

# Curation in EuroGOOS

## Context of curation in EuroGOOS

This section outlines the requirements which have been collected from EuroGOOS on the topic of "Curation" during Task 5.1 of the ENVRplus project.

A description of EUROGOOS was presented in [General requirements for EuroGOOS](#). As explained there, EuroGOOS is not a RI, but the members of its Task Teams belong to communities which could be considered RIs. We have therefore addressed the specific questions to these communities.

## Summary of EuroGOOS requirements for curation

### Detailed requirements

#### 1. The Tide Gauge Task Team and Community

**The Tide Gauge Task Team is not responsible for data curation, but they will compile how this is currently being done by the members of their community.** Until very recently, tide gauge operators have always been responsible for the quality control and processing of the data. For example, in Puertos del Estado, automatic, near real-time quality control, as well as more delayed and detailed yearly quality control of the data are produced and presented in their reports and on their website. There is a similar situation at a national level for most of the national centres. Currently, as part of many European projects, funds have been dedicated to the creation of working groups the purpose of which is to develop quality control and data processing for oceanographic parameters, including sea level, within certain regions. **The Tide Gauge Task team would like to explore how much of this work is duplication, at both national and regional level, to help develop a strategy in order to eliminate it (i.e. homogenize work processes), and to help develop standard methods for quality control and data processing.** Some standards are already in place within the ROOSs. **The Task Team will also explore the different algorithms for quality control and data processing that are necessary for the different new applications of tide gauges, e.g. tsunami warnings, and their lower sampling intervals and lower latency of data transmission (since 2004-2005). This discussion has already started within GLOSS.**

Puertos del Estado do not perform any quality control of the software. Most of their software is old, based on Fortran and C Shell scripts developed by them, and they have been adding elements to it for new applications. They are also using some software which was taken from other organisations. They do not have a complete tool in a unique language, and the RI representative thinks that such a tool would be useful as it would be easier to distribute to other national operators. Puertos del Estado do not perform any quality control of the operating environment, specification or documentation.

The data in Puertos del Estado passes through 3 time series: raw data, raw data flagged with the outputs of quality control, and the first clean time series with processed data, where sometimes the raw data has been interpolated. They keep a backup of these 3 types of data in their institution, and they never discard the original data.

In the case of Spain, in which the observational networks are permanent networks, the government has the commitment to get the funds to maintain the networks in operation (i.e. to ensure that they will be permanently there, that there will be enough personnel for their maintenance, enough funding for spare parts, enough funding for travelling to the stations, etc.). This is also normally the case for other countries. Unfortunately, in recent years this was not possible for some countries, especially in the South of Europe (e.g. the whole network of stations in Italy could disappear in the near future). Even in Spain, during the last 4 years since the economic crisis, they were told to reduce some of their expenses, such as the number of times their stations are maintained from twice to once per year. This is the most critical worry that the Tide Gauge Task Team is dealing with.

In Spain, there is no specific logging system to be used at institute level, but rather one person per network keeps a record of what actions have been carried out on the data, and what errors or other incidents (e.g. change of instrument, maintenance, etc.) took place. Creating a more formal way of doing this has often been discussed within Puertos del Estado, as it is currently difficult for someone other than the person in charge to find out this information. This issue has not been discussed yet within the Task Team, but the RI representative foresaw that it will be an important future point of discussion.

Within Puertos del Estado, when they share their data on international portals, the coordinates (latitude, longitude) and names of the stations are very well defined (for names, they use a 4 letter code for each station, not only for sea level, but also fixed buoys and meteorological stations). For sea level, they also have the tidal datum or tide gauge "zero" of the station, the location of the benchmark which is close to the tide gauge and the distance of this tide gauge benchmark to all the official references in Spain. For this metadata, they use the international standards within the EuroGOOS ROOSs, which were defined within the MyOcean projects. They also have their own internal codes and standards within their national institute. IBIROOS and the rest of the data portals also have their own codes.

Puertos del Estado do not use metadata standards for providing contextualisation or detailed access-level.

**Both the Tide Gauge Task Team and the group in Puertos del Estado would like to work on automating the use of other nearby stations or secondary sensors in the same station for quality control. Nowadays, all the automatic quality control procedures work with a unique time series only. They would like to add in an automatic way the use of time series from the vicinity of a station, even from other types of instrumentation like altimetry.**

#### 2. The FerryBox Community

The following replies were provided in writing by representatives of the FerryBox Task Team.

1. Does the curation cover datasets or also:
  - i. Software
  - ii. Operating environment?

- iii. Specifications/documentation?
2. What is your curation policy on discarding
    - i. Datasets? QC/QA
    - ii. Software?
    - iii. Operating environments?
    - iv. Documents?
  3. How will data accessibility be maintained for the long term? E.g. What is your curation policy regarding media migration? *keeping records of all data on the website*
  4. Do you track with a logging system all curation activities?
  5. What metadata standards do you use for providing:
    - i. Discovery, the Ferrybox data are homogenous, no discovery metadata are needed. For other portals this should be provided by the portals. We use a WFS for discovery- and use-metadata of our own Ferryboxes which can be extended to all routes?
    - ii. Contextualisation (including rights, privacy, security, quality, suitability...)
    - iii. Detailed access-level (i.e. connecting software to data within an operating environment)? Please supply documentation.
  6. If you curate software how do you do it? Preserving the software or a software specification?
  7. If you curate the operating environment how do you do it? Preserving the environment or an environment specification?

## Formalities (who & when)

<b>Go-between</b>	Cristina Adriana Alexandru
<b>RI representative</b>	Begoña Pérez Gómez (Tide Gauge Task Team, working in Puertos del Estado in Spain), Franciscus Colijn, Willy Petersen, G. Breitbach
<b>Period of requirements collection</b>	October - December 2015
<b>Status</b>	