13TH IBERIAN GRID CONFERENCE

IBERGRID 2024

better
software
for
better
science

28-30 OCT
UNIVERSITY
OF PORTO

Contribution ID: **33**   Type: **not specified**

Metadata-powered characterization of DT-GEO Digital Twins

Monday, 28 October 2024 12:00 (15 minutes)

The data management team in the DT-GEO project is following a novel approach for the characterization of the Digital Twins for geophysical extremes being developed in the project. The approach relies on the use of rich metadata to describe the components of Digital Twins (DTCs): (i) the digital assets (DAs) –namely datasets, software-services, workflows and steps within those workflows–, and (ii) the relationships between such assets or entities. This conforms to the structure of CERIF (Common European Research Interchange Format) - a very early example of graph-structured metadata used as the storage form of the European Plate Observing System (EPOS) metadata catalog.

The resulting DT-GEO metadata schema responds to a requirement gathering process carried out jointly with the communities participating in the project. In the first phase, the schema was standarized according to the EPOS-DCAT-AP data catalog vocabulary for the DAs, which resulted in an extended version of the latter. The collected metadata from the communities was then serialized into RDF documents (Resource Description Framework) and exposed through a prototype metadata catalog, based on the EPOS ICS-C data portal, via REST API and web portal interfaces.

At the time of writing the second phase is being undertaken and aims at accommodating the metadata for the workflows operated by the Digital Twins, following again the CERIF data model. As a result of finding EPOS-DCAT-AP inappropriate for the task, the data management team decided to opt for the RO-Crate specification, used to package research objects. Thus, the metadata for the workflow descriptions are serialized into RO-Crate documents (in JSON-LD format) to enable their abstract characterisation, including the aforementioned relationships between DTCs, workflows, and steps within these workflows. This allows to populate the registries and catalogs required by the downstream workflow management system solution, i.e. PyCOMPSs in the DT-GEO case.

The metadata management follows a continuous improvement process where the maintainers of the Digital Twins contribute with new changes, which are then reviewed by the data management team in order to validate and harmonize the metadata values for the sake of interoperability. Throughout this process, the FAIR maturity and quality assurance of the digital assets is key, and thus, each DA is being evaluated through diverse means. In the particular case of the data, a plugin for the DT-GEO prototype catalog was developed for the FAIR-EVA tool, which enables the early evaluation of the FAIRness of the DT-GEO datasets. In the case of the DT-GEO software (code and services), the SQAaaS platform (Software Quality Assurance as a Service) provides accurate reporting about SQA characteristics. Last but not least, the validation of the workflows is done through a continuous integration setup that triggers the workflow execution in a HPC-like environment in the Cloud, once the workflow code has been checked by the SQAaaS service, and the containers that represent the steps in the workflow have been packaged.

Primary authors: Mr JEFFERY, Keith (UKRI); ORVIZ FERNÁNDEZ, Pablo (IFCA-CSIC)

Co-authors: FOLCH, Arnau (Geociencias Barcelona (GEO3BCN-CSIC)); Mr ARCE, David (UPV); PALOMO LLAVONA, Iván (IFCA); Mr PASSMORE, James (UKRI); Ms GIULIACCI, Kety (EPOS / INGV); CARBONELL, Ramon (Geociencias Barcelona (GEO3BCN-CSIC)); Ms PACIELLO, Rossana (EPOS / INGV); BERNARDO, Samuel (LIP); Mr VINCIARELLI, Valerio (EPOS)

Presenter: ORVIZ FERNÁNDEZ, Pablo (IFCA-CSIC)

Session Classification: IBERGRID

Track Classification: Design and implementation of Digital Twins