D-NA2.2.1: Report on the metrics for the assessment of the VERCE platform based on the selected applications (in collaboration with WP4)

19/10/2012

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-NA2.2.pdf
Author(s):	A. Michelini
Location:	http://www.verce.eu/Repository/Deliverables/RP2/
Type of document:	Deliverable
Dissemination level:	Public
Status:	Final
Due date of delivery:	01/10/ 2012
Reviewer:	J.-P. Vilotte
Keywords:	data-intensive, cpu-intensive, HPC, earthquake, seismology, data infrastructure, forward modeling, inversion, metrics, evaluation

Version	*Author*	*Date*	*Comments*
1	A. Michelini (INGV)	11/09/2012	Initial draft for comments
2	A. Michelini (INGV)	24/09/2012	Revised draft after comments from H. Igel and J.-P. Vilotte
3	A. Michelini (INGV)	01/10/2012	Revised version with comments by A. Tosi taken into account

Copyright notice

VERCE, *Virtual Earthquake and seismology Research Community e-science environment in Europe*, is a project co-funded by the European Commission as an Integrated Infrastructure Initiative within the 7th Framework Programme. VERCE began in October 2011 and will run for 4 years.

Contents

Executive Summary

The WP2/NA2 addresses various goals. It is responsible for selecting the existing pilot data-intensive applications and design sound use case scenarios and together with WP8/JRA, WP5/SA1 and WP9/JRA2 analyse and define a priority strategy through the project. Secondly, WP2/NA2 supports and evaluates the "productising" transition of the methods and their implementation performed by WP8/JRA1. It also supports WP5/SA1 and WP9/JRA2 with application requirements for the definition of the workbenches and functionalities. Of major importance for WP2/NA2 is the support and evaluation of the deployment and of the efficiency of the pilot applications and their use case scenarios on the VERCE platform. In collaboration with WP3/NA3, the WP defines and provides improved documentation, best practice guides and tailored training session material and, together with WP4/NA4, defines and provides demonstrators and dissemination material while providing requirement and support to SWP7/A3 and WP9/JRA2 for tailored interfaces of the scientific gateways targeted to the developers and the users.

This deliverable reports the initial development of a metrics to assess the pilot applications deployment and usage of the use cases. The task is to develop evaluation indicators and a strategy to measure the effective improvements attained using the e-science environment provided by VERCE. The proposed evaluation procedure is expected to stimulate continuous interaction between the seismologists and the IT developers and provider of services through the adoption of an objective metrics. The metrics is targeting needs and constraints set by the seismological community—the primary beneficiaries of the VERCE architecture and its services and will serve to the purpose of evaluating the VERCE e-science environment.

In this reporting period, to the goal of defining the metrics above, it has been adapted the so called "iron triangle" type of metrics that is based on *costs*, *time* and *quality* of the target use case or of the service offered. The costs and time metrics provide elements for the evaluation of the effort required in order to join and use the VERCE platform and its use cases. The costs include both effective costs such as hardware and connectivity deriving from assessing and implementing the the VERCE platform and costs resulting from the man power required to implement or develop upon it. The quality indicators, provide guidelines on where additional effort should be put to improve the quality of the software deployed.

There have been proposed about 20 questions that will be compiled by the seismologists involved in the deployment of the use cases. These have been designed to meet the MS9 (M-NA2.4) and provide a measure of the impact of VERCE in the earthquake and seismology community (together with WP4/NA4).

The next steps include the adoption of the proposed metrics to the use cases and the VERCE platform as they are progressively implemented. Given the novelty of the metrics used, we expect it will also need in some cases substantial adjustments to meet the requirement of providing both feedback to the community of practice and, internally, to improve the quality of the provided platform and services.

1 Purpose and structure

"...This task will require the development of evaluation indicators and strategy to measure the effective improvements attained using the e-science environment provided by VERCE. The strategy will draw on the methodology of empirical software engineering. The verification and evaluation process is expected to foster interaction and recursive feedback with the software engineering developers of the JRAs and the platform architects of JRA1 and the service providers of SA1 and SA2. This will assess the goodness and robustness of the VERCE platform and spring improvements of its architecture..." [*VERCE DoW*].

This document provides guidelines on how to perform the evaluation of the e-science environment provided by VERCE. This will be inevitably based on gains obtained in the implementation of the use cases on the e-science environment. The framework is necessarily general since the typology of the candidate use cases to be assessed within VERCE varies. The general guidelines discussed below reflect some general principles to be adopted to measure the implementation. Focus, in fact, is not on the use cases themselves but rather on the manner their implementation is achieved. Overall, application of the metrics presented in this document should provide the community of practice (primarily the seismologists) with elements for judging the gains obtained through the use of the VERCE platform.

The document is organized into an introduction to the problem followed by a description of the basic criteria adopted, a short discussion and conclusions and the next steps to be undertaken.

2 Problem definition

The primary objective is to provide a metrics to estimate the success of VERCE e-science environment from the implementation of the use cases. This is very important since only when there are some measures available it becomes possible to rank the effectiveness of the gains obtained through VERCE. Adoption of a metric system can be also very valuable in the cyclic process of refining the implementation of the use cases and, by so doing, the provision of additional elements for improvement of the VERCE e-science platform.

In general and in order to provide a metrics it is important to define what are the likely targets of the e-science platform and then to evaluate to what degree the targets have been reached or fulfilled using the solutions adopted through the project. Target setting is thus a key part of the estimation.

Given the aims specified in the DoW, the resulting metric should provide elements to improve the overarching architecture through feedback and interaction between the users, the architecture experts and the service providers. Nonetheless, it appears that the improvements obtained through VERCE will be best assessed by the user community (i.e., those that will eventually benefit of the platform when doing their research). It follows that the metrics presented below reflects primarily the requirements set by the *desiderata* of the user community (i.e., the seismologists) and, to some extent, those IT experts that work in the same institutions closely to the formers and who will be tasked to implement parts of the procedures and/or the VERCE platform itself.

If a commercial software approach is followed to make the platform evaluation, three aspects become of relevance—costs primarily in terms of human resources to implement the platform—, time required—for example to run a pre-defined use case or to develop a new one using the libraries made available by the platform service—, and the quality of the results obtained. This approach is also known as the *Iron triangle* [e.g., *Atkinson, R. (1999).* Project management: cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International journal of project management*.]

It is expected that the overarching purpose of the measurement system presented below will be that of helping VERCE to make the right decisions on the type and functionalities of specific services to be implemented. Finally, we think it relevant that only a handful of as much as possible independent measures (no more than 15-20), are to be sought in order to profile the VERCE platform implementation.

3 Candidate metrics

In general, when evaluating software, it is common to rely on matrices which entries provide a profile of the software itself. In VERCE, when evaluating use cases consisting of a mix of workflow procedures, data management and various software, we will also construct a matrix which entries describe the various facets of the use case itself. The required evaluation will be then achieved by comparing the profile results of the use case implemented on the VERCE platform and the same use case implemented outside the VERCE platform.

Adaptation of the iron triangle mentioned above leads to the introduction of estimates based on costs of implementing a given software solution (e.g., number of person months required to develop and/or implement a particular software package, hardware solution required), time required to develop or to modify a use case to obtain a stable installation and the quality measured in the terms of fitting the target requirements expected by the use case itself). The result of the analysis consists of a matrix with entries describing personnel involvement expressed in terms of *person days*, effective costs in Euros, other measurable using an arbitrary scale ranging between 1 and 5 (1-poor, 5-best), and some final yes/no entries that can help to provide an improved profile for example of the openness of the software used.

3.1 Costs:

The costs that can be encountered in the implementation of specific use cases are primarily of two kinds—hardware and personnel. Software costs are not considered since it is expected that all software used within VERCE is of public access and not commercial[1]. Costs nevertheless are relevant when a software solution is to be implemented for a candidate use case or if it is required anyhow by those that want to enter a take part to or take advantage of the overall VERCE platform (e.g., implementation of GridFTP). Thus, the following aspects are thought of relevance to evaluate the costs

1. Overall personnel costs (person days) resulting from the sum of all the involved people to reach the use case production phase (i.e., from the IT engineer developing or improving the development of the software to the community user serving as beta tester of the use case)

2. Potential costs deriving from the adoption of specific hardware (expressed in Euro) for the implementation of particular use cases or software solutions (e.g., hard disks). This may include also connectivity upgrades which may be of particular relevance if transfer speed network upgrades are required.

3.2 Time:

3. Installation—time required to install the necessary software on the target server(s) and by so doing join for example the VERCE platform, or add a use case, develop a workflow, ... (person days)

4. Tuning the installation with the appropriate parameters (person days)

5. Testing the implementation after tuning (days)

3.3 Quality:

6. Quality of the *installation procedure* (1-5). This measurable is expected to provide an overall assessment of how simple the installation has been. For example, installations that run smoothly and errorless without additional intervention (e.g. without the installation of additional libraries)

[1]We note that this may not be always the case since it may be advantageous in some instances to avail of commercial software designed specifically to provide particular services otherwise to be developed ”ex novo” within the project.

will have very high scores. The quality of the installation procedure depends also on the level of installation manual provided.

7. *Friendliness of the software* (1-5). This measurable provides an estimate of "how easy" to use is the use case and its software (e.g., quality of the tutorials and type of examples being provided).

8. *Matching the targets specified by the community* (1-5) (e.g., ability to handle very large data volumes, optimal data staging of large data volumes, ability to request data and download, types of pre-processing, correctness of the pre-processing, ...).

9. *Software performance* (measurements of CPU cycles, up-/down-load performances in Mb/s, etc.).

10. *Gaps of the software for the given use case* (1-5). This measurable attempts to provide a measure of the missing features of the use case. This measure is somewhat subjective and assignment of the score should be based on the relevance that a missing feature of the s/w has in the context of the candidate use case.

11. *Robustness of the target use case procedure* (1-5). This measurable attempts to provide a measure of the reliability of the use case. In assigning the score, elements such as the number of failures/crashes or whether the s/w features automatic re-start after a crash should be considered.

12. *Controlled access* (1-5). This measurable is intended to provide elements on the security of the VERCE platform by examining the level of the use case fine-grained controlled access, (e.g., by user or group through secure password mechanism or other standards including Grid Security Infrastructure (GSI), Kerberos, Shibboleth and how user friendly this all is).

13. *User community behind the use case* (1-5). Existence of a user and of a developer community is extremely relevant for software implementation, improvement and sustainability.

14. Client/server development (yes/no).

15. Metadata sustainability (yes/no).

16. Scalability in terms of data volumes transfers/storage (1-5). It indicates the level of scalability of the software in terms of number of files and file sizes.

17. Computer platforms (1-5). Diversification of computers on which it can be run (i.e., score 5 if it runs on all major computer operating systems).

4 Discussion

The candidate use case metrics presented in the previous section have been designed to provide measures of how the different software implementations fit the VERCE use case functionalities. This evaluation process should be carried out periodically throughout the project since it will enable improvements in both the engineering and the service provider parts of VERCE.

The selected metrics provide elements for the evaluation of the software that are expected to give both the Steering Committee, SC, and the project Executive Board, PEB, rounded information on the features of the software tested on the VERCE platform. The evaluation is based on measurable quantities referring loosely to costs, time and quality to produce an evaluation matrix. Selections will be then made on the basis of the scores resulting in the evaluation matrix compared to similar ones extracted from the implementation of the use cases on the users' platforms.

The proposed 17 questions are expected to profile the software from different perspectives. They include actual costs expressed in terms of person days of human resources required for installation and testing,

and effective costs encountered when the service is implemented on the selected hardware. We note, however, that this last estimate related to the hardware costs is very sensible to the selected solutions and it may not be applicable at this very stage of the Project.

The time measurable provides estimates of both the time (and effort) required by the system engineer in collaboration with the community user to install and tune the software (expressed in person days) and the overall time (days) expected for thorough installation and running the given software solution for the sought candidate service. Note that the human effort time of the overall personnel costs of Q1 results from adding the effort expressed in Q3 and Q4. Q5 is instead expected to provide indications on the effective absolute time estimated to have the software running.

The quality metrics above make an attempt to provide rounded estimates of how the software is provided, how easy is its installation, the size of its community of practice, its main features with a glance at its extendibility and its openness. The scoring adopted suffers of some inherent subjectivity (i.e., the 1-5, 1-poor, 5-best) and generality. Nevertheless, it is thought to be meaningful whenever the requirements of the target candidate services are spelled out clearly thus resulting straightforward what the implemented software is expected to accomplish. Q9, in any case, provides specific measures of the absolute performance of the software (on a given computer architecture). Overall, it is thought that these measures, while to some extent gross and general, can nevertheless provide opportune elements of judgment to make appropriate selections of the software for implementation in the final production environment of VERCE.

5 Conclusions and next steps

The evaluation framework proposed here must be tested against the software so far implemented for the candidate use cases. This will be carried out during the second year of the project by testing the proposed metrics against the implemented use cases. A revised version of this document at the 24 months report deadline will provide important elements and likely modifications based upon the experience matured form application of the framework.

6 Glossary and Links

Definition	Description
2D-surface Wavefields	A collection of "synthetic seismogramms" on every-surface point of the simulated geographical region PRACE http://www.praceproject.eu/
3C	Three component data recorded by seismometers.
ADMIRE	Architectures for Data Intensive Research - http://www.admire-project.eu/
AGU	American Geophysical Union http://sites.agu.org/
API	Application Program Interface, an inter-software communication specification used for accessing functionality or services from programs.
ArcLink	A protocol for data transfer from geographically distribute data archives based on time windows - http://www.seiscomp3.org/wiki/doc/applications/arclink
ATLAS	Automatically Tuned Library Algebra Software
AXISEM	A parallel spectral-element method - http://www.seg.ethz.ch/software/axisem
BADW-LRZ	The Bavarian Academy of Sciences and Humanities - Leibniz Supercomputing Centre - http://www.lrz.de/english/
BDII	Berkeley Database Information Index
BLAS	Basic Linear Algebra Subprograms
CINECA	Consorzio Interuniversitario Cineca
cloud	See cloud computing.
cloud computing	A (business) model for enabling the delivery as a service of shared computing resources such as CPUs, networks, storage and applications to multiple users.
component	One of the computational elements involved in a data-intensive or computational process, such as: application codes, scripts, workflows, services, catalogues, registries, data collections, data resources, functions, gateways, libraries, PEs, PE instances, format definitions and types.
cpu-intensive applications	Compute-intensive applications are those that devote most of their execution time to computational requirements and typically require small volumes of data although they can produce very large to huge data volumes. Compute-intensive is a term that applies to any computer application that demands a lot of computation, such as forward modeling programs for seismic wave propagation or other scientific applications.
cross-correlation	In signal processing, cross-correlation is a measure of similarity of two waveforms as a function of a time-lag applied to one of them.
CREAM	Computing Resource Execution And Management

DAGMan	Directed Acyclic Graph Manager, Pegasus engine for executing workflows on available compute resources.
data	Any digitally encoded information that can be stored, processed and transmitted by computers. Includes text files, database records, images, video sequences and recordings. In seismology, it is customary to refer to data as those digital values acquired by the pair seismometer/data-logger deployed in the field which is recording the ground motion object of the analysis.
data archive	The long-term storage of scientific data and methods.
data integration	The process of combining data residing at different sources and providing the user with a unified view of these data. This process emerges in a variety of situations both commercial (when two similar companies need to merge their databases) and scientific (combining research results from different repositories) domains.
data mining	The process of automatically extracting patterns from data using techniques such as classification, association rule mining and clustering.
data-analysis expert	An expert in building or using knowledge discovery methods in a data-rich environment. In the context of VERCE, for example, they build libraries, such as ObsPy or workflows, such as cross-correlation or visualisers, such as SDX.
data-intensive	An adjectival phrase that denotes that the item to which it is applied requires attention to the properties of data and to the ways in which data are handled.
data-intensive applications	Data-intensive applications are those that are used to process large volumes of data typically of the order of terabytes or petabytes in size and referred to as Big Data. They require large volumes of data and devote most of their processing time to I/O and manipulation of data.
data-intensive computing	Computing that necessitates attention to any relevant property of data, including their volumes, distributed locations, and the heterogeneity of their formats and storage structures.
data-intensive engineer	An expert in designing, providing, tuning, operating and improving the use of computational platforms for data-intensive tasks.
data staging	Indicates the process of moving (stage-in) the data to the site of execution. In some cases, a data stage-out is specified to download the results and data clean-up after execution. In a use case, there may be several such data staging activities, which could occur sequentially or in parallel.

DCI	Distributed Computing Infrastructure
DECI	Distributed European Computing Initiative / DEISA Extreme Computing Initiative
DEISA	Distributed European Infrastructure for Supercomputing Applications
DIRAC	The DIRAC (Distributed Infrastructure with Remote Agent Control) project is a complete Grid solution for a community of users needing access to distributed computing resources. - http://diracgrid.org/
distributed computing	The collective use of distributed resources, including data and applications, to solve a computational problem.
domain expert	A person who is skilled in a particular field of research or decision making. In the context of VERCE, seismologists and later other Earth scientists.
DoW	Description of Work
e-Infrastructure	The ICT element of a research infrastructure, i.e. a distributed collection of data, storage and compute resources, interconnected by digital communications and organised to serve a common research purpose. It includes the hardware, software, middleware, staff, operational procedures and policies needed to make it operate for that purpose, and requires maintenance to function in the evolving digital environment and to meet the changing needs of its user communities.
Earth Model	Assumed one to three dimensional parameter sets of the earthÕs interior on which a simulation is based.
EDGI	European Grid Initiative (http://edgi-project.eu)
EDIM1	Edinburgh Data-Intensive Machine 1, University of Edinburgh experimental architecture for data-intensive computing.
Dispel	Data-Intensive Systems Process Engineering Language, a workflow composition language for data-intensive applications.
Dispel Gateway	The new name of ÒADMIRE gatewayÓ
EGI	European Grid Infrastructure - http://www.egi.eu
EGI ESR VO	Earth Science Research (ESR) Virtual Organisation (VO) in EGI - http://www.euearthsciencegrid.org/content/esr-vo-introduction
EGU	European Geophysical Union, www.egu.eu
EIDA	European Integrated Data Archives infrastructure - http://www.verce.eu//ITCoordinationMeetingFebruary2012/EIDA-Overview.pdf
EMI	European Middleware Initiative - http://www.eu-emi.eu/
EMSC	Euro-Mediterranean Seismological Centre

enactment	"The execution of a workflow on a computational platform; this generally involves coordinated use of multiple and often heterogeneous communication, data and compute resources."
ENVRI	"Common Operations of Environmental Research Infrastructures http://www.egi.eu/about/EGI.eu/EGI.eu_projects/ENVRI.html"
EPOS	"“European Plate Observing System” is an ESFRI approved infrastructure currently in its preparatory phase and funded by the EC (http://www.epos-eu.org). "
EQ	EarthQuake
ESFRI	European Strategy Forum on Research Infrastructures.
ESSL	IBMÕs Engineering and Scientific Subroutine Library
EUDAT	EUropean DATa is a project currently funded by the EC for the development of the Common Data Interface (http://www.eudat.eu).
EUGridPMA	European Union Grid Policy Management Authority. International organisation to coordinate the trust fabric for e-Science authentication in Europe,-http://www.eugridpma.org
Event	data Catalog entry, from an event-catalog (usually < 10lines ascii orQuakeML), including location, magnitude and/or moment values and type of an earthquake.
F2F meeting	Face-to-face meeting
FD	Finite-Difference wave propagation
Forward Simulation	Simulation of seismic wave-propagation, results in synthetic seismograms.
FP7	Seventh Programme Framework
FTS	gLite File Transfer Service - see gLite below
Full-Waveform Inversion	Tomographic inversion of the real seismograms (or differences between real and synthetic seismograms) to determine the underlying earth model.
gateway	A software subsystem, typically at the middleware level, that accepts requests for computational and data-handling tasks. It vets those requests to establish whether they are valid, e.g. are syntactically and semantically consistent, and are authorised. Requests that are not validated are rejected. Requests that are accepted are passed to other software systems, at the same or other locations, for execution. The gateway may partition and translate requests in order to combine heterogeneous services.
gLite	Lightweight Middleware for Grid Computing - http://glite.cern.ch
Globus	Globus - An open source Grid software that addresses the most challenging problems in distributed resource sharing (as per www.globus.org)

Globus Online	A cloud-based, reliable, high performance and secure service for managing file transfers https://www.globusonline.org/
Globus Toolkit	Open source software toolkit used for building grids - http://www.globus.org/toolkit/
GPGPU	General-purpose computing on graphics processing units
GPU	Graphics Processing Unit
GRelC	Grid Relational Catalog Project - http://grelc.unile.it/home.php
grid	A system that is concerned with the integration, virtualisation, and management of services and resources in a distributed, heterogeneous environment that supports collections of users and resources (virtual organisations) across traditional administrative and organisational domains (real organisations).
GridFTP	Grid File Transfer Protocol, an extension of the standard FTP for use with grid computing.
GridSpace2	Provides a Web 2.0-based Experiment Workbench for joint development and execution of virtual experiments by groups of collaborating scientists. - https://gs2.plgrid.pl/
GSI	Grid Security Infrastructure. Also see Globus.
GSI-SSHTerm	A Java based terminal client for accessing the Grid - http://www.grid.lrz.de/en/mware/globus/client/gsissh_term.html
GT	Globus Toolkit
GUI	Graphical User Interface
HDF5	Hierarchical Data Format (HDF, HDF4, or HDF5) is the name of a set of file formats and libraries designed to store and organize large amounts of numerical data (http://www.hdfgroup.org/HDF5/)
high-performance computing (HPC)	Use of powerful processors, high-speed networks and parallel supercomputers for running computationally intensive applications.
IDE	Also known as Integrated Development Environment, a software system designed for supporting software writing, often including a source code editor, a debugger and build automation tools.
IGE	Initiative for Globus in Europe
INCA	Periodic, automated, user-level testing of Grid software and services - http://inca.sdsc.edu/drupal
INGV	Istituto Nazionale di Geofisica e Vulcanologia
INSPIRE	Infrastructure for Spatial Information in Europe, an EU directive aimed at enabling the access, sharing and re-use of spatial data for governance and policy making purposes
IRIS	Incorporated Research Institutions for Seismology (Data-Center)
globalCMT	Global Centroid-Moment-Tensor Project
iRODS	Integrated Rule-Oriented Data-management System - https://www.irods.org/

ISO 20000	The international standard for IT Service management - http://20000.fwtk.org/iso-20000.htm
ITIL	Information Technology Infrastructure Library - http://www.itil-officialsite.com
ITU	International Telecommunication Union
JRA1	Equivalent to Work Package 8 (WP8)
JRA2	Equivalent to Work Package 9 (WP9)
Kepler	Open source scientific workflow management system.
KNIME	Open source system for data mining.
LAPACK	Linear Algebra PACKage
LDAP	Lightweight Directory Access Protocol
LFC	the WLCG File Catalog; part of the gLite middleware see gLite above
LMU	Ludwig-Maximilians-Universitaet Muenchen
LRZ	Leibniz-Rechenzentrum
MAPPER	Multiscale Applications on European e-Infrastructure - http://www.mapper-project.eu
Meandre	Semantic-driven data-intensive workflow execution environment.
metadata	Data that describes data. Metadata may include references to schemas, provenance, and information quality. In Seismology, metadata may also refer to data required in order to sanitise a seismograph's response.
miniSEED	The miniSEED format is a subformat of the commonly used SEED data format used for archiving seismological data.
MKL	Intel Math Kernel Library
MoU	Memorandum of Understanding
MPI	Message Passing Interface
myExperiment	Collaborative virtual research environment for sharing scientific workflows.
MyProxy	Open source software for managing X.509 Public Key Infrastructure (PKI) security credentials. Provide authentication and authorization mechanisms for controlling access to credentials. - http://grid.ncsa.illinois.edu/myproxy
NA	Network activities
NA2	Equivalent to Work Package 2 (WP2)
NCSA	National Center for Supercomputing Applications - http://www.ncsa.illinois.edu/
NERA	Network of European RI for Earthquake Risk Assessmet and Mitigation. EC I3 project, www.nera-eu.org
NERIES	Network of RI for European seismology. EC I3 project ended 2010 www.neries-eu.org
NGI	National Grid Initiatives - http://www.egi.eu/about/ngis
ObsPy	A Python framework for processing seismological data. http://obspy.org/

OGSA	Open Grid Services Architecture supported by Globus. - http://www.globus.org/ogsa
OGSA-DAI	Open Grid Service Architecture Data Access and Integration, an open source product for distributed data access and management.
ontology	In computer science, a formal explicit specification of a shared conceptualisation.
OpenMP	Open Multi-Processing
ORFEUS	Observatories and Research Facilities for European Seismology.www.orfeus-eu.org
PBS	Portable Batch System
PDCA	The Plan-Do-Check-Act cycle - http://labspace.open.ac.uk/mod/resource/view.php?id=346003
Pegasus	Workflow management service, mapping and executing workflows on available compute resources.
PID	Persistent Identifier : A persistent identifier is a permanent, location- independent and globally unique identifier for a resource. Persistent identifiers are generally assigned by agencies who undertake to provide reliable, long-term access to resources. Examples of persistent identifiers include Digital Object Identifiers, Uniform Resource Names, Handles and Archival Resource Keys.
Pilot application	main software routine within a use case (e.g., the cross-correlation analysis in the use case addressing the velocity variations of the Italian peninsula crust properties).
portal	In the context of knowledge discovery, a tool designed for a particular group of domain experts that can be used via their browsers; it enables them to establish their identity and rights, and to pursue conveniently a set of research tasks for which the portal is designed.
PRACE	Partnership for Advanced Computing in Europe - http://www.prace-project.eu/
pre-processing	One or operations performed on the observed data to prepare the latter for the analysis an/or for performing quality control checks.
processing element – PE	A software component that encapsulates a particular functionality and can be used to construct a workflow.
QUEST	QUantitative Estimation of Earth's Seismic Sources and Structure
RAPID	Rapid portals for Seismological Waveform Data - http://research.nesc.ac.uk/node/423
RapidSeis	Portal for interactively running C++ scripts on seismological waveform data Not yet ready for Python.
rdseed	http://www.iris.edu/software/downloads
Real (or observed) Seismograms	Data recorded at one or more seismic stations and made available by data-centers (size of data depends on duration and sampling rate, also meta-data)

Redmine	Project management web application - http://www.redmine.org
registry	A persistent store of definitions and descriptions of data or software components and their relationships accessed by tools and other elements of a distributed research environment. It is intended to facilitate discovery and use of the components.
RegSEM	A Spectral Element Method code to compute seismic wave propagation - http://www.ipgp.fr/ paulcup/RegSEM.html
repository	A store holding software definitions, other shared code and data, that supports distributed concurrent access, update and version management.
Research Infrastructure	The collection of equipment, resources, organisations, policies and community support that enables a particular discipline to conduct research. Normally, this refers to the advanced facilities that enable frontier research, such as the research infrastructures endorsed by ESFRI.
research object	A research item which some researcher wishes to identify. It may be a collection of primary or derived data, code, a workflow, a service, an ontology, a set of metadata, etc. It may be a paper or a talk. Often it is a composition of such elements.
SA1	Equivalent to Work Package 5 (WP5)
SA2	Equivalent to Work Package 6 (WP6)
SA3	Equivalent to Work Package 7 (WP7)
SAC	http://www.iris.edu/software/sac
SAGA	A Simple API for Grid Applications - http://www.saga-project.org/
SAML	Security Assertion Markup Language (SAML) is an XML-based open standard for exchanging authentication and authorization data between security domains - http://saml.xml.org/about-saml
SCALASCA	A software tool that supports the performance optimization of parallel programs by measuring and analyzing their runtime behaviour
science gateway	A consistently presented set of facilities designed to be a convenient working environment for researchers in a particular domain, in this case seismology. It should bring together access to all of the capabilities and resources such a researcher needs: including catalogues of available data and tools, established methods and arrangements for applying them with specified parameters to specified data.
SDX	Seismic Data eXplorer
SEED, mSEED, SAC	Standard seismic data formats
SEC3D	Programme package for simulation of elastic wave propagation in 3D Cartesian earth models

SeisSol	A simulation software based on the Discontinuous Galerkin Finite Element Method - http://www.geophysik.uni-muenchen.de/ kaeser/SeisSol/
SEM	Spectral Element Method wave propagation
SES3D	Programme package for simulation of elastic wave propagation in a spherical section and the computation of Frechet kernels - http://www.geophysik.uni-muenchen.de/Members/fichtner/ses3d
Shibboleth	Standards based, open source software package for web single sign-on across or within organizational boundaries - http://www.shibboleth.net
SHIWA	Sharing Interoperable Workflows for large-scale simulations on Available DCIs - http://www.shiwa-workflow.eu/
SCI-BUS	SCIentific gateway Based User Support (http://www.sci-bus.eu)
SciPy	Scientific Tools for Python
SL5	Scientific Linux 5
SL6	Scientific Linux 6
SLES	SUSE Linux Enterprise Server
SLURM	A high scalable resource manager
SPECFEM3D	A software to simulate seismic wave propagation in sedimentary basins or any other regional geological model
SPECFEM3D	A simulation software code based on the spectral-element method for 3D seismic wave propagation - http://www.seg.ethz.ch/software/specfem3D
STF	Source Time Function. Shape of the original "signal-wave" of the earthquake, calculated by adjoint source inversion (or other methods). The "signature" of the STF is present in the recorded seismograms.
SuperMIG/SuperMUC	The name of a new supercomputer of the LRZ
Synthetic Seismograms	Waveform(time series) calculated in a computer simulation (size of data depends on duration and sampling rate, also meta-data). It is dependent on the solver, the computational grid(mesh), the earth model, the event parameters, and the location of "observation".
Taverna	Open source scientific workflow management system.
The PDCA cycle	The Plan-Do-Check-Act cycle http://labspace.open.ac.uk/mod/resource/view.php?id=346003
Trident	Microsoft workflow management system.
UEDIN	The University of Edinburgh
UMD	Unified Middleware Distribution - http://www.eu-emi.eu/
UNICORE	Uniform Interface to Computing Resources - http://www.unicore.eu/

Use case	In software and systems engineering, a use case is a list of steps, typically defining interactions between a role and a system, to achieve a goal. The actor can be a human or an external system (cf http://en.wikipedia.org/wiki/Use_case). In VERCE it is assumed to represent the entire scientific application (e.g., analysis of the noise cross-correlation of the Italian seismic networks for 6 years period to detect temporal variations of the Crust material properties)
VERCE architecture	A high-level and coherent design for the VERCE e-Infrastructure; it evolves as the seismological goals and digital environment evolve and become better understood. It should guide the development of successive VERCE platforms.
VERCE e-Infrastructure	An envisaged result of VERCE, as an integrated computational and data environment that presents a coherent virtual research environment in which to conduct seismology research and eventually research in other Earth sciences.
VERCE Platform	The current realisation of the VERCE e-Infrastructure at any time in the VERCE project. Initially this is not fully integrated and may only constitute a partial implementation. Nevertheless, it is sufficient both to pursue research identified as priority seismology use cases and to develop and test the design of the VERCE e-Infrastructure. The VERCE platform is an approximation to the VERCE e-Infrastructure. These approximations should converge on the VERCE e-Infrastructure by the end of the VERCE project.
virtual research environment (VRE)	A presentation of (ideally all of) the resources a researcher may need in a consistent and easily used form. These resources include catalogues, data, metadata, libraries, tools, workflows, programs, services, visualisation systems and research methods.
VOMS	Virtual Organization Membership Service - http://www.globus.org/grid_software/security/voms.php
W3C	World Wide Web Consortium, an international community of member organisations and the public that works to define and promote standards for web technologies.
web service	A software system designed to support interoperable machine- or application-oriented interaction over a network.
WLCG	the Worldwide Large Hadron Collider (CERN particule accelerator) Computing Grid - http://lcg.web.cern.ch/lcg

workbench	In this context a work environment for a computationally adept worker, such as a data-analysis expert, a data-intensive engineer or an application developer. It may be an IDE, an advanced editor or a command-line interpreter. It should provide all of the operations those workers need for creating, building, analysing, testing, debugging and making available the seismology and e-Infrastructure components. Many of the tools in a workbench will be familiar and widely used, a few will be specific to VERCE.
workflow	A process of composed data-handling tasks, computational tasks and human interactions intended to implement a research method or established working practice.
WP	Work Package
WP1	NA1
WP2	NA2
WP3	NA3
WP4	NA4
WP5	SA1
WP6	SA2
WP7	SA3
WP8	JRA1
WP9	JRA2
WP leader	The institution that has the responsibility for a certain WP, not the single person. e.g. NA1 leader is CNRS
wrapper	A design pattern where a piece of code allows computational or data-handling components to work together that normally could not because of incompatible interfaces.
X.509	ITU-T (Telecommunication Standardization Sector) standard for a public key infrastructure (PKI) and Privilege Management Infrastructure (PMI) - http://www.itu.int/rec/T-REC-X.509/en
XML	Extensible Markup Language.
XSEDE	Extreme Science and Engineering Discovery Environment - https://www.xsede.org/
XtreemFS	open source distributed and replicated filesystem - http://www.xtreemfs.org
ZigZag	Language used by Meandre for describing the directed graphs that define workflows.