Supporting the research lifecycle of GEO-Geohazard Supersites and Natural Laboratories initiative through HPC and research objects

Stefano Salvi
Elisa Trassati
Istituto Nazionale di Geofisica e Vulcanologia (INGV)
Rome, Italy
[stefano.salvi, elisa.trasatti]@ ingv.it

Marcin Krystek,
Cezary Mazurek, Raul Palma,
Poznan Supercomputing and Networking Center
Poznan, Poland
[mkrystek,mazurek, rpalma]@man.poznan.pl

Jose Manuel Gomez-Perez,
Andrés Garcia-Silva
Expert System
Madrid, Spain
[jmgomez,agracia]@expertsystem.com

Volcanic eruptions are among the most spectacular and dangerous phenomena on Earth, capable of generating disasters at various scales. It is estimated that over 500 million people live within exposure range of active volcanoes and can be affected by their activity. Mega-eruptions can cause global impacts on climate and human health, and induce considerable transnational economic damages. For instance, the eruption of the Icelandic Laki volcano in 1783 generated far-reaching ash and gas clouds which resulted in tens of thousands of victims across Europe. Today the toll would be in the millions.

The assessment and management of volcanic risk is then a matter of global concern, which is constantly addressed at all levels. The first element to evaluate the risk is to develop a sound scientific understanding of the phenomena and processes which are part of the volcanic activity. Then, to effectively manage the risk, the geophysical and geochemical parameters whose anomalies can signal an impending eruption, must be constantly monitored. Thus, volcanoes must be closely observed with a number of instrumental networks whose data should be rapidly analysed to provide early warning. Yet a study of 441 active volcanoes in 16 developing countries revealed that 384 have rudimentary or no ground monitoring, including 65 volcanoes identified as posing a high risk to large populations, as the Nyragongo, in the Democratic Republic of Congo, where a city of 1 million people is located 10 km from the main crater.

A number of Earth Observation satellites is today available to support also volcano monitoring, producing frequent measurements of surface or atmospheric parameters which are useful indicators of volcanic activity, as ground temperature, gas concentrations, ground deformation, etc. Not all of these data are free and open, and should be bought on the market, reducing the possibility for scientists to use them for risk assessment, especially in developing countries. Thus in 2012 the international scientific community decided to establish a partnership, aiming to provide free access to ground and satellite data over a selected number of volcano Supersites where focused scientific research could rapidly produce results for risk reduction.

The Geohazard Supersites and Natural Laboratories initiatives (GSNL) today is a network of 11 Supersites, including volcanoes and seismic areas. The main space agencies of the world support the initiative providing free access to thousands of satellite images and local geophysical institutes put their seismic, geodetic and other data open for the global scientific community. Complex algorithms are used to analyse these data and important information on the volcano activity. For instance, satellite-derived maps showing the millimetric movements of the ground are processed with complex geophysical algorithms, which model the observed ground deformation patterns and help identify the depth and location of the magma sources. To investigate the characteristics of the Earth's crust under the volcano and map its "plumbing system", thousands of seismological recordings are studied using tomographic inversion codes. And when the eruption starts, the path of the large clouds of ash and gas which are emitted up to 12 km into the atmosphere, are constantly monitored and forecasted using dispersion models based on atmospheric and satellite observations. The impacts these clouds can have on the civil aviation is tantamount: during the Eyafjallajökull eruption in 2010, over 100.000 commercial flights were grounded due to possible impacts on airplane's engines functioning, causing over 5 billion Euro economic damage.

All these analyses are based on complex geophysical inversion algorithms, by which a theoretical model based on a set of equations, is calculated tens or hundreds of thousands of times, and compared with large datasets until the model that best reproduces the observations is found. Large computing resources are needed to run these models, and when results are needed for managing a volcanic crisis, as for the ash cloud mapping, these calculations must be carried out in matters of minutes.

In addition to computing power and resources, researchers from the geo-gnsl community, as many other data-intensive science communities, are calling for innovative ways to manage their data, methods and other resources, which can enhance the visibility of scientific breakthroughs, encourage reuse, and foster a broader research accessibility. Furthermore, they are interested in tools and mechanisms supporting the full research lifecycle of their investigations, from the data discovery and selection, to the modelling of geophysical algorithms and creation of computational methods/workflows, to the facilities for reusing and executing such methods (e.g., using cloud/hpc resources) and for managing the results obtained and conclusions derived, up to capabilities enabling the collaboration and sharing of the investigation materials, to name a few. Research Objects (ROs) (http://www.researchobject.org/) are a key technology towards enabling such capabilities. At its core, an ROs is a semantically enriched information unit that encapsulates and describes all the scientific resources related to an investigation, along with the context in which they were used/produced and the people involved. As such, they can address key research communication challenges like preservation, reproducibility, and interoperability, and include metadata that make them uniquely identifiable, processable, and machine readable. Additionally, inspired by software development practices, ROs can keep track and support the research lifecycle by enabling the creation of snapshots, releases and forks.

Snapshots and releases are both immutable copies of the RO at a certain point in time; but while snapshots are meant to be generated to share preliminary results or to preserve versions of an investigation (e.g., when reaching a milestone), releases are meant to be generated when the investigation has been concluded. Forks, on the other hand, provide the means to create a copy of an existing RO, e.g., to test new ideas without affecting the original one, or to start a new research process based on it. This is a key mechanism to foster reuse, which addresses proper accreditation by generating an automatic citation of the source RO. And in addition to foster author accreditation of their respective contributions, ROs enable discussion around the investigation, and ultimately support collaboration.

In this contribution we will present the results of EVER-EST project (H2020-EINFRA-2015-1), in which we created in collaboration with different partners a virtual research environment (VRE) for Earth Science (https://vre.ever-est.eu/), embracing the research object concept and technologies at its core. In particular, the VRE comprises the research object management platform ROHub (http://www.rohub.org/) with several value-added research object services (e.g., enrichment, quality assessment, workflow to RO transformation, etc.), data discovery services, e-learning services, cloud application integration capabilities, virtualized resource access, and different Web user interfaces, among others. The VRE and all its components are deployed at Poznan Supercomputing and Networking Center (PSNC), which provides powerful and scalable resources for the VRE operation.

The VRE was created following a user-centric approach, with scientists from different Earth Science research communities involved in the validation from a very early stage. The supersite initiative, and in particular INGV in Italy, was one of these research communities, and they have been advocating the adoption of the VRE, and the research object concept in general, to other colleagues in their community. As a result, the computing needs of several volcano Supersites are now supported, for some processing applications, by PSNC, which are accessed via the VRE. Powerful virtual machines are provided to the Supersite scientific communities in Ecuador, Chile, Iceland, Italy, which use them on a daily basis to unravel the behavior of their volcanoes. With the constant growth of the number and types of satellite and ground observations, and consequently on the complexity of the models, it is foreseen that in the next decade, the requests for High Performance Computing resources will increase steadily.

Acknowledgements:

The uses cases have been developed Based on research carried out by over a hundred scientists in the GEO-Geohazard Supersites and Natural Laboratories initiative (geo-gsnl.org), chaired by Stefano Salvi, INGV - Istituto Nazionale di Geofisica e Vulcanologia, Italy. We also acknowledge the EVER-EST partners who contributed to the realization of the VRE, and the other Earth Science research communities involved in the validation.