

VERCE Newsletter

Virtual Earthquake & seismology Research Community in Europe e-science environment

Summer 2013 • Issue 2

NEWS & ANNOUNCEMENTS

• **VERCE article in International innovation.** is available at:

<http://www.verce.eu/images/VERCEarticleInternationalInnovation.pdf>

• **VERCE deliverables of the midterm second reporting period** are available at:

<http://www.verce.eu/index.php/resources/deliverables>

• **EPOS latest Newsletter** is available at:

<http://www.epos-eu.org/newsletter/>

• **The European Super-sites Coordination Workshop** was held in Brussels on June, 10-11 2013.

• **Book on data-intensive strategies for all disciplines including the geophysical sciences just published:**

M. P. Atkinson et al., *The DATA Bonanza: Improving Knowledge Discovery in Science, Engineering, and Business*. John Wiley & Sons Inc., 2013, p. 580

DISPEL: describing VERCE's data centric workflows

Dr. Amrey Krause,

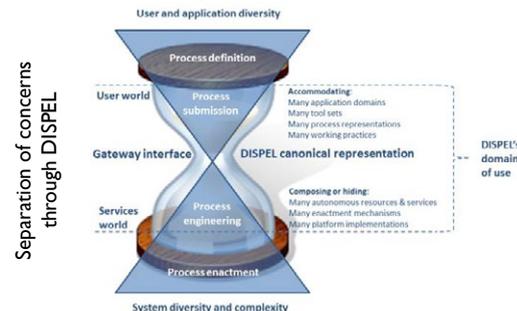
Applications Consultant (at EPCC, The University of Edinburgh)

VERCE is addressing the challenges of data-intensive distributed workflows for the seismology community.

Data-intensive use cases in VERCE are defined as a network of datastreams which are distributed across a number of private and public infrastructures. A typical scenario might combine compute-intensive simulations on HPC resources with the pre- or post-processing of large data volumes on dedicated data-intensive computing infrastructures. The orchestration of distributed workflows presents the seismology community with challenges in many areas, for example security, accounting and data transfer, which VERCE will address by using European data and computing infrastructures such as EUDAT and PRACE. Data flows are defined in a domain specific language, DISPEL, that creates an abstract representation of data stream-based applications at a high level. Through

a separation of concerns, DISPEL offers flexibility for the enactment of a dataflow graph. An enactment gateway validates and optimises a DISPEL data flow and manages the orchestration across a number of execution engines.

The abstract level of DISPEL allows the selection of appropriate execution engines, such as submitting jobs to HPC or Cloud environments, testing on a local cluster or a data intensive computing facility, so shielding the user from low level details.



Malcolm Atkinson
University of Edinburgh, IT Architect

The challenge for an IT architect is to deliver computational power that will enable substantial advances in seismology on an affordable and sustainable technological platform. Seismologists need access to data in their institutions, from their monitoring networks and from their archives. They need to interconnect two patterns of computation: the data-intensive processes where the speed of handling data is the limiting factor, and finite-element models of the physics of earthquakes and wave propagation. This requires new methods of working and new software infrastructure to create and steer compositions that cross the divide between

European grid and HPC infrastructures. To make this affordable and sustainable, we depend on alliances with EGI, PRACE, EUDAT, SCI-BUS and ER-Flow and on fast, reliable digital interconnection from the big computational centres and the seismology archives to the institutional clusters and researchers' mobile computers. For new science to flood from the VERCE computational "building" it needs a balance of familiarity and innovation, coherence in the presentation of its many facilities, consistency with the evolving neighbouring sciences and a stimulating ethos. Not a lot to ask of its builders and designers!



Massimo Cocco
Istituto Nazionale di Geofisica e Vulcanologia

The European Plate Observing System (EPOS) represents a scientific vision and an IT approach in which innovative multidisciplinary research is made possible for a better understanding of the physical processes controlling earthquakes, volcanic eruptions, unrest episodes and tsunamis as well as those driving tectonics and Earth surface dynamics. EPOS has a long-term plan to facilitate integrated use of data, models and facilities from existing (but also new) distributed research infrastructures for solid Earth science. One primary purpose of EPOS is to take full advantage of the new e-science opportunities becoming available. The aim is to obtain an efficient and comprehensive multidisciplinary research platform for the Earth sciences in Europe. This of course will include the availability of computational resources and IT platforms for data massive applications.

Experts Opinion



Summer 2013 • Issue 2

First VERCE training in Liverpool

Andreas Rietbrock

University of Liverpool

A first internal training workshop was organised at Liverpool University on 3-4 September 2012. The motives of this workshop were to get the two communities of ICT and seismology nearer, to make experts and researchers of the above two fields better understand each other needs, and to allow better progress towards project goals. The target group was focused to be project members from the seismological and IT groups working on the VERCE project. The workshop was organised in two parts: presentations and hands-on sessions. Seismology introduction, Use-cases of seismology, data-intensive processes and technologies, for example DISPEL, and technical introduction of GridFTP, iRODS and Unicore were presented to give relevant background information. The exercises, access to HPC resources and usage of DIS-

PEL were provided in hands-on sessions. 27 participants attended this workshop. Selected presentations are available from link of <http://www.verce.eu/index.php/training/tutorials>, but all workshop presentations and hands-on materials had been uploaded in the Redmine Wiki.



Collaboration with another EU project

Michel Campillo

Université de Grenoble, WHISPER



Jean-Pierre Vilotte

Institut de Physique du Globe de Paris, VERCE

WHISPER is a 4-year project supported by an Advanced Grant of the European research council (whisper.obs.ujf-grenoble.fr). WHISPER focuses on the use of the seismic ambient noise to monitor slight changes of properties in the solid Earth. Our passive approach that uses the waves mainly produced by the oceanic swell and not by sophisticated artificial sources, can be applied repeatedly for long periods of time to monitor the evolution of the rocks at depth. We develop new tools to interpret the changes in seismic data in term of spatial-temporal variations of the rocks properties at depth with the objective of understanding natural phenomena associated with dangerous hazards. For example, we found signals that precede volcanic eruptions or landslides and studied the changes that accompany the earthquakes.

The developments in WHISPER rely on the processing and analysis of massive data sets. The new observation networks in seismology can reach several hundreds of sensors recording continuously the 3 directions of ground motions. This results in millions of correlations functions every day that contain a wealth of information on the structure of the Earth and its temporal evolution. When dealing with such large data set, computational time and data management become a major difficulty and we devoted our efforts to make possible to handle present day large data sets. With a distributed storage (Irods) and a grid of clusters, both provided by the Ciment Project at University of Grenoble, the codes of the WHISPER suite succeeded in handling and processing large data sets (~100 TBs), such as several year of continuous recording –

before and after the great Tohoku earthquake (Mw 9.1, 2011, North East Japan) – of the NIED Hi-net network in Japan.

The VERCE initiative will provide the e-science environment that is critically needed for the new scientific and data-intensive computing challenges of WHISPER, and that will leverage those new research methodologies for the seismology research community. Today WHISPER and VERCE are actively collaborating toward an innovative data-intensive e-science environment in support for these data-intensive applications that consist of multiple phases where data acquisition interleaves with data processing and analysis, generating highly parallel and asynchronous data workflows together with massively parallel data access. In particular, VERCE will provide efficient scalable and transparent distributed data storage, data management and data transfer services, together with stream-oriented execution models that enables overlapping data processing and analysis computation with I/O operations, in support for high throughput under heavy asynchronous versioning-based access to data.

The collaboration between WHISPER and VERCE will provide a new data-intensive research paradigm allowing extraction of new scientific information from the rapidly wealth of data – continuous time-series – provided by the dense observation and monitoring networks of the seismology community, with important implications to added societal applications, such as seismic hazard and risk assessment and monitoring, and seismic exploration and exploitation of energy resources.