

Curation in EPOS

Context of curation in EPOS

Complete EPOS report on Curation available at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/42/comments/332/attachments/409/download>

Summary of EPOS requirements for curation

Detailed requirements

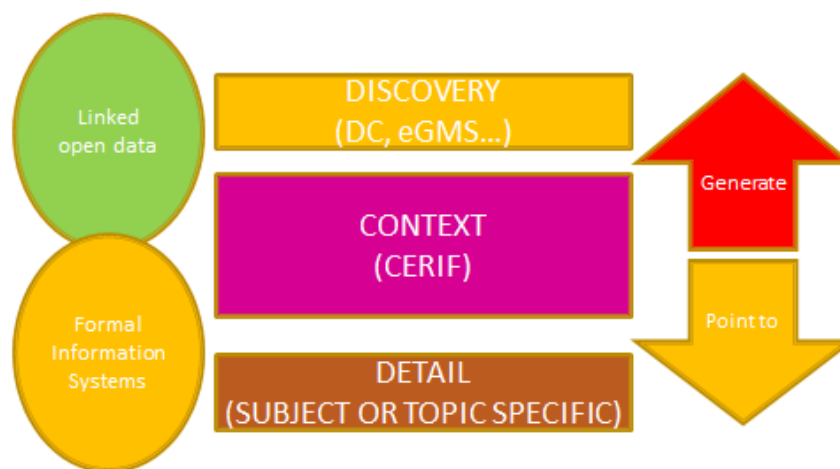
The RI's curation responsibilities will be shared among different communities in EPOS (the suppliers of data, software etc), but some are currently having discussions with EUDAT. AT this stage of EPOS project, curation covers only dataset but they will expand further.

The curation policy for Datasets, EPOS insists in the data management plan, for the communities within EPOS, which has to include: curation, preservation and provenance. For Software and operating environment, EPOS does not have yet a strategy and they are discussing it. In terms of Software, EPOS is considering not to store software but just their specifications.

For providing data accessibility for the long term, EPOS itself is providing the portal and the catalogue for accessing all the information. So, the responsibility for data curation and preservation of data is for the communities within EPOS. EPOS assumes that each of those will have a funding business plan that allows them to do that. In the case of EPOS maintaining the catalogue, they will be moving from the existing EPOS IP project to an ERIC, which will provide the business plan for the preservation of the EPOS catalogue.

Note that, EPOS tracks all curation activities with a logging system.

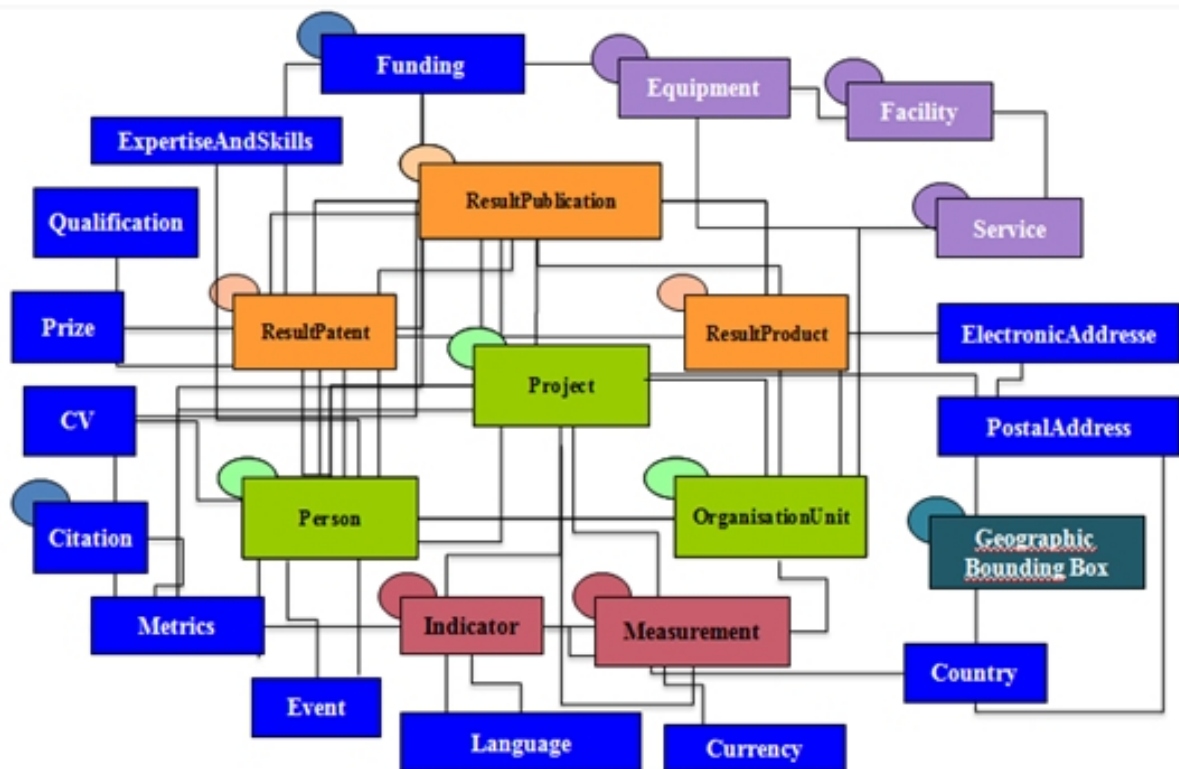
Regarding the metadata standards used by EPOS, it follows a 3-Layer Model (see Figure below): for Contextualisation, EPOS uses CERIF, which is a EU Recommendation to member states now used in 42 countries, allows the use of rich metadata. This provides greatly improved retrieval capabilities, and enables institutions to create linkages between all relevant data — not just research papers, but information about the authors, research datasets and software, funding information, project information, experiments, and even data on the facilities and equipment used by researchers. From CERIF, depending of the community that they are dealing with, they can generate for discovering: DC (Dublin Core), DCAT (Data Catalog Vocabulary), CKAN (Comprehensive Knowledge Archive Network), eGMS (e-Government Metadata Standard – based on DC) and – for geospatial data – INSPIRE. On the details side, there are about 300 metadata standards within EPOS, so there is some intersection with the contextual metadata (CERIF), but most of it is specific to the discipline.



Some information can be found at:

- CERIF documentation [1][2]
- ER Diagram (fully connected graph) [3]
- Model Documentation [4]
- List of all the Entities and their attributes (what is related with the individual entities) and relations (they are the links to other entities) [5]

Each of the boxes of the following diagram is an entity, and each line is a link between them and as the diagram shows, relates together people, projects, organizations, datasets, software, and funding, etc.



EPOS will try to preserve the current software (as long it is operational in the current operating environment), as well write down the specification of the software.

At the current stage, EPOS documents workflows and procedures. However, EPOS is looking to the possibility of using Docker [6] or another container approach to store workflows.

Finally, the steps needed to improve the curation facilities are:

- Automation of processes to record the curation activities
- Rich metadata
- Appropriate user interface.

References:

1. <http://eurocris.org/cerif/main-features-cerif>
2. <http://eurocris.org/cerif/feature-tour/cerif-16>
3. <http://eurocris.org/Uploads/Web%20pages/CERIF1.6/documentation/MImage.html>
4. <http://eurocris.org/Uploads/Web%20pages/CERIF1.6/documentation/Tables.html>
5. <http://eurocris.org/Uploads/Web%20pages/CERIF-1.6/documentation/Tables.html>
6. <https://www.docker.com/>

Formalities (who & when)

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