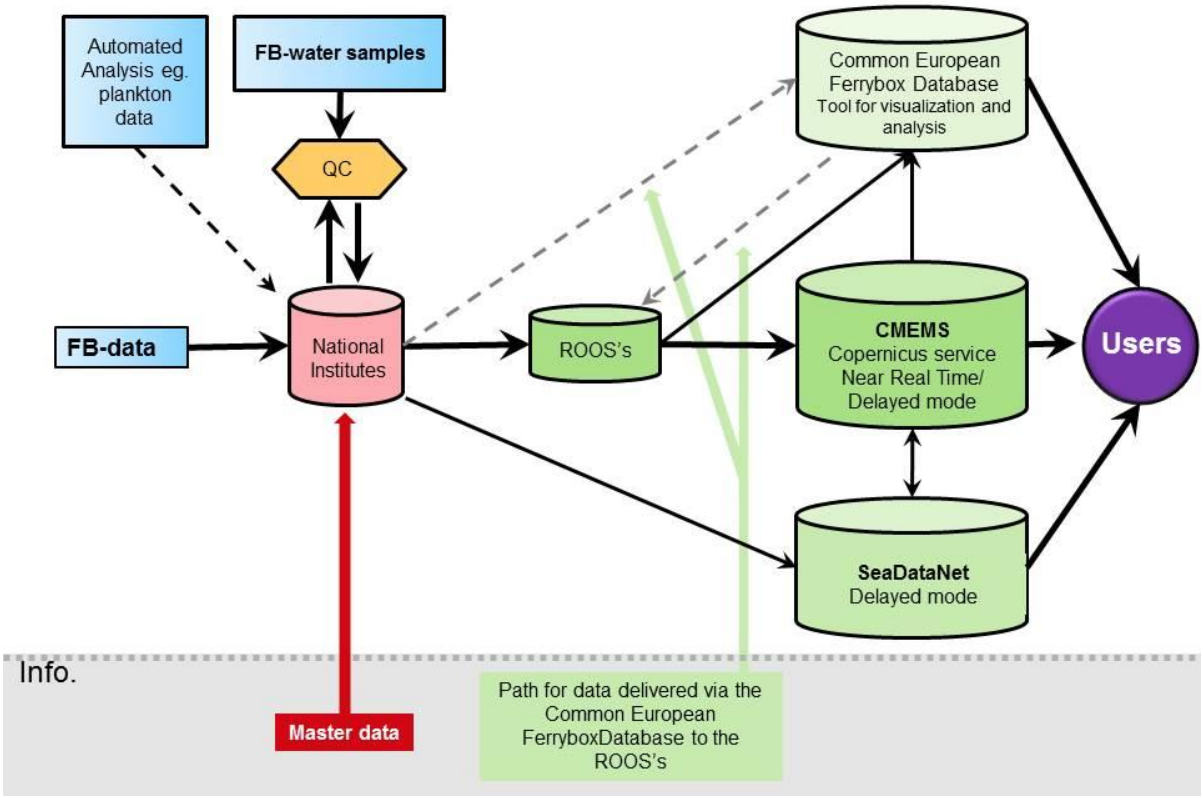

Figure 2 Overview of the FerryBox data roadmap, from data provider to the end user

4.2. Different formats of incoming Ferrybox data

Data from FerryBoxes can be received in different ways and both in near real time and in delayed mode. Common ways are through ftp, mail or collected on board from local servers. It depends on the vessel, the routes and data provider how data is delivered. The dissemination units in In Situ TAC CMEMS handle already a wide range of data formats and converts them to the standardized netCDF format.

While collecting the data, meta data such as type and version of sensors, calibration date and inter-calibrations done, are of interest to link to the data, to be able to use the data properly in future and facilitate comparison with same parameters from other platforms.

4.3. Automatic quality controls

Within CMEMS the dissemination units perform automatic quality controls in near real time to assure the right format of the data sets, e.g. presence of valid time and position stamps and also search for unrealistic values and missing values. All observations get a quality control flag to state if any quality controls have been performed and

to identify the quality status of the data. The quality controls are described in more detailed in the CMEMS Quality Information Document (http://cmems-resources.cls.fr/documents/QUID/CMEMS-INS-QUID-013-030-036.pdf).

For the FerryBox some of the data parameters are dependent on some of the more general quality controls, such as the flow speed through the system and heating of the water sample before reaching the sensors. These two quality controls effect all other marine parameters if they are flagged as bad, the atmospheric parameters on the other hand are not obligatory affected.

For data providers not providing data to CMEMS, the documentation of the undergone quality controls have to be linked to the data set explaining the routines.

Most of the quality controls included in the standard quality controls used by CMEMS are also or will be available in the Common European Ferrybox Database.

4.4. Delayed mode quality controls

Ferrybox data also have to be quality controlled in delayed mode, to be able to see long term trend and validate with water samples and other platforms. This is done with a manually quality control of data and by validation with water samples taken from monitoring program and other research activities (Fig.3). Water samples are important, but also cross-comparing with other FerryBoxes in the area is of interest. For some sensors, like for phycocyanin the calibration of the sensor is sensitive and cooperation around inter-calibrations are most valuable. Here the SeaDataNet is an important source of delayed mode data to use in delayed mode quality controls.

A software, the Ferrybox-toolbox is developed by SMHI within Jerico-NEXT (deliverable 5.12) to use for manually quality controlling FerryBox data. In the toolbox it is possible to load FerryBox data in different formats, just specifying which it is in a set-file. It is also possible to load data from another platform to compare the datasets. It is desirable if a manual quality control can be done twice, both more close in time to when the observation is taken and also after some time to get a wider perspective and to be able to validate with water samples. First a more direct manually check to ensure that everything looks alright and also taking account of other automated platforms, and then later when water samples are analysed to validate the earlier results.

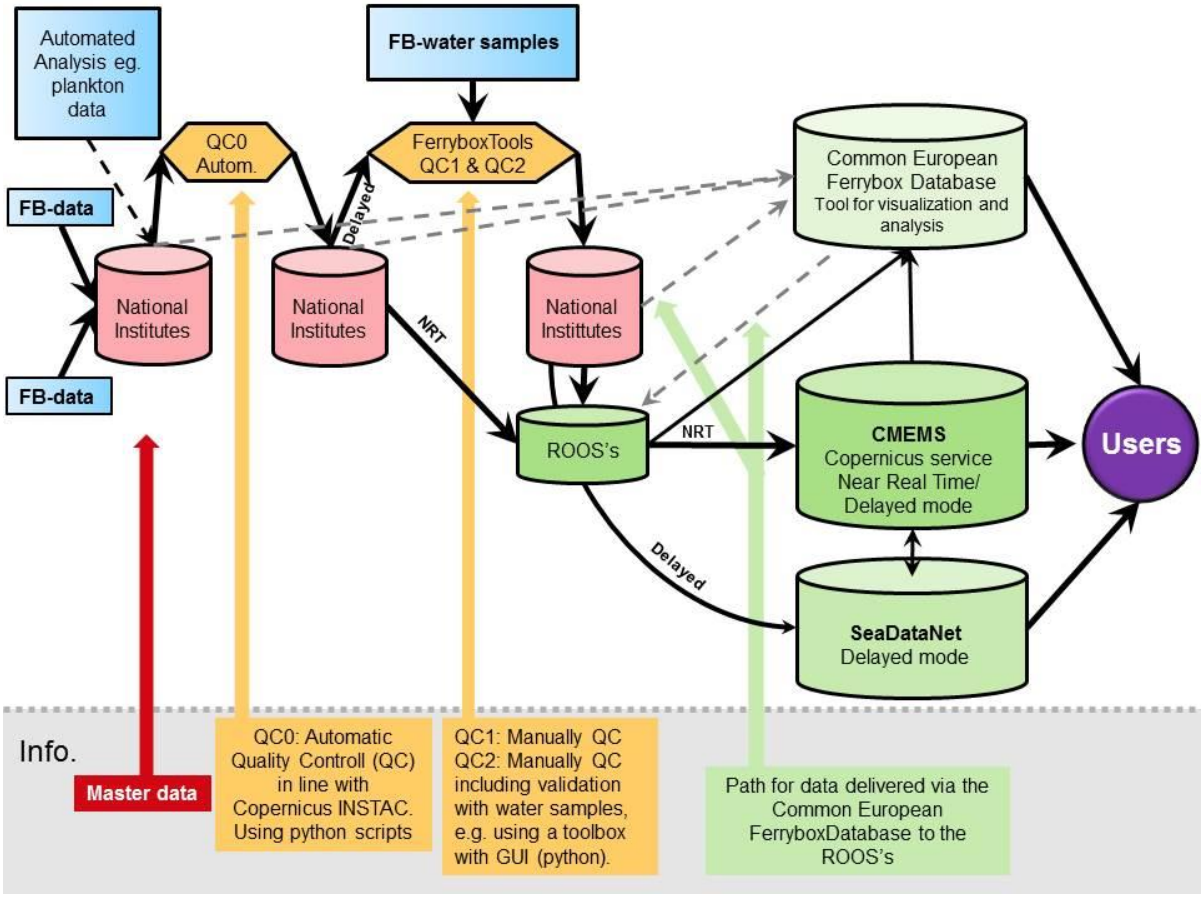

Figure 3 Detailed roadmap over the FerryBox data

5. Conclusions

Like most marine observational data, availability of quality controlled and open access FerryBox data depends on the data providers. Therefor a right forward roadmap for the data from data provider to the end user is important. By providing the FerryBox data in near real time via the ROOS's and then further to CMEMS, a regional contact is established between the data provider and the dissemination unit (the ROOS). Ferrybox data will benefit from the automatic quality controls and standard formats and will be compatible to use with all other types of data that is made available online on the CMEMS site. By using the CMEMS and SeaDataNet track FerryBox data will follow how other open marine observational data in Europe is handled and made available. Data collected by FerryBoxes is not much different than data collected by other operational systems and benefits from use of the existing European infrastructure and capacities of both the operational data nodes and the delayed mode data centres.

An open source software to perform the manual quality controls will be useful for the data providers, or the possibility to store and do the quality control in a Common European Ferrybox Database, where the data also can be provided to the ROOS's if it is flagged with open access.

6. Annexes and references

Copernicus marine Environmental Service http://marine.copernicus.eu/

HZG Ferrybox Database http://ferrydata.hzg.de/index.cgi?seite=start