Geophysical Research Abstracts
Vol. 21, EGU2019-7800, 2019
EGU General Assembly 2019

A web platform for storing, sharing and executing scientific workflows for Natural Risk Assessment: part 2 - description of an interoperable architecture based on open-source components.

Agnès Tellez-arenas, Audrey Hohmann, Adrien Quentin, Caterina Negulescu, and Farid Smaï
BRGM, France, Orléans

Enabling storing, documentation, access and execution of scientific workflows is the aim of the web platform currently designed and developed by the BRGM (French Geological Survey) for the department of Natural Risk assessment. Such a platform shall insure reproducibility, allowing transparency but also improving efficiency by easing collaborative work and sharing results and practices. The scientific scope is risk assessment in the domain of natural hazard (earthquake, landslide and submersion) from the phenomenon modelling to the impact evaluation on exposed elements such as buildings. This web platform initially designed for BRGM experts aims to be in a long-term ambition an open repository for national and international experts of natural hazards. Integration and deployment of new data sets and processes will be as automatic as possible.

The architecture includes a repository of data sets, documents, links and reproducible versioned workflows. It is open and discoverable through interoperable web services, each record described by metadata allowing discovery, evaluation and use. The heart of the architecture is the workflows description and execution mechanism, based on JSON and python descriptions defining acyclic graphs of processes, executed through interoperable web services (OGC WPS). They are executable as a whole or step-by-step, with detailed logs. The execution engine must ensure security, robustness, flexibility and scalability. Used in a multi-users context, it requires asynchronous mechanisms and scheduling of computations. It will be possible for a particular execution of any workflow, to register inputs, parameters and resulting computed data sets, which become new records in the repository. Data sets and workflows will be shareable between groups of users or even publishable as web services when relevant. Connecting heterogeneous scientific codes and input data sources is another challenge. While scientific codes are implemented in various technologies, data to process can be for instance relational georeferenced databases, grids and vectors data provided in various file formats, web services, internal or external data sources, already referenced in the repository or uploaded by a user. In order to enable such an integration, we define normalized code entry points and data connectors. Finally, a web interface offers access to all the functionalities, but main components are also usable through web services or even directly when relevant.