



A Digital Knowledge Preservation Platform for Environmental Sciences

Fernando Aguilar Gómez, Jesús Marco de Lucas, Esther Pertinez, Aida Palacio, and David Perez
IFCA-CSIC, Santander, Spain (aguilarf@ifca.unican.es)

The Digital Knowledge Preservation Platform is the evolution of a pilot project for Open Data supporting the full research data life cycle. It is currently being evolved at IFCA (Instituto de Física de Cantabria) as a combination of different open tools that have been extended: DMPTool (<https://dmptool.org/>) with pilot semantics features (RDF export, parameters definition), INVENIO (<http://invenio-software.org/>) customized version to integrate the entire research data life cycle and Jupyter (<http://jupyter.org/>) as processing tool and reproducibility environment.

This complete platform aims to provide an integrated environment for research data management following the FAIR+R principles:

- Findable: The Web portal based on Invenio provides a search engine and all elements including metadata to make them easily findable.
- Accessible: Both data and software are available online with internal PIDs and DOIs (provided by Datacite).
- Interoperable: Datasets can be combined to perform new analysis. The OAI-PMH standard is also integrated.
- Re-usable: different licenses types and embargo periods can be defined.
- +Reproducible: directly integrated with cloud computing resources.

The deployment of the entire system over a Cloud framework helps to build a dynamic and scalable solution, not only for managing open datasets but also as a useful tool for the final user, who is able to directly process and analyse the open data.

In parallel, the direct use of semantics and metadata is being explored and integrated in the framework. Ontologies, being a knowledge representation, can contribute to define the elements and relationships of the research data life cycle, including DMP, datasets, software, etc. The first advantage of developing an ontology of a knowledge domain is that they provide a common vocabulary hierarchy (i.e. a conceptual schema) that can be used and standardized by all the agents interested in the domain (either humans or machines). This way of using ontologies is one of the basis of the Semantic Web, where ontologies are set to play a key role in establishing a common terminology between agents.

To develop the ontology we are using a graphical tool called Protégé. Protégé is a graphical ontology-development tool which supports a rich knowledge model and it is open-source and freely available. However in order to process and manage the ontology from the web framework, we are using Semantic MediaWiki, which is able to process queries. Semantic MediaWiki is an extension of MediaWiki where we can do semantic search and export data in RDF and CSV format. This system is used as a testbed for the potential use of semantics in a more general environment.

This Digital Knowledge Preservation Platform is very closed related to INDIGO-DataCloud project (<https://www.indigo-datacloud.eu>) since the same data life cycle approach is taking into account (Planning, Collect, Curate, Analyze, Publish, Preserve). INDIGO-DataCloud solutions will be able to support all the different elements in the system, as we showed in the last Research Data Alliance Plenary.

This presentation will show the different elements on the system and how they work, as well as the roadmap of their continuous integration.