

Provenance in EPOS

Context of provenance in EPOS

Complete EPOS report on Provenance available at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/42/comments/318/attachments/376/download>

Summary of EPOS requirements for provenance

Detailed requirements

Complete EPOS report available at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/42/comments/318/attachments/376/download>

EPOS has already data provenance recording, for some of its communities. Provenance in EPOS is always related with the data processing (workflows). Regarding to the question of provenance with data collections, EPOS does pre-caching, which means that users get the data from the services and then they get provenance. At the moment, EPOS does not have provenance related with versioning control in the repositories.

EPOS uses MongoDB and web-services based.

About the standards, Some EPOS communities (Seismo) exports provenance to PROV-XML [1] and it can be exported to PROV-O, but EPOS is not using it yet.

EPOS needs a provenance tracking, and already has one. The information contained is:

- Attribution: Who wants the data and where is the data coming from.
- Processing lineage: How the data is going to be transformed.

EPOS itself wants to know how and why scientists are using the data, for understanding the relevance (quality measure) of the data used. If users use a workflow, then the investigation design is recorded in the prospective provenance data. EPOS needs to record the observation and measurement methods and context, but at the community level. Each community is let free to do it, but EPOS recommends it strongly. Besides, EPO needs to record the processing methods (queries), because EPOS wants to know to how people is using the data. However, the recording-processing methods are not the first priority in EPOS. Finally, EPOS also needs to record failures and downtown failures.

Currently, EPOS uses PROV. Therefore, EPOS's model needs to be complied to that schema (for the relationship). In the long term, EPOS wants to use CERIF data model and ontology.

For provenance tracking, EPOS uses workflows, queries, provenance management APIs, services to store and search and visualizations tools.

VERCE uses provenance already for processing and the set up of the investigation. Seismology community is the most advanced is that sense.

References:

1. <http://www.w3.org/TR/prov-xml/>

Formalities (who & when)

Go-between	Rosa Filgueira
RI representative	Alessandro Spinuso
Period of requirements collection	From September to November 2015
Status	Finished