Data-intensive science gateway for rock physicists and volcanologists.

Rosa Filgueira (1), Malcom Atkinson (1), Andrew Bell (2), Ian Main (2), Steve Boon (3), Philp Meredith (3), and
Christopher Kilburn (3)

(1) University of Edinburgh, School of Informatics, Edinburgh, United Kingdom , (2) University of Edinburgh, School of
GeoSciences, Edinburgh, United Kingdom, (3) University College London, Department of Earth Sciences, London, United
Kingdom

Scientists have always shared data and mathematical models of the phenomena they study. Rock physics and Volcanology, as well as other solid-Earth sciences, have increasingly used Internet communications and computational renditions of their models for this purpose over the last two decades. Here we consider how to organise rock physics and volcanology data to open up opportunities for sharing and comparing both experiment data from experiments, observations and model runs and analytic interpretations of these data. Our hypothesis is that if we facilitate productive information sharing across those communities by using a new science gateway, it will benefit the science.

The proposed science gateway should make the first steps for making existing research practices easier and facilitate new research. It will achieve this by supporting three major functions:

1) sharing data from laboratories and observatories, experimental facilities and models;
2) sharing models of rock fracture and methods for analysing experimental data; and
3) supporting recurrent operational tasks, such as data collection and model application in real time.

We report initial work in two projects (NERC EFFORT and NERC CREEP-2) and experience with an early web-accessible protytpe called EFFORT gateway, where we are implementing such information sharing services for those projects.

1. Sharing data: In EFFORT gateway, we are working on several facilities for sharing data:

*Upload data: We have designed and developed a new adaptive data transfer java tool called FAST (Flexible Automated Streaming Transfer) to upload experimental data and metadata periodically from laboratories to our repository.

*Visualisation: As data are deposited in the repository, a visualisation of the accumulated data is made available for display in the Web portal.

*Metadata and catalogues: The gateway uses a repository to hold all the data and a catalogue to hold all the corresponding metadata.

2. Sharing models and methods: The EFFORT gateway uses a repository to hold all of the models and a catalogue to hold the corresponding metadata. It provides several Web facilities for uploading, accessing and testing models.

*Upload and store models: Through the gateway, researchers can upload as many models to the repository as they want.

*Description of models: The gateway solicits and creates metadata for every model uploaded to store in the catalogue.

*Search for models: Researchers can search the catalogue for models by using prepackaged sql-queries.

*Access to models: Once a researcher has selected the model(s) that is going to be used for analysing an

experiment, it will be obtained from the gateway.

*Services to test and run models: Once a researcher selects a model and the experimental data to which it should be applied, the gateway submits the corresponding computational job to a high-performance computational (HPC) resource hiding technical details. Once a job is submitted to the HPC cluster, the results are displayed in the gateway in real time, catalogued and stored in the data repository, allowing further researcher-instigated operations to retrieve, inspect and aggregate results.

*Services to write models: We have desgined VarPy library, which is an open-source toolbox which provides a Python framework for analysing volcanology and rock physics data. It provides several functions, which allow users to define their own workflows to develop models, analyses and visualizations.

3. Recurrent Operations: We have started to introduce some recurrent operations:

*Automated data upload: FAST provides a mechanism to automate the data upload.

*Periodic activation of models: The EFFORT gateway allows researchers to run different models periodically against the experimental data that are being or have been uploaded