D-NA2.5: Final report on validation and evaluation of enabled applications deployment and use cases (in coordination with WP6 and WP8)

Tuesday 10th November, 2015

Project acronym:	VERCE
Project n :	283543
Funding Scheme:	Combination of CP & CSA
Call Identifier:	FP7-INFRASTRUCTURES-2011-2
WP:	WP2/NA2, Pilot applications and use cases
Filename:	D-NA2.5.pdf
Author(s):	E. Casarotti
Location:	
Type of document:	EU Report
Dissemination level:	Restricted
Status:	Final
Due date of delivery:	30/09/2015
Reviewer:	Malcolm Atkinson, Heiner Igel, Alessandro Spinuso
Keywords:	data-intensive, cpu-intensive, HPC, earthquake, seismology, data infrastructure, forward modeling, inversion

---

Version	*Author*	*Date*	*Comments*
1	E. Casarotti (INGV)	25/09/2015	Initial draft for comments
2	F. Magnoni (INGV)	28/09/2015	comments
3	E. Casarotti (INGV)	29/09/2015	final drafts for internal reviewers
4	E. Casarotti (INGV)	30/09/2015	reply to reviewers
5	E. Casarotti (INGV)	07/11/2015	additional comments
5	F. Magnoni (INGV)	09/11/2015	additional comments

Copyright notice

Contents

Executive Summary

WP2/NA2 led the selection and the design of use case scenarios, evaluating the deployment in the VERCE platform of the pilot applications. In addition, it provided the support of the seismological communities to the WPs with tools, guidelines for user interfaces, suggestions, testing efforts and training sessions.

In detail, the objectives of WP2/NA2 are:

1. select the existing pilot data-intensive applications and design sound use case scenarios;
2. analyze and define a priority strategy during the project with JRA1, SA1 and JRA2;
3. support and evaluate the "productizing" transition of the methods and their implementation performed by JRA1;
4. support SA1 and JRA2 with application requirements for the definition of the workbenches and functionalities;
5. support and evaluate the deployment and the efficiency of the pilot applications and their use case scenarios on the VERCE platform;
6. define and provide, in collaboration with NA3, improved documentation, best practice guides and tailored training session material;
7. define and provide, in collaboration with NA4, demonstrators and dissemination material;
8. provide requirements and support to SA3 and JRA2 for tailored interfaces of the scientific gateways targeted to the developers and the users.

In this report we highlight how NA2 has dealt with these objectives as described in the *Description of Work*.
Overall, it has achieved all of its objectives by closely collaborating with partners and work packages to develop, communicate and drive compelling and challenging use cases.

1 NA2 Activities according to DoW Tasks

In this section, we report a detailed assessment of NA2 tasks described in the DoW:

1. Selection and prioritization of the use cases
2. Validation and evaluation of the pilot applications deployment and usage of the use cases
3. Define and provide, in collaboration with NA3, user-oriented materials
4. Define and provide, in collaboration with NA4, outreach materials
5. Define and support with SA3 the functionalities and content of the scientific gateway

1.1 Task 1: Selection and prioritization of the use cases

This task covered the first phase of VERCE in collaboration with JRA1. Specifically, NA2 designed, collected and commented on surveys aimed to better understand the typology of the use cases offered by the project participants. The scientific topics of the proposed pilot applications ranged from velocity structure monitoring and determination using data- and CPU-intensive approaches, to seismic event detection and location and to rapid tsunami modelling. The prioritization has been based on the typology of the use cases, since many of the applications share common software requirements, on the similarity in HPC or data resources of the test-bed and on ranking the impact of the use case for the community. It can certainly be stated that the presented use-cases and the prioritization reflects the state-of-the-art issues in seismology (and in solid Earth sciences in general).
In particular, two seismological use cases emerged by NA2 survey in order to be included in the VERCE platform. The first consists of enacting 3D forward modelling seismic wave propagation codes such as SPECFEM3D (HPC CPU intensive) and the misfit analysis of the simulated waveforms. The second is the "ambient noise cross-correlation" use case which is strictly a data-intensive use case.
In the Summer 2013, the Steering Committee, in agreement with the review panel, considered to give highest priority to the CPU-intensive use case. Focusing on the forward modelling wave simulation (based on SPECFEM3D) and waveform misfit analysis allowed to make a true step forward in the deployment of a scientific gateway offering a tangible benefit to the seismological community, with planned further developments beyond VERCE within the consolidating EPOS initiative.
The initial data-intensive development as originally planned has been considered not achievable within the timeline of the project. Remarkably, the specific scientific community of "ambient noise" is fragmented, delaying the adoption of authoritative implementations, in contrast with HPC seismology where SPECFEM3D is definitely an authoritative code used by many scientists. In addition, multiple ways of extracting high order information from instrumental noise are emerging so that researchers require to have full control of applications and tools.
Nevertheless, while the original plan was strictly related to the particular case of ambient noise correlation, any data-intensive use case entails the ingestion of a large amounts of data and the execution in a pipeline of a number of processing analyses. VERCE decided to support this high diversity of applications and requested sustaining the development of ObsPy and providing dispel4py, a Python library to describe abstract workflows for distributed data-intensive applications, automatically mapped in multiple environments as Storm and MPI clusters or Jupyter notebooks. This toolbox provides the ability to manipulate seismic traces and releases the seismologists of technical difficulties, thanks to the automatic management of input and output, the provenance-handling capabilities, the possibility to contribute to a shared library of workflows and processing elements and the possibility for the user to select his own comfortable and performant development environment.
The flexibility of dispel4py has been fundamental in the VERCE scientific gateway as tool both for provenance tracking and for seismogram misfit analysis, orchestrating data- and CPU-intensive aspects in the infrastructure.

1.2 Task 2: Validation and evaluation of the pilot applications deployment and usage of the use cases

The evaluation and validation have been carried out in two phases. The initial efforts consisted in the definition of the workflow for the use cases in the framework of a continuous feedback with JRAs, SA1, SA2 and SA3. Processing steps, seismological tools (e.g. Obspy), HPC codes (e.g SPECFEM3D) and frameworks (e.g. iRODS, SCI-BUS) have been evaluated and in some cases refactored for the use in the VERCE platform, sometimes with the contribution of the original developers. In particular, VERCE received additional support from the developers of SPECFEM3D in order to adapt the parameter definition of the code for the scientific gateway. In addition to the coding support, the possibility to interact with the seismological community from the beginning of the project has been crucial to highlight the requirements that users could espect from the VERCE platform. The analysis has been performed based upon metric values, watching costs, installation procedure, licensing and computational demands.
During the second phase of VERCE, the validation and evaluation process concerned principally the scientific gateway. Collaborative evaluation and recursive design of functionalities of the VERCE portal have represented the most important path in order to accomplish the expected results. A deep alliance matured between NA2, SA3 and JRA1, with productive and well-structured face-to-face (code-sprints) meetings between components of the WPs involved in the development of the VERCE Science Gateway. NA2 blended progressively its activities with those of the other work packages, providing guidelines and fundamental seismological point of view, tracking weaknesses and evaluating the implemented solutions, carrying out beta testing efforts and benchmarking results, implementing requests from scientific communities. NA2 contributed to the creation of the initial library of seismological processing elements, ready for use in dispel4py preconfigured pipelines for instrumental noise analysis. Such initial repository will help the researchers to explore the library focusing on the scientific innovation and not on the technical aspects.

1.3 Task 3: Define and provide in collaboration with NA3 users oriented materials

NA2 has always worked closely with NA3 in the efforts to provide materials oriented to the users. Complete details of the activities can be found in the NA3 final report. Here, we highlight the significant contribution of NA2 to the Training Workshops with lectures about general seismology and the use of the scientific gateway, step-by-step how-to presentations about meshes, extensive tutorials about how to obtain and use security certificates in the portal.

1.4 Task 4: Define and provide, in collaboration with NA4, outreach materials

NA2 contributed to the outreach materials provided by NA4 initiatives with:

- some interviews about goals and significance of VERCE;
- the presentation of posters, papers and talks at seismological and computational science international conferences;
- the creation of tutorial movies about the functionalities of the scientific gateways;
- the contribution to the final movie describing the portal and the experience of building the VERCE platform.

The NA4 final report lists all of these contributions.

1.5 Task 5: Define and support with SA3 the functionalities and content of the scientific gateway

As reported in section 1.2, NA2 supported SA3 in the establishment of VERCE scientific gateway. In addition of recursive feedbacks, it has been involved in :

- the definition of the workflow steps for forward wave modelling in the gateway;
- the definition of the portal graphical user interface UX that translates the file-centric user interface of SPECFEM3D in the easier structure of the gateway;
- the creation and validation of the library of meshes and corresponding tomographic velocity models;
- the creation of visualization scripts (e.g. wave propagation animations and Google Earth KML);
- the automatic validation of the mesh provided by users;
- the development of dispel4py processing elements for the production of input files, the visualization, misfit data and synthetics analysis;
- the benchmarks of outputs and performance of dispel4py.
- the estensive test of certificate security steps for credential management from the point of views of the users.

Technical details of the implementation are reported in the SA3 final documents.

2 Fulfillment of the DoW Objectives

The following table contains a schematic summary of the NA2 achievements considering the objectives reported in the *Description Of Work*. As illustrated in the previous section, NA2 deeply collaborated with all the work packages in order to successfully accomplish its specific goals.

DoW Objective	*Achievement*	*Comment*
Select the existing pilot applications and design sound use case scenarios; Analyze and define a priority strategy during the project with JRA1, SA1 and JRA2.	Pilot applications and use cases have been successfully selected, analyzed and prioritized during the first half of the project.	During the second half, continuous evaluation of use case implementations led, in agreement with reviewers' comments, to a shift in the priority. Both for practical and strategic motivations, we have successfully focused on 3D forward modelling wave simulations (based upon SPECFEM3D) and waveform misfit analysis.

Support and evaluate the transition to production of the methods and their implementations performed by JRA1. Support SA1 and JRA2 with application requirements for the definition of the workbenches and functionalities.	Close collaboration with JRA1 and SA1 for the evaluation, the improvement and (in some cases) the refactoring of the wave propagation solvers (e.g SPECFEM3D, Seisol) and other seismological tools (e.g. Obspy, dispel4py PEs). NA2 provided fundamental seismological point of view suggesting functionalities based upon the contribution of selected users from the geoscience community.	Fully completed
Support and evaluate the deployment and the efficiency of the pilot applications and their use case scenarios on the VERCE platform.	NA2 blended progressively its activities with those of the other work packages, providing guidelines and fundamental seismological point of view, tracking weaknesses and evaluating the implemented solutions, carrying out beta testing efforts and benchmarking results, implementing requests from scientific communities.	The recursive feedbacks between seismological community and VERCE has been the guide for delivering an useful tools for scientists.
Define and provide, in collaboration with NA3, improved documentation, best practice guides and tailored training session material.	NA2 actively supported training material and workshops both virtual and in-person. In addition to the activities related to the forward simulations, the meshing process and the science gateway, NA2 contributed to the creation of the initial library of dispel4py seismological processing elements and preconfigured pipelines for instrumental noise analysis.	The activity is still in progress with the inclusion of the functionalities released during the last period.
Define and provide, in collaboration with NA4, demonstrators and dissemination material.	NA2 actively collaborates for outreach, for dissemination material, in particular during seismological international conferences, and for the creation of tutorial how-to movies.	NA2 has been involved in these activities since the first release of the scientific gateway during the second half of the project.

Provide requirements and support to SA3 and JRA2 for tailored interfaces on the scientific gateways targeted to the developers and the users.	NA2 supported SA3/JRA2 in the establishment of VERCE scientific gateway, through collaborative evaluation and recursive design of functionalities. NA2 blended progressively its activities with those of the other work packages, tailoring the interface of the scientific gateway both to the seismological expert and to students.	Reports from training meetings suggest that the interface of the scientific gateway is successfully decreasing the initial burden of HPC seismological simulations. Concerning the portal, the misfit part is a relatively young implementation and it needs to be exposed to a wider audience and cases, possibly under the consideration of EPOS IP.

Final comments

Feedback from attendees to the training workshop or from colleagues and the consideration of EPOS-IP suggest that the seismological community sees the VERCE initiative as an important tool to reduce the initial intellectual and practical barriers encountered when researchers first engage with computational seismology. Furthermore, we remark that VERCE has been an exceptional experience that brought together at the same table scientists with different expertise and backgrounds. Being successful in these heterogeneous initiatives is not inevitable; it is challenging. The creative enthusiasm that computational scientists and seismologists dedicated to VERCE is one of the key results achieved.