

Identification and citation in EPOS

Context of identification and citation in EPOS

Complete EPOS report on Identification and Citation available at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/42/comments/317/attachments/374/download>

Summary of EPOS requirements for identification and citation

Detailed requirements

Identification

EPOS has all types of granularity (content-wise, temporally and spatially) since it has 8 TCS (e.g. seismologists, near-fault observatories, satellite data, volcano observatories, etc.), and each of them has different data products with different granularity. For example, Seismology community has temporally, spatially and content wise granularity, delivering data products (seismic raw data) every day and by region but also by type (e.g. hazard maps).

The data products are stored in the RI in different ways (e.g. databases, static files, other storage solutions), as they depend on the decisions taken by each TCS.

EPOS treats differently the versioning of data depending on the data products. There are cases where EPOS wants to store the previous versions, and others where they are only interested in the last versions. However in most(all) cases the history of the dataset and provenance track should be maintained. EPOS is currently working on how to deal with versioning.

EPOS uses EPIC and DOI, and custom systems, as identifier systems. At the moment, EPOS assigns a PID (by using EPIC) to a digital object (e.g. raw seismic data). Furthermore, for some datasets EPOS is coupling PIDs with metadata by using DOIs. All of the proposed questions fit in the heterogeneous scenarios of EPOS

EPOS is currently experimenting with the use of PIDs for raw sensor data and high-level products. Seismology community also uses PIDs (DOIs) for seismic networks. Again all the scenarios are envisaged in EPOS. Some TCSs might want to assign PID to raw data other to higher-level products.

Stations' DOIs are associated with landing pages with all the information in a human-readable way. But, EPOS will like to have DOIs to point directly to digital objects and/or metadata human and machine-readable.

EPOS operational phase will start after the end of the Implementation Phase project, which will last 4 years. This type of costs (costs associated with PID allocation and maintenance) will be discussed during that period. Maybe they could be associated to the ERIC. Although some of the costs can be left to the TCSs and/or National Infrastructures (NRI layer).

Citation

EPOS scientific communities use data products as an input for modelling and comparison, and they have different ways to refer to datasets, and it depends very much on the maturity of the community. If there are DOIs for datasets, EPOS tries to enforce the use of them for citation and attribution. Note that, currently EPOS does not have a method for referring to specific subsets of data, although they would like to have it.

At the moment, EPOS does not have a strategy for collecting information about the usage of their data products. Currently, they are thinking about metrics for measuring the impact of the data.

References:

1. <http://www.dcc.ac.uk/resources/how-guides/cite-datasets>
2. <http://cera-www.dkrz.de/WDCC/ui/SearchByName.jsp?query=earlinet>

Formalities (who & when)

Go-between	Rosa Filgueira
RI representative	Luca Trani
Period of requirements collection	July to November 2015
Status	Finished