Geophysical Research Abstracts
Vol. 18, EGU2016-6524, 2016
EGU General Assembly 2016

VERCE: a productive e-Infrastructure and e-Science environment for data-intensive seismology research

Jean-Pierre Vilotte (1), Malcolm Atkinson (2), Michele Carpené (9), Emanuele Casarotti (4), Anton Frank (8), Heiner Igel (5), Andreas Rietbrock (6), Horst Schwichtenberg (7), Alessandro Spinuso (3), and the VERCE Team

(1) Institut de Physique du Globe de Paris, Seismology, Paris, France (vilotte@ipgp.fr), (2) University of Edinburgh, UK, (3) KNMI, Netherlands, (4) INGV, Italy, (5) LMU, Germany, (6) University of Liverpool, UK, (7) Fraunhofer-SCAI, Germany, (8) LRZ, Germany, (9) CINECA, Italy

Seismology pioneers global and open-data access – with internationally approved data, metadata and exchange standards facilitated worldwide by the Federation of Digital Seismic Networks (FDSN) and in Europe the European Integrated Data Archives (EIDA). The growing wealth of data generated by dense observation and monitoring systems and recent advances in seismic wave simulation capabilities induces a change in paradigm. Data-intensive seismology research requires a new holistic approach combining scalable high-performance wave simulation codes and statistical data analysis methods, and integrating distributed data and computing resources.

The European E-Infrastructure project "Virtual Earthquake and seismology Research Community e-science environment in Europe" (VERCE) pioneers the federation of autonomous organisations providing data and computing resources, together with a comprehensive, integrated and operational virtual research environment (VRE) and E-infrastructure devoted to the full path of data use in a research-driven context.

VERCE delivers to a broad base of seismology researchers in Europe easily used high-performance full waveform simulations and misfit calculations, together with a data-intensive framework for the collaborative development of innovative statistical data analysis methods, all of which were previously only accessible to a small number of well-resourced groups. It balances flexibility with new integrated capabilities to provide a fluent path from research innovation to production. As such, VERCE is a major contribution to the implementation phase of the "European Plate Observatory System" (EPOS), the ESFRI initiative of the solid-Earth community.

The VRE meets a range of seismic research needs by eliminating chores and technical difficulties to allow users to focus on their research questions. It empowers researchers to harvest the new opportunities provided by well-established and mature high-performance wave simulation codes of the community. It enables active researchers to invent and refine scalable methods for innovative statistical analysis of seismic waveforms in a wide range of application contexts. The VRE paves the way towards a flexible shared framework for seismic waveform inversion, lowering the barriers to uptake for the next generation of researchers.

The VRE can be accessed through the science gateway that puts together computational and data-intensive research into the same framework, integrating multiple data sources and services. It provides a context for task-oriented and data-streaming workflows, and maps user actions to the full gamut of the federated platform resources and procurement policies, activating the necessary behind-the-scene automation and transformation. The platform manages and produces domain metadata, coupling them with the provenance information describing the relationships and the dependencies, which characterise the whole workflow process. This dynamic knowledge base, can be explored for validation purposes via a graphical interface and a web API. Moreover, it fosters the assisted selection and re-use of the data within each phase of the scientific analysis. These phases can be identified as Simulation, Data Access, Preprocessing, Misfit and data processing, and are presented to the users of the gateway as dedicated and interactive workspaces. By enabling researchers to share results and provenance information, VERCE steers open-science behaviour, allowing researchers to discover and build on prior work and thereby to progress faster.

A key asset is the agile strategy that VERCE deployed in a multi-organisational context, engaging seismologists, data scientists, ICT researchers, HPC and data resource providers, system administrators into short-lived tasks each with a goal that is a seismology priority, and intimately coupling research thinking with technical inno-

vation. This changes the focus from HPC production environments and community data services to user-focused scenario, avoiding wasteful bouts of technology centricity where technologists collect requirements and develop a system that is not used because the ideas of the planned users have moved on. As such the technologies and concepts developed in VERCE are relevant to many other disciplines in computational and data driven Earth Sciences and can provide the key technologies for a European wide computational and data intensive framework in Earth Sciences.