D-SA1.5 – VERCE platform: final operation and monitoring report, evaluation of the applications deployment

20/11/2015

Project acronym:	VERCE
Project n°:	283543
Funding Scheme:	Combination of CP & CSA
Call Identifier:	FP7-INFRASTRUCTURES-2011-2
WP:	WP5/SA1, Management and operation of the research platform
Filename:	D-SA1.5.pdf
Authors[1]*:*	S. Hachinger, G. Moguilny, M. Schnell, H. Schwichtenberg.
Location:	http://www.verce.eu/Repository/Deliverables/RP4
Type of document:	Deliverable
Dissemination level:	Public
Status:	Final
Due date of delivery:	30/09/2015 - revised on 20/11/2015
Reviewers:	S.H. Leong, A. Frank (BADW-LRZ), M. Atkinson (UEDIN), M. Ewering (SCAI)
Keywords:	iRODS, INCA, UNICORE, GRAM, GridFTP, gUSE, X.509, GSI, VOMS, DCI bridge, KVM, FDSN, Cloud, Grid, HPC, FUSE

History	*Author*		*Date*	*Comments (and sections)*
1	G. Moguilny	(IPGP)	01/09	Initial skeleton
2	S. Hachinger	(LRZ)	25/09	INCA and KPI generation report (section 3)
3	M. Schnell, H. Schwichtenberg	(SCAI)	28/09	New resources (section 1), iRODS (section 2), Training and user support (section 4)
4	H. Schwichtenberg	(SCAI)	29/09	Summary and conclusions
5	A. Gemünd	(SCAI)	17/11	VERCE Obspy Cloud Service (in conclusion)

[1] Alphabetical order

Contents

List of Figures

List of Tables

Executive Summary

SA1 is responsible for leading the operations and deployment activities of the VERCE platform. The activities of SA1 are aimed at providing a smooth operation of the infrastructure and services, to facilitate the usage and to provide support in case of issues.

The main objectives of the work package are to:

- Operate and manage a Virtual Organisation in coordination with SA3.
- Provide tools and services for unified access to the VERCE research platform combining Data, Grid and HPC resources.
- Provide and manage the distributed research platform.
- Provide and manage a set of application-tailored workbenches and enactment gateways, in coordination with SA2, to support specific use-case scenarios.
- Provide and manage a development testbed on which SA2 and JRA2 develop, integrate and evaluate the next releases of the VERCE platform.

In this document we provide an update of report D-SA1.4.1. In the last reporting period SA1 installed no new services and instead concentrated on:

- daily operation and maintenance of the VERCE infrastructure,
- stabilisation of existing services especially iRODS,
- technical support to the Misfit task force (see SA3),
- preparation of new sites.

Due to the usage of the platform with an increasing number of users, the daily work for the user support was increased. The organisation of the training (see NA3) and technical support to dedicated users are highlighted here. This will be described in the sections about the integration of new resources, the Science Gateway, the iRODS Platform and Monitoring.

SA1 also provided intensive technical support for VERCE components for VERCE developers and users, and worked in the task force for the implementation of the Misfit workflow.

A summary of the activities and developments of SA1 will be given at the end of the deliverable (pages 9 and 11).

1 Integration of new resources

As mentioned in the latest deliverables, all VERCE partners provide compute resources to the VERCE community that are usable under local sites user policies. To access these resources via the VERCE Science Gateway, additional middleware and mechanisms for handling X.509 EUGridPMA certificates have to be installed. Here, the VERCE project is dependent on IT sites and their institutional rules, especially in the case of security. Therefore, the integration is cumbersome, but SA1 provided direct technical support to the local IT groups and connected most VERCE compute partner sites to the Science Gateway.

During this reporting period, a new resource were made available for VERCE at LRZ. In order to ensure enough resources for the large training event at Liverpool as well as covering the contingency of unavailable sites, LRZ provided a second resource as shown in appendix A. This resource supports the Globus middleware, which is commonly used in VERCE and therefore the integration for SA1 was straightforward.

1.1 Operation, Monitoring and Ticketing Service

As described in D-SA1.4.1, SA1 operates the Science Gateway and associated web services, iRODS and provenance services, support portal with documentation and ticketing system, monitoring service (INCA) and the GCMT FDSN service to query seismological catalogues. The services and the resource effort are also described in the previous deliverable. In the last period of the project, the VERCE infrastructure attracted more users and supported them via a professional ticket system. User support, 1st and 2nd level support was delivered by SA1. More details about the number of processed tickets and users can be found in D-SA2.5.

On the other hand, SA1 supports the compute resource provider sites. Here, several issues with local VERCE middleware services have been fixed. Issues have also been detected with the VERCE monitoring and resolved by the daily operator. Another main issue was power losses at the UEDIN and the SCAI sites where most of the VERCE services are hosted. All services came up automatically, but the iRODS virtual machines had been corrupted. SA1 had to deploy the backups. This has shown that in case of power losses, the VERCE infrastructure and damaged virtual machines can be quickly recovered. During the last period, developed and installed iRODS recovery procedures worked well and no VERCE data was lost. From this experience, SA1 improved the startup services to guarantee short interruptions in the event of data-centre failures.

In cooperation with SA2, the data for the key performance indicators (KPIs) were collected and analysed. Some indicators were monitored by INCA directly, as shown in section 3.

1.2 Science Gateway Support

SA1 Science Gateway activities are:

- maintenance and operation of the Science Gateway services and virtual machines,
- technical backend (DCI Brigde) and support of the gUSE middleware (see also SA3 and JRA1),
- Globus and UNICORE interface support at VERCE resource sites and for VERCE developers.

SA1 continued to provide support for the UNICORE integration of **HPC-CINECA-04** during the last reporting period. A deep inspection of the gUSE and UNICORE interface was needed. This showed that the interface in gUSE has to be adapted for use with the requided Grid and Cloud Gateway software. This results in a new branch - a VERCE branch - of gUSE. All specific developments are transmitted to the gUSE community.

Near the end of the project, an updated and extended version of gUSE with full UNICORE support was completed, which will be integrated in the Science Gateway in the follow-up project EPOS-CC, which is a community project inside of EGI-Engage.

Since SA1 is operating and maintaining the Gateway virtual machines, all adaptations with regard to the Misfit workflow originating from the task force have been implemented in this reporting period. SA1 partners are strongly involved in the implementation phase.

The integration of the EGI resources is delayed and will be finished in EPOS-CC.

2 Updated Data platform with iRODS

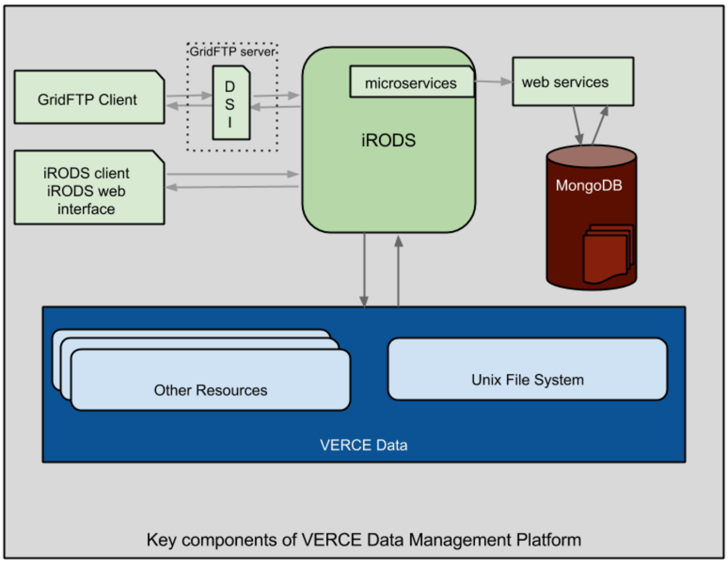

Figure 1 – VERCE data platform with iRODS.

The VERCE data platform builds the foundation for the (meta)-data management of distributed VERCE data. It is maintained and operated by SA1 as the underlying data infrastructure to support the workflow tools and provisioning of application data (see D-SA1.3 / D-SA1.3.1 and D-SA1.4 / D-SA1.4.1).

During the last period, iRODS VERCE infrastructure was mainly updated and renewed. The VERCE iRODS with VERCE adaptations (see deliverables) ran stable including the iRODS Cloud Services at the UEDIN site (see also subsection 1.1). The iRODS storage infrastructure was up 24×7 and was stressed in production. The operating iRODS I/O interface to the Gateway was improved to increase performance. The Misfit use case depends heavily on a large number of small-file transfers, so, first improvement of the iRODS/FUSE mount was with a cache directory. Intensive tests with different file sizes have shown that the solution produces insufficient improvement. Therefore it was decided that SA1 would implement the native iRODS commands within the Gateway interface.

3 INCA monitoring: updates and KPI generation (collab. with SA2)

The INCA monitoring platform, as described in D-SA1.3 and D-SA1.3.1, has received a major security update, hardening it against SSL vulnerabilities (including major ones such as "POODLE"). After a first failure to upgrade, original developers delivered a bug-fixed package which allowed for a smooth installation.

Reporter managers have been successfully installed on HPC-CINECA-04 (GALILEO) with support by

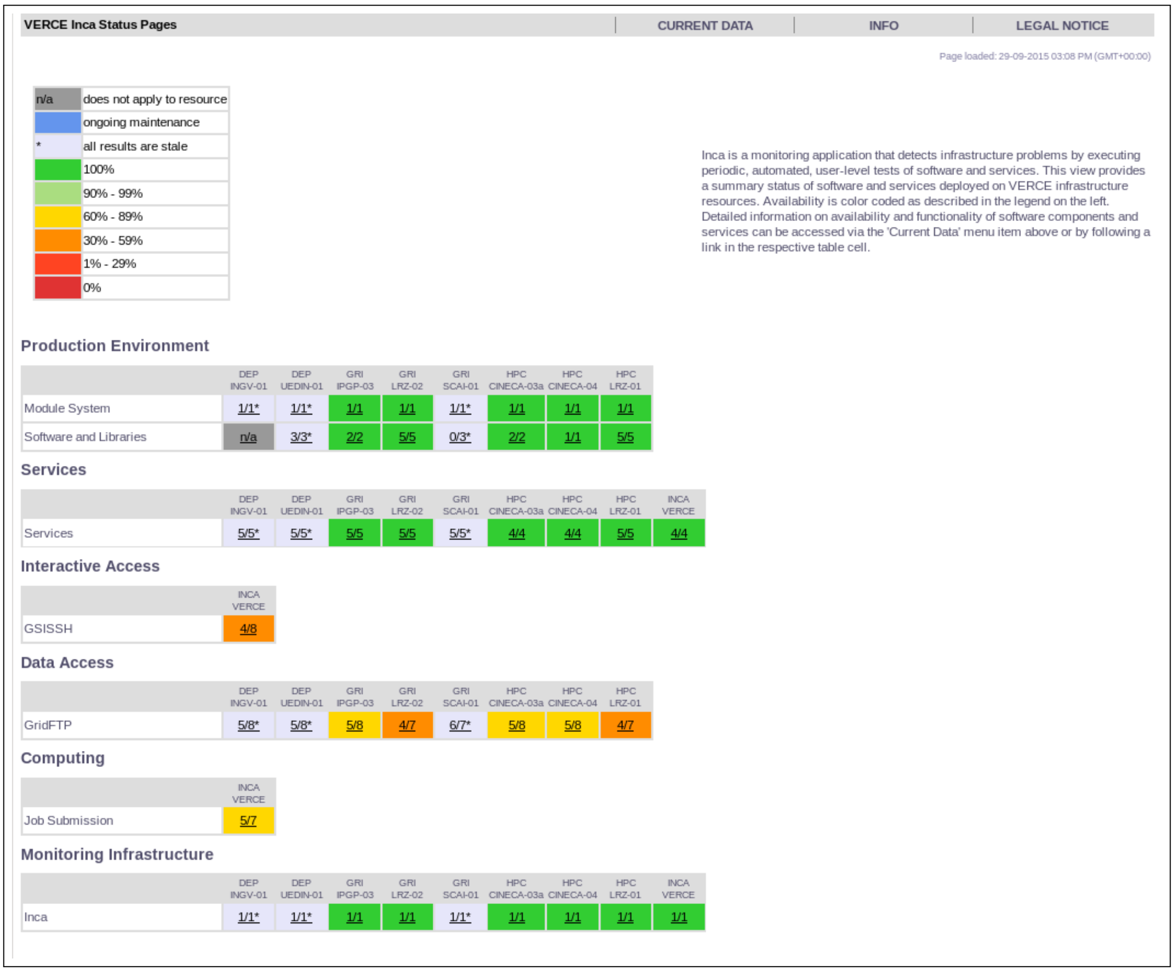

Figure 2 – VERCE INCA status page.

CINECA and maintained on all other sites. Activities as reported in D-SA1.4.1, such as notification of failed services to site administrators, have been continued.

Table 1 – Calculation of KPI metrics for SA2 (updated). The updated formulae are given in a continuous and an actually used (discrete) version, using the symbols: $n(N)OK$ – num. of tests (Not) OK; nTT – total num. of tests; tTI – test interval; $S(t)$ – status of service (0=NOK, 1=OK) at point of time t ; $\Delta t_i = t_{i+1} - t_i$ – test interval no. i.

Metric	unit	Formula D-SA3.1	Formula updated
Service Availability (SA)	%	nOK/nTT	$\int_{t_{\text{begin}}}^{t_{\text{end}}} S(t)\text{d}t / \int_{t_{\text{begin}}}^{t_{\text{end}}} \text{d}t \approx \sum_i S(t_i)\Delta t_i / \sum_i \Delta t_i$
Num. of service interruptions (nSI)	-	$nNOK$	number of status jumps $OK \to NOK$
Time of service interruptions (nSI)	h	$nNOK \times tTI$ (cumulative)	$\int_{t_{\text{begin}}}^{t_{\text{end}}} [1 - S(t)]\text{d}t / \text{nSI} \approx \sum_i S(t_i)\Delta t_i / \text{nSI}$ (average)

In August 2015, INCA data was successfully utilised to calculate key performance indicators (KPIs) in collaboration with SA2 (for the VERCE Portal and iRODS systems, as an exception, the KPIs were evaluated from machine log files as these reach further into the past. Plots of the KPI metrics were prepared by SA1, and are shown/interpreted in D-SA2.5. We note that the evaluation formulae for the KPIs given in D-SA1.3 have been refined (Table 1), assuming that the test time interval (*tTI*) is in general not constant. Also, the number of service interruptions (nSI) we have plotted is now the *actual number of status changes to NOK (not O.K.)*, and we evaluate the time of service interruptions as an averaged number. All numbers are based on integrating the OK/NOK time forward, i.e. the status found in a test is assumed to hold from that test until the next test; changing this assumption is not expected to introduce major changes in the results as the tests are executed frequently (such that failure or success status normally pertains over a larger number of tests).

All in all, we conclude that the INCA system has been of great help not only for status feedback for the technical maintenance teams, but also as a tool to monitor the VERCE platform's performance.

4 Training preparation and end-user support

During this reporting period, two training sessions were held (see http://www.verce.eu/Training/Training2015.php) The arrangement of training sessions is a standard procedure for the SA1 team.

Nevertheless, these training sessions were a challenge because of the number of trainees and the usage of their own laptops. With the support of SA1, most of the trainees installed the security environment to use X509 certificates on their laptops, or used local machines at Liverpool or the LRZ. The distributed VERCE infrastructure hosted at different VERCE sites was stressed concurrently by several users and worked.

Once again, we note that the VERCE infrastructure (or so called platform) is stable and provides the necessary scalability in production.

Conclusion for the Reporting Period

Over the last reporting period, the usage of the VERCE infrastructure has been increased. VERCE infrastructure has been extensively used by the seismic research community and trainees. Especially training sessions with a large number of concurrent users helped to improve and verify the infrastructure. To achieve this require:

- stabilising and performance tuning of the iRODS storage service,
- provision of a professional user support with a ticket system and help desk,
- improved INCA monitoring to produce KPIs,
- elaborated and standardised training support,
- all services including the GCMT FDSN are operational nearly 24×7.

SA1 was also heavily involved in the implementation of the Misfit use-case. As an outcome of this, SA1 people increased their gUSE skills and are now able to support the complete stack from the Gateway down to the VERCE compute resources with UNICORE and Globus.

All achievements in the development around gUSE and UNICORE have been transmitted to the gUSE and UNICORE communities.

As we described in the previous deliverable (D-SA1.4.1), all services are virtualised and may be shifted to other sites. The Obspy Cloud-Docker Service developed by SA1 (D-SA1.3) and shown in figure 3 is available at verce-fsp.scai.fraunhofer.de. It can easily be integrated in the Science Gateway. All services are supported beyond the lifetime of the VERCE project by VERCE partners involved in the EGI-Engage / EPOS-CC and by EPOS itself.

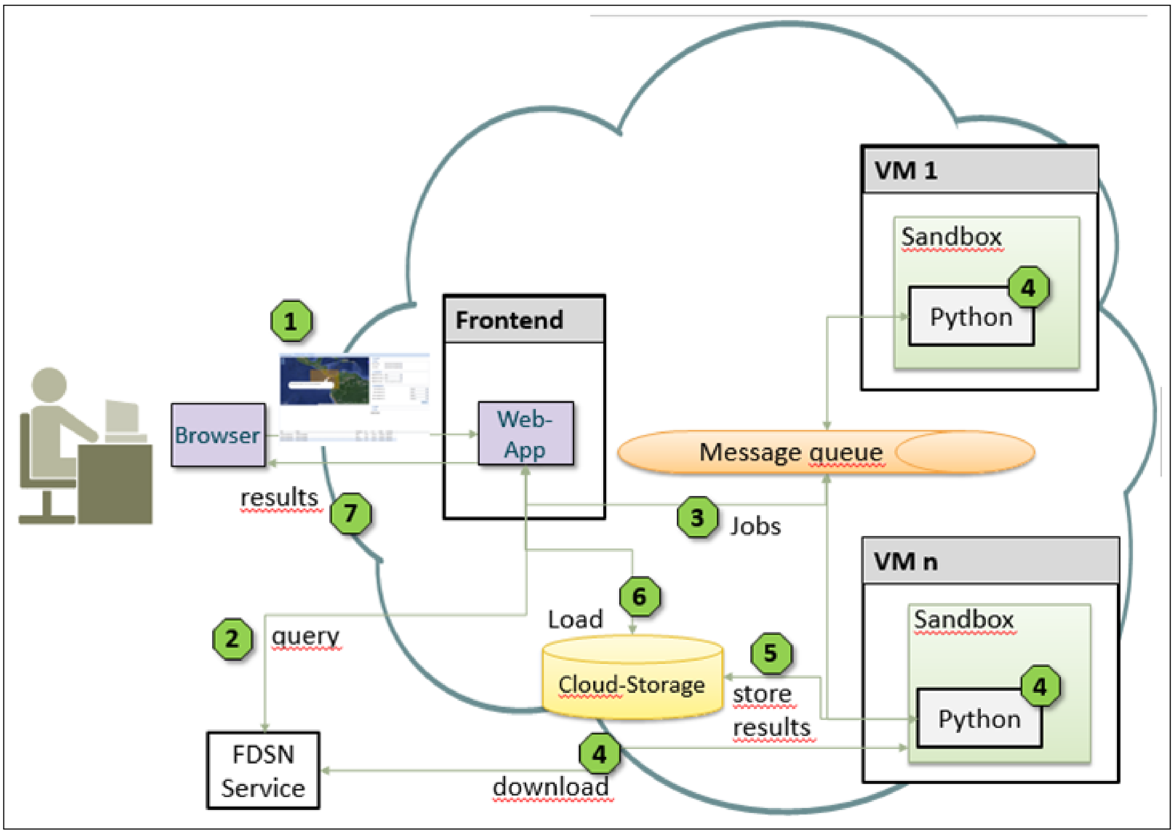

Figure 3 – VERCE Obspy Cloud Service.

Conclusion for the overall project

The objectives of SA1 were to arrange collaborative and federated provision of resources and software systems to support the development work of the other VERCE Work Packages and the seismic research community with production quality facilities. SA1 fully achieved these goals, as set out in more detail in Table 2 below.

Table 2 – SA1 Objectives as described in the DOW.

SA1 Objectives	Achieved	Comment
Operate and manage in coordination with SA3 a Virtual Organisation	A VERCE Virtual Organisation with a significant number of users supported by SA1 and SA3 technical components and services has been established. SA1 operates the complete VERCE infrastructure/platform including all services to support users in using the Gateway of SA3. VERCE users of the portal from Universities and Research, e.g. EPOS, are building the VERCE Virtual Organisation (VO). This user base is supported by SA1 services and maintained by a directory service and a technical Virtual Organisation management tool (VOMS). The latter supports the authorisation. The Science Gateway components of SA3 interact with this service. The administration of different roles of users in the VERCE VO is managed by the tools in the VERCE portal.	Authorisation is one of the crucial issues for end users and is dependent on the resource providers, e.g. PRACE and EGI. This is not harmonised by the European infrastructure. Therefore the implementation of the authorisation components will change in the future to exploit emerging solutions such as eduGAIN.
Provide tools and services for a unified access to the VERCE research platform combining Data, Grid and HPC resources.	A top down support for unified access is realised and supported in VERCE. From a user perspective, SA1 has supported SA3 to integrate unified access for users to login and the data handling including provenance data in the VERCE portal. As backend service, SA1 maintains and supports middleware layers and tools as unified access from the platform to the VERCE resources. The middlewares between the e-infrastructure resource and the research platform are the VERCE Science Gateway using the gUSE layer (see JRA2, SA3, SA1 deliverables). iRODs data, which is building the VERCE storage infrastructure and the resource access layer Globus or UNICORE.	With the storage infrastructure, VERCE iRODS microservices are developed but it was also considered important that the VERCE iRODS infrastructure is still interoperable with EUDAT.

Continued on next page...

Provide and manage the distributed research platform.	A distributed research platform is in place, maintained and operated by VERCE. The platform combines distributed services from e-infrastructures and VERCE partner resources. Even the specific VERCE services are distributed over the different partner sites and can be shifted from one partner to another or into commercial or European Cloud infrastructures.	The EGI Federated Cloud was tested to host VERCE services in future and this will also be part of the future EGI-Engage / EPOS-CC project.
Provide and manage a set of application-tailored workbenches and enactment Gateways, in coordination with SA2, to support specific use-case scenarios.	"Misfit" and "forward simulation" workflows have been implemented by the VERCE work packages. Provenance data from various runs is available and can be provided to users. Variations of the forward simulation have been executed by trainees.	
Provide and manage a development testbed on which SA2 and JRA2 develop, integrate and evaluate the next releases of the VERCE platform.	The development testbed was provided by SA1 based on VERCE local and supercomputer resources from the beginning. Access layers, e.g. Globus, have been installed and operated with the local IT-groups. SA1 followed the PDCA cycle of SA2 by managing and installing the new VERCE components.	

Appendices

A Computing resources available

Table 3 – Hardware resources.

Designation	Resource Type	Short Name	Brief Description	Access Protocols	Accounting	Person in charge
HPC-CINECA-03	Computation	FERMI	IBM-BlueGene /Q Architecture: 10 BGQ Frame with 2 MidPlanes each Front-end Nodes OS: Red-Hat EL 6.2 Compute Node Kernel: lightweight Linux-like kernel Processor Type: IBM PowerA2, 1.6 GHz Computing Nodes: 10.240 with 16 cores each Computing Cores: 163.840 RAM: 16GB / node; 1GB/core Internal Network: Network interface with 11 links ->5D Torus Disk Space: more than 2PB of scratch space Peak Performance: 2.1 PFlop/s	ssh to login node	yes	Piero
HPC-CINECA-04	Computation	GALILEO	IBM NeXtScale Nodes: 512 Processors: 2x8cores Intel Haswell 2.40 Ghz per node Cores: 16 cores/node, 8256 cores in total Accelerators: 2 Intel Phi 7120p per node on 384 nodes (768 in total) RAM: 128 GB/node, 8 GB/core Internal Network: Infiniband with 4x QDR switches Disk Space: about 2PB of scratch space	ssh to login node	yes	Piero
HPC-LRZ-01	Computation	SuperMIG/ SuperMUC	PRACE Tier 0 machine. SuperMUC (Summer 2012). Intel Xeon-Architecture, more than 110,000 cores. 3PFlop/s, 16cores/node with. 2GB/core memory and 40cores/node with 6.4 GB/core memory. 10 TB GPFS (aggregate BW 200 GB/s). linked via fully non-blocking. Infiniband network 1GB Ethernet to archive and backup	ssh to login nodes behind firewall, only accessible from IPs entered in firewall; then port 22 and globus ports range opened.	yes INCA, Dmon	Cerlane
DEP-INGV-01	Computation	selene	48 dual-proc AMD 6136 2.4GHz 8-core 64GB RAM (total 768 cores) 64 dual proc AMD 2378 2.4GHz 4-core 16GB RAM (total 512 cores) 36 TB scratch-grade storage	Account on request, then access via GSIssh (ports 22 and 2222). GRAM via Globus gatekeeper (port 2119).	no	Daniele
DEP-IPGP-06	Computation	S-CAPAD Dell cluster	96 CPU nodes,16 data intensive nodes. 4 GPGPU nodes. 1 SMP node, parallel storage : GPFS on 576 effective TB, a high-performance network with InfiniBand QDR switches.		yes	Geneviève
DEP-UEDIN-01	Computation	EDIM1	Data intensive cluster. 120 nodes, each with 6.24TB of storage. CPU atom dual core, 4GB ram	SSH to login node behind firewall only from IPs within Edin. Informatics or EPCC	Yes, Ganglia	
DEP-SCAI-01	Computation		Intel Xeon 2650 based cluster (4096 cores) with Mellanox FDR Infiniband network and the Fraunhofer parallel file system			André
DEP-ULIV-01	Data	Linux Storage	CPU Intel Xeon dual core 2.8 GHz, 2GB RAM, 2.8TB storage		no	Andreas
GRI-IPGP-03	Computation	EGI Cluster	SiteBDII: sitebdii1.ipgp.fr - ldap://sitebdii.ipgp.fr:2170/mds-vo-name=INSU01-PARISo=grid CREAM-CE: ce1.ipgp.fr - ce1.ipgp.fr:8443/cream-pbs-es DPM: dpm1.ipgp.fr SRM endpoint: srm://dpm1.ipgp.fr:8446/dpm/ipgp.fr/home/verce.eu/ GridFTP endpoint: gsiftp://dpm1.ipgp.fr:2811/dpm/ipgp.fr/home/verce.eu/ 136 CPU RAM: 1GB/CPU Storage: 8TB	gLite job/file operations, ssh to login on User Interface	yes	David
GRI-LRZ-02	Computation	Linux Cluster	> 2800 cores		yes (but only for internal	Cerlane
GRI-SCAI-01	Computation	EGI Cluster	32x Sun Fire X2200 M2 (Dual Opteron 2218, 2x2 cores), 8GB RAM, 250GB HDD 1 x Sun Fire X4600 M2 (Dual Opteron 8220, 4x2 cores), 78GB RAM, 2x140GB SAS HDD 8.5 TB NFS4 on ZFS shared filesystem (homes), 4.5 TB gLite DPM Storage Element (Grid) Connectivity: 2x Gbit Ethernet, Mellanox MT25204 10Gbps InfiniBand.	gLite, Unicore, Globus job/data access. GSISSH login to UI and dev-node (1, 2)	yes	André

Table 4 – Data and Software resources.

Designation	Resource Type	Short Name	Brief Description of Resource	Access Protocols	Installed Software	Accounting	Person in charge
GRI-EGI-01	Computation	EGI Virt. Org.	EGI Earth Science Research VO	gLite operations	gLite software stack	yes (EGI accounting)	André, David
GRI-EGI-02	Computation	EGI Virt. Org.	VERCE.EU VO	gLite operations	gLite software stack	yes (EGI accounting)	André
DAT-KNMI-01	Data	ORFEUS Data Center	Virtual European Broadband Seismograph Network,European Integrated Data Archive	Web services, interactive sessions	ArcLink		
DAT-IPGP-04	Data	COHER-D	IPGP Data Archive, NIED (Japan)	direct access from IPGP network		no	Geneviève

B Software components

Table 5 – Components (recommended by SA2) / resource

	TestBed										
	HPC-LRZ-01	HPC-CINECA-03	HPC-CINECA-04	DEP-UEDIN-01	DEP-ULIV-01	DEP-IPGP-06	DEP-INGV-01	DEP-SCAI-01	GRI-SCAI-01	GRI-IPGP-03	GRI-LRZ-02
Codes											
SeisSol	✓	✓	◒						●		✓
SpecFEM3D	✓	✓	●			◒	●	✓			✓
Scientific libraries											
Matplotlib	✓	●		●		●	✓	✓	✓	✓	✓
ObsPy	✓		●	●		✓	●	✓	✓	✓	✓
Metis	✓	●				●		✓	✓	✓	✓
GMT	✓				●	✓	✓	✓	✓		✓
mpi4py	✓	✓	●	✓		◒	●	✓	✓		✓
NetworkX	✓	✓		✓		✓		✓	✓	✓	✓
h5Py	✓	●	◒			✓	●	✓	✓		✓
dispel4py			◒					✓	✓		
Paraview	✓		●			✓	●	✓			✓
FFmpeg	✓	●	●					✓			✓
Environment / Tools											
iRODS	✓	✓		✓			✓	✓	✓	✓	✓
iRODS-DSI		✓		✓			✓	✓	●	✓	
GSI-SSH	✓	✓					●	✓	✓	●	✓
GridFTP	✓			●			✓	✓	✓	✓	✓
GRAM	✓						✓	✓	✓		✓
JSAGA									✓	✓	
gLite CREAM									✓	✓	
gLite WMS									✓		

◒ Currently being installed ● Installed
✓ Available for VERCE users
An empty cell means not applicable or undecided as of yet