

Enabling Computing on the EPOS ICS-D

Alessandro Spinuso (1), Andrej Mihajlovski (1), Chris Card (2), Daniele Bailo (3), and Valerio Vinciarelli (3)

(1) KNMI - Koninklijk Nederlands Meteorologisch Instituut, Utrecht, The Netherlands, (2) BGS, British Geological Survey, Keyworth, United Kingdom, (3) INGV - Istituto Nazionale Geofisica e Vulcanologia, Roma, Italy

The EPOS-IP project is evaluating strategies to enable user-driven computations exploiting the large wealth of data and services discoverable through its integrated ICS-C catalogue. The actual data is accessible through the web services that are managed by geographically distributed and interdisciplinary RIs. The variety of methodologies and interoperability requirements between data and software suggest the need for identifying and implement general use cases supported by flexible and scalable e-Science solutions. These must be integrated in the EPOS architecture with the preliminary objective of assisting the users in basic tasks, such as resource allocation and data-staging, incrementally accommodating more complex computational scenarios and reusable workflows.

We will present and prioritise the different approaches envisaged for the integration of processing functionalities within the EPOS portal. These will allow users to develop and evaluate new data-intensive methods and to interactively control advanced visual-analytics tools and high-level workflows. Users of the EPOS portal will be able to populate their workspace according to search results of interest, and to selectively manipulate and process data within interactive computational environments deployed onto dedicated resources. Moreover, we envisage that many common operations will be eventually encoded as configurable scientific workflows that will automatically route data across distributed and domain specific processing services. Each scenario will require the communication with remote computational services, in order to dynamically allocate, prepare and access the needed resources. These will be heterogeneous and managed by national and European e-Infrastructures that will constitute the EPOS distributed integrated core services (ICS-D).

The use cases will be introduced together with their technical background. We will explain how we envisage the provision of containerized applications for data visual-analytics and processing, such as the Enlighten tool and Jupyter notebooks, addressing infrastructure orchestration (Docker, Kubernetes) and usability aspects. We will illustrate the approach for the adoption and integration of scientific workflow languages and technologies (CWL, Taverna), including their composition, validation and monitoring interfaces. We will discuss how each of the use cases will be enabled by the interaction with the EPOS ICS-C catalogue and how they will benefit from a flexible provenance management layer. The former will provide information about authorisation, location and access mode of the resources, while the latter will improve the user experience and productivity by tracing and documenting the execution of the methods. Thereby, each component of the EPOS computational framework will contribute to produce and ultimately deliver research data that comply to the FAIR principles (Findable, Accessible, Interoperable and Reusable). The activities will be also presented in the scope of the cooperation with ongoing H2020 initiatives such as VRE4EIC, EOSC-Pilot and the newly funded DARE (Delivering Agile Research Excellence on European e-Infrastructures) project.