

Identification and Citation in EuroGOOS

Context of Identification and Citation in EuroGOOS

This section outlines the requirements which have been collected from EuroGOOS on the topic of "Identification and Citation" 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 Identification and Citation requirements for EuroGOOS

Detailed requirements

The FerryBox Community

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

Identification and Citation

a. Identification

i. What granularity do your RI's data products have:

- Content-wise (all parameters together, or separated e.g. by measurement category)? *available*
- Temporally (yearly, monthly, daily, or other)? *available*
- Spatially (by measurement station, region, country or all together)? *all options are open*

ii. How are the data products of your RI stored - as separate "static" files, in a database system, or a combination? *database*

iii. How does your RI treat the "versioning" of data - are older datasets simply replaced by updates, or are several versions kept accessible in parallel? *mostly replaced by updates, raw data available from the originator*

iv. Is it important to your data users that:

- Every digital data object is tagged with a unique & persistent digital identifier (PID)? *no*
- The metadata for data files contains checksum information for the objects? *no*
- Metadata (including any documentation about the data object contents) is given its own persistent identifier? *No, a persistent identifier would be useless in our case. metadata are available for every transect.*
- Metadata and data objects can be linked persistently by means of PIDs? *There are linked by web-services. In addition we are planning data publications with PIDs.*

v. Is your RI currently using, or planning to use, a standardized system based on persistent digital identifiers (PIDs) for:

- "Raw" sensor data? *No*
- Physical samples? *No*
- Data undergoing processing (QA/QC etc.)? *No*
- Finalized "publishable" data? *Yes, planned*

vi. Please indicate the kind of identifier system that are you using - e.g. Handle-based (EPIC or DOI), UUIDs or your own RI-specific system? *using of DOIs is planned*

vii. If you are using Handle-based PIDs, are these handles pointing to "landing pages"? Are these pages maintained by your RI or an external organization (like the data centre used for archiving)? -

viii. Are costs associated with PID allocation and maintenance (of landing pages etc.) specified in your RI's operational cost budget? -

1. b. Citation

i. How does your "designated scientific community" (typical data users) primarily use your data products? As input for modelling, or for comparisons? *all kinds of usage possible*

ii. Do your primary user community traditionally refer to datasets they use in publications: *no experiences available until now*

- By providing information about producer, year, report number if available, title or short description in the running text (e.g. under Materials and Methods)?
- By adding information about producer, year, report number if available, title or short description in the References section?
- By DOIs, if available, in the References section?

- By using other information?
- iii. Is it important to your data users to be able to refer to specific subsets of the data sets in their citation? Examples:
- Date and time intervals *yes*
 - Geographic selection *yes*
 - Specific parameters or observables *yes*
- iv. Is it important to be able to refer to many separate datasets in a collective way, e.g. having a collection of “all data” from your RI represented by one single DOI? *no*
- v. What strategy does your RI have for collecting information about the usage of your data products?
- Downloads/access *Both*
 - Visualization at your own data portal: *yes*
 - Visualization at other data portals: *No*
 - References in scientific literature: *yes*
 - References in non-scientific literature: *No*
 - Scientific “impact”: *No*
- vi. Who receives credit when a dataset from your RI is cited?
- The RI itself
 - The RI's institutional partners (all or in part, depending on the dataset contents)
 - Experts in the RI's organization (named individuals)
 - “Principal investigators” in charge of measurements or data processing (named individuals), *Yes*
 - Staff (scientists, research engineers etc.) performing the measurements or data processing (named individuals), *Yes*

Formalities (who & when)

Go-between	Cristina Adriana Alexandru
RI representative	Franciscus Colijn, Willy Petersen, G. Breitbach (FerryBox Task Team)
Period of requirements collection	October- November 2015
Status	