Integrated Infrastructure Initiative projects (I3)
Call topics: INFRA-2011-1.2.1

<div style="border:1px solid">

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

VERCE

</div>

Proposed full title:	Virtual Earthquake and seismology Research Community e-Environment in Europe
Proposed acronym:	VERCE
Funding scheme:	Combination of Collaborative Project and Coordination and Support Action: Integrated Infrastructure Initiative (I3)
Work programme:	e-Science Environment
Coordination:	Jean-Pierre Vilotte (vilotte@ipgp.fr) Phone number: +33 (0)6 61 82 71 34

List of participants

Participant no. *	Participant organisation name	Part. short name	Country
1 (Coordinator)	Centre National de la Recherche Scientifique	CNRS-INSU	France
2	University of Edinburgh	UEDIN	United Kingdom
3	Royal Netherlands Meteorological Institute	KNMI-ORFEUS	Netherlands
4	European-Mediterranean Seismological Centre	EMSC	France
5	Istituto Nazionale di Geofisca e Vulcanologia	INGV	Italy
6	Ludwig Maximilians Universität	LMU	Germany
7	University of Liverpool	ULIV	United Kingdom
8	Bayerische Akademie der Wissenchaften	BADW-LRZ	Germany
9	Fraunhofer Gesellschaft e.V.	SCAI	Germany
10	Centro di Calcolo Interuniversitario	CINECA	Italy

Table of Contents

Section 1: Scientific and/or technical quality relevant to the topics addresses by the call

Project summary

The earthquake and seismology research, an intrinsically Global undertaking, addresses both fundamental problems in understanding Earth's internal wave sources and structures, and augment applications to societal concerns about natural hazards, energy resources, environmental change, and national security. This community is central in the European Plate Observing System (EPOS), the ESFRI initiative in solid Earth Sciences. .

Global and regional seismology monitoring systems are continuously operated and transmitting a growing wealth of data from around the world. The multi-use nature of these data puts a great premium on open-access data infrastructures integrated globally. Most of the effort is in Europe, USA and Japan.

The European Integrated Data Archives infrastructure provides strong horizontal data services. Enabling advanced analysis of these data by utilising a data-aware distributed computing environment is instrumental to exploit fully the cornucopia of data, and to guarantee optimal operation and design of the high-cost monitoring facilities.

The strategy of VERCE, driven by the needs of data-intensive applications in data mining and modelling, aims to provide a comprehensive architecture and framework adapted to the scale and the diversity of these applications, and integrating the community Data infrastructure with Grid and HPC infrastructures.

A first novel aspect of VERCE consists of integrating a service-oriented architecture with an efficient communication layer between the Data and the Grid infrastructures, and HPC. A second novel aspect is the coupling between HTC data analysis and HPC data modelling applications through workflow and data sharing mechanisms.

VERCE will strengthen the European earthquake and seismology research competitiveness, and enhance the data exploitation and the modelling capabilities of this community. In turn, it will contribute to the European and National e-infrastructures.

1.1 - Concepts and objectives

The earthquake and seismology research community is inherently international, and address both fundamental problems in understanding Earth's internal wave sources and structures, and augment applications to societal concerns about natural hazards, energy resources, environmental change, and national security. A rich panoply of societal applications has emerged from basic research. The seismology community plays today a central role in hydrocarbon and resource exploration, containment of underground wastes, carbon sequestration, earthquake detection and quantification, volcanic-eruption and tsunami-warning systems, nuclear test monitoring and treaty verification, earthquake hazard assessment and strong ground motion prediction for the built infrastructure, including lifelines and critical facilities. Emerging new applications involve glacier systems, landslide mass movements, ocean wave environment, and other topics relevant to climate and environmental change.

The centrality of the earthquake and seismology community in the solid Earth Sciences engages multiple European and International agencies in supporting the discipline through a number of large scale projects in Europe – e.g. NERA, SHARE, GEM, WHISPER, QUEST

– and outside Europe – e.g. Earthscope, USArray and GEON in the US; the Earth Simulator, the Hi-net and K-net monitoring systems in Japan, and international consortia – e.g. the Comprehensive (Nuclear) Test Ban Treaty Organisation (CTBTO), and the Global Earth Observations System of Systems (GEOSS). **The community is today the central core of the European Plate Observing System (EPOS), a large-scale ESFRI research infrastructure in solid Earth Sciences that entered its preparatory phase in 2010.**

Global and regional seismic networks, are continuously transmitting a rapidly growing wealth of data from around the world. **These tremendous volumes of seismograms,** i.e., records of ground motions as a function of time arising from both natural and human-made energy sources distributed around the world, **have a definite multi-use attribute.** Seismic data recorded for any particular purpose – e.g., monitoring nuclear testing or earthquake hazard analysis – intrinsically provide signals that are valuable for multiple unrelated uses. This places a great premium on data resources together with a growing commitment to the effective exploitation of these data.

The earthquake and seismology community has for decades pioneered the prevailing philosophies of global, open-data access and sharing. Creation of internationally integrated – within the Federation of Digital Seismic Networks (FDSN) – massive on-line and open-access distributed data resources, housing hundreds of Terabytes, and the adoption of standards for data services, has enabled proliferating discoveries and new societal applications by the community at a dramatic pace *driven by new data*.

To exploit the full potential of this rapidly growing European and Global data-rich environment, and to guarantee optimal operation and design of the high-cost monitoring facilities, the earthquake and seismology data-driven research has entered a fundamental paradigm shift. Data-intensive research is rapidly spreading in the community. Data analysis and data modelling methods and tools, required for revealing the Earth's interior physics, cover both a wide range of time scales and spatial orderings. **Large volumes of time-continuous seismograms contain a wealth of hidden information about the Earth's interior properties and wave sources, and their variation through time.** Mining, analyzing and modelling, this cornucopia of digital data will reveal new insights at all depths in the planetary interior and at higher resolution than is possible by any other approach.

To accelerate data-intensive research, the earthquake and seismology community is facing new challenges in terms of data management – e.g. efficient strategies for data query, access and movements from data sources – and in terms of methods and tools – e.g. data integration, physics-based distributed data analysis and cpu-rich data modelling – e.g. imaging/inversion and simulations – as well as in terms of data-oriented computational resources.

Data-aware distributed e-infrastructures resources

The last decade has seen the emergence of a living ecosystem of European e-infrastructures, middleware and core service, built on top of the European network infrastructure provided by GÉANT and the NRENs. The European e-infrastructures are providing already extended computing and data storage capabilities, and services.

- The European Grid Infrastructure (EGI and the associated NGIs) provides distributed computing and storage resources. The Unified Middleware Distribution (UMD) – including gLite, UNICORE, Globus and ARC – enables the shared use of various ITC resources across multiple administrative domains and organisational models;
- The European High Performance Computing (HPC) infrastructure (PRACE and DEISA2), is an association of European entities providing a pyramid of National HPC resources and services, and an emerging persistent pan-European level of front-end resources and HPC services. The HPC services are based on the Globus Toolkit and UNICORE middleware.

The earthquake and seismology community, and the VERCE partners of the consortium, has pioneered the use of the Grid infrastructure – as it was evolving from the European Data Grid (EDH) to the European and National Grid infrastructures (EGI and NGIs) – and is actively participating with the EGI and EGI-Inspire projects within the Earth Science Virtual Community of Research (VCR). At the same time, the earthquake and seismology research community is increasingly using leading-edge HPC capabilities. The partners of the VERCE consortium have been actively involved in DEISE-DECI projects, employing substantial computational resources at distributed European supercomputer centres. They are also involved in the European Exascale Computing Initiative (EESI) through EPOS.

Even if the boundaries between HPC and Grid are increasingly ill-defined, for historical and practical reasons these e-infrastructures do not yet provide the seamless interactive services and resources that would best enable advanced seismology research.

Distributed data infrastructure in seismology

In the last decades, through a number of coordinated European and Global initiatives, an internationally integrated seismology infrastructure of distributed data resources and data services has been established under the umbrella of the international Federation of Digital Seismic Networks (FDSN) with in particular

- In Europe, the European Integrated Data Archives (EIDA) and ORFEUS acting as the European consortium of the FDSN
- In US, the Data Management Centre of the Incorporated Research Institutions for Seismology (IRIS-DMC), and IRIS acting as the US consortium of the FDSN
- In Japan, within the National Institute of Earthquakes Disaster (NIED) and the Japan Agency for Marine-Earth Science and Technology (JAMStEC), JAMStEC being the Japanese consortium member of the FDSN.

For decades, within the FDSN, the community has pioneered an open access policy to the data and a distributed data architecture model. **The community is organized with international standards for distributed Data Management Systems (DMS) – data and interchange formats – e.g. data, meta-data, ontologies – as well as internationally standardised distributed data query and access protocols – e.g., for example ArcLink and NetDC – together with enabling service-oriented architecture comprising a set of Web 2.0 services for data handling and data processing, integrated into community portals.**

In Europe, the underlying network infrastructure layer and services – as provided by GÉANT and the NRENs – has been instrumental for supporting the seismology international model of a loosely connected system of distributed data archives across Europe. **As the amount and the complexity of available data are becoming even more overwhelming, extension of the European seismology data infrastructure capabilities – in terms of architectural model and data services – will be instrumental in advancing the European earthquake and seismology research to a very competitive status.** This evolution could be fostered rapidly by the forthcoming EUDAT initiative, in which the earthquake and seismology community is participating through EPOS.

Datascopes methods

European earthquake and seismology community is playing a leading role in the development of new physics-based methods of data exploration, data visualisation, data integration, data analysis, and data modelling methods – designed here as 'datascope' - of increasing scale and complexity. This is supported by a number of ongoing European projects – e.g., WHISPER, QUEST, NERA. A wide range of these datascopes have been

developed as operational software and made accessible to the community at large. The VERCE partners are world leaders in this field.

- **Data exploration and visualisation, data integration of large-scale distributed data sources; and distributed data analysis methods**, **are now crucial for exploring and extracting new information from the large distributed volumes of seismograms in the European and International data resources**. Recent breakthroughs in theory and methods, now allow every byte of continuous seismological data to be scanned to detect and extract earthquake events; and to be mined to extract the coherent information contained in every part of the signal, even the background seismic "noise" previously dismissed, using pair-wise or higher correlations. This opens entirely new approaches for the imaging of the wave sources and structures, the investigations of environmental changes, and the monitoring of volcanic and earthquake hazards. Data analysis and data integration applications are rapidly increasing in scale and complexity. Today, seismic noise correlation techniques are mining hundreds of terabytes of data, which need to be efficiently accessed and integrated from distributed data sources into distributed storage infrastructures. In turn, seismic noise correlation techniques are producing petabytes of new information that must be searchable for further use. This is a challenging issue for data-aware distributed computing and data management architecture.
- **Fully three-dimensional physics-based simulation of the Earth's internal wave sources and wave propagation (forward problem) are crucial for generating synthetic seismograms.** These simulations open new approaches for ground motion prediction with enormous practical benefits for seismic risk estimation through improvement in seismic hazard evaluation. **Full waveform inversion methods provide a new high-resolution multi-scales 3D and 4D imaging of the Earth's interior with enormous implications in terms of energy resources and environmental management.** Thousands of simulations of the forward and adjoint 3D wave propagation problems must be performed as an iterative optimisation. Forward and adjoint 3D states need to be stored at each of the iterations. While High Performance Computing capabilities make it possible for the first time to perform high-resolution multi-scale imaging of the Earth's interior, it raises challenging scalability, memory and I/O optimisation problems.
- **Data analysis applications at one level produce new data sets which can be used in turn by other data analysis or data modelling applications, repeatedly across the range of scales and types of information required.** In this process more than raw data must be communicated and the overall concept of "data fitness for use" must be considered. Information about the sensitivity of models on particular data may represent high-value added information in many applications. Thus, enabling a rich bi-directional exchange of both data and metadata between processes and scales is becoming a critical issue in enabling earthquake and seismology progress

<u>VERCE e-Science environment for data-intensive research:</u>

The existing European e-infrastructures provide a uniform network of resources, including data storage and a large body of expertise in the digital library community, with globally consistent generic services. **As the scale and complexity of the seismological data-intensive applications rapidly increase – both in terms of distributed data sources and in terms of computational methods – they require additional bespoke and specialised services and tools that are well integrated to provide a service-oriented environment conducive to new science.**

To reap the benefits, the VERCE project aims to significantly improve the exploitation of the data and of different data-aware computing capabilities – provided by the European e-infrastructures – by delivering a service-oriented e-Science environment through three categories of output: a comprehensive framework, an architecture and a set of productised data-intensive applications – with their demonstration use cases – that illustrate

how this e-Science environment can be used by the community at large to improve data-intensive applications.

The overall objective is to harness the core services provided by the European e-infrastructure with a collection of tools, services, data-flow and work-flow technologies, encapsulating relevant parts of the European e-Infrastructures and of the community data infrastructure. These will be delivered as a research-led overarching research Platform-as-a-Service (PaaS) to the seismological community and beyond to the earth sciences community within the EPOS Research Infrastructure.

The PaaS will

- Improve data-flows and workflows across the Grid and HPC components of the framework and support orchestration – an important issue for complex data-intensive applications using both distributed data analysis and HPC data modelling.
- Act both as an integrative framework (from the e-infrastructure providers side) and as an e-Science environment (from the seismological community side).
- Provide a resilient computing and data e-science environment – with well defined APIs and protocols – that expects and exploits the rapid advances in technology, data handling and methods.
- Perform as an incentive 'intellectual ramp', with a scientific gateway – providing facilitators to guide the researchers and interfaces for users – that let them engage incrementally with the tools, techniques and methods of data-intensive research so that they can meet their own needs when they choose.

The VERCE project is user-centric and led by the 'productisation' needs of a core of open source pilot data-intensive applications – and use cases – selected on the base of their scientific importance and their support from the earthquake and seismology community.

- Data exploration and data analysis applications will address critical issues regarding query and access, data movements and data integration strategies of distributed large volume data sets; distributed analysis of data and analysis of distributed data algorithmic mapping.
- Data modelling applications will address critical issues of scalability and memory – and I/O – complexity on High Performance Computing architectures.
- Data analysis and data modelling coupling will address critical issues of data interchange between Grid and HPC components requiring well-defined and efficient data exchange interfaces and orchestrating workflow technologies.

The e-Science environment relies heavily upon the architecture defined to match the diversity of user and developer requirements. The VERCE architecture will include:

- a separation, via a canonical form for the data-intensive process definition, between diverse and extensible domain of data-intensive tools and a range of data-intensive enactment platforms;
- a model for describing all of the components participating in data-intensive processes that supports the range of data-intensive tools, data-intensive enactment optimisation and adaptation to changes in data sources and services;
- data-intensive gateways that mediate requests in the canonical form and hide the transformation, delegation and heterogeneity of the distributed underlying data, computing resources and services;
- efficient direct data paths for delivering results and monitoring data-intensive process enactment.

A common ontology for data integration and data analysis applications will be developed based on the ongoing developments of the ADMIRE project.

VERCE Objectives

The objectives of VERCE can be summarized as

- Provide to the Virtual Earthquake and seismology Research Community in Europe, a data-intensive e-Science environment – based on a service-oriented platform – integrating a number of specialized services, tools, data-flow and work-flow engines, to support the data-intensive applications of this community and beyond to the EPOS community,
- Provide a service-oriented architecture and a framework wrapping the seismology data-infrastructure resources and services with a set of distributed data-aware Grid and HPC resources provided by the European e-infrastructures and the community.
- Productise a core of pilot data-intensive applications and use cases of the Virtual Earthquake and seismology Community of research in Europe that exemplify the power of the platform architecture and its capabilities,
- Deliver a scientific gateway providing a unified access, management and monitoring of the platform services and tools, domain specific interfaces supporting the co-evolution of research practices and their supporting software,
- Deliver an 'intelectual ramp' providing safe and supported means for researchers and users of the community at large to advance to more sophisticated data use through tailored interfaces and facilitators integrated within the scientific gateway.
- Deliver a 'research-methods ramp' through a toolkit of a training programs for data-intensive research – composed as a set of training session material, demonstrators, and best practice guides – based on tailored use-case scenarios and productised data-intensive applications of the community.
- Provide a collaborative environment between the earthquake and seismology research community and the computer scientists, system architects and data-aware engineers, fostering the emergence of 'research technologists' with sustained mastery for data-handling methods and a thorough understanding of the research goals.

1.2 - Progress beyond the state-of-the-art

Despite continuous progress in data-intensive methods, and their software implementation, the earthquake and seismology community is far from taking full advantage of the rapidly increasing cornucopia of data and of the continuously evolving resources and capabilities provided by the European e-infrastructure. In part, this results from insufficient links between the earthquake and seismology research community and the IT community experts.

As the wealth of data is rapidly growing, while at the same time the scale, and the complexity of the earthquake and seismology data-intensive applications are rapidly increasing, more computing and data capability integration is needed. This will be instrumental in advancing the European earthquake and seismology research to a very competitive status and in providing the capabilities for the effective exploitation of this cornucopia of data.

Today data-intensive applications – and their software implementation - have been developed and shared by a number of the community research groups in Europe. The computational and data requirements of these existing applications are diverse. Access to national and European HPC and Grid resources, in terms of support and expertise, differ based on national or regional priorities and existing resources. In contrast to the number of currently supported Pan European research projects and networks of the seismology community, there is no accompanying programme to provide an environment integrating computational and data resources, and supporting the 'productisation' of the data-intensive applications of the community.

The VERCE project intends to fill this gap and significantly improve this state-of-the-art by delivering an e-Science environment for the data-intensive applications of the earthquake and seismology research community and beyond to the EPOS Research infrastructures.

The VERCE project will deliver three categories of output: a framework, a service-oriented architecture and a productised set of data-intensive applications and use cases that demonstrate how this environment can be used by the community.
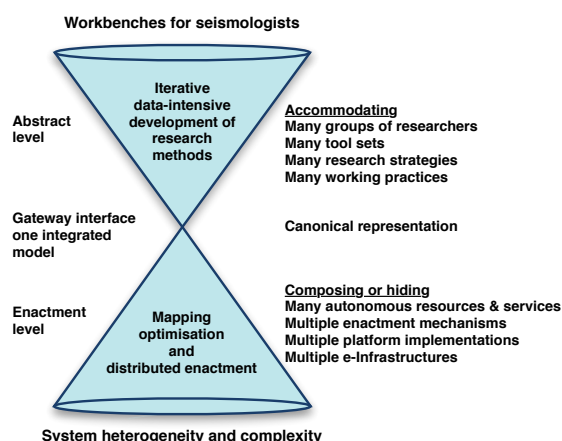
Workbenches for seismologists

Iterative data-intensive development of research methods

Abstract level

Accommodating
Many groups of researchers
Many tool sets
Many research strategies
Many working practices

Gateway interface one integrated model

Canonical representation

Enactment level

Mapping optimisation and distributed enactment

Composing or hiding
Many autonomous resources & services
Multiple enactment mechanisms
Multiple platform implementations
Multiple e-Infrastructures

System heterogeneity and complexity

Figure 1 Platform architecture concerns

The e-Science environment will be built on a stratification of concerns:

- An abstract top layer where users and developers of the data-intensive applications communicate with a collection of services, tools, data-flow and work-flow engines – integrated into a number of workbenches – belonging to an abstract layer that allows a separation via a canonical form between diverse and extensible domain tools and range of data-intensive enactment platforms;
- An enacting layer where a set of integrated distributed gateways and registries, mediate and delegate requests to an enacting platform providing instantiation, services and APIs – hiding the transformations and heterogeneity of the underlying distributed computational and data resources services;
- An underlying layer of distributed heterogeneous components and services, which include the computing and data resources and core services provided by the evolving European e-infrastructures and by the global seismological data infrastructure.

The VERCE framework will provide to the community a coherent model in which data exploration, data integration, data analysis and data modelling are integrated and well supported; a robust and efficient underlying distributed data and computing set of resources provided by the European e-infrastructures and the seismology data infrastructure.

The VERCE architecture will draw initially on the ADMIRE architecture provided by UEDIN and be adapted and introduce to the HPC and data-centre contexts of the VERCE community, drawing on MAPPER and IGE – through LMU and BADW-LRZ – and OMII-UK/Europe – through UEDIN – experience.

A crucial contribution of VERCE will be to bring to the earthquake and seismology community the range of experts needed in order to take an integrated view of all the stages of the project and it will ensure that research practices of the community co-evolve with new technologies, methods and their supporting data-intensive software.

1.2.1 - Datascopes and use cases

In order to focus the effort and strengthen the links to the earthquake and seismology community at large, a core of important community applications for the community has been identified. It covers a wide range of the earthquake and seismology community in data integration, data mining, and data modelling. The selected core applications are briefly

described below; more details on the applications and of their usage scenarios can be found in Annex I at the end of the document.

Data-analysis applications

Seismic waveform exploration (RapidSeis), which consists of a combination of the Seismic Data eXplorer (SDX) developed at ULIV, the NERIES portal developed at ORFEUS/KNMI and Rapid technology developed at UEDIN, provides web- enabled tools to explore, visualise and analyze seismic waveform data. RapidSeis gives the user an easy point of access to explore seismic data stored at data centres. It also provides a development environment for developing, deploying, and sharing of seismic analysis tools. The development of this application was supported by the FP6 NERIES project and the development of the prototype of SDX embedded in NERIES web portal was funded by the UK's Joint Infrastructure Services Council as the RapidSeis project. In the current project we will transfer the same functionality into the ADMIRE framework to enable the scientist to easily write their own data analysis and reduction procedures without worrying about the underlying VERCE infrastructure. We will also provide a basic set of analysis modules typical in many seismic applications, such as: filtering, decimations, rotation, among others; these modules can be used directly or act as templates for further specialisation. *Status: free of license, ready.*

Waveform High Resolution Locations (WaveHRL), this application, developed at INGV, is an innovative modular procedure that performs sequentially and automatically the following analyses on very large volumes of continuously recorded data: seismic event detection and first location estimation, waveform picking using standard and cross-correlation techniques on detected clusters of events, earthquake refined location techniques using non-linear global inversion and/or double difference re-location techniques. Orchestration of these different phases can be extended to more complex data mining analysis that combine Grid-serial and Grid-parallel processes. The development of this application has been carried out through Italian national projects and by NERA. *Status: free of license, ready.*

Tsunami Risk (TSUMAPS), this application, developed at INGV, exploits a coupling between an operational, real-time data analysis application – earthquake location, magnitude and tsunami potential (through source duration) used at the INGV seismology data centre – as input to an open source tsunami propagation code extensively used at INGV to forecast in near real-time (<5 min) tsunami heights in front of Italian coasts after a submarine earthquake. The calculation of explicit tsunami propagation is then initialized on an HPC environment. The application generates maximum wave height forecastes along the target coastlines. Tsunami travel-time maps will complement wave height information. The calculations can be replicated to form a set of scenarios, for the treatment of the uncertainties with a Bayesian approach, provided that the input seismic parameters are supplied in the same fashion. The forecasted TSUMAPS will be disseminated in the project NERA to the centres that are responsible for tsunami warnings. *Status: free of license, ready.*

Imaging and Monitoring based on Correlations of Ambient Seismic Noise (IMCASN). This innovative application was developed by at ISTerre (former LGIT) and IPGP and its full development is supported by the ERC project WHISPER. The application consists of empirical reconstruction of deterministic waveforms (Green functions) from correlations of records of ambient seismic noise and of their further exploration with spatio-temporal imaging algorithms. Digital continuous records by networks of multiple seismic stations are used as input information. These data are pre-processed and then inter-correlated for all possible station pairs (including auto-correlations). The volume of input information may be very significant (several TB). The first step's output consists of correlations that are individually smaller than the original data but whose number scales as the square of number of stations producing a large result data set. On the other hand, computation of correlations is easily distributable. Therefore the IMCASN is a "data mining" grid-based application. *Status: free of license, ready.*

Data modelling applications

Non-linear Inversion for Kinematic Earthquake Source (NIKES), this application, developed at INGV, is a modular procedure to provide complete spatio-temporal description of the earthquake source by jointly inverting different kinds of ground motion data using a two-stage non-linear technique: a non-linear global inversion stage, using a "heat-bath" simulated-annealing algorithm; a statistical analysis—or appraisal stage—of the model ensemble to compute an average model and its standard deviation, as well as the correlation between the model parameters. The workflows of this parameterised application involve a combination of Grid-serial and Grid-parallel processes, some being computational demanding. *Status: free of license, ready.*

The next two applications are challenging applications for High Performance Computing with increasing memory complexity.

Visco-elastic anisotropic 3D wave propagation simulation at regional scales (SENUM3D, SPECFEM3D, DG3D-SEIsSOL), these applications, developed at CNRS (IPGP-ISTerre-Pau) and LMU, are based on spectral element and discontinuous Galerkin methods. The problem is formulated in time and space. The spatial discretisation is based on local high-order nodal or modal representation of the solution, while time approximation schemes may be of different order using classical or more evolved simplectic formulations. Complex geological media are discretised via 3D unstructured meshes. Classical domain decomposition techniques lead to natural parallel formulation with eventually complex communication fabrics. Actual applications involve parameterised large-scale and very long running simulations that make these simulations challenging for HPC, with important scalability problems on massively parallel multi-core/multi-processors architectures. These applications were supported by the FP6 SPICE Training Network project and currently by the new FP7 QUEST Training Network project. *Status: GNU General Public License, ready, Gordon Bell Award (SPECFEM3D), DEISA extreme computing initiative (DG3D-SEIsSOL).*

High resolution imaging based on 3D full waveforms inversion (SES3D), this application developed at LMU, is based on iterative local optimisation methods enabled with an adjoint-based formulation for the calculation of the cost-function. The problem is classically formulated in space and time. Classical seismology applications involve large volumes of digital continuous records from networks of multiple seismic stations and multiple seismic sources. Thousands of forward and adjoint wave simulation sub-problems, using the above methods, are needed during the optimisation iterations, and at each iteration large 3D forward and adjoint states need to be stored. This application is a challenging application pushing the limit of high performance computing capabilities. Scalability, memory complexity and fast access to large volumes of data are important aspects of this application. SES3D was supported by the FP6 SPICE Training Network project and is currently supported by the FP7 QUEST Training Network project. *Status: free of license, ready*

During the project, new applications and developers group will increase the set of core pilot applications broadening the applications spectrum and enlarging the engaged community.

1.2.2 - Community developers involvement

The developers of all selected core applications have all participated in the project and expressed a strong interest to be involved in the process. All these codes are currently released to the wider earthquake and seismology community through the open-source library repositories of ORFEUS and the QUEST project.

A Community Of Practice (COP) will be established within the project representing the developers of the domain applications. They will provide usage scenarios, workflows and benchmarks for automatic testing.

During the project the developers feed-back will be continuously monitored and analysed to improve and re-orient some of the developments and services.

1.2.3 - *Computational harness of the pilot applications*

With very few exceptions, the core applications and codes have been developed and written by earthquake and seismology scientists with an emphasis on the physics, but with a much smaller emphasis on using the latest technologies available from the data integration, data mining and computing science communities. The codes and tools are written in different computer languages and are dependent on different libraries and computational platforms. The proposal aims at improving the situation in a number of ways.

- *Refactoring:* identify potential re-usable data and computation oriented components that can be extracted by refactoring existing methods and software implementations; and then improve their interfaces.
- *Re-engineering:* indentify in these re-usable data and computation components those that need re-engineering, improvements to algorithms – or data and computational strategies modifications – to improve their performance and to better exploit the capabilities of the different HPC and Grid resources via the VERCE platform;
- *Workflow development:* analyze and identify the granularity of the different data and computation process elements and of the data exchange components of the pilot applications and use-case scenarios;

Specific issues of the data exploration and data analysis pilot applications are related to: complex data queries, distributed and heterogeneous data access, data movement and data integration, as well as complex data preparation and analysis. Another issue is the bottleneck of the I/O and network bandwidth that has to be addressed through parallel analysis of data or analysis of parallel data models.

Specific issues of the data modelling pilot applications are related to: scalability on multi-core architectures and GPU adaptation, memory complexity and fabrics. Another issue will be their mapping on the service-oriented architecture.

1.2.4 - *Workflow tools environment*

The diversity and the increasing scale and complexity of data-intensive applications in earthquake seismology requires the development of workbenches and gateways that support a coherent and consistent model in which data exploration, data integration and data analysis processes integration can be handled efficiently.

Integration and optimisation will be facilitated by the development of workflow patterns and efficient workflow engines that enable enactment optimisation and automated adaptation to changes in the data resources and service infrastructures.

The coupling of different data mining and data modelling techniques to achieve new innovative simulations and imaging methods requires a large element of coordination, structured data management and resource scheduling to be performed across the Grid and HPC infrastructures. A workflow orchestration tool will greatly facilitate the integration process and interactive monitoring.

VERCE will explore and adapt the approach and tools already developed by UEDIN in a number of supported European projects, e.g., the Advanced Data Mining and Integration Research for Europe (ADMIRE) project, which supports data-intensive workflows and OGSA-DAI, which provides data access and integration standards.

In the past decades, a wide range of workflow management systems have been established to support the construction and management of workflows; for example, the Open Source Kepler workflow management system adopted in the Geosciences Network GEON project in

the US, offering graphical tools for workflow composition, used on HPC and Grid environments internationally.

1.2.5 - *Grid and HPC resources*

The deployment of an underlying data-aware distributed computing layer needs to be analysed from a strategic and feasibility point of view. The keywords for the services are "Pragmatic" and "Smart, Simple and Fast": Pragmatic, as the services will use and adapt existing technologies; Smart, as the services will have to overcome all kinds of obstacles so that developers and data resources are not confronted by them; Simple, because the use and integration in the working processes of the domain developers and data resource managers must be simplified; Fast, because the services must be intuitive and responsive to use for both the data resource managers and scientists.

The storage and computing resources will include:

- A number of high value computational and data resources existing in the seismological partners' sites of the consortium already – or soon – included in the EGI/NGIs infrastructure, and that are locally managed,
- A number of additional Grid resources of the EGI/NGIs infrastructure open to VERCE consortium through the Earth Sciences Virtual Community of Research (VCR),
- A number of accesses to HPC computational resources will be provided to the project by the HPC centres of the VERCE consortium (LRZ, UEDIN-EPCC, CINECA).

The Grid resources are operated within the EGI/NGIs infrastructure with the core services of UMD. The HPC resources are operated by the HPC centres of the VERCE consortium; they will provide access and core services based on Globus and UNICORE.

A VERCE Virtual Organization (VO) – linked to the Earth Sciences VCR (EGI-Inspire) – will be created. The VO will provide a framework that leads to more inter-working and holistic, coherent organization of the distributed Grid and HPC resources and of the users' activity within the VERCE e-science environment – with global services like membership management, registration, authorization, monitoring and authentication. The operation and management of the VO will be undertaken by the VERCE consortium.

A VERCE VO, will also be a natural framework interfacing the VERCE platform to other existing specialised earthquake and seismology platforms through interoperable scientific gateways and portals. The ETH Zurich is, in the context of NERA and SHARE and of the work related to GEM, implementing a comprehensive web service-oriented architecture. This will provide access to seismic risk databases, robust computational engines based on OpenSHA; that have been developed by the SCEC and the USGS (US), for seismic hazard and risk assessment. A set of web-based systems is provided to control the computational engines and to deliver results from the seismic hazard calculations. OpenSHA can be used in a Grid and HPC environment.

Access and support for porting earthquake and seismology applications to High Performance Computing infrastructures is currently done individually by some national HPC centres which have earthquake and seismology researchers as users. These national resources generally have the drawback of requiring application proposals to be from a national research user. By integrating HPC infrastructures access to the VERCE platform through a number of bridges, linked to LMC, EPCC, and CINECA centres, VERCE will foster community applications and allow them to run on a number of different platforms, improving their robustness and allowing the exploration of different optimisation strategies. The current DEISA mechanisms for compute resource sharing will be further exploited to secure resources also outside these providers. The development of a consistent work program to productise and optimise on a number of platforms a set of pilot applications software will be an important achievement

with regards to the roadmap of a European High Performance Computing infrastructure and the involvement of the earthquake and seismology community in this roadmap.

1.2.6 - *Data infrastructure and services*

Enabling the European Integrated Data Archives (EIDA) infrastructure, with new capabilities in terms of data access, data integration and data transportation, together with new data-intensive capabilities will be instrumental in advancing the scientific exploitation of the increasing wealth of data and information in earthquake and seismology research.

Data sources will be wrapped using the Open Grid Services Architecture-Data Access and Integration (OGSA-DAI) standards. These standards ensure a uniform approach to data access and integration, and are sufficiently flexible to allow the selection of appropriate transportation mechanisms. The project will provide adaptation of ArcLink as well as explore new technologies for the low-level transport mechanism (e.g., technologies developed within EUDAT, if funded). Currently, OGSA-DAI already includes other transportation mechanisms such as GridFTP, and other mechanisms can be added easily through its open architecture design.

1.2.7 - *User support*

Appropriate support mechanisms helping the earthquake seismology community to use the service capabilities of the VERCE platform must be smart, simple and fast.

User support will be provided through a single web-enabled interface, providing a central point of contact while exploiting the distributed user-support services and pool of expertise associated with the ORFEUS and the support services of the European e-infrastructure. This interface will be integrated to the scientific gateway hosted by ORFEUS.
A number of web-based software solutions will be investigated, such as the Service Administration from UEDIN-EPCC (SAFE), a state-of-the-art technology developed by UEDIN.

1.2.8 - *Progress indicators*

There are three phases of the VERCE project. The first is the resources integration and enabling core services phase.

This will be achieved by the timely deployment of the services supporting:

- Service administration facility (user access and registration point)
- Grid services
- European Integrated Data Resources
- HPC access and bridges

This alone will be a major achievement of VERCE.

The second is the successive integration and deployment of the releases of the VERCE platform services, tools, data and software engines that progressively address the requirements of the data analysis and the data modelling applications.

The metrics will be defined to encompass:

- the 6-month release cycle of the VERCE research platform,
- the increasing number of seismology application codes operational on that platform,
- the monitoring information recording how those applications used the platform, and
- users' evaluations of the platform's usability and capability.

The second is the adaptation and 'productisation' of core pilot applications to the e-science environment provided by VERCE, drawing on Grid and HPC resources from the European

14

e-infrastructures. Progress in this area is more easily quantified. Metrics will be defined to encompass:

- Number of codes successfully deployed on the VERCE infrastructures
- Number of codes successfully adapted to the new architecture paradigm
- Demonstrated code performance improvements on a given infrastructure

The third is a standardisation and final release of the research platform. It builds to some extent on the developments in the first and second phase but actual developments start in parallel. The aim is to develop and get broadly adopted a set of standards for the services and tools, and their APIs.

1.3 - Methodology to achieve the objectives of the project

The VERCE project is composed of three different phases, which can be partly developed in parallel at the beginning of the project, but come increasingly interconnected as they are fully integrated during the project. All of these phases contribute to the development and use of the e-Science environment that will empower the Virtual Research Community. An overview of this architecture is provided in the figure 3; it also shows which work packages are responsible for which elements in or activities surrounding the architecture.

- The first phase is the initial set-up of the underlying layer of distributed computational and data resources provided by the project partners, including the community European integrated data infrastructure linked to the ORFEUS consortium. This phase will involve the deployment of Grid middleware and services, and integration of bridges to the HPC infrastructures. The core services are provided mainly by gLite, Globus and UNICORE, but will be extended during the project. This phase will create the earthquake and seismology research Virtual Organisation (VO), linked to the ESR VCR, and provide unified access to the Grid and HPC resources for the project members and the earthquake and seismology community at large. This phase will involve coordination with the EGI-NGIs infrastructure and the HPC centres.
- The second phase will be a continuous six-monthly cycle of successive release of the VERCE platform. Based on the pilot application requirements and the platform monitoring, the architecture and the services and tools will be updated with carefully selected new components and technologies. These new components are evaluated, integrated and released in a new platform version that is then deployed as the new research platform. Pilot applications are made operational on the new platform and monitored together with the services and tools. This will allow us to identify gaps, issues or improvements that constitute the new requirements for the next version of the platform. A 6-month cycle is adopted in VERCE to achieve agility and rapid delivery.
- The third phase is the 'productisation' of the pilot applications and of the use case scenarios. This phase will lead to (1) refactoring, re-engineering and optimization of the software implementation of the pilot applications; mapping these components to the platform and evaluating possible gaps or architectural issues. This is part of the requirements gathering for the design of the next version platform.
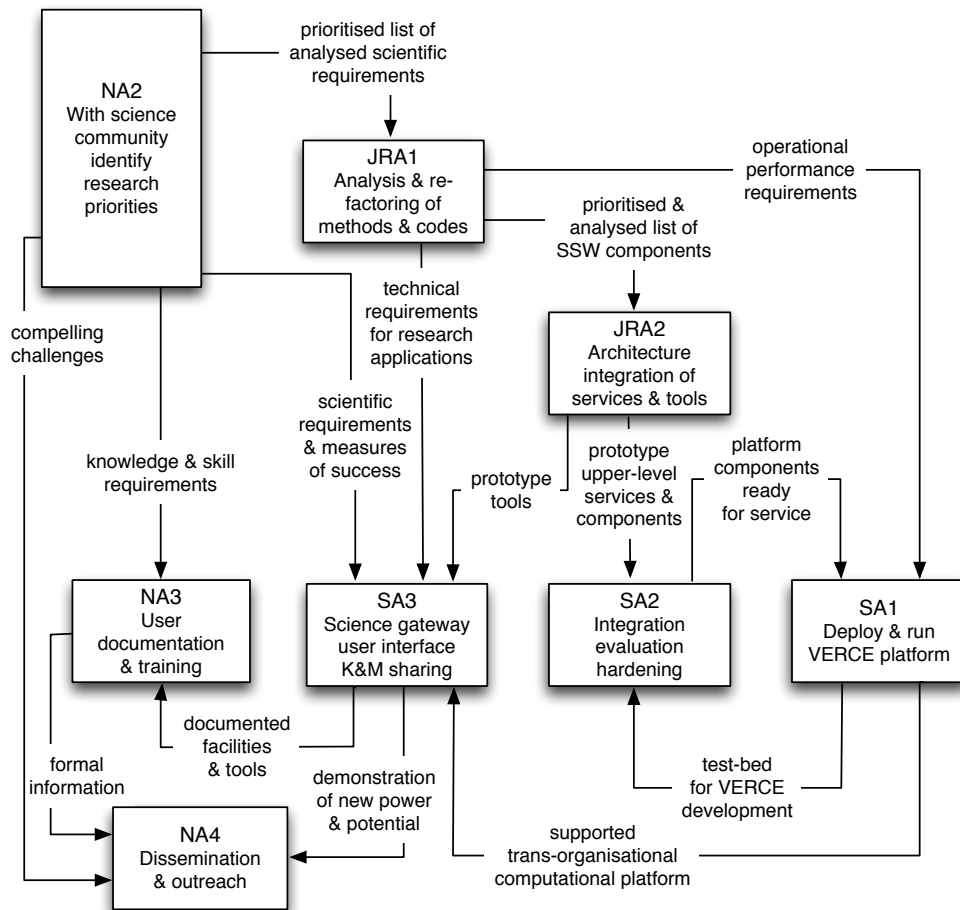
Figure 2 Work packages relations in VERCE

The project consists therefore of a set of related work packages. A high level of coordination and monitoring is needed to insure the timely delivery of the different components. Emphasis has been put on providing sufficient management structures and resources at the different levels of the project. The VERCE project is therefore structured into a set of networking, services and research activities according to the following organisation:

Networking activities
- NA1 - Management
- NA2 - Pilot data-intensive applications and use cases
- NA3 - Training and user documentation
- NA4 - Dissemination and public outreach

Service activities – infrastructure and deployment operation
- SA1 - Management and operation of the research platform
- SA2 – Integration and evaluation of the platform services
- SA3 – Scientific gateway, user interfaces and knowledge, and method sharing

Joint Research Activities
- JRA1 - Harnessing data-intensive applications
- JRA2 – Architecture and tools for data analysis and data modelling applications

The contents of the work packages within these activities are outlined below.

1.4 - Networking activities

1.4.1 - *Overall strategy and general description.*

The networking activities will run for the duration of the project. The NA activities are horizontal orchestration activities which emphasize a user-centricity evident throughout the project. These activities will lead seeding changes in the collaborations across seismology, computer sciences and the data-aware engineers of the VERCE consortium.

Network activities are divided between a management and coordination work package (NA1), providing the alignment of interests encompassing the community of users in research and education, data and e-infrastructure providers; a pilot applications and use case scenarios work package (NA2) allowing user validation of the architecture, tools and services; and finally, the training and user documentation work package (NA3), which, together with dissemination and public outreach work package (NA4), will build up 'intellectual ramps' to encourage and expand the user community, thereby improving the sustainability of the e-science environment.

The network activities strategy is to:

- Smooth the path from theoretical research, through proof of concept and into a sustainable and dependable research e-science environment,
- Create and share data-intensive methods, tools and practice – named here as 'datascopes'- for exploiting the data and revealing new insights,
- Foster the emergence of 'research technologists' who support the researchers of the scientific community with a thorough understanding of the research goals,
- Build the so-called 'intellectual ramps' by providing education and training to foster the adoption of data-intensive methods and practice by a wider community of users in the earthquake seismology community and beyond.

The networking activities are essential for the success of the project, as the NAs

- Define and monitor the project strategies,
- Carry out the project plan and achieve the defined objectives,
- Identify the needs of the scientific users, promote and seek user feed-back on the developed services and tools,
- Validate the implemented services and tools, based on well-defined real scientific user applications,
- Establish a Community of Practice (COP) representing the scientists and the data resources,
- Ensure that researchers developing new 'datascope' methods work closely with researchers using the methods to prevent technological centricity,
- Derive synergy by working with evolving European e-infrastructures, e.g., EGI/NGIs, PRACE, and if funded EUDAT,
- Provide a seeding contribution to the e-science environment, the ESFRI EPOS-PP in the solid Earth Sciences community
- Drive and draw on related European research initiatives such as ENVRI, if funded
- Develop and share coordinated activities with other related European projects in the community, e.g., QUEST, WHISPER, NERA, SHARE, and international projects of the international seismology data community in the US (IRIS-DMC) and Japan (Jamstec and NIED),
- Investigate additional sources of funding for the project through contacts with other decision makers.

The NA activities are intimately related to all other service and joint research activities of the VERCE project.

NA1: Management

NA1 provides the administrative support and the management for the project as a whole. The management of VERCE concerns the coordination and the integration of all the project activities. In particular, ensuring and promoting channels of communication within and between joint research, service and networking activities will be instrumental to guarantee the full integration of the different parts of the project.

The CNRS, through the Institut de Physique du Globe de Paris (IPGP) will act as coordinating partner. UEDIN will have an IT Coordinator (ITC) deputy role. The coordinator will be assisted by a **Project Management Office (PMO)**, which will handle administrative matters, including financial and legal services. The PMO is provided by the CNRS.

To harmonise the e-science environment development with the ESFRI EPOS-PP requirements, and to enforce the ties to the earthquake and seismology community at large, the coordination and management will be led by an earthquake and seismology partner together with an IT partner with a deputy role, and a steering group involving a strong technology support.

A Project Executive board (PEB) will be set up and meet remotely by video conferencing tools on a weekly basis or whenever a particular need arises. This group will be providing the technical and scientific management on a day-to-day basis. The PEB will consist of the Project Coordinator, the IT Coordinator acting also as the architecture coordinator (JRA2), the enabling application coordinator (JRA1), the platform service coordinator (SA1/SA2), the outreach officer (NA3/NA4), the scientific gateway coordinator (SA3) and the user application coordinator (NA2). The PSG will also organize quarterly face-to-face meetings, to include all the work package leaders and key personnel on the project. The location of this meeting will rotate through the participant institutes

In addition, an Advisory Board (AB) and a User Board (UB) will be set-up with leading representatives of the European e-infrastructures and of the VERCE RIs ecosystem, with a special mention to the ESFRI EPOS-PP, according to the detailed description of the management structure provided in section 2.1 of this document. Both AB and UB will meet once a year.

The PEB, in coordination with the Advisory Board, will drive synergy with a number of other European projects:

- The e-infrastructure projects EGI/NGIs, PRACE/DEISA2 and if funded EUDAT,
- EPOS – the ESFRI infrastructure of the solid Earth Sciences – which just entered the preparatory phase (i.e., EPOS-PP) and, if funded, ENVRI,
- The existing related European seismological projects like NERA, QUEST and SHARE,
- The associated projects ADMIRE, through UEDIN, and MAPPER, through BADW/LRZ,
- The seismological international data infrastructure consortia: IRIS-DM, in the US; Jamstec and NIED, in Japan,

Other related projects like D4Science-II, GENESI-DC, SIENA, gSLM …

Management will liaise with application developers, data resources managers and users of the earthquake and seismology community to seek user feedback and monitor additional capabilities and computational resources for improving the exploitation of the data resources at the European level and physics-based data modelling
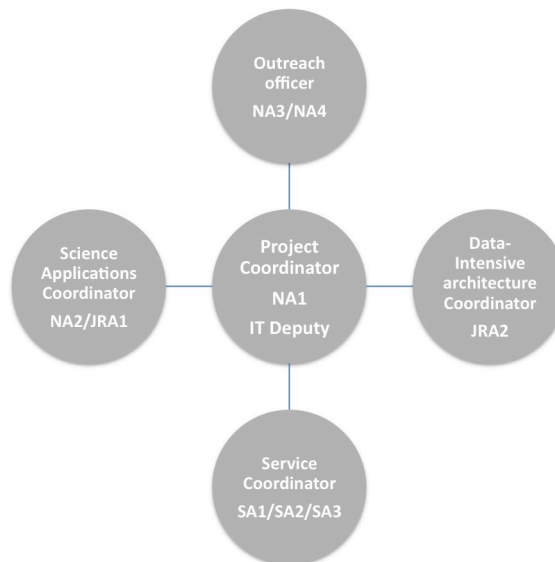
Figure 3 Management and coordination

Dependencies between the different components will increase during the project, in particular for the integration of the applications within data and workflows and the definition of tailored workbenches. This shall require a horizontal monitoring and collaboration between the JRAs, SAs, NA2, NA3 and NA4. This will be implemented with a task leader to report, evaluate risks and propose contingency plans to the management team.

The Project Steering committee will also investigate additional sources of funding for the project through contacts with decision-makers; explore calls and opportunities for furthering project collaborations within project members' national initiatives; and endeavour towards strategic actions with stakeholders involved in similar technologies and infrastructures.

NA2: Pilot data-intensive applications and use cases

The NA2 activity is strategic for VERCE – it selects the pilot data-intensive applications and it defines sound use case scenarios based upon challenging research problems that make effective use of the data. These pilot data-intensive applications are named here 'datascopes'.

The selection of the 'datascopes', and the incremental priority strategy to be adopted during the project, shall be based upon

- The evaluation of their scientific impact by the domain specialists,
- The analysis with JRA1 and SA1 of their complexity and issues in terms of new methods, services and tools,
- The identification with JRA1 of generic and reusable processes shared between different 'datascopes' in terms of data flow (gathering, preparation, integration and processing movements) and processing workflows.

It is incumbent on NA2 to emphasize user-centricity throughout the project and to ensure a research-led e-science environment. The 'datascope' challenges are driving the VERCE project towards harnessing greater capabilities in data, computation and collaboration, many of them lying in new and existing inter-disciplinary areas.

Rallying this effort can be difficult as it does not benefit a specific group – it depends on sustained leadership raising interest and attracting commitment. The seismological researchers and application providers in NA2 are domain specialists and expert practitioners involved in the development and the distribution of new data analysis and data modelling methods in seismology and related fields. They work in synergy with associated European

seismological projects: the ERC project WHISPER and the ITN project QUEST. They have already expressed their commitment to release their applications to the wider EU earthquake and seismology community under an Open Source model and licensing policy. They have all expressed their commitment to be actively involved in the process of building up 'datascopes' from their applications.

NA2 will ensure research practices co-evolve with new methods and their supporting software. Important data-intensive research questions and new datascopes also demand changing methods and practice. This is the space that gains most through collaboration between domain specialists (NA2), computer scientists (JRA1&JRA2), and data-intensive engineers (SA1&SA2), not just because new algorithmic approaches are needed but also because new ways of assembling the components are greatly needed.

It is incumbent on NA2 to ensure that researchers developing new datascopes work closely with researchers using the methods to prevent technology centricity. This will be achieved through collaboration between NA2 and NA3 for designing tailored training session material based on the pilot applications and use case scenarios, application documentation and best practice guides; and NA4 for providing demonstrators and for improving dissemination material toward new potential users of the solid earth sciences community; and SA3 for the design of interfaces to the scientific gateway tailored for users and developers of the datascopes.

During the project, the pilot applications and software will transition to an in-house 'product' that the group will support and use repeatedly. This 'productisation' of the methods and their implementation – that may require re-factoring and re-engineering, improved architecture and interfaces, workflow and data flow optimisation – will be performed by JRA1, JRA2 and SA2. During transition the pilot application is particularly vulnerable to diverging from research requirements. NA2 will ensure a research 'mind-share' niche embedding domain users and contributors with members of the transition team

Finally, the tooling to work provided by the VERCE e-science environment must improve the development and the execution of these applications in quality, speed and scale and hence will accelerate the process of research. It is the task of NA2 to observe, measure and evaluate how the e-science environment (services, tools, collaboration techniques) provided by VERCE improves the implementation and the efficiency of the datascopes, the 'intellectual velocity', the production and the quality of the results. The NA2 strategy will draw on the methodology of empirical software engineering.

The NA2 objectives are:

- Select the pilot applications and design the use case scenarios in coordination with the JRAs and the SAs, together with a community of practice (COP)
- Provide documentation to the other SAs and JRAs
- Support and validate with JRA1 the 'productisation' of the methods and their implementation (re-engineering, re-factoring, improved interfaces and workflow and data flow optimisation). Thereby additional innovations cycles will be initiated.
- Validate the application and the use case scenarios integration and deployment on the VERCE platform in coordination with SA2, SA1 and JRA2. Thereby additional innovations cycles will be initiated.
- Provide contribution and support to NA3 for the definition and the design of training session material, best practice guidelines and documentation.
- Provide contribution and support to NA4 for the selection of demonstrators and the design of dissemination material
- Provide requirements and support to SA3 for tailored interfaces of the scientific gateway for users and developers.

NA3: Training and user documentation

The strategy is to provide a two-layered approach of training and education activity:

1. The first layer provides the users and developers of the earthquake and seismology community with a tailored incentive set of training sessions, documentation, and best practice guides for learning how to best use the e-science environment of VERCE for their data-intensive research; and beacons of good practice for developing new datascopes and their supporting software.
2. The second layer provides a set of actions to leverage the awareness and knowledge of the community at a time of rapid changes in the data and computational infrastructures capabilities, and of new challenges both in terms of methods and algorithms for the scalability of data-intensive software.

The NA3 activity draws on the synergy and the collaborations with NA2 and JRA1 for research-led training applications and use cases; SA3 for the scientific gateway tools and tailored training interfaces; and SA1 for providing tailored training workbenches.

This training and user documentation activity program will draw on a survey of the education and training needs, which will be conducted with NA2 and JRA2.

The training may be organised by an annual "summer school/boot camp" which serves the dual purposes of (a) training VERCE's participants and key ambassadors for VERCE, who are already committed to the cause, and (b) developing and refining training material that can be made available for self-paced learners, who will use the material when they recognize the value of VERCE's approach to research – possibly as a result of the outreach by NA4.

Training sessions

NA3 will provide a user-centric training program driven by the earthquake and seismology data-intensive problems, and designed to enable the researchers to develop incrementally the required knowledge and skills to best use the e-science environment services and tools provided by VERCE for their data-intensive research. This will be composed of

- A comprehensive set of 1-2 days training sessions, organised around a number of selected incentive training scientific scenarios and applications. These will be designed based upon the actual pilot applications and use case scenarios of NA2;
- A companion guide for users of the e-science environment;
- A set of test cases;
- A set of best practice guides based on beacons of the pilot data-intensive applications deployment and adaptation made by NA2 and JRA1.

The training sessions will be integrated in the scientific gateway provided by SA3, through a number of tailored interfaces, and deliver to the external community.

The set of training sessions will incrementally introduce researchers to tools and services of increasing complexity. This may include

- Using the VERCE e-science environment: how the e-science environment works, how to access and submit jobs, …
- Data handling within the VERCE e-science environment: how to explore and visualise data sets, how to gather, transform and integrate data, how to parallelize data flows …
- Software development within the VERCE e-science environment: how to re-factor and 'productise' data-intensive codes, how to improve their robustness and transportability, how to improve code architecture by making use of software engines …

- Software optimization within the VERCE e-science environment: how to improve efficiency through well-designed data-analysis and data modelling algorithms, how to improve data flows and workflows, how to optimize the distilling and visualisation of the results...

The aim is to provide an 'intellectual ramp' to the solid Erath science community and beyond, and improve their use of the data. A safe ramp is particularly necessary because many researchers need to adopt new data-intensive computer-enabled methods at a time when methods and technology are changing rapidly. It will also become the most common point of first contact with the VERCE e-science environment and beyond with the European e-infrastructures.

To avoid duplicated effort and to foster outreach activity for potential users in the research community, the strategy of the training programme will draw on

- *An active coordination with the European ITN network QUEST, run by LMU and to which a number of seismological partners of the consortium are already contributing.* Training sessions will be organized during the annual workshops of QUEST, and as dedicated workshops of QUEST. This will foster the education in new data-intensive applications and methods of young students and researchers; improving the co-evolution of their research practices and of the e-science environment provided by VERCE; increase the use of data during their education. The collaboration with QUEST will ensure a dissemination of VERCE technologies to a wide range of academic and industrial users.
- *An active synergy with other European seismological projects of the EPOS ecosystem like the ERC project (ERC) WHISPER run by the CNRS (LGIT), and the projects NERA run by KNMI-ORFEUS and SHARE run by ETH Zurich.* Training sessions will be provided and tailored for these projects to enlarge the user community and expose them to the new e-science environment capabilities provided by VERCE.

More technical training sessions, covering all aspects regarding the deployment and the management of the services and tools of VERCE, will also be organised at the sites of the partners of the consortium to support and leverage the knowledge across the technical and research staff in these sites. This will foster the emergence of research technologists.

In addition, the NA3 activity will promote the incorporation of best practice guides into related academic courses and postgraduate activities across the consortium.

Leveraging awareness and knowledge

NA3 will organizes and promote access to training courses to leverage the awareness and knowledge of the community at a time of rapid change in the data and computational infrastructures technology, and in parallel software development and optimisation.

This training can encompass the following topics:

- Fundamental and advanced concepts of grid and high performance computing,
- Exploiting the European Grid and HPC e-infrastructures in Europe
- Message passing programming with MPI;
- GPU programming techniques for visualisation, data processing and simulation;
- Cloud computing for data-intensive applications;
- Parallel scalability and optimisation on large-scale multi-core architectures;
- Green computing challenges and issues;
- Data-centric infrastructures and architectures …

To avoid duplicated effort and to facilitate interdisciplinary training and sharing, NA3 will advertise and promote the participation of the VERCE user community to such training

courses provided by the national HPC centres partners of the VERCE consortium, the EGI and NGIs training activity as well as training activity in related data-intensive projects in Astronomy & Astrophysics. In doing so, the NA3 activity will foster the contribution and the visibility of the earthquake and seismology community in these general training activities.

NA4: Dissemination and public outreach

VERCE is designed to provide a direct benefit to the entire earthquake and seismology community, as well as to the broader, solid earth sciences communities as a whole, as structured by ESFRI EPOS-PP. In addition, VERCE shall also organize and ensure a large, coordinated public outreach program to disseminate information and educational material about the benefits and advances made by the project activities.

The public dissemination of project information and results will play a crucial role in developing community support, as well as in identifying potential partners from outside the consortium, and to raise interest from students, media, and the general public. It is expected that dissemination will provide a general picture of the project.

In general, the first contact with the VERCE project will be through public dissemination activities. Therefore the outreach of the project will strongly depend on the quality and effectiveness of the NA4 activities. Information and material geared towards the specific targeted audiences will be provided by participating domain experts.

The NA4 objectives are as follow:

- Define the dissemination methods and message content for reaching the target audiences
- Distribute information on the goals and benefits of using the VERCE e-science environment,
- Provide and make available through multiple channels informative material adapted to the target audiences,
- Keep the earthquake and seismology communities, as well as the broader earth sciences communities and other interested parties, informed of progress, new developments, improvements and functionalities within VERCE,
- Disseminate demonstration cases through the scientific gateways, as they are prepared by NA3 and NA2,
- Provide printed material for distribution at conferences, workshops, and through other channels
- Disseminate outreach materials to national actors
- Gather and maintain usage statistics for outreach and project activities

NA4 activity will define a comprehensive Public Dissemination Plan. Specific activities will be carried out to reach:

- National public services, policy-makers and governments
- Industry actors in the Earth Sciences, and in Computing and Information technologies
- Domain scientists, as well as members of the broader earth sciences communities
- Mass media

The dissemination strategy of VERCE will benefit from the extensive and diverse exposure available through the EMSC and ORFEUS organizations. Due to its role providing real-time earthquake information services, the EMSC web site alone attracts nearly a million visitors a month from over 150 countries. Project information material will also be made widely available for further dissemination via the EMSC members and ORFEUS participants, which together comprise more than 100 institutes.

Each target audience has its own information requirements, and it is essential that VERCE project information material be provided that appropriately targets the specific audience groups. Policy makers and governments, as well as the general public, must clearly recognise the national and international value brought forth by the project and its scientific and technological advances. The general public as well as the broader solid earth sciences communities will benefit from educational and informative materials produced by the domain experts demonstrating the advances achieved through the project technology and core applications.

The core earthquake and seismological communities will be informed of the developments within the VERCE framework and application deployment through the NA4 dissemination channels. This will have the dual benefit of expanding project awareness within the core communities, as well as providing an opportunity to identify and recruit new users and applications for the framework. These dissemination channels will include a coordinated project web site, Internet news feeds (RSS, Twitter, etc.) of updates and new results, newsletters, and leaflets and other publicity items distributed at technical and domain conferences and workshops and through affiliate institutions.

For the public, the first contact with VERCE will be mainly through the dissemination tools and activities, likely via the project web site. To this end, the web site must be quickly and easily understood, clear, and well organized, so that the promotional and informational materials can be quickly and easily discovered. These materials must focus on raising awareness about the project and the scientific and public value of the ongoing developments within the VERCE framework and applications.

For the VERCE partners, the dissemination platform, developed in conjunction with NA3 and SA3, will be integrated within the scientific gateway and provide a unified communication system encompassing project information and news, system documentation and tutorials, and a comprehensive and integrated "help desk" to handle user questions and problems.

Education tools and material

NA4 activity will support and disseminate a number of educational tools for a new generation of Earth scientists; to introduce graduate and post-graduate students and scientists to new research data mining and data modelling tools and methods, to provide new teaching material at the University and School levels. This involves the production of simple tailored interfaces integrated within the scientific gateway. They should combine: demonstrator educational applications, basic tools and services. This dissemination through EMSC and ORFEUS web sites and portals will promote VERCE to a broad community.

Internal and external communication tools

NA4 disseminations tools will be used to improve communication within the project. A close coordination with NA3 activity will support the dissemination of on-line training courses and a various number of training and support material.

NA4 will survey and evaluate the use of new possible communication and outreaching technologies, like for example the Hub technologies. This will allow the assessment of remote collaboration and communication systems, which ultimately could be integrated within VERCE scientific gateway.

Promotional material

NA4, in conjunction with NA3 activities and participating domain experts, will design and provide a range of promotional material at different levels: non-technical public, policy-makers, and targeted and broader scientific communities. Promotional material will be used to prospect new scientific communities for potential users, as well as to build project awareness

and support. These resources will be available on the project web site and distributed via affiliate institutions and at conferences and workshops.

Resources

Because of the targeted audiences, the main resources will be a communication officer with experience in the scientific community that will work closely with the NA3 effort to convert scientific information into informative dissemination materials (leaflets, newsletters, etc.) targeting the different audiences, and a web developer in charge of developing and maintaining the web site, in coordination with NA3, to provide the targeted information in an easily-accessible web format.

Dissemination events planning and organisation

NA4 will support stands to relevant external events and conferences in different scientific areas during the whole project extension, and the representatives and the activity scientific leaders of the project will attend relevant conferences. NA4 will also support dissemination activities, in the framework of the major European e-infrastructures.

A task of NA4 will be to identify the most relevant events in the areas concerned by the project, with the goal to produce a high visibility of the VERCE activities in those events. A wise planning and prioritisation will assure that travel funds reverberate in a good image of the project.

Indicators and metrics

Metrics will be implemented on all the on-line tools to characterise the number of users and their geographical origin. Publications in media will also be accounted for, as well as the number of printed materials produced and distributed. These, and other system usage metrics collected within the VERCE framework, will in turn be made available in order to characterize the impact of the project.

Industry and policy-makers relations

As part of the integral vision of dissemination activity, VERCE will try to establish bridges between the technologies of the project and major industry across Europe. To cover this target, specific dissemination material will be created and targeted for senior industrial managers and policy-makers.

Efforts will be done also to launch collaborative actions with a number of industry forums associated with the European e-infrastructures (EGI, PRACE), which brought together companies of different sizes scopes, software houses and service providers. Efforts will done to drive synergy with related projects like SHARE and GMES. Finally VERCE will improve awareness among the industry, in particular the oil and exploration industries, promoting them to participate and producing technologies transfers.

1.4.2 - Timing of the NA work packages and their components
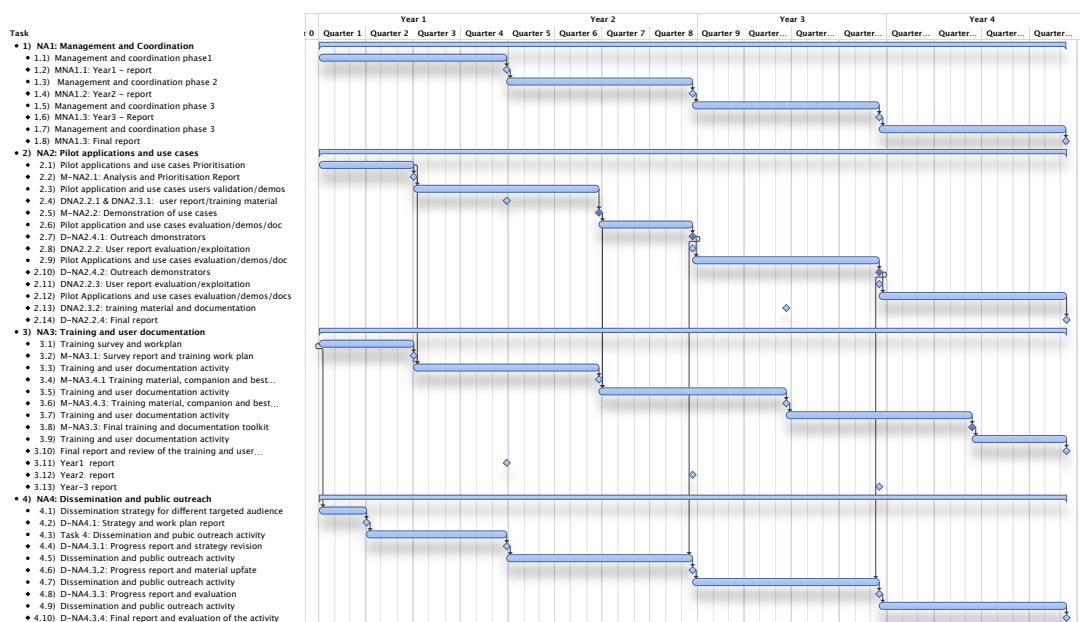
1.4.3 - List of the Networking Activity Work Packages

WP	Work package title	Type of activity	Lead	Lead participant short name	P/M	Start	End
NA1	Management	MGT	1	CNRS	30	M1	M48
NA2	Pilot applications and use cases	COORD	5	INGV	36	M1	M48
NA3	Training and user documentation	COORD	7	ULIV	49	M1	M48
NA4	Dissemination and Public Outreach	COORD	4	EMSC	28	M1	M48
	TOTAL				143		

1.4.4 - List of the NA deliverables

number	Deliverable name	WP	Dissemination	Expected date
D-NA1.2.1	Mid-year brief progress reports	NA1	R, PU	PM6
D-NA4.2	Project website and internal information tools	NA4	R, PU	PM6
D-NA2.2.1	Year-1 Pilot applications and use cases report	NA2	R, PU	PM12
D-NA2.3.1	Tailored training applications documentation, best practice guides (users),	NA2	R, PU	PM12
D-NA3.2.1	Year-1 Training and user documentation	NA3	R, PU	PM12

	report			
D-NA4.3.1	Year-1 Dissemination and public outreach report	NA4	R, PU	PM12
D-NA1.2.2	Mid-year brief project progress report	NA1	R, PU	PM18
D-NA2.2.2	Year-2 Pilot applications and use cases report	NA2	R, PU	PM24
D-NA2.4.1	Demonstrators for dissemination activities	NA2	D, PU	PM24
D-NA3.2.2	Year-2 Training and user documentation report	NA3	R, PU	PM24
D-NA4.3.2	Year-2 Dissemination and public outreach report	NA4	R, PU	PM24
D-NA1.2.3	Mid-year brief project progress reports	NA1	R, PU	PM30
D-NA2.3.2	Tailored training applications documentation, best practice guides (users),	NA2	R, PU	PM30
D-NA2.2.3	Year-3 Pilot applications and use cases report	NA2	R, PU	PM36
D-NA3.2.3	Year-3 Training and user documentation report	NA3	R, PU	PM36
DNA2.4.2	Demonstrators for dissemination activities	NA2	D, PU	PM36
D-NA4.3.3	Year-3 Dissemination and public outreach report	NA4	R, PU	PM36
D-NA1.2.4	Mid-year brief progress reports	NA1	R,PU	PM42
D-NA2.2.4	Year-4 Pilot applications and use cases report	NA2	R, PU	PM48
D-NA3.2.2	Year-4 Training and user documentation report	NA3	R, PU	PM48
D-NA4.3.3	Year-4 Dissemination and public outreach report	NA4	R, PU	PM48

Work package number	NA1	Start date or starting event:		M1			
Work package title	Management						
Activity type	MGT						
Participant number	1						
Participant short name	CNRS						
Person-months per participant	30						

Objectives

The aim of this work package is to ensure the effective management of the VERCE project. The objectives are to:
- Establish procedures for tracking and controlling the progress of the project;
- Establish procedures for taking decisions effectively and quickly;
- Establish quality control procedures with respect to all outputs and deliverables;
- Ensure the project proceeds according to administrative, financial and legal principles defined by European and national regulations;
- Ensure that participants conform to their obligations under the contract and the consortium agreements;
- Coordinate the efforts of all work packages of the VERCE project;
- Coordinate and harmonise activities with ESFRI EPOS-PP and seismological European projects, EGI and the NGIs, PRACE and DEISA2, and with EUDAT if funded;
- Investigate additional sources of funding through national initiatives and contacts with decision makers at the national level.

Description of work

Provide the technical and scientific management of the VERCE project
The Project Executive Board directs the strategic orientation of the project and safeguards the interests of all participants. It directs the project to the workplan, establishes the quality control criteria, assesses the performance of the project and takes appropriate action if performance is found to be lacking. It ensures that the work packages interact effectively and convenes the scheduled weekly and quarterly meetings.

Facilitate coordination and harmonisation with relevant external projects
The Advisory Board and the User Board will ensure that liaison and cooperation with other projects takes place in the most efficient manner, and that the project responds appropriately to requests for cooperation form other projects.

Establish and run the Project Management Office (PMO) to support the management of the VERCE project

The PMO handles the administrative, financial and legal services, maintains a project document repository, tracks and reports on project progress, informs on quality control requirements of outputs and deliverables, communicates project matters to all partners in the consortium, and coordinates requests for information from other projects.

Deliverables (brief description) and month of delivery

M-NA1.1.1 (PM12), **M-NA1.1.2** (PM24), **M-NA1.1.3** (PM36), **M-NA1.1.4** (PM48) – Annual Review Reports

D-NA1.2.1 (PM6), D-NA1.2.2 (PM18), D-NA1.2.3 (PM30), D-NA1.2.4 (PM42) – Mid-year brief progress reports

Work package number	NA2	Start date or starting event:		M1			
Work package title	Pilot applications and use cases						
Activity type	MGT						
Participant number	1	5	6	7			
Participant short name	CNRS	**INGV**	LMU	ULIV			
Person-months per participant	7	15	7	7			

Objectives

- Select the existing pilot data-intensive applications and design sound use case scenarios; Analyze and define a priority strategy during the project with JRA1, SA1 and JRA2.
- Support and evaluate the 'productising' transition of the methods and their implementation performed by JRA1
- Support SA1 and JRA2 with application requirements for the definition of the workbenches and functionalities
- Support and evaluate the deployment and the efficiency of the pilot applications and their use case scenarios on the VERCE platform
- Define and provide in collaboration with NA3 improved documentation, best practice guides and tailored training session material
- Define and provide in collaboration with NA4 demonstrators and dissemination material
- Provide requirement and support to SA3 and JRA2 for tailored interfaces of the scientific gateways targeted to the developers and the users

Description of work

Task 1. Selection and prioritization of the use cases:

It selects a number of existing pilot applications based upon their users community, scientific impact. It defines scientific use case scenarios based on actual research case. With the help of JRA1, SA1 and JRA2, it analyze and prioritizes the pilot applications and use case scenarios in terms of there complexity and requirements (services and tools). This prioritization will be based on the typology of the use cases since many of the applications do share common software requirements (e.g., same core routines, common parts of the workflow), and/or require similar HPC or data resources of the test-bed. The prioritization will also follow from ranking the impact of the use case for the community. Special attention will be also paid to the ongoing developments in the enlarged community of practice (COP). This will result in documentation of the pilot applications and user case scenarios, as well ass in requirements for new services, tools, software engines and tailored workbenches.

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

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.

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

This includes documentations, best practice guides and tailored training session material based upon the pilot applications and the use case scenarios; together with a survey of the earthquake and seismology community needs in terms of training and user documentation.

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

This includes demonstrators and dissemination material to enlarge the community of users.

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

This concerns in particular tailored interfaces for the users and the developers of the pilot applications. Evaluate of the VERCE the scientific gateway with regard to user access and documentation.

Deliverables (brief description) and month of delivery

D-NA2.2.1 (PM12), D-NA2.2.2 (PM24), D-NA2.2.3 (PM36): D-NA2.2.4 (PM48): Reports on the progress and user evaluation of the 'productisation' and deployment of the pilot applications, and use case scenarios, and new application requirements

D-NA2.3.1 (PM12), D-NA2.3.2 (PM30): Documentation on tailored training session applications and best practice guidelines

D-NA2.4.1 (PM24), DNA2.4.2 (PM36): Demonstrators for the dissemination activities

Milestones

M-NA2.1 (PM6): Selected scientific pilot applications and use case scenarios, and their prioritization.

M-NA2.2 (PM18): Demonstration case of scientific exploitation of a pilot application deployed on the VERCE platform

Work package number	NA3	Start date or starting event:		M1			
Work package title	Training and user documentation						
Activity type	MGT						
Participant number	1	2	3	5	6	7	8	10
Participant short name	CNRS	UEDIN	KNMI	INGV	LMU	**ULIV**	BDW LRZ	CINECA
Person-months per participant	4	4	10	4	4	**15**	4	4

Objectives

Provide users and developers of the earthquake and seismology community with

1. A comprehensive and incentive set of user-centric tailored training sessions and documentation for learning how to best use the seismological e-science capabilities developed in VERCE for their data and computing intensive research;
2. Good practice guides based on the development and adaptation of new 'datascopes' methods, and their supporting software, during the project;
3. A set of incentive actions to leverage the awareness and knowledge of the community at a time of rapid changes in the data and computational infrastructures capabilities utilising existing training programs of the European and national e-infrastructures.

The aim of this work package is to provide 'intellectual ramps' for earthquake and seismology researchers to engage incrementally with the tools, techniques and methods of data-intensive research.

Description of work

Task 1. Provide a training and user documentation work plan:
The work plan will be based on a community survey, in coordination with NA2 and JRA2, of the education and training needs for the VERCE project. This survey will be conducted in synergy with the ESFRI EPOS-PP project together with the related European seismological projects WHISPER (ERC), QUEST (ITN), NERA and SHARE. The outcome of the survey will define the needs for training programs provided by the different partners of the consortium;

Task 2. Provide and run annually a comprehensive set of training sessions:
1-2 days training session will be provided centred a round selected scientific scenarios and applications chosen in coordination with NA2 and JRA1. This encompasses training sessions provided and targeted toward scientific users in synergy with QUEST and WHISPER and also training session toward the technical staff at the different sites of the VERCE platform. We will also engage the wider Earth Science community by staging a set of broader training sessions on e-science applications.

Task 3. Organise and publicise training materials for the scientific users of the e-science environment provided by VERCE:
In this task we will compile the user documentation provided by the different pilot applications and use cases. The scope and contents will be defined in coordination with NA2 and JRA1 upon beacons of 'datascopes' methods based on the pilot applications and use cases

of NA2. This will lead to a unified set of documentation for the pilot applications, a companion guide, and best practice guides.

Task 4. Organise and package the training material as a comprehensive 'toolkit':
Additionally, the material will be organized in coordination with NA2 and JRA1, as well in an annual "summer school/boot camp". This material will be made available through the scientific gateway and will serve a dual purpose for (a) researchers and technical staff associated with the VERCE consortium; (b) self-paced learners who will use the material to apply the VERCE's approach to their own research.

Task 5. Support and promote incentive initiatives to leverage the awareness and knowledge of the earthquake and seismology community as well as the broader Earth science community:
We will actively engage with ongoing training programs of e-infrastructures providers – e.g. PRACE, DEISA2, EGI and the NGIs, some of those are not very well known in the Earth Science community. We actively publicise these training programmes and imbed them in our own training strategy. By doing so we will build 'intelectual ramps' for Earth scientists into the broader e-science community and disseminate the VERCE infrastructure and knowledge to the wider community.

Deliverables (brief description) and month of delivery

D-NA3.2.1 (PM12), D-NA3.2.2 (PM24), D-NA3.2.3 (PM36), D-NA3.2.4 (PM48) – Annual reports on the training and user documentation activity including the catalogue of the training sessions delivered through the scientific gateway and the next year work plan.

Milestones

M-NA3.1 (PM6):   Training survey report and work plan
M-NA3.4.1 (PM18):   Companion guide, best practice guides, training material
M-NA3.4.2 (PM30):   Companion guides, best practice guides, training material
M-NA3.3 (PM42):   Training and user documentation final toolkit

Work package number	NA4		Start date or starting event:			M1			
Work package title	Dissemination and Public Outreach								
Activity type	MGT								
Participant number	2	3	4	5	7	8	9	10	
Participant short name	UEDIN	KNMI	**EMSC**						
Person-months per participant	4	9	15						

Objectives

- Define the dissemination methods and message content for reaching the target audiences
- Distribute information on the goals and benefits of using the VERCE e-science environment,
- Work in close collaboration with NA3 to provide and make available through multiple channels informative materials adapted to the target audiences,
- Keep the earthquake and seismology communities, as well as the broader earth sciences communities and other interested parties, informed of progress, new developments, improvements and functionalities within VERCE
- Disseminate demonstration cases through the scientific gateways, as they are prepared by NA3 and NA2,
- Provide printed material for distribution at conferences, workshops, and through other channels
- Disseminate outreach materials to national actors
- Gather and maintain usage statistics for outreach and project activities

NA4 activity will define a comprehensive Public Dissemination Plan and specific activities that will be carried out to reach:

- National public services, policy-makers and governments
- Industry actors in the Earth Sciences, and in Computing and Information technologies
- Domain scientists, as well as members of the broader earth sciences communities
- Mass media

Description of work

- A first step will define the target users and public; and produce a work plan for the dissemination and public outreach strategy of VERCE, defining the possible approach to address the target users. Each target audience has their own information requirements, and it is essential that VERCE project information material be provided that appropriately targets the specific audience groups.

- A second step will be to perform the dissemination and public outreach according to the work plan. This include:

- ✓ Education tools and material: support and disseminate a number of educational tools for a new generation of Earth scientists to raise awareness about the wealth of information contained in the European integrated data resources. This involves the production of simple tailored interfaces integrated within the scientific gateway. They should combine: demonstrator educational applications, basic tools and services. This dissemination through the VERCE scientific gateway and the EMSC and ORFEUS web portals will promote VERCE to a broad community.

- ✓ Internal and external communication tools: support in coordination with SA3 the dissemination of on-line training courses and a various number of training and support material through the scientific gateway; support a number of dissemination channels including a coordinated project web site, internet news feeds (RSS, Twitter, etc.) of updates and new results, newsletters, and leaflets and other publicity items distributed at technical and domain conferences and workshops and through affiliate institutions. In coordination with NA3 activities, design and provide a range of promotional material at different levels: non-technical public, policy-makers, and targeted and broader scientific communities. These resources will be available on the project web site and distributed via affiliate institutions and at conferences and workshops.

- ✓ Dissemination events planning and organisation: support, prioritise and plan stands to relevant external events and conferences in different scientific areas during the whole project extension; support dissemination activities, in the framework of the major European e-infrastructures.

- ✓ Indicators and metrics: define and implement metrics on all the on-line tools to characterise the number of users and their geographical origin. Publications in media will also be accounted for, as well as the number of printed materials produced and distributed.

- ✓ Industry and policy-makers relations: establish bridges between the technologies of the project and major industry across Europe drawing on a number of industry forums initiatives associated with the European e-infrastructures providers; improve awareness among the industry, in particular the oil and exploration industries, promoting them to participate and producing technologies transfers.

The activities in the second step could be carried out in parallel.

Deliverables (brief description) and month of delivery

D-NA4.2 (PM6) – Project website and internal information tools for SA3
D-NA4.3.1 (PM12), D-NA4.3.2 (PM24), D-NA4.3.3 (PM36), D-NA4.3.3 (PM48) Progress reports of the dissemination and public outreach activities and potential revision of the work plan

Milestones

D-NA4.1 (PM6) – Dissemination and Public outreach work plan

1.4.6 - NA efforts for the full duration of the project

Participant short name	NA1	NA2	NA3	NA4	Total person months
CNRS	**30**	7	4		41
UEDIN			4	4	8
KNMI			10	9	19
EMSC				**15**	15
INGV		**15**	4		19
LMU		7	4		11
ULIV		7	15		22
BADW-LZR			4		4
SCAI					
CINECA			4		4
Total:	30	36	49	28	

1.4.7 - List of the NA milestones

Milestone number	Milestone name	WP	date	Means of verify.
M-NA3.1	Training survey report and work plan	NA3	PM6	R, PU
M-NA2.1	Prioritisation of Pilot applications and use cases	NA2	PM6	R,PU
M-NA1.1.1	Dissemination and Public outreach work plan	NA1	PM12	R, PU
M-NA3.4.1	Companion guides, best practice guides, training material	NA3	PM18	R, PU
M-NA2.2	Demonstration case of the scientific exploitation of a pilot application on the VERCE platform	NA2	PM18	D,PU
M-NA1.1.2	Annual project review Report	NA1	PM24	R, PU
M-NA3.4.2	Companion guides, best practice guides, training material	NA3	PM30	R, PU
M-NA1.1.3	Annual project review Report	NA1	PM36	R, PU
M-NA3.3	Training and user documentation final toolkit	NA3	PM42	R, PU
M-NA1.1.4	Annual project review Report	NA1	PM48	R, PU

1.5 - Service activities and associated work plan

1.5.1 - *Overall strategy and general description*

The services activities will exploit to the full the facilities of the evolving ecosystem of e-Infrastructures, including three existing European infrastructure core services:

1. the European Grid Infrastructure (EGI and NGIs), providing middleware and core services included in the Uniformed Middleware Distribution (UMD) based on gLite, UNICORE, Globus, and ARC;
2. the European HPC infrastructure (PRACE and DEISA2), providing services based on the Globus Toolkit and UNICORE middleware; and
3. the European Integrated Data Archives infrastructure (EIDA) of the seismological community providing open access to data with international standards for data, metadata, and exchange formats; distributed data management systems (ArcLink) and distributed access protocols; together with a service-oriented model for data access and exchange.

The overall strategy is to enhance the core services of the e-infrastructure providers with a tailored collection of tools, services and software engines – defining the VERCE e-Science environment – that encapsulates relevant parts of these e-Infrastructures and delivers them as a research-led data oriented Platform-as-a-Service (PaaS) to the seismological community and beyond.

The service activities can be described as:

1. Providing and managing a research platform and supporting a virtual organization (VO) comprising its user and provider community. The research platform includes the additional set of distributed computational and data resources required by seismology researchers. It will be shared across the VERCE consortium and will exploit the evolving European e-infrastructures (EGI/NGIs, PRACE) and the seismological data infrastructures (EIDA). (SA1)
2. Enhancing and expanding the capabilities of the research platform to meet the priority requirements of the seismology researchers. The enhancements will include integrating services, tools and software engines for the development and optimization of data-intensive seismology research. Corresponding enhancements to the enactment gateways will deliver the new mappings onto the underpinning resources components (Grid, HPC) required to deliver the new capabilities and to exploit new e-Infrastructure facilities. The resulting collection of tools, services, software and enactment gateways yielding access to the e-Infrastructures defines the VERCE e-Science environment. (SA2)
3. Instantiating and operating incremental platform revisions to provide an improved production e-Science environment for research users and a development platform for the integration of new components by SA2 for the next incremental evolution cycle of the platform. (SA1)
4. Providing a scientific gateway – defined as a set of portal web 2.0-services – integrating web-oriented interfaces and allowing a unified access, authentication and management for the VERCE platform together with tailored interfaces for users and research-application developers. (SA3)

The service-activity strategy is to start working on application deployment as early as possible in the project. The service activity will therefore follow a twin track approach:

- The fast track adapts, integrates and deploys a minimal set of initial components defining the first release of the VERCE platform (provided by JRA2 and SA2), to enable a first set of data-intensive applications.
- The deep track, sustained throughout the project, undertakes similar work for the higher-level tools and service components that are required to realize the full and integrated VERCE platform. VERCE will meet the advanced requirements of the data-intensive applications (NA2, JRA1) through repeated cycles of top-down (JRA2, JRA1) and bottom-up (SA1, SA2, NA2) evaluation of the platform and incremental revision and releases that incorporate new services and tools pioneered by JRA2 and integrated by SA2.

The service platform will also improve data- and work-flows across the Grid and HPC components of the research platform. This is important for complex innovative data analyses; they require orchestrated data- and work-flows where the data are either the result of distributed data analyses or of HPC simulations.

The service activities will draw heavily on the implementation of the ADMIRE software, which can be deployed rapidly using OGSA-DAI, and the MAPPER architecture through UEDIN and BADW-LRZ, closely cooperating with the MAPPER partner LMU.

To avoid duplicated effort and facilitate interdisciplinary research and sharing, the service activities will influence and draw on:

- The infrastructure core services providers, eased by the participation of UEDIN, BADW-LRZ and CINECA in PRACE; of BADW-LRZ, UEDIN, SCAI and CNRS in EGI/NGIs and EGI-InSPIRE; and of KNMI, EMSC as European consortia of the international seismological data infrastructure;
- Projects like IGE and OGSA-DAI, eased by the active contribution of UEDIN, BADW-LRZ in those projects.
- Other European infrastructure projects like EUDAT and ESFRI/ENVRI – if funded – as well as D4Science-II, SIENA, gSLM, and GENESI-DC.

The service activities are organised as follow:

- SA1 will manage the VERCE VO and the associated underlying distributed computational and data VERCE platform. It deploys – with support from SA2 – successive versions of the VERCE platform, and operates them as a production platform for researchers and as a development testbed for SA2 developing the next release. SA1 organizes and delivers specialized workbenches for the use cases developed by NA2 and JRA1, and for the training activities in collaboration with NA3 and SA3. The specialisation is achieved by integrating and tailoring existing services and by establishing more advanced capabilities as services. SA1 supplies technical documentation to SA3 and NA3.
- SA2 integrates the services, tools, software engines and enactment gateways provided by JRA2 on the development testbed provided by SA1. It then evaluates with JRA2, JRA1, SA1 and NA2 the proposed release to ensure it is fit for purpose. These integrated workbenches draw on the underlying layer of core services – provided by the Data, Grid and HPC infrastructures – together with other technologies emerging from contemporary developments. SA2 manages the releases of those services that are research-ready into the current platform version. When released, the new version is delivered to SA1 for deployment on the operational platform. SA2 supplies technical documentation to SA1 and JRA2.
- SA3 integrates – with the support of SA2 and SA1 – and operates the scientific gateway – a set of portal web 2.0 services, downloadable clients and virtualisation mechanisms – with various interfaces to ease the use and the access to the functionalities of the testbed platform for the researchers of the developers of the consortium and the users beyond. SA3 supports the user community by providing

Web 2.0 services to encourage method sharing and collaborative development, unified management and access to the VERCE platform. SA3 supplies technical documentation and tailored interfaces based on the use cases to NA3 and NA4.

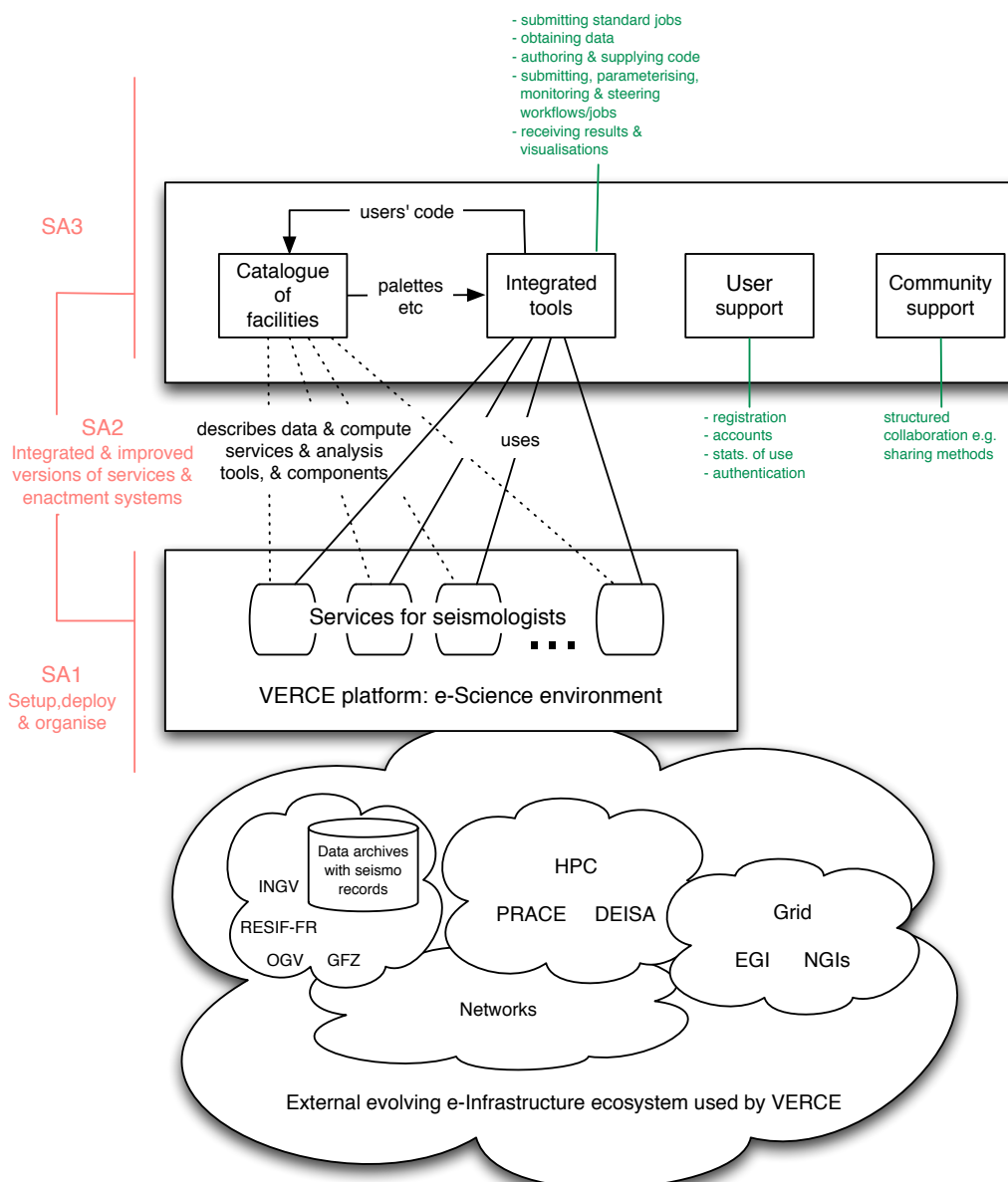Figure 4 Services activities within the e-science environment of VERCE

SA1: Management of operation of the research platform

The main tasks of the SA1 activity are:

The main tasks of the SA1 activity are:
- Integrate and manage the underlying distributed data and computational resources of the VERCE platform provided to the consortium using the services provided by the European e-infrastructures.

- Deploy across the distributed data and computational resources successive versions of the platform, provided by SA2, and deliver it as a research platform to the consortium with specialized workbenches.
- Operate the research platform, exploiting the European e-infrastructures central core services
- Define and provide a deployment testbed to SA2 for the assembly and testing of incremental versions of the platform.
- Provide temporary tailored testbed platforms to JRA1 for specific scientific use cases. They are based on a temporary agreement in place for that research, e.g. the use of a particular HPC facility, that cannot be made generally available – the resource was allocated to a particular research team and the arrangement is not sustained after the end of that team's allocation.
- Deployment of the applications on the research platform with NA2 and JRA1.
- Supply technical documentation to NA3, SA3, JRA1 and JRA2

Distributed data and computational resources

Large-scale, distributed data-analysis applications imply complex data integration flows, which require efficient distributed computing and storage facilities with mechanisms to bring computation as close as possible to the data. Large-scale modelling (simulation and inversion) implies front-end HPC capabilities with efficient I/O and memory fabrics and high bandwidth.

The VERCE framework is built upon a diversity of distributed computational and data resources – integrated within the European Grid and HPC e-infrastructures – provided by the partners of the VERCE consortium and shared in the consortium.

The data and computing resources will include:
- High value computational and data resources existing in the seismological partners' sites of the consortium which are already or will become nodes of the EGI/NGIs infrastructure, and that are locally managed,
- Additional Grid resources of the EGI/NGIs infrastructure are available to VERCE consortium via the Earth Sciences Virtual Community of Research (VCR),
- Access to HPC computational resources will be provided by the HPC centres of the VERCE consortium (LRZ, UEDIN-EPCC, CINECA).
- Data resources provided by the international seismological distributed data archives infrastructure (EIDA, FDSN-IRIS) will be provided through the European ORFEUS/KNMI consortium.

Early in the project lifetime the resource usage policy will be agreed by consensus by all the partners of the project based on the existing policies of the e-infrastructures.

These resources should ideally be operated as an integrated system along the lines developed by PRACE and EGI. However the core services legacy solutions, for historical and practical reasons, require dividing these resources into components.
- The Grid resources operated in the EGI/NGIs framework are mainly based on the gLite middleware as well as UNICORE, ARC and Globus. These resources are installed and administered locally at the partner's nodes. Central operation management services are provided through different instance levels like EGI/NGIs and VCRs. Support is provided centrally by the VCRs and EGI/NGIs, and locally at the sites.
- The HPC resources are operated by the national and European HPC centres. Access and support are provided at the national HPC centres through national resource allocation mechanisms as well as the new allocation mechanism currently established by the PRACE AISBL. PRACE is providing core services based on the Globus

Toolkit and UNICORE for sharing distributed HPC resources and supporting scientific research communities.
- The European Seismological Data Resources are already integrated at the European level (EIDA). They are organized into a service-oriented model providing – through the European consortia ORFEUS and EMSC – a comprehensive data access, data exploration and transfer. Integration into the Grid infrastructure requires installation and operation of Open Grid Services Architecture-Data Access (OGSA-DAI), as well as improved data transportation mechanisms.

The approach adopted in the SA1 work package emphasizes an inclusion strategy within the VERCE framework that leads to more inter-working and holistic, coherent science environment, which will shape the practical tasks.

SA1 will develop, with the HPC partners of the consortium, a set of tailored activities for HPC resources to:
- Secure access to the HPC resources through the commitment of HPC centres of the VERCE consortium. These centres are all actively contributing to PRACE.
- Receive CPU cycles on the HPC resources, through the commitment of the HPC partners (BADW/LRZ, EPCC and CINECA) of the consortium to provide some of those resources, and through the established resource allocation mechanisms and initiatives for complementary resources.
- Provide HPC support to the application and platform development and deployment, as well as the testbed provisioning through the commitment of the HPC .

SA1 will in parallel develop a set of tailored activities for the Grid resources of the testbed:
- Include the Grid resources within the EGI/NGIs framework by installing of core services oriented to providing access by publishing the available services, assuring common operation, including a consistent monitoring and accounting, providing a common interface to other components at the respective sites.
- Provide a Grid Appliance server, defined as a single server running all services needed by a local site to connect its resources. This will draw on the developments of the project EUFORIA and the expertise of BADW-LRZ and UEDIN. This virtual-ization of the service machines required to join the EGI – like Computing Element and Storage Elements, BDII and Monitoring Box – will definitely help the partners.

Virtual Organization and Virtual Research Community

The VERCE Virtual Organization (VO) will provide a framework for membership management, registration, authorization, monitoring and authentication, which will lead to a coherent organization of the users activities on the distributed Grid and HPC environments.

The consortium includes institutes experienced in resource provision, operating, support and usage of Grid and HPC resources. In particular, SCAI is leading the Earth Science Heavy User Community (VCR) efforts of EGI-InSPIRE and is the ESR VO administrator; BADW/LRZ, CINECA and UEDIN (EPCC) are actively involved in DEISA2.

From a practical and technical point of view, the adopted strategy will draw on:
- The ESR VO: before VERCE VO exists, users can apply for ESR VO and can use existing ES resources
- A VERCE VO: at the early stage of the project, a VERCE VO - linked to the Earth Sciences VCR (EGI-InSPIRE) – will be created. The VERCE partners will commit first resources to the VO, in parallel other sites will be contacted for resources (compute and storage SEs). The local sites may run VO core services - e.g. like VOMS, VMS, AMGA (for metadata if needed), Argus Authorization Service (new in EMI), SLCS (if needed) – and additional VERCE services.

The VO management, in coordination with SA3, involves activities such as providing the interface such that the decisions made within the VO can be properly transmitted to the operation team, and implemented by the VERCE providers in a timely manner. This includes in particular interfaces for e.g. users registration and authorization, management of the resources (assign/share/change) among the different project groups, users support; storing user information and resources allocation into database provided and operated by the VO operation team.

The VO operation involves activities to ensure access to the requested resources within a reasonable timeframe (propagation of data) and software and tools required by the VO (monitoring). This implies the provision of services and tools such that the VO can use the infrastructure on demand and interfaces to update the information to e-infrastructure providers, as well the tools necessary for the individual software developments in JRA1 and JRA2 (bug tracker, repositories, etc.).

For sake of simplicity, the operational scheme can be considered initially as:

- Operate and manage a Virtual Organization. The management activity will be done in coordination with SA3, and will provide tailored interfaces to integrate within the scientific gateway. This draw on the expertise of SCAI and BADW-LRZ.
- Provide tools and services for a unified access to the VERCE research platform combining Data, Grid and HPC resources. This draw on the expertise of BADW-LRZ and UEDIN.
- Deploy, operate and manage the distributed research platform. This involves successive deployment of incremental versions of the VERCE platform of services and tools, provided by SA2 and JRA2, in coordination with the Grid resources sites and the HPC centres. This draw on the expertise of BADW-LRZ, UEDIN, SCAI and CNRS.
- Provide and manage a set of application-tailored workbenches and enactment gateways, in coordination with SA2, as part of the VERCE platform and in support to the use-case scenarios defined by NA2 and JRA1. Each workbench will support a development environment providing software engines (work and data flow).
- Provide and manage a development testbed – a tailored sub-set of the VERCE platform - on which SA2 and JRA2 develop, integrate and evaluate the next releases of the VERCE platform. The development testbed is defined in collaboration with SA2 and JRA2.

The operation and the management to the European Integrated Data Resources is already provided by the European consortia ORFEUS (KNMI) and EMSC, with the contribution of CNRS and INGV.

Several members (UEDIN, CNRS, BADW/LRZ, SCAI) have a long contribution history to the Grid and HPC European infrastructures – EGEE, EGI/NGIs, PRACE and DEISA – and middleware development and operation – gLite, ARC, Globus Toolkit, UNICORE, OGSA-DAI, OMII-UK, OMII-europe, SSI, NextGrid and BEinGRID

Others (UEDIN, LMU in cooperation with BADW-LRZ) are leading groups in the development of data-intensive platform services integrating Grid and HPC resources – through projects such as ADMIRE and MAPPER.

Mechanism and service-oriented task

For the organization purpose a hierarchy of responsibility and coordination must exist in the SA1 activity. This hierarchy is composed of two layers. The top layer contains a single platform management body named Platform Operation Team (POT) that is responsible for the

management and coordination of the platform operations. The resource providers compose the bottom layer.

The **Platform Operation Team**, lead by CNRS, BADW-LRZ and UEDIN, will coordinate all aspects of the platform operations and management. The POT will interact with other project activities to coordinate the platform operations, provide and request services, give and obtain feedback to improve the platform integration and operations quality. The POT coordinates the management services needed to operated the platform. At the lower level, this will make use of the basic services already provided and supported by the e-infrastructures.

Networking support for data centres access

In collaboration with the data centres part of the European Integrated Data resources (EIDA), SA1 will investigate specialized protocols, services and tools for data search, data access and transfer of large volume of data on distributed storage resources based on Open Services-Data Access.

SA1 will develop synergy and collaborations with NRENs and GEANT3 to explore the potential use of new hybrid model enabling lightpaths in parallel with IP-based services.

Execution plan

The activity will start by establishing the management bodies, the structure and the mandates

The objective is to start working on application deployment as early as possible in the project. The service activity will therefore deploys a minimal set of components defining, the first release of the VERCE platform, to enable a first set of data intensive applications.

First 3 months	Planning of the testbed and assessment of the service requirements in coordination with NA2, SA2, JRA1, and JRA2
And Months 3- 6	Initial set-up phase: Integration of the data and computing resources of a few pilot sites; deployment of a first version of the platform service. Set-up of a middleware repository including auto-build facilities to support site installations. Support to the first applications deployment.
Months 6-12	Full deployment of the research platform by progressively adding sites according to the plan, deployment of the second release of the platform
Months 12-48	The operation activity of the platform will work closely with SA2 and JRA1, with an incremental (6-monthly) cycle of platform releases and provisioning of a tailored development testbed platform to SA2.

Grid resources

CNRS - IPGP	**Paris**	
	Serial Resources: for central services production	
	Processors	34 CPUs
	Storage	1 TB
	Connectivity	1 Gb/s
	Metadata server	Grelc
	Cluster Resources	

	Processors	64 servers IBM x3550 (melianxx) bi-pro Intel Xeon E5420 @2.5 Ghz (12MB L2) : 512 cores 8GB RAM (PC2-5300 CL5 ECC DDR2) and disk SATA II (160 GB)
	Interconnect	Interconnect Myrinet-2000
	Scheduler	LSF
	Compilers	Portland, Pathscale, Intel
	MPI	MPI-MX
	Debugger	DDT
	QoS storage	NIAS AXIOM 500 (Pillar) 59TB to be upgraded in 2010 to 118TB
INGV	**Roma**	
	Cluster Resources	
	Processors	48 HP ProLiant DL485G2 servers, dual AMD 2378 @2.4GHz with 16GB RAM 32 HP ProLiant DL485 servers, dual AMD 2214HE @2.2GHz with 8GB RAM 512 cores Each server has a 72GB 15k SAS disk.
	Interconnect	Gigabit Ethernet + DDR Infiniband (on 32 nodes)
	Scheduler	PBS GridWorks
	Compilers	Portland, Pathscale, Intel, Absoft
	MPI	OpenMPI, Intel MPI, MPICH
	Debugger	GDB
	QoS storage	Nexsan 10TB on NFS, HP MSA2000 12TB on Lustre
CNRS - LGIT	**Grenoble**	
	Cluster Resources	
	Processors	64 servers IBM x3650 bi-pro Intel Xeon E5420 @ 2.5 GHz/1333MHz (12MB L2) with 16 GB RAM,. 512 cores
	Interconnect	Infiniband
	Scheduler	OAR
	Compilers	Intel
	MPI	openMPI
	Debugger	DDT
	QoS storage	10 Tb SAS storage (IBM EXP3000). 100Tb facility provided by the ERC-WHISPER project (2010)

HPC resources

Among the three associated centres – BADW/LRZ, CINECA and UEDIN (EPCC) – the VERCE platform will have access to an extensive range of leading edge HPC resources. In addition JRA1 will provide assistance to the VERCE users in preparing HPC resource applications proposal so as to increase their chances of acceptance to national HPC resources and European initiatives of PRACE.

UEDIN, UK (http://www.ed.ac.uk)

HECToR	Cray XT6 system: 20 XT6 cabinets (464 compute blades each with 4 nodes of 2 12-core AMD Opteron 2.1GHz Magny Cours processors. Each 12-core socket is coupled with a Cray SeaStar2 routing and communications chip, to be upgraded late-2010 to Cray Gemini interconnect. 59.4 Tb total system memory, over 360 Tflops theoretical peak performance) Cray XT5h system: XT4 component (33 XT4 cabinets comprising 3072 quad-core compute nodes of AMD 2.33 GHz Opteron processors, 2.17 GB/s point-to-point bandwidth, 6μs latency between 2 nodes); and X2 vector component (28 x 4-vector processor compute nodes, each node sharing 32 GB memory, each processor capable of 25.6 Gflops, 16 GB/s point-to-point bandwidth, average MPI latency 4.6μs) http://www.hector.ac.uk/
Eddie / Eddie Mk2	Mark1 cluster: 118 IBM x3550 servers, each with 2 Intel Harpertown quad-core processors, all nodes connected by Gigabit Ethernet network, 60 of which are also connected by SDR Infiniband network; Mark2 cluster: 128 IBM iDataPlex servers, each with 2 Intel Westmere quad-core processors, nodes connected by Gigabit Ethernet network, 32 of which are also connected by QDR Infiniband network. From 07/2011 Eddie will consist of 250 IBM iDataPlex servers each with 2 Intel Westmere six-core processors (http://www.ecdf.ed.ac.uk)
(under construction)	120 nodes (of 1 Atom, 4 GB RAM, 3*2 TB HDD and 256 GB SSD) which will be used for parts of applications that are disk I/O intensive.

CINECA, Italy (http://www.cineca.it)

HPC scientific	IBM P6-575 Infiniband Cluster (5376 IBM Power6, 4.7 GHz, 21TBytes memory, Infiniband x4 DDR, 101 Tflops peak performance) IBM BlueGene MPP (1024 nodes IBM PowerPC 0.85GHz, 4096 cores, 4TBytes memory, 14 Tflops peak performance.
	IBM PLXiDataPlexCluster, 2192 cores Intel Xeon X5570Nehalemat 2.93 GHz: (274 nodes quad core dual processors). Peak performace 24 Tflop/s, 8 Tbyte Memory, Infiniband with 4 QDR switches

BADW-LRZ, Germany (http://www.lrz-muenchen.de)

HLRB-II system	SGI Altix 4700 platform (Intel Itanium 2 dual core 1.6 GHz, 9728 cores, 39TBytes memory, 62.3 Tflops peak performance, NUMALink 4, 64GB/s bidirectional, 1-5 μs MPI latency. http://www.lrz.de/services/compute/hlrb/
RVS1 Remote Visualisation System	Sun x4600 server (8 dual-core Opterons, 128 GB RAM), 4 Nvidia Quadro FX5500 graphics cards -1GB RAM) http://www.lrz.de/services/compute/visualisation/

Linux Cluster	Cluster with 2800 cores (AMD Opteron, Intel Xeon, Nocona, Montecito, Nehalem), 7.2 TByte 20 Tflops peak performance http://www.lrz.de/services/compute/linux-cluster/

SA2: Integration and evaluation of the platform services

The main objective of the SA2 work package is to enhance and expand the capabilities of the VERCE research platform by integrating the application codes, client services and tools according to the requirements of the seismology and other earth science research communities. The required functionalities, in particular the results of the JRA1 and JRA2 efforts, will be added to the enactment gateways to fully exploit the capabilities of the appropriate e-Infrastructure facilities. To this aim, SA2 manages the release process for the VERCE platform by selecting and preparing the tools for research-production use and to support SA1 with the deployment of these tools on the e-Infrastructures.

To fulfill this objective two main activity areas within SA2 can be identified:
- Assembly of the tools, services and simulation codes and
- Management of the release process.

Although the assembly activity can be considered to be an integral part of the release process, it is described as a separate activity since it will encompass the main part of this work package.

Release management

A set of procedures for the release process will be defined and managed in SA2 to efficiently perform the required integration and evaluation activities. In order to achieve a permanent extension and improvement of the evolving VERCE platform while ensuring a continuously high quality of service the application of the methodology of the PDCA cycle recommended in the framework of IT service management in accordance to ISO/IEC 20000 seems appropriate. In sync with SA1 and SA3 SA2 will conduct release cycles with a frequency of six months.

Plan: The components to be integrated into the VERCE platform selected by JRA1 and JRA2 based on the requirements of NA2 will be analyzed in order to identify any missing functionalities which are required for using the target e-Infrastructure or the user portal. SA2 will analyse and provide functionalities not yet covered by JRA1, JRA2 or SA1 e.g. by leveraging on developments or tools (e.g. SLCS, AHE, AdHoc, GridSAFE, globus-online) from other sources like the EU projects DEISA, EGEE, ADMIRE, ViroLab, or the Open Source Community In close cooperation with the other work packages a prioritisation will be conducted in terms of impact to the current environment, importance to the scientists, complexity of the integration and availability of infrastructure services and resources in order to obtain a clear defined set of complimentary components to be deployed in the new release. For the simulation codes, the computational performance and scalability will be evaluated and classified on the VERCE resources to identify the best-suited infrastructure or resource on which to install the code. The release shall be agreed and authorized by all relevant parties, including measures to reverse or remedy the roll out if unsuccessful. Therefore a planning activity will include joint meetings with JRA1 and JRA2 as well as SA1 and SA3 to achieve a common agreement on the target enhancements of each release cycle.

Do: After the identification of the necessary integration the necessary change actions will be applied to the codes and tools. It should be noted that not only the existing tools or services provided by the European e-Infrastructures are considered, but also convenient tools developed by other partners or commonly used standards (e.g OGF standards such as SAGA or DRMAA) might be included. This will be elaborated further in the section "Tool assembly".

46

Check: The completed tool set of the release cycle is then provided to SA1 to be deployed on the development testbed provided and operated by SA1. Extensive testing is conducted in order to evaluate if the new or enhanced services provide the required functionalities with an appropriate quality. If the required quality is met, the complete set of software components is prepared for deployment on the e-Infrastructures. The deployment itself is again carried out by SA1 with the support from SA2.

Act: The services integrated into the VERCE platform are operated by SA1 in production mode. While in operation, SA1 will continuously monitor the provided services in terms of availability, stability, robustness, performance, etc. and forward this data to SA2 for smaller bug fixes realizable in the next minor release as well as a long-term evaluation indicating possible areas of enhancement for future major releases. This general service quality information will be discussed with JRA1 and JRA2 for improvement, enhancement, and/or expansion of services and will be included in the planning of subsequent release cycles. In addition, the subsequent release cycles will take full account of contemporaneous technical and scientific developments as well as scientists' priorities. By combining these two sources of intelligence in the planning cycle that drives the VERCE enhancements, we will deliver new research capabilities to VERCE researchers at the earliest possible moment.

Tool assembly

The primary flow of candidate components for incorporation in the integrated research platform will come from JRA2, which will have proposed them in response to JRA1's prioritised technical requirements list. The candidate components are then enhanced by adding supporting tools and other technology according to the requirements of the infrastructures – provided by SA1 – or the user support – provided by SA3. All components will be maintained with version control on a common software repository.
The tools to be integrated into the VERCE platform can be categorized as follows:
- data services (access, transfer and integration) and data-flow engines for seismology distributed data resources,
- compute services (numerical codes, job submission, scheduling) for seismological simulations
- services supporting software engines,
- enactment gateways and workbenches,
- unified management and access tools to the e-Infrastructures
- tools to orchestrate workflows across the VERCE platform.

Within SA2 specialized teams will be set up according to the experiences of the project partners involved, which should then take the responsibility for the respective category of tools proposed for inclusion in the next release. These teams will also provide the necessary technical documentation to SA1 as well as user related information to SA3.

In order to provide the VERCE platform already early in the project, the first release cycle (M1-M6) will be based on an initial portfolio of software components (fast track). It must provide the basic functionalities including secure user access, resource allocation, job submission, and data management.
In terms of the functionalities to be added by SA2 the following tools are currently envisioned as candidates for this initial portfolio:
- Globus Toolkit 5 (www.globus.org) for reliable data transfer, job submission and secure access/single sign on;
- UNICORE (www.unicore.eu) for job submission and workflow management;
- OGSA-DAI (www.ogsa-dai.org.uk) for distributed data access and management of data within the European e-Infrastructures and selected seismology data centres;
- ADMIRE (www.admire-project.eu) running enactment gateways on and deploying encapsulated enactment environments across these e-Infrastructures as well as seismology and Cloud infrastructures;

- SAGA (saga.cct.lsu.edu) on the primary e-Infrastructures used by VERCE so that subsequent application, tool and workbench development is portable across evolving e-Infrastructures;
- The RAPID portal generation system (research.nesc.ac.uk/node/494) deployed at sites hosting VERCE science gateways;

Since the service environments f the European and national e-Infrastructures and the international agreements and standards are rapidly evolving in this field, the above list will be reviewed and updated immediately after project start (M1). New requirements (if any) will be gathered from the seismologists and will be taken into consideration as well.

Most of the tools defined above are expected to receive support from either the partners involved and/or from EU projects. VERCE will receive strong support for the Globus components from the IGE project, which is coordinated by BADW-LRZ. Via the German Gauss Centre for Supercomputing (GCS) BADW-LRZ is closely linked to the Research Centre Jülich (FZJ), home of the UNICORE core developer group. UEDIN is leading the ADMIRE project and is strongly involved in the development of the RAPID system.

With respect to the IT service management release process
SA2 will consult the gSLM project, which is coordinated by the LMU, to carry out the IT service management release process in the most effective and efficient manner. SA2 will also work closely together with other projects which are constructing science platforms directed towards the requirements of other scientific virtual communities (e.g. the MAPPER platform developed for the Fusion, Material Science, Computational Biology, and Engineering Communities). In addition SA2 will constantly track any upcoming activities concerning the integration of the European e-Infrastructures (e.g. the EGI UMD) in order to avoid redundant project effort by quickly adjusting the VERCE platform to the developments in these areas, if possible.

Execution plan

The main objective of the SA2 work package is to integrate and evaluate the application codes, client services and tools to be offered by the VERCE platform and thereby select and prepare them for research-production use in the next release deployed by SA1. The primary flow of candidate components for incorporation in the integrated research platform will come from JRA2, who will have proposed them in response to JRA1's prioritised technical requirements list.

SA2 will be responsible for completing their adaptation to the latest operational platforms, as well as their installation on key VERCE data and computing sites. This will, for the first time, provide seismology researchers with an e-Infrastructure that spans the observatories, the seismic-data warehouses and both HTC and HPC e-Infrastructures.

Because the environment of services from European and national e-Infrastructures and the international agreements and standards are rapidly evolving in this field, the above list will be rapidly reviewed and updated in the period immediately after project start (M1).

The activity will start by establishing the release process and the product teams as well as selecting the set of tools for the first release.

In accordance with SA1 the objective is to start working on a production release as early as possible in the project. The service activity will therefore release a minimal set of components to enable a fast roll-out to the users.

First 3 months	Definition of the release management process. Establishment of product teams. Planning of the initial phase of the VERCE platform selection and the components to be integrated.
Months 3- 6	Initial integration phase: Integration of these first components of the VERCE platform with the underlying core services provided by the gLite, Globus Toolkit and UNICORE middleware. First release of the initial VERCE platform (fast track).
Months 6-12	Support to SA1 for the first deployment of the research platform and selected pilot applications. Second release cycle of the VERCE platform based of the new selected components provided by JRA1 and JRA2
Based on the Months 12-48	Continue the integration activity with an 6-monthly cycle of platform releases on the research platform. Support for the deployment and adaptation of the pilot applications and use case scenarios on the platform. Evaluate and monitor the quality of service of the different components based on the use case scenarios of the pilot applications.

SA3: Scientific gateway, user support and management

The SA3 work package will be responsible for an integrated science gateway with easily understood user interfaces, knowledge and method sharing support. It will develop and operate a scientific gateway that (a) enables seismology researchers to use all the facilities provided by the VERCE project and (b) provides a computational environment that encourages effective use of those facilities. The former will be achieved by developing an integrated view of VERCE facilities whilst providing a variety of modes of interacting with the VERCE e-Science environment to suite the full spectrum of its users. The latter will be achieved by user-management, in collaboration with user support activities provided by NA3, and by community support in conjunction with NA4.

As shown in the VERCE WPs relationships, SA3 will be the central service activity in the VERCE project, as it will be the integration point for all provided features; it provides the face of the VERCE project towards the seismology researchers, presenting all the work done in the other JRAs and SAs. SA3 will receive science requirements from NA2 and technical requirements from JRA1. Services and prototype tools will be delivered by JRA2. SA1 will provide SA3 with the access to the DCI platforms (Grid, Cloud, HPC). Results will be disseminated by NA4.

There will be four elements to VERCE's scientific gateway provided by SA3:
1. A catalogue of VERCE facilities
2. An integrated set of tools to support the full range of activities
3. A user support system
4. A community support system

These four elements are depicted in figure x VERCE-SA coherence and explained below.

1) The catalogue of VERCE facilities will allow providers to announce their facilities, such as available data sets, analysis services, etc. and allow researchers to search for and understand the new services, data, analysis components and tools. Tools will use the catalogue, built initially on the ADMIRE registry or similar technology, to populate the pallets of authoring tools and to validate the consistency of workflows and other requests.

2) An integrated set of tools to support the full range of activities, from researchers using standard service to advance applications developers engineering new services for their own or their community's use. These will include: submission of pre-packaged jobs; requests for selected data; authoring and supplying user-defined code, e.g. a particular analysis or

correlation algorithm; parameterisation, submission, monitoring and steering of workflows or jobs; receiving and visualising results; and diagnostics of analysis failures.

3) The user support system will include low-costs guest use, registration and extensive use, accounting and property ticketed bug-reporting, advice and new requirements handling.

4) The community support system will encourage data and code sharing among consenting researchers and hence facilitate more productive collaborative behaviour. As the myExperiment endeavour has shown this requires a carefully constructed mechanism for controlling the extent of sharing for each item, e.g. to no one, to my group, to this group except for this person, etc. It will also provide a well-refined and strongly encouraged mechanism for attribution or annotation by the community.

As far as possible, these four elements will reuse elements, code or systems developed elsewhere (e.g. from ADMIRE, NERIES, NERA, RapidSeis, OpenKnowledge, OpenSocial and myExperiment).

SA3 will have a balanced release schedule with mayor releases every six months. This to keep pace with the development of new facilities developed by the JRAs whilst not be too disruptive for researchers already using the facilities provided by SA3. In between the mayor releases there will be minor releases (bug fixes, etc.).

1.5.2 - *Timing of the SA work packages and their components*

Task

Year 1 | Year 2 | Year 3 | Year 4

0 Quart... Quart... Quart... Quarte... Quart... Quart... Quarte... Quarte... Quart... Quart... Quarte... Quarte... Quart... Quart... Quarte... Quarte...

- **1) SA1: Management and operation of the research platform**
  - 1.1) Set-up phase
  - 1.2) M–SA1.1: Initial resources integration
  - 1.3) Platform management, operation and deployment
  - 1.4) M–SA1.3.1: use case deployment demonstrator
  - 1.5) M–SA1.2.1: deployment platform release 1
  - 1.6) Platform, management, operation and monitoring
  - 1.7) M–SA1.2.2: deployment platform release 2
  - 1.8) Platform management, operation and monitoring
  - 1.9) M–SA1.2.3: deployment platform release 3
  - 1.10) Platform management, operation and monitoring
  - 1.11) M–SA1.2.4: deployment platform release 4
  - 1.12) Platform management, operation and monitoring
  - 1.13) M–SA1.32: Use case deployment demonstrator
  - 1.14) M–SA1.2.5: deployment platform release 5
  - 1.15) Platform management, operation and monitoring
  - 1.16) M–SA1.3.3: Use case deployment demonstrator
  - 1.17) M–SA1.2.6: deployment platform release 6
  - 1.18) Platform management, operation and monitoring
  - 1.19) M–SA1.2.7: deployment platform release 7
- **2) SA2: Integration and evaluation of the platform services**
  - 2.1) Initial platform components integration
  - 2.2) M–SA2.1: Initial platform release
  - 2.3) Evaluation, integration new platform components
  - 2.4) M–SA2.2: release platform version 1
  - 2.5) Evaluation, integration new platform components
  - 2.6) M–SA2.3: release platform version 2
  - 2.7) Evaluation, integration new platform components
  - 2.8) M–SA2.4: release platform version 3
  - 2.9) Evaluation, integration new platform components
  - 2.10) M–SA2.5: release platform version 4
  - 2.11) Evaluation, integration new platform components
  - 2.12) M–SA2.6: release platform version 2
  - 2.13) Evaluation, integration new platform components
  - 2.14) M–SA2.7: release platform version 6
  - 2.15) Evaluation, integration new platform components
  - 2.16) M–SA2.8: final platform release
- **3) SA3: Scientific gateway: user interface, management**
  - 3.1) First set-up phase:
  - 3.2) M–SA3.1: Initial scientific gateway
  - 3.3) Develop, operate, monitor the scientific gateway
  - 3.4) M–SA3.2: scientific gateway release 1
  - 3.5) Develop operate, monitor the scientific gateway
  - 3.6) M–SA3.3: scientific gateway release 2
  - 3.7) Develop, operate, monitor the scientific gateway
  - 3.8) M–SA3.4: scientific gateway release 3
  - 3.9) Develop, operate, monitor the scientific gateway
  - 3.10) M–SA3.5: scientific gateway release 4
  - 3.11) M–SA3.6: scientific gateway release 5
  - 3.12) Develop, operate, monitor the scientific gateway
  - 3.13) Develop, operate, monitor the scientific gateway
  - 3.14) M–SA3.7: scientific gateway release 6
  - 3.15) Develop, operate, monitor the scientific gateway
  - 3.16) M–SA3.8: final scientific gateway release

Risk Management

The key roles in the service activities are system engineers who will deploy the services on the existing sites resources and the HPC centres support. Lack of cooperation with the sites administrators and the HPC centres are the risks for the operational deployment of project services. Steep learning curve and data storage usability and reliability are the main risks for the service users.

No	Risk identified	Probability	Impact	Contingency plan
1	SA1: Lack of cooperation with local cluster system administrators	Low	High → DSA1.*	Several partners will offer clusters to spread risks, also several applications will be explored to ensure the project does not rely on all services in operation.
2	SA1 & SA2: Unable to get technology to work at all the	Low	High → DSA1,.*	UEDIN has experience with several alternative

	partners		and DSA3.*	sets of Grid technology. Scientific gateway will be made independent of browsers, but may request specific version if necessary.
3	SA2: Lack of cooperation with HPC system engineers	Medium	Medium → DSA2.* & DSA1.*	Several HPC centres are involved to spread risk, moreover, medium-to-large clusters are available at partner sites.
4	SA1 & SA2: Steep learning curves to adoption of technology for users	Low	High → DSA1.* & DSA3.*	Customised "canned" workflows can be provided to allow even the most novice users to run analyses.
5	SA3: Partner responsible for web portal does not deliver (the primary entry point for the Virtual Organisation)	Low	Medium → DSA3.*	Other partners have experience in the development of scientific gateways and can take over effort.

1.5.3 - *List of the service activity work packages*

WP	Work package title	Type of activity	Lead part.	Lead participant short name	P/M	Start	End
SA1	Management and operation of the research platform	SVC	1	CNRS	95	M1	M48
SA2	Integration and evaluation of the platform services	SVC	6	BADW-LRZ	81	M1	M48
SA3	Scientific gateway, user interfaces and knowledge and method sharing	SVC	3	KNMI-ORFEUS	45	M1	M48
	TOTAL				221		

1.5.4 - *List of the SA deliverables*

number	Deliverable name	WP	Dissemination	Exp. date
D-SA1.1	Initial research platform and software repository report	SA1	R, PU	PM6

D-SA2.1.1	Report-0 on the first platform components selected	SA2	R, PU	PM6
D-SA2.1.2	Report-1 on the platform components selected	SA2	R, PU	PM12
D-SA2.2.1	Report on the release management process	SA2	R, PU	PM12
D-SA3.1.1	Year-1 report on Scientific gateways services	SA3	R, PU	PM12
D-SA1.2	Research platform operation and monitoring report 1	SA1	R, PU	PM18
D-SA3.3.1	Integrated Scientific Gateway: features available	SA3	R, PU	PM18
D-SA3.4.1	Scientific Gateway: usage and user feedback	SA3	R, PU	PM18
D-SA2.1.3	Report-2 on the platform components selected	SA2	R, PU	PM18
D-SA1.3.1	Applications deployment and monitoring report	SA1	R, PU	PM24
D-SA2.2.2	Update on performance indicators and quality measurement	SA2	R, PU	PM24
D-SA2.1.4	Report-3 on the platform components selected	SA2	R, PU	PM24
D-SA3.1.2	Year-2 report on Scientific gateways services	SA3	R, PU	PM24
D-SA1.4	Report on the Grid Appliance server	SA1	R, PU	PM30
D-SA2.1.5	Report-4 on the platform components selected	SA2	R, PU	PM30
D-SA3.1.3	Year-3 report on Scientific gateways services	SA3	R, PU	PM36
D-SA1.5	Research platform operation and monitoring report 2	SA1	R, PU	PM36
D-SA2.1.6	Report-5 on the platform components selected	SA2	R, PU	PM36
D-SA3.3.2	Integrated Scientific Gateway: features available	SA3	R, PU	PM36
D-SA3.4.2	Scientific Gateway: usage and user feedback	SA3	R, PU	PM36
D-SA2.1.7	Report-6 on the platform components selected	SA2	R, PU	PM40
D-SA1.3.2	Applications deployment and monitoring report	SA1	R, PU	PM44
D-SA1.6	Final research platform operation and application monitoring	SA1	R,PU	PM48
D-SA2.1.8	Integrated platform components	SA2	R, PU	PM48
D-SA2.3	Final report on the status and achievement of SA2	SA2	R, PU	PM48
D-SA3.1.4	Report on Scientific gateways services	SA3	R, PU	PM48
D-SA3.3.3	Integrated Scientific Gateway: features available	SA3	R, PU	PM48
D-SA3.4.4	Scientific Gateway: usage and user feedback	SA3	R, PU	PM48

1.5.5 - Description of the SA work packages

SA1 work package: Management and operation of the research platform

Work package number	SA1	Start date or starting event:			M1		
Work package title	Management and operation of the research platform						
Activity type	SVC						
Participant number	1	2	3	5	7	8	9	10
Participant short name	**CNRS**	UEDIN	KNMI	ULIV	INGV	BADW-LRZ	SCAI	CINECA
Person-months per participant	25	10	7	7	7	16	8	15

Objectives

The main objectives of this work package is to:

- Operate and manage in coordination with SA3 a Virtual Organization;
- Provide tools and services for a 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.

Description of work

Task 1 - Integrate and operate the distributed data and computational resources (All)

This task deals with setting up, manage and maintaining the set of distributed data and computational resources provided by the partners of the consortium and underlying the VERCE platform. These resources combine Grid and HPC resources. HPC resources are provided by the HPC centres of the consortium and operated by them within the PRACE/DEISA2 infrastructure ecosystem; Grid resources are provided by a number of distributed high value resources within the seismological partners of the consortium. These are – or will be at the early stage – included and operated within the EGI/NGIs ecosystem. The underlying core services include components based on gLite, Globus Toolkit and UNICORE and will evolve in synergy with EMI. The task will support a pragmatic inclusion strategy within a framework that leads to more inter-working and holistic, coherent e-science environment:

- Identify, setup and integrate the computing and storage resources for the testbed (All).
- Manage the distributed data and computing resources making use of the core services provided by the evolving HPC and Grid infrastructure ecosystem with a layered hierarchy of responsibility with at its top a central Platform Management Operation (PMO) (CNRS, BADW/LRZ, SCAI).

- Operate and Manage a Virtual Organization (VO) – in synergy with the Earth Sciences VCR and the ESR (VO) - with global services like membership management, registration, authorization, monitoring and authentication (SCAI, CNRS, BADW/LRZ) – provide a Grid Appliance server.
- Define and provide tools for unified access and service provision (BADW/LRZ, UEDIN).

Task 2: Operate and manage the platform (CNRS, BADW/LRZ, UEDIN)

This task deals with the deployment of the VERCE platform of services and tools – provided by SA2 and JRA2 – and its management and operation during the project. This will be coordinated by the PMO – in conjunction with SA2 and JRA2 – and may require solutions, sourced from within the e-infrastructures (HPC, Grid) ecosystems when they are available.

This task will deliver

- a research platform to the consortium and beyond,
- a deployment testbed – a subset of the research platform – to SA2 for the development, integration and evaluation of successive incremental releases of the platform,
- tailored workbenches in support of specific use case scenarios defined by NA2 and JRA1, including development environments (software engines …), and tailored gateway services.

A first minimal version of the platform will be delivered early in the project to enable applications to be deployed rapidly.

Task 3: Support the deployment of applications and use cases (CNRS, BADW/LRZ, CINECA)

This task deals with supporting services – identified with SA2, JRA1 and JRA2 – for the adaptation and deployment of the pilot applications and the use case scenarios on the platform. For HPC applications, the support will be provided through the support services of the HPC centres; for the Grid applications, this will be supported at the level of the PMO and the local sites. This task will also contribute to the evaluation of the platform and the monitoring of the different services of the platform required by the pilot applications.

Task 4: Supply technical documentation (CNRS, BASW/LRZ, SCAI, UEDIN)

This task will supply technical documentation to NA2, NA3 and JRA1 (users and developers), and to JRA2 and SA3 (technical). To avoid duplicate task and facilitate sharing, this task will draw on the existing support services provided by the e-infrastructures providers. It will be coordinated by the PMO.

Deliverables (brief description) and month of delivery

D-SA1.1 (PM6): Initial research platform operation status report

D-SA1.2 (PM18): First research platform operation and monitoring report – document describing the experience with the platform deployment and assessment of the platform capabilities, reliability and usage.

D-SA1.3.1 (PM24): Applications deployment and monitoring report

D-SA1.4 (PM30): Report on the Grid Appliance server

D-SA1.5 (PM36): Second report on the research platform operation status report

D-SA1.3.2 (PM44): Applications deployment and monitoring report

D-SA1.7 (PM 48): Final report on the research platform use and experiences

Milestones

M-SA1.1 (PM6): Initial production service platform and software repository setup. The initial research platform will be monitored by a server installed at BADW-LRZ. The software repository will be accessible via a savannah server which will run at CNRS – IPGP

M-SA1.2 (PM24) Release of the Grid Appliance on the VERCE platform. The Grid Appliance will be released in the savannah server of the project. At least one of the core services sites (Badw-LRZ, CNRS, SCAI) will be running the service on the Grid Appliance already.

M-SA1.3.1 (PM14), M-SA1.3.2 (PM32), M-SA1.3.3 (PM44) Use case demonstration for the platform capabilities using one of the NA2 and JRA1 adapted pilot applications

M-SA1.2.1 (PM12), M-SA1.2.2 (PM18), M-SA1.2.3 (PM24), M-SA1.2.4 (PM30), M-SA1.2.5 (PM36), M-SA1.2.6 (PM42): 6-month release of the research platform deployed
M-SA1.2.7 (PM48): Final release of the research platform deployed

SA2 work package: Integration of the platform and deployment of pilot applications

Work package number	SA2	Start date or starting event:			M1			
Work package title	Integration and evaluation of the platform services							
Activity type	SVC							
Participant number	1	2	3	8	9			
Participant short name	CNRS	UEDIN	KNMI	**BADW LRZ**	SCAI			
Person-months per participant	20	18	10	25	8			

Objectives

The objective of SA2 is to integrate and evaluate the upper-level services provided for the VERCE architecture, delivered to the consortium by JRA2. Integration and evaluation of the enhanced services will be performed in close collaboration with SA1.

- Define a standard set of procedures for integration and evaluation activities (e.g., plan-check-do-act) and the appropriate software-engineering processes.
- Analyse and rank codes and tools in terms of ease to install and deploy.
- Check performance and scalability of application codes on VERCE platforms to identify the best-suited platforms for integration and deployment.
- Evaluate existing services, tool and middle-ware offered by participating infrastructures.
- Integration of essential higher-level development and optimisation tools, e.g. ADMIRE, for data exploration, data integration, workflow construction and optimisation.
- Assist SA1 to deploy application codes, client services and tools on the test-bed and finally on the production infrastructure.
- Analyse the statistical data and user feedback collected by SA1 and SA3 and assess the quality of data services and tools and perform corrective measures (including re-evaluation and re-integration) and/or enhancement or expansion of services and tools if necessary.
- Support to NA2 for the evaluation of the VERCE platform based upon the selected use cases.

Description of work (possibly broken down into tasks) and role of partners

Task 1: Coordination and management of the integration efforts

This task will start by accepting the codes and tools from JRA1 and JRA2. It will then form the teams that will take care of integration, testing and assessment of the respective tools and then coordinate the processes necessary to perform the integration. This task will coordinate the different integration teams as well as the deployment with SA1 and the feedback to NA2, JRA1 and JRA2. It will also do the overall planning to streamline the efforts done in Task2.

One of the main efforts is to ensure the adherence to the quality improvement cycle of all integration processes.

Task 2: Assembly of tools, services, and simulation codes

This task will start by accepting the codes and tools from JRA1 and JA2. It will then form the teams that will take care of integration, testing and assessment of the respective tools and then coordinate the process necessary to perform the integration.This task contains the integration teams consisting of member of all partners involved in this work package. These teams are established according to the competences and experiences of the partners and the tools in question.

The teams will add missing functionalities or interfaces, tools or services necessary for using the European e-Infrastructures or standards commonly used in these areas (e.g OGF standards such as SAGA or DRMAA). They will fill the identified gaps by appropriate measures (API or interface changes, library substitution, wrapper scripts, etc.)

Bugs and issues as listed in the bug tracking system of VERCE that relate to the integration layer of the VERCE platform will also be handled by this task.

Task 3: Testing

During the assembly the tools are already deployed on the testbed for standalone and integrated testing. The testing phase of a release is finalized by regression and stress tests. After a successful testing the codes and/or tools are prepared for deployment on the production infrastructure. In addition, support will be provided to SA1 in establishing a continuous testing and monitoring environment for the VERCE platform.

Task 4: Definition of performance indicators and assessment of the quality of service

In order to ensure a high quality of service the deployed codes and tools, which are operated and monitored by SA1, are continuously assessed and evaluated by this task. If the quality in terms of stability, robustness, performance and security are not met, corrective actions are triggered. This can happen either within SA2 for smaller adjustments (bug fixes) or with the involvement of JRA1 and JRA2 for significant changes. Task 4 will also feedback to NA2 in order to better align the VERCE platform with the requirements of the users.

Deliverables (brief description) and month of delivery

D-SA2.1: Report on the first components selected for the initial platform and its integration on the VERCE framework of resources (PM6), with update about the releases every 6 months. (PM12, 18, 24, 30, 36, 42, 48)

D-SA2.2: Report on the release management process (PM12); update on the the performance indicators and quality measurement (PM24)

D-SA2.3: Final report on the status and achievements of the SA2 activities (PM48).

Milestones

M-SA2. 1: Initial version (fast track) of the VERCE platform release (PM6)

M-SA2.2 (PM12), **M-SA2.3** (PM18), **M-SA2.4** (PM24), **M-SA2.5** (PM30), **M-SA2.6** (PM36), **M-SA2.7** (PM42), **M-SA2.8** (PM48): 6-month releases of the VERCE platform components.

SA3 work package: Scientific gateway user interface K&M sharing

Work package number	SA3	Start date or starting event:		M1				
Work package title	Scientific gateway, user interfaces and knowledge and method sharing							
Activity type	SVC							
Participant number	1	2	3					
Participant short name	UEDIN	**KNMI**	EMSC					
Person-months per participant	15	20	10					

Objectives:

SA3 develops and operates a scientific gateway that enables seismology researchers to easily use all of the facilities made available by the VERCE project and provides a computational environment that encourages effective use of those facilities. This is done by:

- Providing a catalogue of VERCE facilities
- Providing an integrated set of tools to support the full range of activities
- Providing a user support system
- Providing a community support system

Objectives:

SA3 develops and operates a scientific gateway that enables seismology researchers to easily use all of the facilities made available by the VERCE project and provides a computational environment that encourages effective use of those facilities. This is done by:

- Providing a catalogue of VERCE facilities
- Providing an integrated set of tools to support the full range of activities
- Providing a user support system
  Providing a community support system

Deliverables

D-SA3.1: Yearly reports on Scientific Gateway services (PM12, PM24, PM36, PM45)

D-SA3.3: Integrated Scientific Gateway, providing the features available (PM18, PM36, PM48)

D-SA3.4: Reports on Scientific Gateway usage and user feedback (PM18, PM36, PM48)

Milestones (brief description)

M-SA3.1.1 (PM6), M-SA3.1.2 (PM12), M-SA3.1.3 (PM18), M-SA3.1.4. (PM24), M-SA3.1.5 (PM30), M-SA3.1.6 (PM36), M-SA3.1.7) (PM42), M-SA3.1.8 (PM48): 6- month release of the scientific gateway

1.5.6 - SA efforts for the full duration of the project

Participant short name	SA1	SA2	SA3	Total person months
CNRS	**25**	20		45
UEDIN	10	18	15	43
KNMI	7	10	**20**	37
EMSC			10	10
INGV	7			7
LMU				
ULIV	7			7
BADW-LZR	16	**25**		41
SCAI	8	8		16
CINECA	15			15
Total:	95	81	45	

1.5.7 - List of the SA milestones

Milestone number	Milestone name	WP	Date	Means of verify.
M-SA2. 1	Initial version (fast track) of the VERCE platform	SA2	PM4	P, PU
M-SA1.1	Initial production platform and software repository setup.	SA1	PM6	P, PU
M-SA3.1.1	Release-1 Scientific gateway	SA3	PM6	P, PU
M-SA2.2	Release-1 of the platform integrated components	SA2	PM12	P, PU
M-SA1.2.1	Release-1 of the deployed research platform	SA1	PM12	P, PU
M-SA3.1.2	Release-2 Scientific gateway	SA3	PM12	P, PU
M-SA1.3.1	Use case demonstration for the platform capabilities	SA1	PM14	D, PU
M-SA2.3	Release-2 of the VERCE integrated components	SA2	PM18	P, PU
M-SA1.2.2	Release-2 of the deployed research platform	SA1	PM18	P, PU
M-SA3.1.3	Release-3 Scientific gateway	SA3	PM18	P, PU
M-SA2.4	Release-3 of the VERCE integrated components	SA2	PM24	P, PU
M-SA1.2.3	Release-3 of the deployed research platform	SA1	PM24	P, PU
M-SA1.2	Release of the Grid Appliance server	SA1	PM24	P, PU
M-SA3.1.4	Release-4 Scientific gateway	SA3	PM24	D, PU
M-SA2.5	Release-4 of the VERCE integrated components	SA2	PM30	P, PU
M-SA1.2.4	Release-4 of the deployed research platform	SA1	PM30	P, PU
M-SA3.1.5	Release-5 Scientific gateway	SA3	PM3	P, PU

				0	
M-SA1.3.2	Use case demonstration for the platform capabilities	SA1	PM3 2	D, PU
M-SA2.6	Release-5 of the VERCE integrated components	SA2	PM3 6	P, PU
M-SA1.2.5	Release-5 of the deployed research platform	SA1	PM3 6	P, PU
M-SA3.1.6	Release-6 Scientific gateway	SA3	PM3 6	P, PU
M-SA2.7	Release-6 of the VERCE integrated components	SA2	PM4 2	P, PU
M-SA1.2.6	Release-6 of the deployed research platform	SA1	PM4 2	P, PU
M-SA3.1.7	Release-7 Scientific gateway	SA3	PM4 2	P, PU
M-SA1.3.3	Use case demonstration for the platform capabilities	SA1	PM4 4	D, PU
M-SA2.8	Final release of the VERCE integrated components	SA2	PM4 8	P, PU
M-SA1.2.7	Final release of the deployed research platform	SA1	PM4 8	D, PU
M-SA3.1.8	Final release of the Scientific gateway	SA3	PM4 8	D, PU

1.6 - Joint Research Activities and associated work plan

1.6.1 - *Overall strategy and general description*

The JRA activities draw on the pilot data-intensive applications, and their scientific use case scenarios, provided by NA2 based on their scientific impact in the earthquake and seismology research community and beyond in the related fields of exploration geophysics seismic engineering and Earth Sciences. These data-intensive applications cover a wide range of methods: exploration and visualization of large scale distributed data resources; large scale distributed data analysis and data mining; large scale HPC data-oriented simulation, imaging and inversion that requires an effective use of the Data, Grid and HPC infrastructure capabilities.

The goal of the JRA activities is to take these applications and methods through to datascopes capabilities aligned with the research questions and practise of the community. These are the driver for the JRA enabling activities that can be divided in

1. Harnessing the methods and their software implementation to increase their scope; re-factoring to improve their mobility and maintenance; re-engineering to increase their performance; improving documentation, user interfaces and parameterisation to facilitate their sharing and adoption by a wider users community. (JRA1)
2. Harnessing greater capability in data, computation and collaboration with a framework of Data, grid and HPC resources, a research-led service-oriented architecture and platform of services and tools. (JRA2)

This tooling to work should improve the execution of these methods and their software implementation in quality, speed and scale.

The pilot applications and use cases scenarios will draw on the ERC project WHISPER for complex data analysis case and the ITN project QUEST for the HPC data modelling cases. The synergy there will be facilitated by active contribution of the seismological partners of the consortium to these projects. This strategy should facilitate the transition from proof-of-concept demonstration to dependable research e-science environment in the earthquake and seismology community and beyond in the ESFRI-EPOS community in solid Earth sciences.

An outcome of the JRA1 will be technical requirements for other WPs. For example, it will establish a prioritised list of technical issues to be addressed by the architecture and high-level services and tools that are the remit of JRA2; and of codes to be ingested by SA2, adapted and made into operational form on the platform.

In turn, JRA2 will revise the architecture definition and the service components keeping a realistic balance between wrapping and describing existing facilities, services and tools, with re-engineering codes from application and services to fit with the architecture.

The coordination between JRA1 and JRA2 will identify representational views, standard assemblies of tools, services and codes, and their descriptions that should be developed by SA3. JRA1 establishes criteria that will need to be met by SA1 before scientific goals can be met.

JRA1 and JRA2 will provide NA4 with the roadmap so that it can explain how the e-science environment is being improved and what new power that will deliver to users. That roadmap will also alert NA3 to internal and external training requirements.

Testing and Verification

In coordination with SA2, testing and validation of the adaptation of the applications and of the architecture will be a continuous effort and will make of the benchmark cases provided by NA2.

The use cases scenarios will form the baseline for measuring the performance improvements and to guarantee that the correct functionalities are maintained. This involves coordination efforts with SA2 and SA1.

This evaluation will be instrumental for the JRA1 and JRA2 developments.

Resources for HPC applications

High Performance Computing resources will be provided for the JRAs in part by the HPC centres of the consortium; in part nominally through the national resource allocation procedures of the national HPC centres. Other procedures for CPU cycles sharing will be investigated by VERCE during the project with the support of the HPC centres of the consortium.

JRA1: Harnessing data-intensive applications

The critical role of JRA1 is to take the NA2 pilot data-intensive applications methods, and their software implementation, through to 'productisation' by providing these applications with a computational harness that enables them to be used in combination with the model and architecture delivered by JRA2.

The critical role of JRA1 is to analyse the existing software and research methods, and to identify how these can be best supported and improved in terms of architecture, robustness and efficiency. By identifying patterns that recur, technical requirements can be specified. Frequently, the algorithms and computational strategies may require modification to better exploit the modern distributed computational environment that is delivered by VERCE.

A first list of potential pilot applications has been identified - and provided in section 2.2.1 and in Annex I of this document. These pilot applications span a number of methods and software implementation - from exploration of large scale distributed data resources, complex distributed analysis and data integration, and HPC data modelling (simulation imaging and inversion) - and generic of the needs of the data-intensive research challenges of the community.

To achieve this objective during the project, a generic methodology cycle can be briefly summarized as follows:

1. Analyze and prioritized a list of pilot data-intensive applications in coordination with NA2
2. Analyze, Improve and Evaluate – refactoring, reengineering and workflow translation – the pilot applications methods
3. Validate the VERCE model and architecture based on the mapping of the pilot applications and use cases scenarios implementation to the VERCE platform
4. Improve the monitoring and operational criteria of the pilot applications
5. Improve the use of the applications by the community

The outcome of JRA1's analyses will be technical requirements for other WPs.

For example, it will establish a prioritised list of requirements and technical issues to be analysed with JRA2, solution may imply a new cycle of re-engineering from the application

implementation (codes and components) to the services, or a refactoring from the application implementation to the use case scenario, to fit with the existing version of the architecture.

It will identify required presentational views, standard assemblies of services and their descriptions that should be developed by SA3. It will establish operational and monitoring criteria that will need to be met by SA1 before scientific goals can be met.

It will provide NA4 with the roadmap so that it can explain how the e-Infrastructure is being improved and what new power that will deliver to researchers. That roadmap will also alert NA3 to internal and external training requirements.

The strategy is to start as early as possible in the project, and JRA1 will support JRA2 and SA2 in the definition of the first version of the platform in terms of application requirements.

Testing and verification

Testing and evaluation is contiguous to the JRA1 activities. This requires coordination with NA2 and SA1.Testing will make use of the benchmarks and the use case scenarios provided by NA2.

Data analysis applications

Redundant data movements should be avoided and the processing platforms should have a balanced capacity for computation and data movement. A key objective in refactoring and reengineering the data-analysis applications is to optimize their data flow and to recognize the cases where their algorithm can be recast into incremental algorithms that use linear access patterns as far as possible. Another related issue for distributed data analysis is to indentify parallel speedup either through data parallelism pipeline processing or task parallelism pipeline processing.

Once the methods have been decomposed into their logical stages, it is useful to calculate the principal Ahmdahl numbers (e.g. ratio of I/O in bits to CPU cycles) and then target those parts of the computations to computer architectures that are designed to match those characteristics. These matches have to be weighted against the cost of the data movements.

JRA1 will benefit from the expertise of the DMI experts in UEDIN.

HPC modelling pilot applications

Capability computing is technologically challenging. The HPC data modelling applications require large amounts of memory and process extensive data sets. They are dependent on a bandwidth and latency fabric that connects the nodes. A critical aspect is scalability and memory complexity. Identification of memory hierarchy strategies and investigation of techniques for restructuring critical parts of the application implementation code will be quite instrumental.

A key issue is here is to have access to large-scale HPC infrastructures in order to perform such an analysis. High Performance Computing resources will be provided for the JRAs in part by the HPC centres of the consortium; in part nominally through the national resource allocation procedures of the national HPC centres. Other procedures for CPU cycles sharing will be investigated by VERCE during the project with the support of the HPC centres of the consortium.

JRA1 will benefit from the HPC expertise in the HPC centres partners in the consortium (BADW/LRZ, CINECA and EPCC).

Workflows and data flows

The data analysis and data modelling software architecture can take advantage of the exploitation of workflows and data flows systems.

Workflow engines like the Open Source Kepler workflow management system adaopted by GEOsciences Networl (GEON) project on the US, can also be exploited to couple distributed data analysis and HPC data modelling. This coupling will require a large element of coordination, structured data management and orchestration scheduling to be performed across the Grid and HPC components of the VERCE framework. Defining tailored interfaces for data interchange between two computing environment is an important issue.

Risk and Contingencies

The main risk identified is the difficulty to port some pilot applications implementation software to the VERCE platform. These software have been developed and implemented by the earthquake and seismology researchers and may require significant refactoring and reengineering.

Even with delays in the JRA2 deliveries, The pilot applications implementation software will have immediate added value to the earthquake and seismology research community.

A second risk is related to workflow orchestration management systems that can require quite evolved implementation and technology developments. Close collaborations with the GEON project in the US and the Euforia project in Europe, through some of the VERCE partners, will be important here.

A third risk is the shortage of CPUs cycles granted for the project. This must require an active coordination with the national HPC centres and an active strategy through the national application programs.

JRA2: VERCE architecture and tools for data-intensive applications

The JRA2 activities lead the development of a high-level architecture, and prototypes new distributed, High-level integration services and tools to enable new methods for seismology research. It delivers prototypes to SA2.

The main objectives are

1. Define the VERCE architecture and prototype critical components and services - The VERCE architecture will draw heavily on the ADMIRE (Advanced Data Mining and Integration for Europe) architecture and its implementation. Drawing on MAPPER and OMII-Europe experience, an analysis will determine how best to adapt ADMIRE and introduce it into the HPC and data-centre contexts of the VERCE partners.
2. Identify and adapt existing seismology data resources and analysis tools for integration with the architecture - Existing analysis tools will be wrapped as services to allow integration with other tools and to ensure that full advantage is taken of ADMIRE's advanced features such as data streaming and automatic optimisation of workflows. Progressive introduction of incremental and parallel algorithms will increase the power available to researchers.
3. Select and adapt a toolset for the development, parameterisation, extension and optimisation of scientific data-intensive workflows
4. Select and adapt technologies for the VERCE scientific gateway to facilitate adoption by the user community

JRA2 will work in coordination with JRA1. A realistic balance is needed between wrapping and describing existing facilities, services and code, with re-engineering code from applications and services to fit with the architecture also defined and revised by JRA2, but largely implemented by other WPs.

In particular, as JRA2 carries out those processes it will also clarify and refine the architecture definition, so that the WPs are better able to adopt it, e.g. describe their components in the terms necessary for the tools to facilitate use of those components in workflows and applications. To allow existing tools to be re-used protocols and transformations will be needed to arrange their dataflow and control flow.

The principal focus is on the higher-level tools and services that are typically composites of many components and core services already available within the partners of the VERCE and within the European HPC and Grid e-infrastructures consortium, or ingested via JRA1 and SA2 as described above.

There are also many projects developing software in other domains, e.g. for INSPIRE directive compliance, or in environmental organisations, such as, LIFEWATCH, DataONE and IPCC. These components and services may be relevant to seismology research. In which case, JRA2 would need to identify mechanisms to support their ingest and use. It is inconceivable that all the good software components and modules will be created within VERCE – therefore it is essential that the architecture and tools support imports when they are needed.

The initial architecture will draw heavily on the ADMIRE architecture and implementation. That is already capable of running on a variety of e-Infrastructures. A capability that is being extended by IGE.
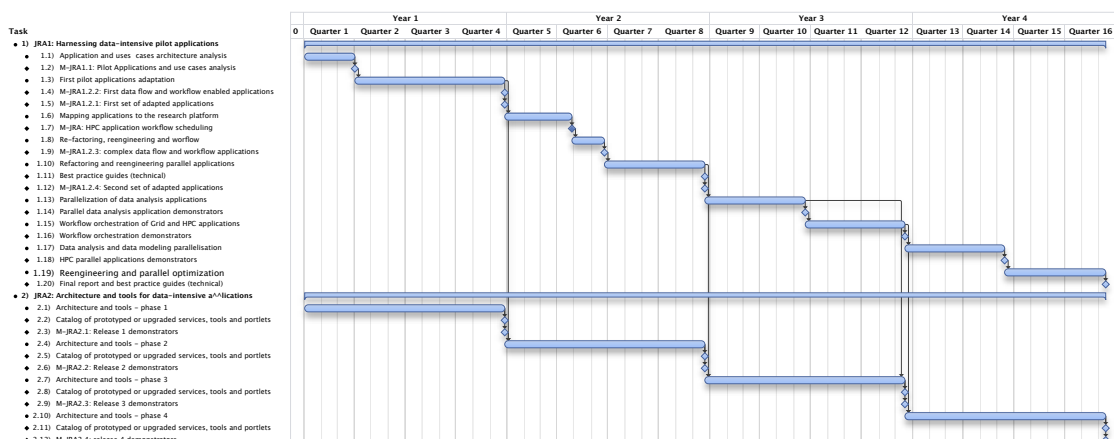
To avoid duplicated effort and to facilitate interdisciplinary research and sharing, JRA2 will make best use of existing and developing technologies of the VERCE partners, and will negotiate with and use existing and developing EU e-Infrastructures, such as those organised by EGI/NGIs, PRACE

It will be further extended by the contemporaneous EUDAT project (if that is funded) and it already provides virtual machines to exploit Cloud architectures and to facilitate computation close to the data in archival data centres and observatories. The architecture makes extensive use of existing standards and will track emerging standards to facilitate integration, e.g. into global research infrastructures and projects such as ENVRI.

UEDIN will provide the ADMIRE architecture, tools and optimisation framework. Analysis will be needed to decide how best to adapt it and introduce it into the HPC and data-centre contexts of the other partners. Drawing on MAPPER and OMII-Europe experience, an analysis will determine how best to adapt ADMIRE and introduce it into the HPC and data-centre contexts of the VERCE partners. This will be driven by a series of rapid, incremental integration steps, each conducted as a pair-wise collaborations and contained within a period of three months.

A six-monthly review and refinement of the architecture will ensure that it supports an agile evolution of the VERCE facilities and hence the rapid delivery of new capabilities to the research community. The researchers and research-application developers will be brought into the process, both to gather a better understanding of existing methods and the potential for new methods in seismology, and to harness their effort in bringing in data resources, tools and services. This will be facilitated by portals and tools presented by SA3 for them to describe and harness these components, and then to reap the benefits by using them. A critical mass will need to be established first by SA2 and SA1, in order to make it worth researchers using the system (e.g. see experience with bio-catalogue). The NA3 team will need to train the VERCE partners and then external researchers to do this work.

1.6.2 - Timing of the JRA work packages and their components

Task	Year 1				Year 2				Year 3				Year 4				
	0	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Quarter 5	Quarter 6	Quarter 7	Quarter 8	Quarter 9	Quarter 10	Quarter 11	Quarter 12	Quarter 13	Quarter 14	Quarter 15	Quarter 16

- 1) JRA1: Harnessing data-intensive pilot applications
  - 1.1) Application and uses cases architecture analysis
  - 1.2) M-JRA1.1: Pilot Applications and use cases analysis
  - 1.3) First pilot applications adaptation
  - 1.4) M-JRA1.2.2: First data flow and workflow enabled applications
  - 1.5) M-JRA1.2.1: First set of adapted applications
  - 1.6) Mapping applications to the research platform
  - 1.7) M-JRA: HPC application workflow scheduling
  - 1.8) Re-factoring, reengineering and worflow
  - 1.9) M-JRA1.2.3: complex data flow and workflow applications
  - 1.10) Refactoring and reengineering parallel applications
  - 1.11) Best practice guides (technical)
  - 1.12) M-JRA1.2.4: Second set of adapted applications
  - 1.13) Parallelization of data analysis applications
  - 1.14) Parallel data analysis application demonstrators
  - 1.15) Workflow orchestration of Grid and HPC applications
  - 1.16) Workflow orchestration demonstrators
  - 1.17) Data analysis and data modeling parallelisation
  - 1.18) HPC parallel applications demonstrators
  - 1.19) Reengineering and parallel optimization
  - 1.20) Final report and best practice guides (technical)
- 2) JRA2: Architecture and tools for data-intensive a^^lications
  - 2.1) Architecture and tools – phase 1
  - 2.2) Catalog of prototyped or upgraded services, tools and portlets
  - 2.3) M-JRA2.1: Release 1 demonstrators
  - 2.4) Architecture and tools – phase 2
  - 2.5) Catalog of prototyped or upgraded services, tools and portlets
  - 2.6) M-JRA2.2: Release 2 demonstrators
  - 2.7) Architecture and tools – phase 3
  - 2.8) Catalog of prototyped or upgraded services, tools and portlets
  - 2.9) M-JRA2.3: Release 3 demonstrators
  - 2.10) Architecture and tools – phase 4
  - 2.11) Catalog of prototyped or upgraded services, tools and portlets
  - 2.12) M-JRA2.4: release 4 demonstrators

Risk Management

 In the RTD activities we have identified the following risks. Uncertainty in how well current codes and software of the seismology community can be adapted to suit Grid and HPC computing infrastructures. The bridging of data infrastructure, grid infrastructure and HPC facilities brings risks in terms of dependencies on homogeneous software across resources, specifically software for coordination of computing and software for security and authentication. The volume of data involves (several 100s of terabytes) brings risks in terms of insufficient resources to transport data via current network connections to partners, to process all data with current computing resources and to store results on current available storage.

No	Risk identified	Probability	Impact	Contingency plan
1	JRA1 : Current codes/software unsuitable for HPC/Grid	Medium	High → DJRA1.* & DJRA2.*	Several application scenarios will be explored to spread the risk of unsuitability of current codes/software to HPC/Grid.
2	JRA2: Bridging between data and Grid, and Grid and HPC not possible	Low	Medium → DJRA1.* and DJRA2.*	Several data, Grid and HPC providers are included to allow choosing appropriate providers to suit technology. Application scenarios have two stages: 1. Data mining and 2. Data modelling; completion of stage 1 is already considered a leap forward.
3	JRA2: Data volumes to large to handle with current resources	High	High → DJRA2.*, DJRA2.* & DJRA2.*	UEDIN has 10 years experiences in grid and HPC computing and a network of contacts who have run integrated, computationally steered applications in combinations including EGEE, NG, OpenScience Grid and tera Grid. First, move data-reduction

67

					closer to data sources. Second, judiciously deploy data compression stages before data transfers. Third scale down the application scenarios. Fourth, add caching of data near computing resources. Fith, move data closer to processing via physical media transfers.
4	JRA1 & 2: Workflow orchestration tools not suitable for scientific applications	Low	Medium → DJRA1.*, & DJRA3.*	Several open source scientific workflow systems exist as possible alternatives to the proposed solution. UEDIN has experience in the use of a number of these in the e-Science community.

1.6.3 - List of the RTD work packages

WP	Work package title	Type of action	Lead	Lead participant short name	P/M	Start	End
JRA1	Harnessing data-intensive applications	SVC	6	LMU	125	M1	M48
JRA2	VERCE Architecture and Tools for Data Analysis and Data Modelling Applications	SVC	2	UEDIN	80	M1	M48
	TOTAL				200		

1.6.4 - List of the JRA deliverables

Deliver. number	Deliverable name	WP	Dissem ination	Date
D-JRA1.2.1	Annual report on architecture and software improvements of the pilot applications	JRA1	R, PU	PM12
D-JRA1.3.1	Annual report on the validation of the VERCE model and architecture	JRA1	R, PU	PM12
D-JRA2.1.1	Annual revision of the VERCE architecture.	JRA2	R, PU	PM12
D-JRA2.2.1	Annual catalogue of prototyped or upgraded services, tools and code packages that are available to SA2 and to research-application developers	JRA2	R, PU	PM12
D-JRA2.3.1	Annual catalogue of new or upgraded portlets in the scientific gateway tuned to the requirements of researchers	JRA2	R, PU	PM12
D-JRA1.2.2	Annual report on architecture and software improvements of the pilot applications	JRA1	R, PU	PM24

D-JRA1.3.2	Annual report on the validation of the VERCE model and architecture	JRA1	R, PU	PM24
D-JRA2.1.2	Annual revision of the VERCE architecture.	JRA2	R, PU	PM24
D-JRA2.2.2	Annual catalogue of prototyped or upgraded services, tools and code packages that are available to SA2 and to research-application developers	JRA2	R, PU	PM36
D-JRA2.3.2	Annual catalogue of new or upgraded portlets in the scientific gateway tuned to the requirements of researchers	JRA2	R, PU	PM36
D-JRA1.2.3	Annual report on architecture and software improvements of the pilot applications	JRA1	R, PU	PM36
D-JRA1.3.3	Annual report on the validation of the VERCE model and architecture	JRA1	R, PU	PM36
D-JRA2.1.3	Annual revision of the VERCE architecture.	JRA2	R, PU	PM36
D-JRA2.2.3	Annual catalogue of prototyped or upgraded services, tools and code packages that are available to SA2 and to research-application developers	JRA2	R, PU	PM36
D-JRA2.3.3	Annual catalogue of new or upgraded portlets in the scientific gateway tuned to the requirements of researchers	JRA2	R, PU	PM36
D-JRA1.2.4	Final report on architecture and software improvements of the pilot applications	JRA1	R, PU	PM42
D-JRA1.3.4	Final report on the validation of the VERCE model and architecture	JRA1	R, PU	PM42
D-JRA2.1.4	Annual revision of the VERCE architecture.	JRA2	R, PU	PM42
D-JRA2.2.4	Final catalogue of prototyped or upgraded services, tools and code packages that are available to SA2 and to research-application developers	JRA2	R, PU	PM42
D-JRA2.3.4	Final catalogue of new or upgraded portlets in the scientific gateway tuned to the requirements of researchers	JRA2	R, PU	PM42

1.6.5 - *Description of the JRA work packages*

Work package JRA1: Harnessing data-intensive applications

Work package number	JRA1	Start date or starting event:		M1			
Work package title	Harnessing data-intensive applications						
Activity type	RTD						
Participant number	1	2	5	6	7	8	10
Participant short name	CNRS	UEDIN	INGV	**LMU**	ULIV	BADW LRZ	CINECA
Person-months per participant	25	12	20	30	15	8	15

Objectives

The main objectives are:

- Analyze and prioritized the data-intensive pilot applications and use cases scenarios provided by NA2
- Productise the pilot applications implementation software through refactoring, reengineering and workflow optimisation
- Map the pilot applications and use cases scenarios on the platform of services and tools provided by JRA2
- Improve the evaluation and use of these applications

The role of JRA1 is to take the NA2 pilot data-intensive applications, and their software implementation, through to 'productisation' by providing these applications with a computational harness that enables them to be used in combination with the model and architecture delivered by JRA2.

It will establish a prioritised list of technical issues to be addressed by the architecture and high-level services that are the remit of JRA2. It will deliver initial analyses and lists of components and tools that should be ingested by SA2. It will identify required presentational views, standard assemblies of services and their descriptions that should be developed by SA3. It will establish operational criteria that will need to be met by SA1 before scientific goals can be met. It will provide NA4 with the roadmap so that it can explain how the e-Infrastructure is being improved and what new power that will deliver to researchers. That roadmap will also alert NA3 to internal and external training requirements.

Description of work

Task 1 - Analyze and prioritized a list of pilot data-intensive applications

Starting with the list provided by NA2, JRA1 will analyse the research methods and their

software implementation, as well as the proposed use case scenarios in technical details for the computational scientists and the service-oriented architectures experts to actually use them. Based upon their complexity and adaptation to the service-oriented architecture, JRA1 will provide a prioritized list of the pilot applications and use cases scenarios.

Task 2 - Analyze, Improve and Check the methods and their implementation

During the project, the JRA1 experts will ensure continuous number of adaptation and optimization cycle to productise the pilot applications on the platform
- *Refactoring:* identify potential re-usable data and computation oriented components that can be extracted by refactoring existing methods and software implementation; improve their interfaces.
- *Re-engineering:* indentify in these re-usable data and computation components those who need reengineering, algorithm - or data and computational strategies modifications - to improve their performance and better exploit the capabilities of the different HPC and Grid VERCE resource components of the platform;
- *Workflow development:* analyze and identify the granularity of the different data and computation process elements and of the data exchange components of the pilot applications and use cases scenarios; translate them into data- and/or work-flows; and plug them into software and dataflow engines provided by JRA2, e.g. like Kepler, the ADMIRE workflow or others.

Task 3 - Map the pilot applications and use cases scenarios implementation to the VERCE platform

In coordination with SA2, evaluate and identify deficiencies of the architecture or of the lower enacting platform when no place for an application component, or a necessary connection between components is found in the architecture

- architecture of the system including gateways and the underlying core services
- expressing and implementing the application processes as workflow on the lower enacting platform, e.g. OGSA-DAI, SAGA …
- scheduling the workflow components across Grid and HPC infrastructures.

This will provide a prioritised list of requirements and technical issues to be analysed with JRA2, solution may imply a new cycle of re-engineering from the application implementation (codes and components) to the services, or a refactoring from the application implementation to the use case scenario, to fit with the existing version of the architecture.

Task 4 - Improve the evaluation and use of the applications

- Monitoring criteria that need to be achieved by SA1 to evaluate their efficiency of the improvements;
- A set of inputs to facilitate their use by research application developers and users: application documentation (NA3, SA3), understandable user-interfaces (SA3), domain-oriented training material (NA3), outreach material in terms of achievements and roadmap (NA4);
- A set of inputs and technical documentation to SA2 to make them into operationally robust codes - and to SA1 to deploy them - on the platform;
- Operational criteria that will need to be met by SA1 to support their scientific exploitation by NA2

Deliverables (brief description) and month of delivery

D-JRA1.2.1 (PM12), D-JRA1.2.2 (PM24), D-JRA1.2.3 (PM36), D-JRA1.2.4 (PM48): Annual report on architecture and software improvements of the pilot applications
D-JRA1.3.1 (PM12), D-JRA1.3.2 (PM24), D-JRA1.3.3 (PM36), D-JRA1.3.3 (PM48): Annual report on the validation of the VERCE model and architecture

Milestones

M-JRA1.1 (PM3): Architecture analysis of the pilot application and their prioritisation
M-JRA1.3 (**PM3):** Applications requirements for the first VERCE model and architecture
M-JRA1.2.1 (PM12) First set of pilot applications adapted
M-JRA1.2.2 (PM12) Data flow and worflow enabled applications
M-JRA1.4.1 (PM12) Best Practice guide
MJRA1.3.1 (PM16) HPC workflow scheduling
MJRA1.2.3 (PM18) Complex workflow and data flow demonstrators
MJRA1.2.4 (PM24) Second set of adapted and parallelised applications
M-JRA1.2.5 (PM30) distributed data analysis and distributed analysis of data demonstrators
M-JRA1.3.2 (PM36) Workflow orchestration of Grid and HPC applications
M-JRA1.2.6 (PM42) HPC and data Parallel optimization demonstrators
M-JRA1.4.1 (PM12), M-JRA1.4.2 (PM36), M-JRA1.4.3 (PM48): Best practice guides

Work package JRA2: VERCE architecture and tools for data analysis and data modelling application

Work package number	JRA2	Start date or starting event:		M1			
Work package title	VERCE Architecture and Tools for Data Analysis and Data Modelling Applications						
Activity type	RTD						
Participant number	1	2	5	6	7	8	10
Participant short name	CNRS	**UEDIN**	KNMI	ULIV	BADW LRZ		
Person-months per participant	5	37	15	15	8		

Objectives

Based on the prioritised scientific requirements captured in NA2, and the resulting technical requirements defined in JRA1, JRA2 will lead the development of the VERCE e-Infrastructure comprising distributed, high-level integration services and tools to enable new methods for seismology research. Its objectives are to:

1. Define the VERCE architecture and prototype critical components and services
2. Identify and adapt existing seismology data resources and analysis tools for integration with the architecture
3. Select and adapt a toolset for the development, parameterisation, extension and optimisation of scientific data-intensive workflows
4. Select and adapt technologies for the VERCE scientific gateway to facilitate uptake by the user community

JRA2 will deliver prototype upper-level services and components to SA2 for evaluation, hardening and integration

Description of work

Task 1 - Define the VERCE architecture and prototype critical components and services

To avoid duplicated effort and to facilitate interdisciplinary research and sharing, JRA2 will make best use of existing and developing technologies of the VERCE partners, and will negotiate with and use existing and developing EU e-Infrastructures, such as those organised by EGI/NGIs, PRACE and the proposed EUDAT project. The architecture will also take account of technological developments and standards, particularly those adopted by relevant global research initiatives.

The VERCE architecture will draw heavily on the ADMIRE (Advanced Data Mining and Integration for Europe) architecture and its implementation. ADMIRE already facilitates computation close to the data in archival data centres and observatories, provides virtual machines to exploit Cloud architectures, and is capable of running on a variety of e-Infrastructures. OGSA-DAI (Open Grid Services Architecture-Data Access and Integration) is used globally to orchestrate multi-infrastructure distributed data access, data integration and data flow.

At the start of VERCE, ADMIRE will be operational on the Fujitsu platform and on Globus

GTK5. These are similar to the e-Infrastructures used by DEISA/HPC and the gLite e-Infrastructure coordinated by EGI. SA2 and SA2 will be responsible for completing their adaptation to the latest operational platforms, as well as their installation on key VERCE data-repository sites. This will, for the first time, provide seismology researchers with an e-Infrastructure that spans the observatories, the seismic-data warehouses and both HTC and HPC e-Infrastructures. The encapsulated virtual machine for ADMIRE can be deployed safely anywhere – it was designed to exploit Cloud infrastructures.

Enactment of scientific workflows is initiated via gateways; these will need to be deployed at strategic locations and made accessible via the facilities deployed by SA3. The function of a gateway is to validate requests and then partition and transform them to distribute the work appropriately and efficiently over the heterogeneous underlying structure. Initially, the thrust of JRA2 will be to extend the processing-element descriptions to embrace the full range of VERCE components and the optimisation strategies to cover the additional diversity of e-Infrastructure. Subsequent work in JRA2 will deliver greater integration of these e-Infrastructures and more dynamic deployment and optimisation, to yield greater control and agility to VERCE scientists.

Drawing on MAPPER and OMII-Europe experience, an analysis will determine how best to adapt ADMIRE and introduce it into the HPC and data-centre contexts of the VERCE partners. This will be driven by a carefully directed series of rapid, incremental integration steps, each conducted as a pair-wise collaboration and contained within a period of three months. A six-monthly review and refinement of the emerging VERCE architecture will ensure that it supports an agile evolution of its facilities and hence the rapid delivery of new capabilities to the research community. This architecture will draw on and technically integrate the e-Infrastructure developments in all work packages.

Task 2 - Identify and adapt existing seismology data resources and analysis tools for integration with the architecture

Existing analysis tools will be wrapped as services to allow integration with other tools and to ensure that full advantage is taken of ADMIRE's advanced features such as data streaming and automatic optimisation of workflows. Progressive introduction of incremental and parallel algorithms will increase the power available to researchers.

A key requirement is to establish protocols and plug-ins for tools that seismologists routinely use on their workstations.

VERCE partners, as members of EIDA, use agreed protocols for accessing and transporting data, based on ArcLink and NetDC. JRA2 will devise strategies for integrating these with other high-volume data protocols and standards drawing on experience with OGSA-DAI, which includes GridFTP, and DataMINX (both developed under OMII-UK and other projects) that use standardised data-transfer-request protocols for accessing and transforming bulk sets of files in data archives. We will also explore the use of CERN's Monalisa and Data Turbine.

Task 3 - Select and adapt a toolset for the development, parameterisation, extension and optimisation of scientific data-intensive workflows

A wide range of workflow management systems have been established to support the construction and management of scientific workflows, such as the Open Source Kepler workflow management system adopted in the US Geosciences Network (GEON) project, which offers graphical tools for workflow composition on HPC and Grid environments. Researchers often have their favourite tools and we anticipate a continuation of the rapid global development of new tools and workflow systems. Therefore, the VERCE e-Infrastructure will adopt standards that will facilitate the import of and integration of externally defined workflow systems and tools.

The process of optimisation of workflows starts with running basic versions of the workflows with small samples of data. We monitor all the individual steps in a workflow in terms of processing speed, and data transfer rates between steps. We identify the major bottlenecks and ameliorate them, by parallelisation and dynamic computation placement, to facilitate scale up of computational investigations. Workflows become inherently more complex to orchestrate as the distribution, scale-up and refinement progress to meet new research requirements. This is handled by the proposed mechanism as further monitoring is performed automatically. Several iterations will be applied, where each time the workflow is made more efficient by distributing work, by reducing data transport, or by balancing available compute resources with data speeds.

Task 4 - Select and adapt technologies for the VERCE scientific gateway to facilitate adoption by the user community

Once workflows have been implemented, executed and optimised, they will be made available on the VERCE scientific gateway – the community web portal. JRA2 will investigate whether the facilities of myExperiment, or at least its technology, can be reused to meet this goal.

To facilitate the visualisation of the output of these workflows we will embed existing visualisation software in the gateway. This software, SDX, is developed by ULIV and an early prototype of a web portal confirms that integration of SDX with existing portal software to drive the scientific gateway is feasible.

Each workflow will have an associated and domain-specific interface on the gateway, which will be produced using Rapid, a technology developed by UEDIN. Rapid allows fast generation of portlets that adhere to the industry standard for portal interfaces as well as to HPC and Grid standards for job execution. A prototype where this was demonstrated successfully was delivered as part of a collaboration between UEDIN, ORFEUS and ULIV.

The submitted jobs will, as is traditional, be parameterised by choices of target data, numeric controls and specification of result destinations. More innovatively, researchers will be able to supply their own code elements to perform specific steps in the workflow. This will be achieved by providing an authoring environment for researchers to define these code elements. These will then be automatically catalogued and installed in the relevant computational sites, and used when selected in the scientific workflows. Feedback during enactment will help researchers improve their code, diagnose faults and steer analyses and models. Later versions will include recovery and clean up after failures or researcher-triggered termination

---

Deliverables (brief description) and month of delivery

D-JRA2.1.1 (PM12), D-JAR2.1.2 (PM24) and D-JRA2.1.3 (PM36) – Annual revision of the VERCE architecture.

D-JRA2.2.1 (PM12), D-JAR2.2.2 (PM24), D-JRA2.2.3 (PM36) and D-JRA2.2.4 (PM48) – Annual catalogue of prototyped or upgraded services, tools and code packages that are available to SA2 and to research-application developers.

D-JRA2.3.1 (PM12), D-JAR2.3.2 (PM24), D-JRA2.3.3 (PM36) and D-JRA2.3.4 (PM48) – Annual catalogue of new or upgraded portlets in the scientific gateway tuned to the requirements of researchers.

1.6.6 - JRA efforts for the full duration of the project

Participant short name	JRA1	JRA2	Total person months
CNRS	20	25	25
UEDIN	12	**37**	49
KNMI		15	15
EMSC			
INGV	20		20
LMU	**30**		30
ULIV	15	15	7
BADW-LZR	8	8	16
SCAI			
CINECA	15		15
Total:	120	80	

1.6.7 - List of the JRA milestones

Milestone number	Milestone name	WP	Date	Means of verify.
M-JRA1.1	Architecture analysis of the pilot applications and their use cases – Prioritized list	JRA1	PM3	R, PU
M-JRA1.3	Pilot applications requirements for the initial VERCE model and architecture	JRA1	PM3	R, PU
M-JRA1.2.1	First set of pilot applications adapted	JRA1	PM12	O: code running, PU
M-JRA1.2.2	Data flow and workflow enabled applications	JRA1	PM12	O: code running, PU
M-JRA1.4.1	Best Practice guide	JRA1	PM12	R, PU
M-JRA2.1	Delivery of VERCE architecture and catalogue versions with a demonstration of their capability.	JRA2	PM12	R, D, PU
MJRA1.3.1	HPC workflow scheduling	JRA1	PM16	D, PU
MJRA1.2.3	Complex workflow and data flow demonstrators	JRA1	PM18	D, PU
MJRA1.2.4	Second set of adapted and parallelised applications	JRA1	PM24	O: code running, PU
M-JRA1.4.2	Best practice guides (technical)	JRA1	PM24	R, PU
M-JRA2.2	Delivery of VERCE architecture and catalogue versions with a demonstration of their capability.	JRA2	PM24	R, D, PU
M-	Distributed data analysis and distributed analysis	JRA1	PM3	D, PU

JRA1.2.5	of data demonstrators		0	
M-JRA1.3.2	Workflow orchestration of Grid and HPC applications	JRA1	PM36	D, PU
M-JRA1.4.2	Best practice guides (technical)	JRA1	PM36	R, PU
M-JRA2.3	Delivery of VERCE architecture and catalogue versions with a demonstration of their capability.	JRA2	PM36	R, D, PU
M-JRA1.2.6	HPC and data Parallel optimization demonstrators	JRA1	PM42	D, PU
M-JRA1.4.3	Best practice guides (technical)	JRA1	PM48	R, PU
M-JRA2.4	Delivery of VERCE architecture and catalogue versions with a demonstration of their capability.	JRA2	PM48	R, D, PU

Section 2: Implementation

2.1 - Management structure and procedures

NA1 provides administrative support and management for the VERCE project as a whole. In particular coordination and integration between the different research and service activities is an NA1 responsibility and emphasis will be put on promoting and ensuring communication within research and services activities to guarantee full integration of the activities.

The project will bring together Earthquake and Seismology scientists, data mining, Data resources, Grid and HPC Computing experts over 5 European countries. **Their common aim is to create and operate a Data and Computational infrastructure driven by the needs and priorities of the user community.**

In the one hand, the domain experts are committed to define what is being to be deployed, on the other hand the European seismological data centres and the European Grid and HPC e-infrastructure providers are liable to best respond to the demand of different communities.

VERCE will have strong synergy with EPOS, the ESFRI research infrastructure in solid Earth Sciences, as well as with a number of earthquake and seismology European projects, e.g. QUEST, WHISPER, NERA, SHARE. VERCE will also established links to the European e-infrastructure initiatives, e.g. EGI-NGIs, PRACE, EGI-Inspire. This will be greatly facilitated by the involvement of the VERCE partners in those projects. Verce will strive to develop strong synergy with related international projects, noticeably in the USA, e.g., through existing collaborations with IRIS-DMC and SCEC.

The dissemination activities are of strategic importance to introduce state-of-the-art Data-intensive research, Data, Grid and HPC infrastructure technologies and services to the earthquake and seismology community, and beyond to the solid Earth Sciences community. At the same time, the knowledge dissemination events will contribute to further disseminate the project and the benefits of the VERCE e-Science environment. This is seen as a first step in the process of bringing more new communities into the project, as well as fostering the e-infrastructure development of EPOS. An efficient coordination in the areas of dissemination (NA4) and dissemination of Knowledge (NA3) are thus needed to achieve these goals in either of the activities.

The CNRS, through the Institut de Physique du Globe de Paris (IPGP) will act as coordinating partner. UEDIN, through the National e-Sciences Centre (NeSC) will have an ITC deputy role. The coordinator will be assisted by a **Project Management Office**, including the Project Financial Manager, will be located at the CNRS-INSU. The PMO will handle administrative matters including financial and legal services for EU projects.

To harmonise the e-science environment development with the ESFRI EPOS-PP requirements, and to enforce the ties to the earthquake and seismology community at large, the coordination and management will be led by an earthquake and seismology partner together with an IT partner with a deputy role, and a steering group involving a strong technology support.

The VERCE management structure

The following instances describe the constituents of the VERCE management structure, their roles and relationship in the project:

- **A Project Steering Committee (PSG)**: has the final authority of the project,

consisting of one representative from each partner and chaired by the Project Coordinator. The MB is responsible for ensuring that the EU contract is properly executed and in accordance with a Consortium Agreement. The Consortium Agreement will cover all aspects of the relations between partners, their responsibilities, liabilities, ownership of IPR, licensing, and exploitation issues and will also address conflict resolution methods. It will be signed once the project has been selected and prior to the EU contract coming into force. The MB will meet once a year. The MB will need to monitor all the milestones in the projects and flag critical path of developments.

- **A Project Executive Board (PEB)**: is the principal executive organism of the project. It will be set up and meet remotely by video conferencing tools on a weekly basis or whenever a particular need arises. This group will be providing the technical and scientific management on a day-to-day basis. The PEB will consist of the Project Coordinator, the IT Coordinator acting also as the *architecture coordinator* (JRA2), the *enabling application coordinator* (JRA1), the *platform service coordinator* (SA1/SA2), the *outreach officer* (NA3/NA4), the *scientific gateway coordinator* (SA3) and the *user application coordinator* (NA2). The PEB will also organize quarterly face-to-face meetings, to include all the work package leaders and key personnel on the project. The location of this meeting will rotate through the participant institutes

- **User Board (UB): t**he earthquake and seismology application experts, internal to the project (NA2), will compose this board. The UB will meet remotely with the project coordinator and the ITC coordinator on a quarterly basis. The UB will be represented in the PEB with a position, The special position of UB in the management structure reflect the community-driven strategy of VERCE.

- **Advisory Board (PAB):** three international experts, external to the project, will be appointed to make up this independent board. This board will provide constructive criticism on the development of the infrastructure, its strategic direction and on the quality and relevancy of the services offered and provide advice on the adoption of best practice from any comparable infrastructure elsewhere in the world. The Advisory Board will attend the SC meeting once per year.

- **The Project Coordinator** acts as the primary point of contact with the European Commission, receives feedback on research results from each work package, ensures the project maintains effective progress towards the project objectives based on these results, produces any required project management reports, ensures that deliverables are produced according to the planned schedule and delivered to the Commission and project reviewers as required, and resolves disputes between project partners as and when these arise. He will convene regular management and technical meetings, monitor progress on each work package, collate deliverables, and maintain good contact with each site, in addition to producing the annual report.

- **Work Package Leaders (WPL)** will ensure the follow-up of the activities, keeping in direct contacts with internal and external participants. The WPL coordinate their tasks via the standard communication procedures (mailing lists, phone- or video-conferences, face-to-face meetings). WPL will report to the PEB, which will summarize on-going issues and report in turn to the PEB for approval. All essential decisions will be discussed in the PEB with particular emphasis on potential conflicts. The PSC will be the final authority within the Consortium for critical decisions concerning the Project's development.

An initial selection of the Project Executive Board is foreseen as follow:

Project coordinator: CNRS (IPGP)

Deputy ITC coordination: UEDIN (NeSc) acting also as Architecture and Framework coordinator
Dissemination and Outreach officer (including NA3 and NA4): EMSC
Enabling application Coordinator: LMU
Platform service coordinator: BADW-LRZ
Scientific Gateway coordinator: KNMI (ORFEUS)
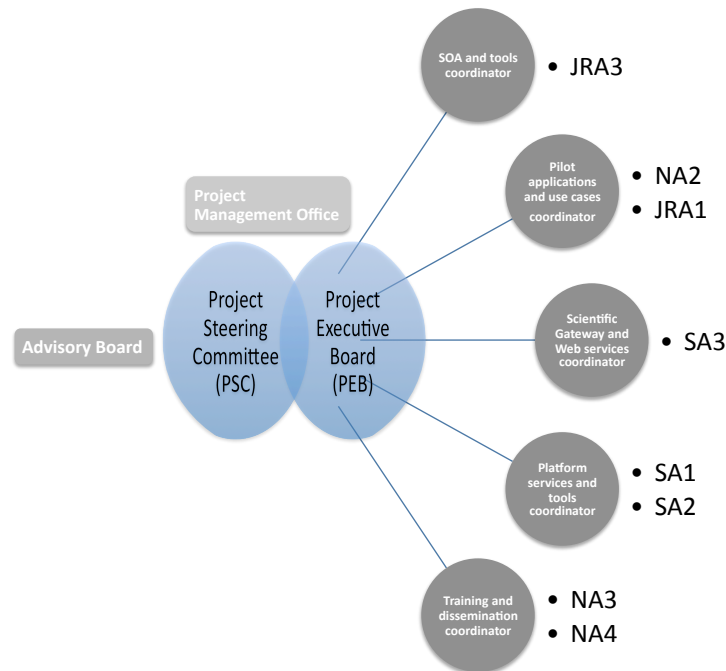User Board coordinator: INGV

Figure 5 The VERCE Management Structure

All tasks within the project are defined with an associated set of milestones and deliverables. A monitoring of the progress, reporting and deliverance of these constitutes a major part of the continuous assessment of progress.

A quarterly report will be issued reflecting the current status of the project and highlight any critical issues that need to be resolved for the continued progress.

All more software-related development tasks are done in close proximity with an entity playing a customer role (NA2). For code adaptation and optimization, the "customer" is the code developer(s) and work is done in continuous communication with the code developer(s).

As the project advances in time there will more and more dependencies between the project activities requiring horizontal monitoring and collaboration. This is reflected in the PEB composition, but ad-hoc monitoring activities will be implemented as needed across the different JRAs and SAs.

In summary the management of NA1 activity should

- Provide administrative and management services for the consortium as a whole
- Provide global coordination between and within the Joint Research and Service Activities
- Coordinate the efforts of activity NA4 (Training and user documentation) and NA3(Outreach and public dissemination)
- Promote coordination with EPOS, and the related projects, e.g. QUEST, WHISPER,

NERA, SHARE …
- Promote the coordination with European e-infrastructures: EGI-NGIs, PRACE
- Promote and extend the coordination with the HPC centres and investigate additional CPU's cycles
- Promote International coordination with related projects in Data and Computational Infrastructures initiatives, noticeable in the USA e.g. IRIS-DMC, GEON, SCEC and Data-Net, and in Japan , e.g. JAMSTEC, NIED, Earth Simulator ..
- Investigate additional sources of funding for the project through contacts with national agencies and national research programs.

Escalation procedure:
Possible issues between partners and conflicts arising from the management will be resolved by the management structure using conflict resolution rules, as described below. For issues between members of the consortium, the partners will have a means of escalating their concerns to higher level oversight, up to the PSC, if necessary.

Main management activities:
In greater detail the three main management activities are: Scientific and Technical Project Coordination;
Operational Management; Quality and Assessment

Scientific and Technical Project Coordination
Coordination of the work plan: the aim will be to ensure that the work is performed according the overall strategy; assign the required budget and effort, ensure the availability of resources, detect any potential pivotal and risk factors, the expected deliverables and milestones are accomplished with the maximum quality level and the work is implemented within the requested start date and the expected end date. This will include follow-up of agreed milestones and ensuring each participant fulfil their commitment to each WP in direct collaboration with the WPL. Activities will also ensure the implementation of decisions taken by the PSC on the overall organization of the project, informing the PSC of any risk eventuality, enabling them to implement contingency plans to ensure the activities are performed within budget, on schedule, and to the highest quality. This activity refers to the work carried out by the Project Coordinator as well as the work of other participants as far as they are involved in the PSC, the PEB and their work as leaders coordinating activities and WP. It also comprises work of the PMO for co-ordination purposes supporting the PSC, the PEB and the Project Coordinator.

Operational Management

Day-to-day management: coordinated by the Project Board and the Project Management Office, the activities will involve follow-up and monitoring project work plan and time schedule; liaison with the EC Project Officers; timely administration of project reports to be delivered by the Consortium and transmitted to the EU Commission; risk identification and assessment; contingency plans proposals to the project; communication with the Consortium; resolution of conflicts among partners, reporting of unresolved issues and conflicts, as informed by the WPL; meetings organization and meetings' minutes production; control of effort and resources spent by the Consortium, justification and adequacy to the work carried out; implementation of corrective actions; management of financial and legal issues concerning the EC Contract and the Consortium Agreement.

Periodic reporting: This will be organized the PMO in collaboration with the PSC (Coordinator and Work Package Leaders) to ensure that periodic reporting is performed in the most efficient and pragmatic way, according to Commission guidelines, and keeping all the participants of the consortia informed on matters that can influence the outcome of the project. This will entail that all participants follow an established standardized format to ensure a consistent flow of information and will guarantee that the periodic reporting to the project Steering Committee and EU Commission.

Financial and legal management: the Project Financial Manager will ensure that all budgetary actions are performed correctly and within the rules and regulations established by the European Commission and the Consortium Agreement. Monitoring the activities linked to consortium-level financial and accounting, allocation of funding received from the EU Commission and obtaining audit certificates. This includes establishing a good operating practice for financial management adapted for the financial system for each participant country, to ensure that the received funds are correctly distributed, accounted for and that financial statements and audits are received. Project Manager and all Partners will be responsible for ensuring agreement with and constant update of legal issues pertaining to the project, such as the Consortium Agreement.

Communication strategy: The definition of an effective communication strategy for the project will be done by the Dissemination and Outreach Officer, representing NA3 and NA4, and being in charge of the internal and external dissemination. The communication strategy will be based on a "key messages" agreed upon by the management structure and targeted for the selected audience (data providers, research community, public opinion, etc.) to maximize the impact of the communication and dissemination. An example of key message may be "e-Science data-intensive environment to extract new information from earthquake and seismology scientific data". A web-site and the web-services services will be integrated within the community gateway with private (consortium) and public access by SA3 in coordination with NA3-NA4. The web-site will publish updated information about the vision, objectives, and activities: meetings, training and documentation material, educational tools, publications of the project. All project documentation (whether managerial, legal or technical) will be maintained through a centralized electronic repository, and accessible to all consortium members on an open basis. Templates will be available to support the financial administration, scientific and technical communication, elaboration of the work plan and the budget. The PEB will ensure a proper communication flow between partners and an Intranet structure will be set up to support management activities, communication and exchange of information between partners.

Internal communication: The participants will be made aware of the resources available and required effort and actions needed to complete the task on time, to high quality and within budget. All participants will be informed of the assigned actions, timelines, milestones and deliverables to be factored into the work. They will also be informed of the appropriate communication channels and contingency plans to use if external adverse risk enters into the project, so that the work plan can be adjusted properly. It simultaneously permits the participating parties to access via the project web portal all information regarding the real-time status of the project and to communicate with each other all information pertinent to the evolution of the project. Templates will be available to support the financial administration, scientific and technical communication, elaboration of the work plan and the budget. Moreover the system may be used to compare delivered work and actual spending of budget with the project plan. This flow of well-managed information will contribute to performing the specific management tasks.

External communication The goals of external communication is to disseminate knowledge and good practice about the technical framework of the project, to support the user community, to facilitate cooperation with main stakeholders (scientific community, data resources providers, standardization bodies) and to present a favourable image of the scientific data infrastructure and its innovative services to potential and actual users and to society at large. A variety of channels may be used for external communication from face-to-face meetings, print media and the Internet. Part of this work will be done by NA3 and NA4. The community web portal will be public with updated information about activities: vision, objectives, meetings, publications, site descriptions, etc.. Technical reports, training material, documentation..., will be posted in the Documentation section of the web site. For publication of the work results, the PSC will ensures that no conflict is created by disagreements between partners. The access to confidential documents will be restricted to registered users. Any

publicity material produced by partners, and based on project templates, should be routed to NA4 for clearance. Communication items must include recognition of financing by the European Union. Thus, all items must include a statement such as the following in a highly visible area (e.g., press release, cover page, top or bottom of a poster): "Project financed in part by the European Commission's Seventh Framework Programme.". These items must also include the FP7 relevant official logos. Possible liaison with other scientific domains besides solid Earth Science realm, or with other communities which develops generic framework and tools will be actively explored. As it concerns the communication toward citizens it could occur by the participation in special events organized at local or national level and publication in non-specific journal to make aware the public about the Data and Computing infrastructure services and attract young people to those emerging technologies.

Knowledge Management: The compilation of all the material will be integrated into a comprehensive and accessible package. It will be available to all participating parties through the web-based and secured communication and management electronic system. The MB will prepare and monitor a series of template documents, in accordance with the scientists and the web designer to ensure that information is passed from the owner to the distributor in the most efficient way. Containing all information about the project and impacting information, such as available materials, Applications and codes, standard operating protocols, reports and publications. As the project progresses this information will be continually updated to ensure the project runs at critical mass and there is no information missing which could have an impact on the project, including data and information generated external to the project, ensuring timely use of knowledge management tools.

Intellectual property evaluation, protection and exploitation: The Outreach Officer will ensure that the capacity for the project to generate Intellectual Property will be evaluated prior to the commencement of the project and will establish an IP strategy, based on the interests of all participants

Quality and Assessment
The internal assessment of project progress is a principle management task. This activity refers to the work of the SC regarding quality assurance and assessment of the fulfilment of the objectives of the project. The results of this assessment activity will be normally in the form of proposals raised to the PEB, and subsequently implemented as part of the operational management activities. A hierarchical communication structure is put in place that ensures an efficient project progress control. This structure involves the WPL that are organized in the PEB, the Project Coordinator. WPL provide regular progress reports which have to include also concerns about potential deviations from the work plan (time schedule). The Project Coordinator, assisted by the PMO, reviews the reports with regard to the overall project objectives, and they keep the PSC informed accordingly. The essential parts of the reports concerning the progress towards the objectives will be part of the regular management reports to be provided as deliverables. In the case of deviations from or delays to the work plan, adequate correction measures will be identified by the Coordinator, assisted by the PSC. Ultimately, PSC makes the decision on the implementation of adequate correction measures.

Risk management: This activity specifically includes risk management, which comprises risk identification, evaluation and follow-up until risk is managed out or has to be implemented in the plan. It also comprises the design of appropriate contingency plans to mitigate potential negative effects (in the case of threats) or maximize potential benefits (in the case of opportunities). Regular assessment activities will have to be undertaken by the Project in order to closely follow-up the degree of achievement of objectives, as the time schedule is tight for the planned objectives, so as to be able to promptly propose strategic re-orientation or amendment of the Implementation Plan.

Review process: Milestones and Deliverables will be reviewed. The review process for documents will be as follows: the document will be reviewed by at least two people not involved in writing the document, if this is not possible within the project team, external

reviewers will be used. The reviewers will write review reports and send this to the authors of the document. The authors will answer to the reviewers' questions and remarks and change the deliverable document accordingly. The document will then be offered to the Project Board, which will approve or reject the document. If rejected, a meeting will be set-up for comments resolution where MB, reviewers and authors will discuss further improvements of the document by the authors. After improvement, the document is again offered to the PB. When approved, the document will be delivered to the EC project coordinator.

2.2 - Individual participants

CNRS - INSU

The *CNRS* (National Centre for Scientific Research) is the principal funding organization for fundamental research in natural sciences in France. CNRS carried out research in all fields of knowledge through its 8 institutes and with its two national institutes helped with a department of legal affairs of 20 people. It contributes to the application and promotion of research results, and develops scientific information. The CNRS employs about 25 000 scientists and technical and administrative research personnel in Universities and specialized CNRS research centers in France. The annual operating budget, of the order of €300 million is used for specific targeted research programmes which are defined in 4 year cycles. Total budget is about 2 500 M€.

The CNRS play a major role in the development of the national HPC and Grid infrastructures. The Institut du Développement et des Ressources Informatiques (IDRIS) of the CNRS is part of the Grand Equipement National de Calcul Intensif (GENCI) a hosting member of PRACE. IDRIS has long been also a key member of DEISAs and through GENCI. The CNRS is also leading the French NGI "France Grille" through the Institut des Grilles (CNRS-IdG).

The *National Institute for Earth Sciences and Astronomy* (INSU) as one of the two national institutes of CNRS, elaborate, develop and coordinate research and projects in astronomy, Earth and Planetary Science, including ocean and atmosphere Sciences and space Sciences. Within its field, INSU undertakes and coordinates research performed at national and international levels by CNRS and other public French research bodies. It develops, pilots and implements instruments for Very large facilities; Platforms (networks of geophysical and geodesic instruments and borehole measurements); Observatories (astronomy, seismology, oceanography, meteorology, geophysical hazards, land use); Observation services; Software facilities (databanks, community models, centres for satellite data analysis). INSU has a strong involvement in many European EC funded Projects in various fields and in infrastructure projects following national and EC roadmap. The French partners associated to the project are all associated with CNRS and INSU. For the Institut de Physique du Globe de Paris (IPGP) and the Institut des Sciences de la Terre (ISTerre) de Grenoble, the involvement in the project will be under the administrative responsibility of CNRS-INSU which is acting as coordinator of the project. INSU is also coordinating the French contribution to the EPOS ESFRI infrastructure project.

The Institut de Physique du Globe de Paris (IPGP) is a research and educational institution. Its mission is to achieve research and provide education in the fields of gesociences the institute is also in charge of monitoring seismic and volcanic hazard in the French observatories (Antilles, Réunion). The Institute conducts research in all the fields of studies of the solid Earth (geophysics, geochemistry and quantitative geology) by combining observations in land and at sea, from laboratory analysis to the construction of conceptual, analogical and numerical models. In solid Earth Sciences, the Institute play a leading role in data analysis and HPC simulations and imaging methods. The Institute is part of IdG and has been actively involved in a number of Grid projects, e.g. DataGrid, EGEEs, EELAs. IPGP together with the Institut Pierre et Simon Laplace (IPSL) maintain a Grid node of EGEE and of the National Grid. IPGP is involved in the Earth Sciences Virtual Organization of the

EGEE-III project. The Institute is also part of the European Data Archives (EIDA) infrastructure and of the ORFEUS European consortia. IPGP is participating in a number of EU projects, such as NERIES and QUEST.

Tasks in the project

IPGP-CNRS will be the coordinator of the project and work leader of SA1. CNRS-IPGP will coordinate the Management and Operation of the research platform (SA1) and is involved in the NAs and JRAs activities. IPGP, as co-PI of the WHISPER project, will provide Data analysis and also HPC simulations applications. IPGP will provide a horizontal monitoring of the activities. IPGP, as node of the French NGI "France Grille" and ins part of the CNRS IdG. IPGP will provide Grid resources and expertise. IPGP will be part of the VO management and operation.

Key personnel

Prof. Jean-Pierre Vilotte joined IPGP in 1994. He received his PhD in Earth's Sciences from the University of Montpellier in 1982. He worked for several years in the Department of Civil Engineering at the University of Swansea (Wales) and then in the Ecole Normale Supérieure de Paris (ENS ULM). In 2007, he has been invited research professor at the Earthquake Research Institute of the University of Tokyo, then in 2008, he has been visiting Miller research professor at the department of Earth and Planetary Sciences of the University of California Berkeley. His research activity is focused on 3D simulation of earthquake dynamics and wave propagation in complex media, as well as on HPC parallel computing and numerical algorithms. He has been the former head of the seismology department in IPGP and is currently leading the Computing and Data analysis centre of IPGP. He is involved in the EGEE-III project and in the French National Grid (IdG). He is participating to the ITN Quest project and will be coordinating the VERCE project.

Prof. Nikolai Shapiro joined the IPGP in 2005. He received his PhD in Earth's Sciences from Université Joseph Fourier at Grenoble in 1996. He worked during several years first in the Universidad National Autonoma de Mexico and then in the University of Colorado at Boulder where he participated in various research projects in the area of observational seismology and seismic tomography. At IPGP he assumed a position as head of the seismology department. After his arrival to France, Dr. Shapiro was awarded with a "Chaire d'Excellence" COHERSIS from the French Agence Nationale de Recherche (ANR). This grant was largely devoted to the research on methods for noise-based seismic tomography and to developing of infrastructure for intensive data mining and processing. Similar research is now continued in scope of the ERC project WHISPER for which Dr. Shapiro is a co-PI on the IPGP side. Dr. Shapiro is also leading the project FOSFORE aimed at distribution of French instrumental seismological data and is the member of the Executive Committee of ORFEUS.

The *Institut des Sciences de la Terre* (ISTerre) is part of the Grenoble Observatory (CNRS-INSU), an institution regrouping the Earth and Planetary Sciences in Grenoble. LGIT is a joint laboratory of Université Joseph Fourier, CNRS, IRD (Institut de Recherche pour le Développement) and Laboratoire Central des Ponts et Chaussées (National Laboratory for Public works). The Joseph Fourier University is a public High Education Institution in the field of science, technology and medicine. It is one of the major actors that make the city of Grenoble an internationally recognised scientific centre. International facilities such as the Synchrotron and the Langevin Institute (neutron reactor), international laboratories and research centres attract scientists from all over the world and give this scientific city a remarkable concentration of top level scientific researches. Six international schools, more than ten master courses taught in English and research internship opportunities in laboratories also contribute to a stimulating melting pot of international students and young scientists. As a result, more than one third of the 1500 PhD students at the Joseph Fourier University are international students.

With more than 50 researchers, the ISTerre offers a vast range of expertise from dynamics of the Earth's core to active tectonics and investigations of natural hazards. ISTerre has a known expertise in seismology and computational methods for wave propagation, seismic imaging and earthquake source dynamics. The 'Waves and Structures' team is composed of 7 faculties and researchers, seismologists and physicists, working on methods of imaging in seismology and acoustics. The 'Seismic Risk' team is composed of 12 faculties and researchers, seismologists and earthquake engineers, working on seismic hazard assessment with experimental and numerical methods. The ISTerre is leading the ERC project WHISPER (continuous monitoring of the continuously changing Earth), and involved in the projects NERA and SHARE. It is also involved in the European Training network QUEST and the ERASMUS Mundus MEES.

Tasks in the project

ISTerre-CNRS will participate to the NA2, SA1, JRA1 and JRA2 activities. ISTerre, as PI of the WHISPER project, will provide data analysis applications. ISTerre will also provide HPC application. ISTerre is a member of the CIMENT Grid part of the French National Grid (IdG) and will provide Grid resources and expertise. ISTerre will be a Grid node in the VERCE VO and will participate to the VO mangement.

Key Personnel

Prof. Michel Campillo, is full professor at Université Joseph Fourier and a senior member of Institut Universitaire de France. He his the head of the research group 'Ondes et Structures' (Waves and structures). He is a seismologist presently involved in different aspects of monitoring of elastic properties of the solid Earth. His research is supported by an ERC advanced grant (WHISPER-2008-adG).

Dr. Emmanuel Chaljub, obtained his PhD at IPG Paris and is a permanent researcher in the Seismic Risk team of LGIT since 2005. He has contributed to the development of the Spectral Element Method for Global Seismology and is currently working on numerical prediction of strong ground motion. He is the head of the Grenoble University High Performance Computing Centre CIMENT*, a distributed network of HPC clusters offering a total of about 1500 cores delivering a peak performance of 18 Tflops/s. CIMENT is leading an active research activity on Grid computing and operates a local grid with access to more than 2200 cores distributed in the Rhône-Alpes region.*

UEDIN

UEDIN will draw on the strengths of EPCC (70 staff), the UK National e-Science Centre (NeSC, 45 staff) and the School of Informatics (200 staff), each of which is an international research leader. UEDIN hosts the e-Science Institute and directs the UK's HPC services: HPCx (£53 million 2002-08) and HECToR (£113 million 2007-13). Through the OGSA-DAI project UEDIN brings experience of building data access and integration (DAI) services that run on five platforms and are used by over 1600 registered users worldwide, including major projects such as NAREGI in Japan, and CaBIG in the USA.

UEDIN takes leading roles in a wide range of grid and application projects:
* *EDG, EGEE I-III, EGI, ICEAGE*: led research into high-bandwidth protocols and led training;
* *NextGRID*: coordinated the project to design new data-grid architectures and services;
* *DGEMap*: design of a European developmental gene expression map;
* *FireGRID*: research into building safety and emergency response;
* *BEinGrid*: data grid technologies for commercial pilot projects;

- *OGSA-DAI, OMII-UK, OMII-Europe, SSI*: providing production quality, sustainable, distributed software to support research;
- *QCDGrid*: data grid to pool QCD simulation results;
- *eDIKT2*: distributed systems to support biomedical and chemistry research;
- *Rapid-OMERO*: computational systems for microscopy of cells;
- *Hazard Forecasting*: determining the forecasting power of brittle failure events for the earth sciences;
- *ADMIRE*: developing an architecture for the integration and analysis of distributed data;
- *DEISA, DEISA2*: integration of European HPC centres.

Tasks in the project

UEDIN participates in the project through the School of Informatics, NeSC and EPCC. It will have an ITC deputy role, coordinate the Architecture development (JRA3) and participate in all SA and JRA activities. UEDIN brings ADMIRE, DMI, HPC and GRID experiences and expertise.

Key personnel

Prof Malcolm Atkinson FRSE obtained his PhD from Cambridge. He is the UK e-Science Envoy, Director of the e-Science Institute, and UK representative at e-IRG. Leader of European projects since FP2, currently architect of the ADMIRE project.

Dr Mark Parsons obtained his PhD from the University of Edinburgh. He is the Commercial director of the EPCC and NeSC, Vice Dean for e-Research, and leader of six European projects. He is member of the Globus Management Committee.

KNMI

The Royal Netherlands Meteorological Institute (KNMI) provides data, knowledge and information in the fields of weather, climate and seismology. The KNMI is an agency under the Dutch Ministry of Transport, Public Works and Water Management with about 450 employees of which 200 are scientists with key expertise in physical and dynamical meteorology, oceanography and geophysics. A substantial part of KNMI's research is carried out in scientific national and international programmes and networks. The seismology research group is specialized in applied research, like seismic hazard, observational technologies, induced seismicity and infrasound. The group is the national information centre for earthquakes and related phenomena and hosts the ORFEUS staff and services. KNMI's climate research acts as national research and information centre for climate and climate change. Its research is aimed at providing advice to the Dutch government and society in respect to climate change impacts and thus focussed on observing, understanding and predicting the climate system. Meteorological research pursues optimal and international quality of basic meteorological observations, implement scientific and technological developments in operational practice and provide efficiently meteorological data and knowledge to a wide internal and external group of users. An important part of the KNMI is its IT infrastructure department where, among others, information and observation technology research aims at innovating KNMI's ICT infrastructure and observation technology.

The Observatories and Research Facilities for European Seismology (ORFEUS), founded in 1987, is the non-profit foundation that co-ordinates and promotes digital, broadband (BB) seismology in the European-Mediterranean area. Its governance consists of an international board of directors representing the 14 corporate founders (countries) that provide the basic funding to the foundation, an Executive Committee, controlling the daily operations, and a staff, hosted by the KNMI. ORFEUS represents currently over 65 academic and research organizations registered as paying participants from 28 different European countries.

ORFEUS core activities consist of a) Coordinating data exchange within Europe and the Mediterranean countries, including data archiving, quality control and providing advanced data access services and facilities to the scientific community. b) Coordinating activities and information dissemination aimed at promoting data access to the scientific community, including web services, workshops/meetings, and coordinating Research Infrastructure related projects. Within the EC-project NERIES, ORFEUS set up a SOA with web services and a portal in collaboration with the EMSC, providing seismological data.

ORFEUS, member of the International Federation of Digital Seismograph Networks (FDSN), operates under the auspices of the European Seismological Society (ESC). ORFEUS coordinates the Virtual European Broadband Seismic Network (VEBSN) and the European Integrated waveform Data Archive (EIDA). ORFEUS coordinates and manages the I3 project NERIES (RII3-CT-2006-026130) and coordinated the MEREDIAN (EVR1-CT2000-40007) (18 countries) and the EMICES (EVK2-CT2001-80002) (5 countries) project and participated in several earlier EC projects. ORFEUS has been a driving force behind and core element of the European Plate Observing System (EPOS) RI initiative. EPOS is currently on the ESFRI2008 roadmap.

The KNMI and ORFEUS will operate as a Joint Research Unit (JRU) within the project as the ORFEUS operational staff and the KNMI are strongly integrated, as documented in an agreement since 1997.

Tasks in the project

KNMI – ORFEUS is a European seismological consortium and is running the Community gateway. KNMI has expertise in web-services and Data management. KNMI will coordinate the SA3 activity and participate to JRA2, NA3, NA4 ans SA2. KNMI has also a long expertise in Grid management and is linked to the Dutch National Grid . KNMI will participate to SA2 and JRA2. KNMI will participate to the VO management and will be a Grid node in the VERCE VO

Key Personnel

Dr. Torild van Eck, Ph.D. is currently senior scientist at the Seismology Division of the KNMI, since 1997 Secretary General of ORFEUS and since 2007 also secretary of the Federation of Digital seismic Network (FDSN). He is author and co-author of more than 30 publications in references international journals in the fields of seismic hazard, seismo-tectonics, earthquake kinematics and dynamics and seismological infrastructure, and reviewer for international scientific journals and funding agencies. He has been PI of several large international projects, a.o. coordinator of the EC-projects MEREDIAN (EVR1-2000-40007) and EMICES (EVK2-CT2001-80002), project manager of NERIES (RII3-CT-2006-026130) and co-writer of the EPOS proposal for the ESFRI roadmap. He is coordinating the NERA project and contribute to the SHARE project. Notable achievements within these projects are the VEBSN, the EIDA and the data web services (SOA) at ORFEUS and the joint seismic portal with EMSC and numerous coordinating global and European workshops and meetings.

Wim Som de Cerff, MSc. computer science from Twente University, is a senior researcher and works at the the R&D Information and Observation Technology division of the KNMI. Wim works for 12 years now at KNMI, and will be the coordinator of the ES-G project. Wim has over 5 years experience in leading software development and research projects developing data centres (KNMI Operational Data Centre, Netherlands Sciamachy Data Center). He has participated in EU DataGrid project (IST-2000-25182), leading WP9.4 "demonstration of EO applications" and in the DEGREE (IST- 2005-5.2.5.4) project, leading the work package on science applications and requirements (WP1)

The European-Mediterranean Seismological Centre (EMSC) is a scientific non-profit NGO created in 1975 to provide rapid earthquake information at the Euro-Med and global scales. The EMSC comprised of 85 seismological institute members from 55 countries, whose scientific activities are related to seismic hazard and earthquake monitoring (alert system and Euro-Med seismological bulletin) and promoting research and collaboration within the region. The EMSC offers a unique source of information for international authorities in order to meet society's need for protection, enhance scientific progress and improve general information while also providing a back-up service to national seismological institutes.

The EMSC is the European organization responsible for aggregating and providing trans-national access to seismic event parametric data from all participating European and Mediterranean seismological networks. In addition, the EMSC gathers and distributes other diverse seismological products and macro-seismic information, including felt maps, scientific field reports, and contributed pictures and witness accounts. This seismological information is provided through several channels, including prepared seismic bulletins and publications, web search interfaces, and through online web data services.

The EMSC also provides a real-time information service on earthquakes in the Euro-Med zone and worldwide (according to predefined criteria). Earthquake information is sent via email, SMS, or fax, within 20-25 minutes from the earthquake occurrence. Over 8000 users (among them institutions such as ministries and civil protection agencies, relief teams, citizens etc) have registered to use this free service. The EMSC also collects data from observatories and institutes to publish the reference bulletin in Euro-Med seismology.

The EMSC encourages scientific cooperation in the Euro-Med zone, and is a part of major Unesco initiatives such as Reducing Earthquake Losses in the Extended Mediterranean Region (RELEMR). It leads other European projects such as Earthquake and Earthquake Risk in Western Mediterranean Region (EERWEM). The EMSC also plays a key role in the integration of the scientific community, notably through the development of major IT projects including the EC project NERA and SHARE, in collaboration with ORFEUS.

Tasks in the project

EMSC has long experience in Data management and web-services. EMSC is running the community gateway with ORFEUS. EMSC will be act as Dissemination and Public Outreach Officer. EMSC will coordinate NA4 and participate to NA3 and SA3.

Key Personnel

Dr. Rémy BOSSU, obtained is PhD in seismology from the University of Grenoble. Since 2002, he is Secretary General of the European-Mediterranean Seismological Centre, and his activities include research project initiation, definition and management (EC); fund raising; international cooperation (involved in GMES and GEOSS); evaluation of seismic monitoring capacities, and implementation of data exchanges. He has led the EC projects FP6:EERWEM and FP5:EPSI, and is on the management committee of the EC FP6:NERIES project. Prior to his role as Secretary General, Dr. Bossu worked as a Seismologist at the Laboratory for Seismic and Geophysical Risks, the Institute for Nuclear Safety, and at the Applied Seismology and Rock Physics Lab, researching seismic hazard assessment, site effects, earthquake monitoring, and micro-seismicity.

Dr. Linus Kamb, obtained a M.Sc. in Computer Sciences from the university of Utah. he is Geophysical IT Specialist and Software Project Manager at EMSC with 15 years experience in the IT sector. He is currently working on web services and portlet applications at the European-Mediterranean Seismological Centre. At his previous position, he spent 7 years working on several technology projects at the Data Management Centre of the Incorporated

Research Institutions for Seismology (IRIS-DMC), on projects ranging from specifying standard XML document structure and schema definitions; developing SOAP and REST-style web data and query services; web applications for querying data, metadata, and product archives; stand-alone data management applications; and participating in research projects, including the SCEC CME and EarthScope projects. Prior to that, he worked at a technology start-up where he designed and developed the communication protocols and graphical user interface elements for a distributed database synchronization application, and in a research group where he developed a kernel-level, distributed inter-process communication system for a research operating system.

Marie-Line Nottin, is Communication Officer in the European-Mediterranean Seismological Centre. Ms Nottin has 4 years of experience in Communications, including Corporate Communications, External Communications, and Sponsoring. Ms Nottin held communications officer positions at Textilot and Volvo Information Technology, where she was responsible for a broad range of corporate communications and publicity materials.

INGV

The *Istituto Nazionale di Geofisica e Vulcanologia* (INGV) is a governmental research Institute created in 1999 by merging different large pre-existing research institutions: the Istituto Nazionale di Geofisica (ING), the Observatorio Vesuviano (OV), the Istituto Internazionale di Vulcanologia (IIV), the Istituto di Ricerca per il Rischio Sismico (IRRS) and the Istituto per la Geochimica dei Fluidi (IGF). INGV is the reference institution for the National Civil Protection Department for the monitoring of seismic areas and active volcanoes in Italy and leads the official programs for seismic and volcanic hazard assessment promoted by the Italian government. INGV is the most relevant Italian Public Research Institution in Earth Sciences, with a particular emphasis on geohazards, and it is the reference institution for the Italian Ministry of Education, Research and University on these fields. INGV is a leading institution for fundamental research on the physics of the Earth interior, the earthquake generation process and volcano dynamics (INGV was ranked number one institution in volcanology and number three in seismology, based on scientific publications during year 2005, in a worldwide enquiry by ISI Thomson). More recently, new programs were developed and focused to study climate changes and environmental science. INGV has a long-standing experience in project coordination at national and European level, including the most recent successful EPOS (European Plate Observing System) proposal to update the European Roadmap for Research Infrastructures released by ESFRI (European Forum for Research Infrastructures). INGV has been partner or coordinator of several EC and international projects such as EXPLORIS, NERIES, SAFER, SESAME, SPICE, MEREDIAN, TRANSFER, GEOSTAR, NEAREST, QUEST, SHARE, NERA, EPOSetc.

Tasks in the project

INGV has long expertise in data analysis and data modelling applications in seismology. INGV will act as User Board coordinator. INGV will provide data analysis and data modelling applications. INGV is coordinating NA2 and participating to NA3, SA1, and JRA1. INGV will provide Grid resources to the project and will be a Grid node in the VERCE VO.

Key Personnel

Dr. Massimo Cocco is Senior Researcher, Seismologist at the Istituto Nazionale di Geofisica e Vulcanologia (INGV). He is the coordinator of the EPOS (European Plate Observing System) Preparatory Phase, included in the updated ESFRI roadmap, which aims to integrate research infrastructures on solid Earth science (www.epos-eu.org). His research activities include the analysis of strong-motion recordings, propagation of seismic waves, real-time

seismology for hazard mitigation; modelling of seismogenic processes; mechanics and interaction between active faults; dynamics of rupture propagation during large earthquakes.

Dr. Alberto Michelini: degree geology (1983, University of Trieste) and PhD in Geophysics (1991, UC Berkeley). He has been employed as researcher at the University of Trieste from 1993 to 1997, "primo ricercatore" at OGS from 1997 to 2003 and senior researcher at INGV since 2003. He is author of about 50 publications in international journals. President of the ORFEUS ExeCom since 2008.

Dr. Alessio Piatanesi: Researcher. Degree in Physics and PhD in Geophysics. He has 15 years experience in seismological research, including forward and inverse modelling of tsunami, geodetic and strong motion data for source determination. He is co-author of about 45 peer-review papers. He was involved in several EU project on tsunami reasrch (GITEC, GITEC-TWO, TRANSFER).

LMU

The *Ludwig-Maximilians-Universität* (LMU) München is one of the leading research universities in Europe, with its president, Prof. Dr. Huber, currently chairing the League of European Research Universities. LMU recently received an additional €180 million by the German state to support outstanding graduates, establish research clusters and implement a comprehensive institutional strategy through 2011. To ensure that LMU Munich maintains its high level of research, particular emphasis has been placed on promoting exceptional graduates and emerging young academics from all over the world by offering a broad variety of training options: Besides the traditional individual professorial supervision of PhD candidates, LMU Munich runs a growing number of structured doctoral programs. National and international candidates can apply for one of our multidisciplinary Research Training Groups funded by the German Research Foundation or the international Masters and doctorate programs within the Elite Network of Bavaria (LMU Geophysics runs such an elite graduate college called THESIS). In addition, LMU has very strong links within the international training networks supported by the European Commission. Since 2002 LMU Munich has been participating in 17 Research Training Groups of the Marie Curie program. Seven of those networks have been coordinated by LMU scientists. Moreover, LMU profits from Munich's exceptionally diverse academic environment – with three universities and eleven Max-Planck Institutes –, allowing them to develop even beyond their specific professional qualifications. The geophysics section of the Department of Earth Sciences has in the past years developed a local computing infrastructure with version-controlled code development, limited support staff for software support and the 170-core Linux cluster dedicated to Earth science applications. Munich Geophysics is partner of the Munich Center for Advanced Computing (MAC) teaming up with the technical university in the area of computational science for wave simulation and inversion. LMU seismology had several projects in the DEISA-DECI initiative.

Tasks in the project

LMU, as PI of the QUEST project, will be the Optimization Software coordinator and the leader of JRA1.

Key personnel

Prof. Heiner Igel is professor of seismology since 1999. He coordinated the EU-MCRTN Spice Project focusing on computational wave propagation and will coordinate the ITN QUEST as of December 2009 for four years (focus is computational solution to inverse problems). He is Executive Secretary for Theoretical Seismology of the European Geosciences Union (EGU).

Dr. Martin Käser is Emmy-Noether Scholar of the German Research Foundation and is running the research group that developed the discontinuous Galerkin method for wave propagation and dynamic rupture

ULIV

The *Department of Earth & Ocean Sciences at the University of Liverpool* is consistently rated very highly by national reviews of both teaching and research. It is a top department for geology, geophysics and oceanography, well regarded for its excellence. In the 2001 & 1996 Research Assessment Exercises conducted by the University Funding Council, the Department was awarded a Grade-5 rating, which means "Research quality that equates to attainable levels of international excellence in some sub-areas of activity and to attainable levels of national excellence in virtually all others". The Department embraces interdisciplinary approaches to the study of the Earth and its ocean and climate systems, their development in the geological past, their functioning and their interaction with human activity, and conducts both fundamental and applied research within the disciplines of Geology, Geophysics and Ocean Sciences

Key Personnel

Prof. Andreas Rietbrock MSc (Geophysics), PhD. Since graduation (1992), worked in Munich, Berlin, Paris and Kobe, on a range of seismological topics (see example publications below). At Liverpool research areas in (i) high precision relocation of subduction zone seismicity, (ii) numerical modelling of guided waves in subduction zones, (iii) earthquake hazard, (iv) data mining and real time seismology applications. He has coordinated the Joint Research Activity on "Data mining and real time seismology" in the NERIES project and has developed numerous program in Seismology for data analysis (e.g. GIANT/PITSA).

Dr Andrew Heath B.Sc. (Chemistry), Ph.D. D-o-B. He Obtained his PhD in 1989 (Crystallography, Leicester) and then worked for Silicon Graphics Inc. Came to Liverpool in 1991. He has authored several large scale software systems including TransGen, a fault flow modelling system incorporating significant 3D visualisation, which is used by several major oil companies including Shell, Petrobras and Total. He has developed a large variety of software tools with Earth Science application e.g. created the distributed visualisation system MantleVis in 2001 to image Earth's mantle within a 128 node PC-cluster. He has worked with the Seismology Group for 4 years creating a variety of tools including the seismic data analysis system SDX. Some of these tools form Liverpool's deliverables within the NERIES project. Most recently he coordinated a collaborative pilot project, RapidSeis, which successfully prototyped a web-based plug-in code system that modified the functionality of a 'virtual' processing application.

Tasks in the project

ULIV has along experience in data analysis and data visualization. ULIV will provide applications in data analysis, data exploration and visualization. ULIV has long been collaborated with UEDIN to lower barriers to uptake and to develop user oriented web-services. ULIV will be coordinating NA3 and will actively participate to JRA2, JRA1, and SA1 activities.

BADW-LRZ

The *Leibniz Supercomputing Centre* (Leibniz-Rechenzentrum, BADW-LRZ) is part of the Bavarian Academy of Sciences and Humanities (Bayerische Akademie der Wissenschaften, BADW). The BADW-LRZ is member of the Gauss Centre for Supercomputing, the alliance of the three national supercomputing centres in Germany (JSC-Jülich, HLRS-Stuttgart, LRZ-Garching) and a central site for large scale data archiving and backup. It has been an active player in the area of high performance computing for over 20 years and offers computing

power on several different levels. It is offering its services to universities in Germany and to publicly funded research institutions,like Max-Planck and Helmholtz Institutes. LRZ is operating a nearly ten thousand cores SGI Altix 4700 with a peak-performance of more than 60 TeraFlop/s as well as a large number of general purpose and specialised clusters. BADW-LRZ has more than 150 employees, about 30 of them are working in national or international research projects.

BADW-LRZ supports the adaptation of appropriate algorithms to its supercomputer architectures. This is carried out in close collaboration with international centres and research groups, especially with the Competence Network for Technical and Scientific High Performance Computing (Kompetenznetzwerk für Technisch-Wissenschaftliches Hoch- und Höchstleistungsrechnen, KONWIHR). It participates in education and research together with other research institutions in the world. BADW-LRZ operates a powerful communication infrastructure called Munich Scientific Network (MWN) and a competence centre for data communication networks. MWN connects all research institutions in the greater Munich area an offers a 10GE internet uplink. It interconnects nearly 70.000 systems and has 120.000 users. Besides that BADW-LRZ provides general IT services for the universities in Munich and for the Bavarian Academy of Sciences and Humanities (BAdW).

BADW-LRZ has a long-standing expertise in security, network technologies, IT-Management, IT-operations, data archiving, high performance- and grid computing. It is internationally known for research in these areas. BADW-LRZ is a major player in several European founded projects, like DEISA, eDEISA, DEISA2, PRACE, PRACE-1IP, EGI-InSPIRE, 100GET (100 Gigabit Ethernet Technologies), it leads the consortium of the IGE project, and participates through the German Research Network (Deutsches Forschungsnetz, DFN) in Geant, Geant2 and Geant3. BADW-LRZ is operating a Tier-2 centre for the Large Hadron Collider Computing Grid within the LCG project. It is also a member of the German NGI called D-Grid and is a member in numerous international grid consortia.

Tasks in the project

BADW-LRZ has a long-standing expertise in HPC and Grid computing research, the provisioning of HPC and Grid ressources as well as training activities in these fields. BADW-LRZ will lead the SA2 work package and will actively participate in the work packages NA3, SA1, JRA1 and JRA2

Key personnel

Prof. Hans-Joachim Bungartz holds the chair of Scientific Computing in the informatics department of Technische Universität München (TUM) and he is a member of the board of directors of the BADW-LRZ. He holds diploma degrees in mathematics and informatics and a Ph.D. in informatics (all from TUM). Prior to his current position, he had professorships in Augsburg and Stuttgart. Hans-Joachim Bungartz is director of both the Bavarian Graduate School of Computational Engineering and the Munich Centre of Advanced Computing. Furthermore, he currently serves as the head of the IT infrastructure board of the German Research Foundation (DFG), as a member of the steering committee of the Stuttgart supercomputing centre (HLRS), and as a member of the new scientific advisory board of the Gauss Centre for Supercomputing (GCS). His main fields of research are algorithmics and informatics-related aspects of scientific computing, high-performance computing (HPC), and HPC applications, in particular related to fluid flow.

Dr. Anton Frank joined the BADW-LRZ in 2004. He received his doctoral degree in computer science from Technische Universität München (TUM) for his work on efficient algorithms and coupled systems, concluding a research and teaching position at the computer science department of the TUM, where he participated in various research projects in the HPC area. At BADW-LRZ in his position as assistant group leader of the Distributed Resources Group Dr. Frank was responsible for establishing and operating a sustainable

infrastructure for grid services which included the involvement in a couple of grid projects (DEISA, PRACE, D-Grid). In addition, he is now coordinating the EU project involvement at the BADW-LRZ.

CINECA

CINECA - Consorzio Interuniversitario is a nonprofit consortium, made up of 44 Italian universities, the National Institute of Oceanography and Experimental Geophysics - OGS, the CNR (National Research Council), and the Ministry of Education, University and Research (MIUR). Established in 1969, today it is the largest Italian supercomputing centre and one of the most important worldwide with an HPC environment equipped with cutting-edge technology, the most advanced hardware resources and highly-qualified personnel which cooperate with researchers in the use of the technological infrastructure, in both the academic and industrial fields. Some of the most important Italian industries (e.g. the Italian Oil and Gas Company ENI, Toro Rosso, Centro Ricerche Fiat, etc.) use CINECA's HPC facilities. CINECA's aim is to enable researchers to use HPC systems in a profitable way, as well as to stay tuned with newest technology advances in HPC, to understand and satisfy the future users' needs of computational resources.

The present supercomputing facility is based on an integrated infrastructure of more than 15.000 processors. The state of the art HPC system is an IBM SP Power 6 supercomputer with 5376 cores and 21 Tbyte of RAM memory (168 nodes, with 32 cores and 128 GB of RAM per node). The peak performance is 101 TFlop/s. The interconnection network is based on DDR Infiniband technology and the storage accounts for 1.8 Petabytes of storage. Besides that, CINECA runs a IBM Blue Gene P with 4096 cores and a peak performance of 14 Teraflops. The data centre of CINECA is equipped to provide business continuity technology. Besides the national scientific supercomputing facility CINECA manages and exploits the supercomputing facility of ENI that is an integrated HPC facility of up to 10.000 cores. The entire infrastructure is completed by a Virtual Theatre, where viewers can experience a semi-immersive Virtual Reality experience.

HPC Group in CINECA has a long experience in cooperating with the researchers in parallelising, enabling and scaling up their applications in different computational disciplines, covering condensed matter physics, astrophysics, geophysics, chemistry, geophysics, earth sciences, engineering, mathematics and bioinformatics, but also "non-traditional" ones, such as biomedicine, archaeology and data-mining. CINECA has a wide experience in providing education and training in the different fields of parallel computing and computational sciences.

At European level CINECA has a several years deep experience, due to his involvement in several projects, financed by the EU since the 3rd FP. In FP7, CINECA coordinates HPC-Europa2, it is a main partner in DEISA2 and has been appointed to represent Italy as hosting Partner in PRACE. Moreover, CINECA is partner in SMART-LM, PLUG-IT, EUROPLANET, BEINGRID and some National and Regional projects.

Tasks in the project

CINECA has a long expertise in HPC and Grid computing. CINECA is participating in JRA1, SA1 and NA3. CINECA is one of the Italian National HPC actor.

Key Personnel

Dr. Giovanni Erbacci holds a laurea degree in Computer Science from the University of Pise. Since 1999, he heads the CINECA's Supercomputing Group, in charge for enabling and scaling HPC applications, co-operating both with academic and industrial institutions. Moreover he is responsible for the CINECA's training and education activities in the field of HPC. He is active in different EC projects since the IV FP. Actually he is in charge for the scientific activity in HPC-Europa 2, he is responsible for the DEISA Application Task Force in CINECA, as well as for coordinating the CINECA activities in PRACE 1IP related to PetaScaling computational application, HPC Ecosystem and future HPC Technologies.

94

Giovanni is the author or the co-author of more than 40 papers published in journals and conference proceedings and is a member of ACM.

Dr. Giuseppe Fiameni has a laurea degree in Computer Science from the University of Bologna (Italy). In 2004 he joins CINECA as member of the middleware and infrastructure group. Over the years he has collaborated to many important European projects maturing a strong experience in distributed systems, authentication and authorization mechanisms, computational resource management and quality assurance models definition. Currently, he is leading the Technology Watch task (WP6) in PRACE-1IP.

SCAI

The *Fraunhofer-Gesellschaft* is the largest organization for applied research in Europe. It operates more than 80 research units, including 59 Fraunhofer Institutes, at different locations in Germany and currently employs around 17.000 people, the majority of which are qualified scientists and engineers. Fraunhofer undertakes applied research of direct utility to private and public enterprise and of wide benefit to society. Fraunhofer-Gesellschaft carries out publicly funded pre-competitive research. This forms the basis of contract research projects conducted for customers. Private-sector earnings enable the organization to finance a major proportion of its budget through its own means.
Fraunhofer serves customers of the industry & service sector as well as public administration by developing technological innovations and novel systems solutions. Fraunhofer-Gesellschaft is a fixed element in the German and European research landscape and maintains close ties with other institutions engaged in fundamental research. Fraunhofer-Gesellschaft acts autonomously in defining its own strategic orientation, on which it bases its planned research activities. The Fraunhofer Institutes work in close association with one another. They form permanent alliances, such as the recent Cloud computing alliance, or pool their expertise in ad hoc interdisciplinary collaborative networks dedicated to specific projects.

The *Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)* conducts research and offers customer services in the research areas numerical software, simulation engineering, optimization, distributed computing, bioinformatics and virtual material design. At present, SCAI operates several clusters of computers with varying hardware and operating systems, including Microsoft Windows HPC based clusters, clusters with GPU-computing hardware and a number of linux based clusters of different complexity, including high-speed interconnect fabrics and different storage and data archival systems. Together with the Institute for Numerical Simulation of the University of Bonn, SCAI has recently been selected by NVIDIA to be a 2010 CUDA Research Center. The institute is experienced with several middlewares and development environments in the field of Grid computing, specifically gLite, Unicore and Globus. Projects and experiences range from the development of tools and operating of systems to the realization of complex applications and integrating service platforms. In *EGI-InSPIRE*, SCAI continues this work with its activity in the Heavy User Communities workpackage and its Operating tasks in the German NGI (*NGI-DE*). SCAI has lead the European funded Grid project *SIMDAT* and was and is a partner in other European and national funded Grid projects like *BEINGRID*, *ProGrid*, In the *PartnerGrid* project, an OSGi compliant software platform was used to offer capabilities such as a workflow engine, privilege management and a standardized interface to external compute and storage resources (e.g. Grid or compute clusters). *SmartLM is* providing generic and flexible licensing virtualization technology for new service-oriented business models across organization boundaries. At the present time, SCAI is extending its research investments in infrastructure virtualization and Cloud computing and using the competence achieved in the above mentioned projects to provide its customers solutions for next generation software services.

Tasks in the project

SCAI has long expertise in Grid and HPC computing. SCAI is leading the Earth Science Heavy User Community efforts of EGI-InSPIRE, and is the ESR VO Administrator. SCAI is Regional Operating Center for Germany & Switzerland since EGEE, and is a EGI Resource Provider

SCAI will actively participate to SA1 and SA2 activities. SCAI is one of the German National Grid actor and is contributing to the EGI operation.

Key personnel

Horst Schwichtenberg is head of the information and communication infrastructure of Fraunhofer SCAI. He received his diploma in Mathematics from the University of Bonn. His expertise in High Performance Computing and numerical simulation goes back more than twenty years. He is currently leading the activities of Fraunhofer in the European project EGI. The work of his team includes IT infrastructure management and virtualisation, service platforms, web services and all facets of distributed computing. He has been involved in a number of collaborative projects of both German and European scale about distributed infrastructures and services.

Steffen Claus received his diploma in Computer Science at the University of Bonn. His thesis analysed interactions of workflow scheduling systems with the underlying Grid infrastructure. At SCAI he is working in the area of High Performance Computing (HPC) and Software as a Service (SaaS). He developed and is maintaining SaaS solutions for optimization software on top of Windows HPC on the basis of the Windows Communication Foundation in .NET and is thus an experienced developer in the field of Web Services. Due to his activity in benchmarking simulation codes in Linux and Windows, he is experienced in application deployment and the requirements of scientific software on middle- and hardware.

Andre Gemünd studied Computer Science at the University of Bonn. He is engaged in the field of Grid computing since 2005, starting with development work on Unicore extensions. He has since successfully participated for SCAI in the general project work of the European projects EGEE-II, EGEE-III, DEGREE, BeinGRID and EGI.

2.3 - Consortium as a whole

The VERCE project bridges together world leading complementary skills in

- Data-aware distributed Grid computing and infrastructure technologies
- High Performance Computing and infrastructure technologies
- Data Integration and Data Mining architecture and framework technologies
- Service-oriented scientific gateways and web-services
- Data management and data services technologies
- Visualisation and Workflow environment and technologies

The earthquake and seismology community is represented in the project through several European leading Institutes with domain experts in data exploration, data visualisation, data analysis and data modelling. The community is also represented through a core of applications and their developers. Although not directly funded through the VERCE project, the earthquake and seismology application users form an essential part of the project and its motivation. Fundamental to e-science data and computing infrastructures is the recognition that domain application needs and requirements drive the form and function of e-infrastructure technologies and architecture. This provide the path to lower the barrier to uptake and infuse usability into the project outcomes. Usability, flexibility and resilience in the definition of the architecture and the methodologies is regarded as a strategic asset for VERCE.

The project brings together leading European HPC centres (BADW-LRZ, CINECA, EPCC) that are all committed to the project. Through the CNRS, a active collaboration with the French IDRIS NHPC is also foreseen.

The data-intensive data analysis and data modelling needs of the earthquake and seismology research community requires an efficient and resilient service-oriented data-aware distributed computing e-science environment, integrating into a single framework the Data archives, HTC and HPC infrastrures through a service-oriented architecture and an overarching platform of services and tolls. The expertise and the technologies are well established within the consortium VERCE. Most of the partners have dual or multiple expertise and skills, bringing further strengths and benefits to the e-science environment promoted in the VERCE project.

The majority of the partners have successfully worked together before in a number of European projects and are already involved in the European Grid and HPC e-infrastrures, as well as in the National Grid Infrastructures. Contacts with other FP7 projects and initiatives are well established. In the earthquake and seismology community, partners are all part of EPOS-PP project, the ESFRI initiative of the solid Earth Science community leaded by INGV. The partners are also directly involved in the ERC WHISPER, RTN Quest and the I3 SHARE and NERA projects, insuring a strong synergy with these projects and theitr community.

France	CNRS - IPGP	Long term data archiving, authenticity control, data federation, data access, VO and Grid management, authentication and authorization, Grid resources, user support, data integration, data management, data analysis and data modelling applications, Grid and HPC computing, training
	CNRS-LGIT	Grid management, Grid resources, data integration, data management, data analysis and data modelling applications, Grid and HPC computing
	EMSC	Long term data archiving, authenticity control, authentication and authorization, data federation, data access, (component) meta-data, SOA web-services, interoperability and standards, scientific gateway and portal, dissemination and public outreach
United Kingdom	UEDIN	HPC resources, Grid and HPC management, data management, authentication and authorization, user support, data access and transfer, data integration and data mining, workflows environment and process streaming, SOA web-services, interoperability and standards, code optimization, Grid and HPC computing, training
	ULIV	Discovery, data exploration and data visualisation, data analysis applications, data integration, SOA and web-services, workflows, Grid computing
,Holland	KNMI-ORFEUS	Long term data archiving, authenticity control, authentication and authorization, data federation, data access, (component) meta-data, SOA web-services, interoperability and standards, VO and Grid management, user support, scientific gateway and portal, Grid computing, dissemination and public outreach
Italy	INGV	Long term data archiving, authenticity control, data federation, data exploration and visualisation, data analysis and data modelling applications, data integration, Grid resources, Grid and HPC computing
	CINECA	HPC resources and HPC management, Grid management, user support, authentication and authorization, interoperability and standards, data management, Grid and HPC computing, HPC development environment, Code optimization, workflows,

		visualisation, training
Germany	LMU	Data modelling applications, HPC computing, code optimization, training
	BADW-LRZ	HPC resources and management, VO and Grid management, data management, authentication and authorization, user support, monitoring and accounting, interoperability and standards, SOA and web-services, Grid and HPC computing, HPC development environment, code optimization, workflows, visualization,
	SCAI	VO and Grid management, authentication and authorization, Grid and HPC computing, code optimisation, workflows, SOA web-services

From the HPC and Grid side, expertise is available to improve and adapt the existing applications, a resource that has not been available in the community before. Service-oriented data integration and data exploration of large distributed data sets discovery is of crucial importance for the data-intensive research of the community. Integration of scientific workflows within a service-oriented architecture is a new challenging and incentive perspective in the earthquake and seismology community.

Through the partners of the consortium, VERCE will drive and draw on a rich ecosstem of related projects, as briefly illustrated:

HPC:
DEISA and PRACE: UEDIN through EPCC, BADW-LRZ, CINECA

GRID:
EGI: BADW-LRZ, UEDIN, SCAI, CNRS through IPGP
EU Grid projects: SCAI, BADW-LRZ, CINECA and CNRS are all involved in EGI-Inspire
NGIs: CNRS through IPGP and LGIT, SCAI, BADW-LRZ, UEDIN, CINECA

Research Infrastructures projects:
EU ADMIRE project (Advanced Data Mining and Data Integration Research in Europe): UEDIN
EU MAPPER project (Multiscales APPlications on European e-infRastructures): LMU with BADW-LRZ

Earthquake and seismology:
EPOS-PP: INGV, KNMI, CNRS, LMU, ULIV
QUEST, RTN EU project: LMU, CNRS(IPGP-LGIT)
NERA, RI3CT EU project: KNMI, EMSC, CNRS(IPGP-LGIT), INGV, ULIV, LMU
WHISPER, ERC project: CNRS(IPGP-LGIT)
SHARE, RI3CT project: KNMI, CNRS(LGIT), INGV

2.4 - Resources to be committed

The VERCE project will provide to the earthquake and seismology community a single service-oriented architecture and framework integrating Data, Grid and HPC infrastructures.

Data infrastructure:

The earthquake and seismology data resources are already structured and integrated within the European Integrated Data Archives (EIDA) infrastructure, providing well-defined standards for data-exchange, meta-data and ontology. This data infrastructure is enabled with standardized distributed access protocols, e.g. the ArcLink technology, and harmonized data

management system (DMS). The ORFEUS and EMSC consortia are operating a community portal, or scientific gateway, integrating tailored web-services enable data access and data distribution interfaces for data-specific query applications. These resources will be made available to the VERCE project.

HPC infrastructure:

The High Performance Computer infrastructure in Europe represent large investment already allocated in the national countries of the partners. They are considered as strategic national resources. Access to the HPC centre resources is in general only granted through established national selection mechanisms. The VERCE project will actively seek computational time through strategic liaison between the earthquake and seismology community and the National HPC centres partners of the consortium. In this respect, the HPC capabilities and support environment of these centres will fuel the VERCE project.

Grid infrastructure:

From the Grid side, the situation is somewhat different. The VERCE project will integrate a number of large existing data and computing resources provided by the VERCE partners (INGV, CNRS). These resources will complement the project. The VERCE project will operate and manage these resources in the VO environment. Access to complementing resources within the Earth Sciences VO will also complement the project.

The VERCE project is budgeted at ~4.8 M€ and runs for a 4 year period (48 months). The VERCE will therefore in complete phase EPOS, the ESFRI initiative of the solid earth Sciences community, during its preparatory phase. The total effort consists of 560 funded person months corresponding to an average of ~ 11 FTE/yr complementing unfunded partners commitment to the project. All partners are EC partners and, except for man power, the budgeted costs cover also for internal circulation across the partners sites of the consortium, project meeting, training, dissemination and public outreach activities, minor hardware and consumables. The travel budget is less than 5% of the budget in support to project meetings and on-site technical support.

Networking activities:

The networking activities are divided in a user-driven activity with the pilot data-intensive applications and use cases (NA2), and a dissemination activity articulated by the training and user documentation - NA3 activity – knowledge based dissemination - and the dissemination and public outreach activity (NA4).

Together these activities implies 113 person/month over the full project at ~ 730 k€ total and allow for travels and part of the tutorial and dissemination material costs, the remaining being provisioned in the management.

Service activities:

These activities will deliver the required data-oriented platform of services, tools combining data, grid and HPC resources. It is articulated in three interacting parts: one deals with the VERCE platform deployment, management and monitoring (SA1), the second with the integration and evaluation of the services and tools components of the platform with the European e-infrastructures environment (SA2); the last one with the integration of web-services and web-interfaces of the scientific gateway (SA3). All together this activity involves 221 person/month for the duration of the project at ~2.3 M€ including user support and travels.

Joint Research activities:

These activities are essential and are to be considered an one off investment and momentum pulse allowing the earthquake and seismology, and beyond the EPOS community, to gain of the quantum change in e-science environment architecture and methods required for its new data-intensive research challenges. This activity is articulated in: a software enabling part – JRA1 - dealing with the architecture and engineering of the software implementation of the data-intensive applications and their adaptation to a service-oriented architecture model; an architecture and methods enabling part – JRA2 - together with the data-oriented tools components, to effectively exploit the distributed data sources and the Grid and HPC capabilities provided by the EU e-infrastructures. All together this activity involves 200 person/month for the duration of the project at ~ 1.5 M€.

Section 3: Impact

3.1 - Expected impacts listed in the work programme

Global and regional earthquake and seismology monitoring systems, exploiting state-of-the-art digital ground motion sensors, are continuously operated and transmitting a rapidly growing wealth of data from around the world. The multi-use nature of these data puts a great premium on data archive infrastructures that are well integrated, open access and global, as pioneered by the earthquake and seismology community. Most of the effort is in Europe, USA and Japan.

In Europe, based upon the GEANT network infrastructures, the earthquake and seismology community is structured with the European Integrated Data Archives (EIDA) infrastructure and strong "horizontal" data services. On top of this distributed data archive system, the community has recently, within the EC project NERIES, developed an advanced SOA based web services and a unified portal system.

Enabling the Data infrastructure with a data-intensive environment on top of Grid and HPC computing infrastructures is becoming instrumental to fully exploit this cornucopia of data, and to guarantee optimal operation and design of the high-cost monitoring facilities.

To ensure the earthquake and seismology community can take full advantage of these data and the underlying computing infrastructure such a data-intensive environment must be delivered to the doorstep of every scientist in this area of research. The most successful approach currently available is through scientific gateways in the form of web portals. They offer several advantages over other approaches, most notably that the scientists themselves are insulated against the complexities of the underlying hardware, network and software configurations.

The deployment strategy of VERCE is driven by the needs of user applications in data exploration, data visualization, data integration, data analysis and data modelling. These applications will be delivered through the existing community web portal with the enhancing functionalities required by the community around each application. The union of all applications through one scientific gateway based on a shared underlying infrastructure will allow the whole community to push the frontiers of discovery in earthquake and seismology.

VERCE rely on an architecture and framework that take into account the scale and the diversity of these data-intensive applications. The project will provide a comprehensive architecture and framework integrating the European seismological Data infrastructure with Grid and HPC infrastructures, taking into account the scale and the diversity of the data-intensive applications.

- A first novel aspect is a service-oriented architecture and framework based on well equipped workbenches with: a tailored tools sets supporting the description, the communication and streaming about the application processes and their integration into re-usable workflows; a level distributed gateways enabled with powerful enactment technology and data communication services, and a data-aware distributed computing platform integrating Data, Grid and HPC infrastructures that delivers the level of performance, efficiency and resilience required by these data-intensive applications.
- A second novel aspect is the coupling between HTC data analysis and HPC data modeling applications through workflow and data sharing mechanisms between the Grid and the HPC infrastructures. The project will be preparing for activities, mainly data challenges, to promote a shared use of both infrastructures.

The VERCE project completes well the strategic vision within the European seismological research and e-science community.

The project builds on the portal and web services developments pioneered in NERIES project and is linked to the European Plate Observing System (EPOS), the ESFRI project of the solid earth science community, as well as a number of European projects in the earthquake and seismology community, e.g., WHISPER, QUEST, NERA, SHARE, and in the earth science community, e.g. GEM.

The project will also draw on other EU initiatives - e.g. Advanced Data Mining and Integration Research for Europe (ADMIRE: EU/FP7), and Open Grid Services Architecture-Data Access and Integration (OGSA-DAI: EU/FP6, UK/EPSRC, UK/DTI), through UEDIN; and Multiscale APPlications on European e-infRastructures through LMU and BADW-LZR.

The ORFEUS and EMSC partners in VERCE, will as the European consortia in seismology, guarantee the involvement, presence and visibility of the project in the community.

From the point of view of computational and data requirements, we rely on the consolidation of computing-based and data-based research across Europe. For this purpose the project will follow the recommendations of international bodies such as e-IRG.

From the point of view of High Performance Computing, the partners of VERCE (CINECA, EPCC, BADW-LRZ) will guarantee the presence of this project in the discussions and roadmap of current and future supercomputer facilities at the European level as well as coordination with PRACE.

From the point of view of the Grid-infrastructure in Europe, some members of the project (UEDIN, SCAI, BADW-LRZ, CINECA, CNRS, KNMI) have an active role in the EGI-NGIs roadmap and previously in EGEE. In particular, VERCE contains KNMI, CNRS and SCAI, which are leading the deployment of the Earth Sciences VO in EGEE.

The sustainability of the earthquake and seismology Grid resources is linked to that of the EGI-NGIs related national project. VERCE is linked with National Grid Infrastructures of France (France Grille), Germany (D-GRID), Italy (Grid.it), the Netherlands, and the UK e-Science program.

The VERCE project also fits well within the international strategy of the seismological and e-science communities, an important aspect since the earthquake and seismology research is intrinsically a Global and international undertaking.

The project VERCE has strong links to prominent data infrastructure organisations in the USA (IRIS-DMC) and in Japan (NIED and JAMSTEC) through the ORFEUS consortium and the FDSN, as well as with similar projects such as SCEC and GEON in the USA and GeoGrid in Japan. This will guarantee strong synergy and interoperability on a global scale. In

particular, collaborations with IRIS-DMC will allow VERCE to explore potential interactions with the DataNet-US programme.

The expected impacts of the VERCE Virtual Community and its underlying architecture are as follows.

Impact on the earthquake and seismology scientific community

VERCE will

> lay the basis for a transformative development in data exploitation and modelling capabilities of the earthquake and seismology research community in Europe and consequently have a significant impact on data mining and analysis in research;

> have a significant impact on the European collaboration and strengthen the European earthquake and seismology research competitiveness;

> establish and nurture close collaborations between the research community and ICT experts in data mining, data engineering and distributed and HPC computing engineering, opening new visions on data mining and analysis;

> improve and optimise strategic data analysis and data modelling software targeted to Grid and HPC computing capabilities;

> remove the barrier to uptake by the earthquake and seismology community of integrated Data, Grid and HPC infrastructures and technologies, by enabling the applications in an existing community portal, in the form of self-contained tailored widgets;
> establish a Virtual Organization in terms of scalable and sustainable community infrastructures that can contribute to the European and National e-infrastructures and to the consolidation of a computing-based research space across Europe.

Impact on the earth science community

VERCE will
> provide a data-intensive Platform as a Service (PaaS), defining an e-science environment architecture and services able to be extended and tailored to the needs of a larger solid earth sciences research community. This will be an important contribution in support of the e-science strategy of the EPOS-PP project;

> enable innovation through the multi-use of the seismological data and re-use of workflow patterns, strengthening the potential for cross-field collaborations within EPOS and beyond with the Earth sciences in particular.

Impact on European data infrastructure and e-infrastructure

VERCE will

> enable data infrastructure with Grid and HPC infrastructures through a service-oriented architecture and framework, increasing the exploitation and the impact of data on research;

> enhance worldwide interoperability in data infrastructures and "horizontal" services by establishing common European data-intensive service infrastructures, and enhanced data communication and data transfer mechanisms;

strengthen dissemination effort, involving research and the e-infrastructures side of the earthquake and seismology European community, to improve the awareness of decision makers, including governments and other national funding agencies, of the impact of data on research and of the need for data development in Europe;

organize and increase the visibility of the virtual earthquake and seismology community, through integrated strategic applications enabled with a community of practice, to impact in the development of European front-end data and computing e-infrastructure

Impact on the European society and economy

VERCE will

improve data analysis and data modeling capabilities for societal applications, where earthquake and seismology has a central role, in natural hazards and risk assessment, energy resources, environmental change and national security;

improve research efficiency, reduce costs with a better and faster research development and production; an effective and efficient use of research infrastructures, funding and outcome through integration of data and computational resources;

lay the basis for a European e-science environment for teaching and education of young students and life-long learners in earth sciences by integrating virtual research environments with data collections, computational resources, learning objects and materials.

3.2 - Dissemination and/or exploitation of project results and management of intellectual property

From past beacons of European projects, dissemination is identified as an important and strategic component of the VERCE project.

The dissemination activities will run during the whole duration of the project and beyond:

- at the beginning, a dissemination activity will raise attention to the project;
- during the project, the dissemination will inform the target audiences about the project activities and progresses;
- in the final stage, the dissemination activity will spread the project results and experiences.

A strategy of coordinating dissemination of knowledge (NA3), and dissemination and public outreach (NA4) is thus needed to achieve the goals in these activities. While dissemination and public outreach will provide an accurate image and awareness of the project, knowledge dissemination will provide the information to the user community in and outside the project promoting education and scientific benefits in using an e-infrastructure architecture combining data, Grid and HPC infrastructures.

A first targeted audience is the international earthquake and seismology community and beyond the Earth Sciences community. These communities will be made aware of VERCE activities and objectives as soon as possible. The dissemination strategy of VERCE will benefit from the ORFEUS and EMSC partners, which as the European consortia in seismology, will guarantee the presence and visibility of the project in these communities at the European and International levels. The dissemination will also benefit from the strong

synergy between VERCE and QUEST, the new EU Marie Curie Initial Training Network in seismology, and EPOS, an ESFRI infrastructure initiative of the solid earth sciences.

A second targeted audience is the ICT community involved in the European Data and Computing technologies and infrastructures and roadmap. These communities will be made aware of the VERCE activities, and of the earthquake and seismology community needs in general, to foster new synergies and collaborations. The dissemination strategy of VERCE will on the one hand benefit from the HPC centres of the project, and their implications in the DEISA and PRACE dissemination initiatives, on the other hand from the partners of VERCE already committed in the European and National Grid initiatives and in the dissemination of related technologies.

A third targeted audience is Industries and Public actors in seismic hazard and risk management, energy resources and environmental changes to establish bridges between the emerging technologies in VERCE and the major actors in Europe. The dissemination strategy of VERCE will on the one hand benefit from the large and diverse exposure of the EMSC and ORFEUS consortia providing real time earthquake information services, on the other hand from the strong synergy between VERCE and other European projects like SHARE, WHISPER and QUEST. This dissemination activity will improve awareness among industry and public actors, encouraging them to collaborate with the project and fostering technologies transfers.

A scientific gateway will be integrated within the community portal managed by ORFEUS and EMSC, and will provide a Hub environment with constant high quality information flow to the user community and the public. Furthermore the Dissemination activity will select among the most important conferences or events in the related fields (one or two per year) and send experts to present the results of the work done within VERCE. These participations will be important occasion for dissemination and cross-fertilization. Earth Sciences will be specifically targeted and valuable examples in point are: The European Geosciences Union (EGU), the European Seismological Commissions (ESC) General Assembly, the the American Geophysical Union (AGU), the International Association of Seismology and Physics of the Earth's Interior (IASPEI), and the International Union of Geodesy and Geophysics (IUGG).

Other important dissemination and exploitation frameworks will the International programmes and initiatives which interest Earth and Environmental Sciences such as: GMES and GEOSS.

Management of Intellectual property

The Communication officer together with the Project Executive Board and the Project Management Office will ensure that the capacity of the project to generate Intellectual Property will be evaluated prior to commencement of the project and will establish an IP strategy, based on the interests of all partners. This will assign the aims, milestones and deliverables of the project and determine where the greatest potential to produce information will occur. This will consider

- the need to safeguard intellectual property rights,
- the benefit of disseminating the information in order to avoid duplications of research effort and work in the earthquake and seismology community,
- the confidentiality and the legitimate interests of the partners.

All the pilot applications codes in the VERCE project are available under an Open Source License or even license free. All the codes and software developed during VERCE will be made available under an Open Source compatible license. The PEB and the PSC will investigate which license is best for each software. Likely candidates are: European Union

Public license 1.1 GNU General Public License 3.0; Mozilla Public License 1.1; Apache License 2.0. The overall strategy in VERCE will be not to merge software or licenses.

The intellectual rights around the applications software that make uses of the middleware remains the right of the communities concerned.

All documents produced by the project for general distribution are covered by copyright establishing ownership to partners of the project, which allows each partner to copy and distribute the documents as required. In particular, training material is intended to be used by other parties and outside groups.

The vast majority of all deliverables and documentation associated with milestones in the project are intended for public consumption.

Section 4: Ethical issues

The VERCE consortium will pursue the general policy of the earthquake and seismology community of open data and open source code and software. Therefore the consortium does not foresee to have to secure patents or grant licenses. Neither does the consortium expect commitment that may hamper exploitation.

	YES	PAGE
Informed Consent		
• Does the proposal involve children?		
• Does the proposal involve patients or persons not able to give consent?		
• Does the proposal involve adult healthy volunteers?		
• Does the proposal involve Human Genetic Material?		
• Does the proposal involve Human biological samples?		
• Does the proposal involve Human data collection?		
Research on Human embryo/foetus		
• Does the proposal involve Human Embryos?		
• Does the proposal involve Human Foetal Tissue / Cells?		
• Does the proposal involve Human Embryonic Stem Cells?		
Privacy		
• Does the proposal involve processing of genetic information or personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)		
• Does the proposal involve tracking the location or observation of people?		
Research on Animals		
• Does the proposal involve research on animals?		
• Are those animals transgenic small laboratory animals?		
• Are those animals transgenic farm animals?		
• Are those animals cloned farm animals?		
• Are those animals non-human primates?		
Research Involving Developing Countries		
• Use of local resources (genetic, animal, plant etc)		

• Impact on local community		
Dual Use		
• Research having direct military application		
• Research having the potential for terrorist abuse		
ICT Implants		
• Does the proposal involve clinical trials of ICT implants?		
I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL	**YES**	

## Annex I: List of Acronyms

ADMIRE Advanced Data Mining and Integration Research for Europe
www.admire-project.eu

AGU American Geophysical Union
www.agu.org

ArcLink Distributed Data Archive Protocol
http://geofon.gfz-potsdam.de/geofon/new/arc_inf.htm

CTBTO Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization
http://www.ctbto.org

DEISA Distributed European Infrastructure for Supercomputing Applications
http://www.deisa.eu

D-Grid Deutsch Grid-Initiative
http://www.d-grid.de/index.php?id=1&L=1

Earthscope A National Science Foundation project
http://www.earthscope.org/

Earth Simulator Supercomputing project targeted at creating a Virtual Earth
http://www.jamstec.go.jp/es/en/index.html

EELA E-science grid facility for Europe and Latin America
http://www.eu-eela.eu/

EGEE Enabling Grids for E-sciencE
http://www.eu-egee.org/

EGI-NGIs European Grid Infrastructure - National Grid Infrastructures
http://collaborating.eu-egee.org/index.php?id=529

EGU European Geosciences Union
http://www.egu.eu/

EIDA European Integrated Data Archives
http://www.webdc.eu/webdc_arc_eu.html

EPOS European Plate Observation System, an ESFRI initiative
www.epos-eu.org

ESC European Seismological Commission
http://www.esc-web.org/

FDSN Federation of Digital Seismograph Networks
www.fdsn.org

FTS A CERN File Transfer System
http://egee-jra1-dm.web.cern.ch/egee-jra1-dm/FTS/

GEM Global Earthquake Model
http://www.globalquakemodel.org/

GeoGrid Global Earth Observation Grid
http://www.geogrid.org

GEON GEOsceinces Network
http://www.geongrid.org/

GEOSS Global Earth Observation System of Systems
http://www.ieee-earth.org/GEOSS

IASPEI International Association of Seismology and Physics of the Earth's Interior
http://www.iaspei.org/

IRIS	Incorporated Research Institutions for Seismology http://www.iris.edu/	
IRIS-DMC	IRIS Data Management Center http://www.iris.edu/hq/programs/dms	
JAMSTEC	Japan Agency for Marine-Earth Science and Technology http://www.jamstec.go.jp/	
MEREDIAN	Mediterranean-European Rapid Earthquake Data Information and Archiving Network http://www.orfeus-eu.org/Organization/Projects/Meredian/meredian.html	
Monalisa	MONitoring Agents using a Large Integrated Services Architecture http://monalisa.cern.ch/FDT/	
NAREGI	National Research Grid Initiative http://www.naregi.org/index_e.html	
NERIES	Network of Research Infrastructures for European Seismology www.neries-eu.org	
NIED	National Research Institute for Earth Science and Disaster Prevention http://www.bosai.go.jp/e/	
OGSA-DAI	Open Grid Services Architecture - Data Access and Integration http://www.ogsadai.org.uk/	
OpenSHA	Open Seismic Hazard Analysis http://www.opensha.org/	
ORFEUS	Observatories and Research Facilities for European Seismology www.orfeus-eu.org	
QUEST	QUantitative Estimation of Earth's Seismic Sources and Structure http://www.quest-itn.org/	
SAFE	Resource management on advanced computing facilities http://www.epcc.ed.ac.uk/projects/grid-safe	
SCEC	Southern California Earthquake Center http://www.scec.org/	
SHARE	Seismic hazard harmonization http://www.share-eu.org/magnoliaPublic/SHARE.html	
SPICE	Seismic wave Propagation and Imaging in Complex media http://www.spice-rtn.org	
Teuta	Tool for UML Based Composition of Scientific Grid Workflows http://www.par.univie.ac.at/~pllana/papers/pqf_ags05.pdf	
USArray	A continental-scale seismic observatory http://www.iris.edu/USArray/	
USGS	U.S. Geological Survey http://www.usgs.gov/	
VEBSN	Virtual European Broadband Seismic Network www.orfeus-eu.org/Data-info/vebsn.html	
WHISPER	Towards continuous monitoring of the continuously changing Earth http://mesoimage.grenoble.cnrs.fr/spip.php?article84/	

Annex II: Example of DMI applications

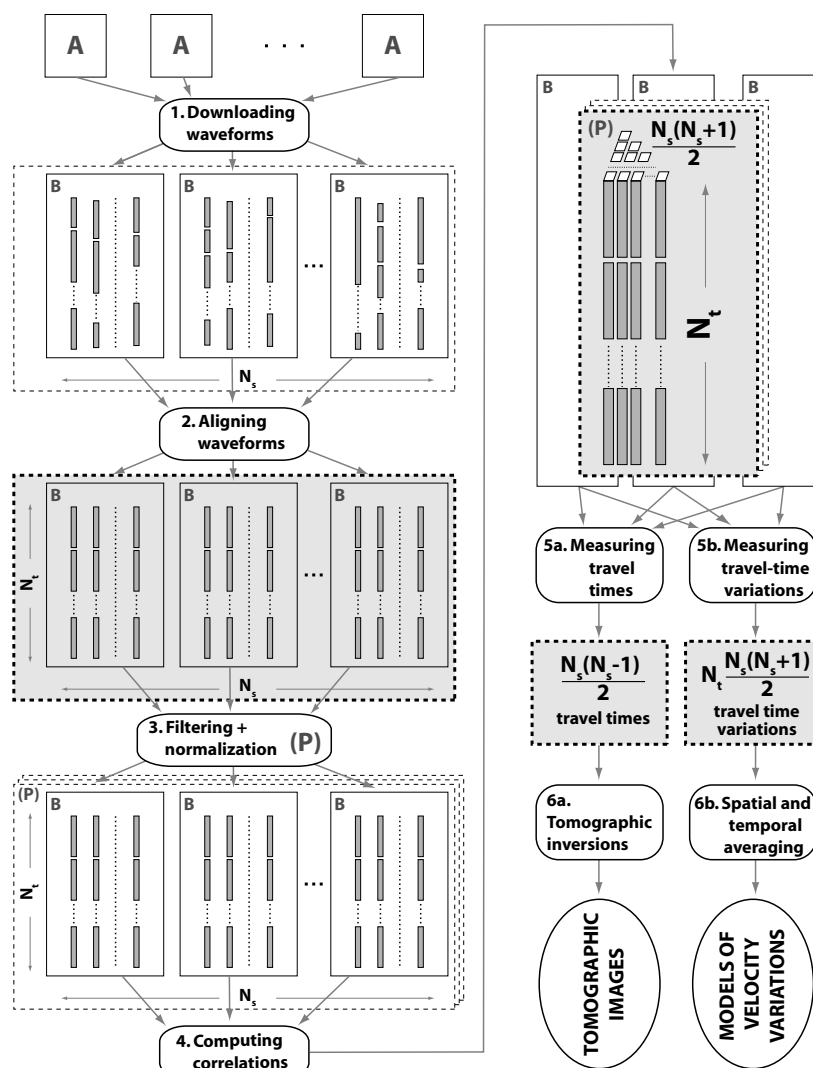**A.2.1 Seismic data exploration and visualization (SDX)**

This application, developed at the ULIV by A. Rietbrock in collaboration with UEDIN (J. van Hemert), provides web-enabled   tools to explore, visualize and analyse seismic wave waveform data: a user interface provides control and visualization; a processing engine supplies waveform data to custom analysis and processing in the form of plugins with simple programming interfaces. SDX is exposed through a web portal in two ways: an Editor Portlet, allowing user to construct data processing plugins; an Excecutor Portlet, providing the interface to other existing plugins. Orchestration of the application workflow currently involve mainly running waveforms transformations. The portlets are directly embedded within the ORFEUS community gateway and linked to the European community Data Resources. Data and computing are orchestrated via the Rapid technology developed at UEDIN. The development of this application was supported by the FP6 NERIES project and the development of the prototype of SDX embedded in NERIES' web portal was funded by the UK's Joint Infrastructure Service Council as the RapidSeis project.
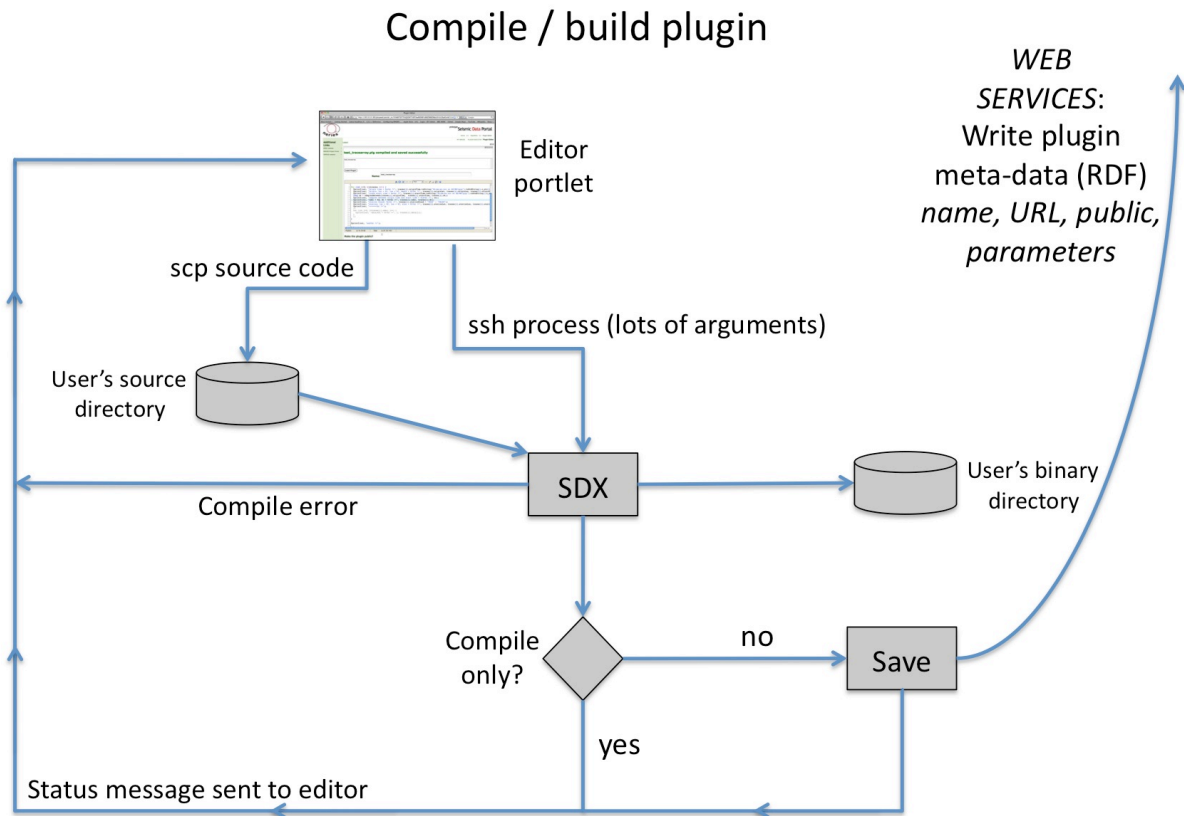
A.2.2 Imaging and Monitoring based on Correlations of Ambient Seismic Noise (IMCASN),

This workflow corresponds to two main applications using correlations of ambient seismic noise: noise-based imaging (e.g., Shapiro et al., 2005) and noise-based monitoring (e.g., Brenguier et al., 2008). The workflow consists of 6 main steps. Four initial steps as well as most volumetric databases are commune for applications of type 'imaging' and 'monitoring'. During these steps, several databases are created on the distributed data processing platform that includes a few separated data storage and processing servers. Distribution of the databases between data servers will be optimized relative to the network configuration to minimize fluxes between the original datacenters, the data storage and processing servers and the users. Important constraint is the global visibility of all databases across the distributed data processing platform. Some of these databases are required only temporary and can be deleted after terminating corresponding data processing. Main databases (indicated with gray boxes), that include: (1) database of re-organized waveforms, (2) databases of noise correlations and (3) databases of noise based travel-time and time-variation measurements can be sheared between different projects and, therefore, it is preferable to keep them quasi permanently on the data processing platform and to upgrade with processing new data.

Size of initial datasets used in applications related to correlation of ambient seismic noise can easily reach several terabytes (e.g., noise based tomography of Europe with using several years of records by hundreds of stations of Virtual European Broadband Seismic Network).

A   A   · · ·   A

1. Downloading waveforms

B   B   B

N_s

2. Aligning waveforms

B   B   B

N_t N_s

3. Filtering + normalization   (P)

(P)   B   B   B

N_t N_s

4. Computing correlations

B   B   B

(P) $\dfrac{N_s(N_s+1)}{2}$

N_t

5a. Measuring travel times

5b. Measuring travel-time variations

$\dfrac{N_s(N_s-1)}{2}$ travel times

$N_t \dfrac{N_s(N_s+1)}{2}$ travel time variations

6a. Tomographic inversions

6b. Spatial and temporal averaging

TOMOGRAPHIC IMAGES

MODELS OF VELOCITY VARIATIONS

Step 1. Downloading waveforms from seismological datacenters (A) (e.g., ORFEUS, GEOFON, GEOSCOPE, INGV) to the data storage servers (B) of the distributed data processing platform. This step results in database of raw time series for N_s stations with a-priori arbitrary segmentation in time.

Step 2. Aligning waveforms. In this step, the time segmentation of waveforms is changed. For all N_s station, seismograms are concatenated and then cut to be organized in a set of N_t equal-length segments. Typical length of this segments is ~24 hours for broadband data and a few hours for high sampling rate data. This step results in a database of organized waveforms that can then be shared and complemented between different projects.

Step 3. Filtering and normalizing the waveforms. Depending on goals of the project, seismograms must be filtered and normalized in a certain way (e.g., Bensen et al., 2007). Therefore, this step and following databases depend on a number of parameters indicated as **P** that include type of normalization and filter type and corner frequencies. This step results in an intermediate database of normalized and filtered waveforms.

Step 4. Computing correlations between all possible station pairs (including auto-correlations) for all N_t time windows. This step results in a database of correlations that includes $N_t \cdot N_s \cdot (N_s+1)/2$ time series. The kept length of this time series can be significantly shorter that the length of correlated waveform segments reducing the volume of the correlation database. At the same time, the number of correlations grows as $\sim N_s^2$ making this database potentially very large for datasets with many stations. Moreover, for the same set of initial waveforms, several correlation databases with different filtering and normalization parameters **P** can be required. Correlation databases can be shared and complemented between different projects. Note that advanced processing can be applied to improve the quality of the signals extracted from noise records when using large networks. Among them the higher order correlations (correlations of correlation, such as the C3 function of Stehly et al;, 2008) are signals which have to be generated and stored.

Step 5. Measuring travel times for tomographic inversion or travel time variations for temporal monitoring. For tomography, only cross-correlations are used and they are averaged over the whole time period resulting in $N_s \cdot (N_s-1)/2$ travel time measurements. For monitoring, relative time variations **dt/t** are measured with method like 'seismic doublets' (e.g., Poupinet et al., 1984; Brenguier et al. 2008) or 'stretching' (e.g., Hadziioannou et al., 2009) for all station pairs and time windows resulting in $N_t \cdot N_s \cdot (N_s+1)/2$ scalar measurements. At this step, time series are used to obtain a small number of scalar measurements and, therefore, the size of databases is significantly reduced. Measurement databases can be shared and complemented between different projects.

Step 6. Applying inversion and/or averaging algorithms to obtain final tomographic images or models of media temporal variations. The noise correlations make it possible to generate huge quantities of equivalent seismograms for local paths inside a network. It provides an improved coverage, leading to high-resolution imaging. When applied to actual data in complex geological settings (e.g. Stehly et al., 2009), the lateral variations that are revealed are far larger than the ones assumed in the classical perturbative imaging techniques. A corollary of the development of noise-based measurements is the need for precise 3D numerical modeling of wave propagation in complex lithospheric models, defined as reference regional models shared by several groups.
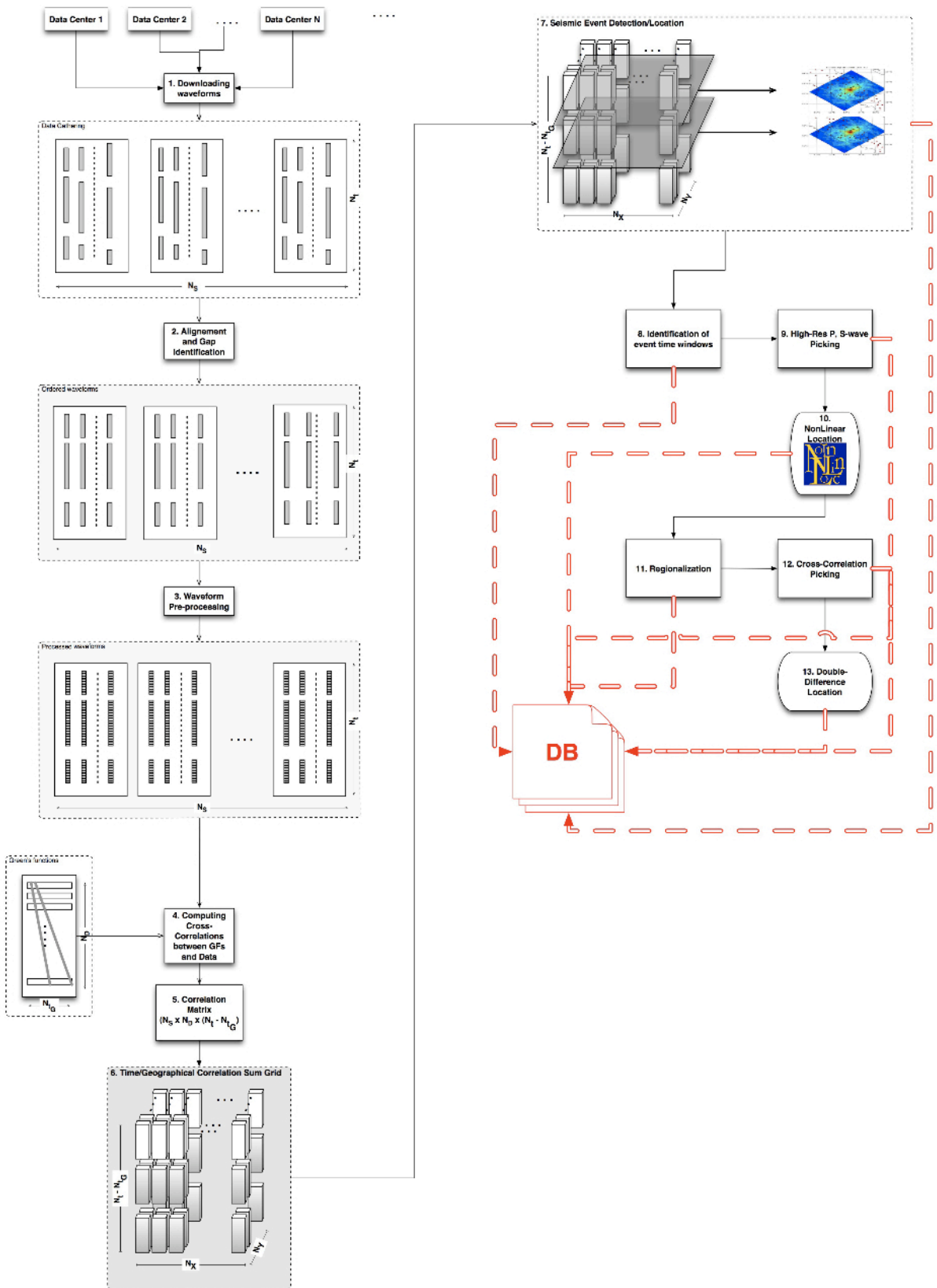
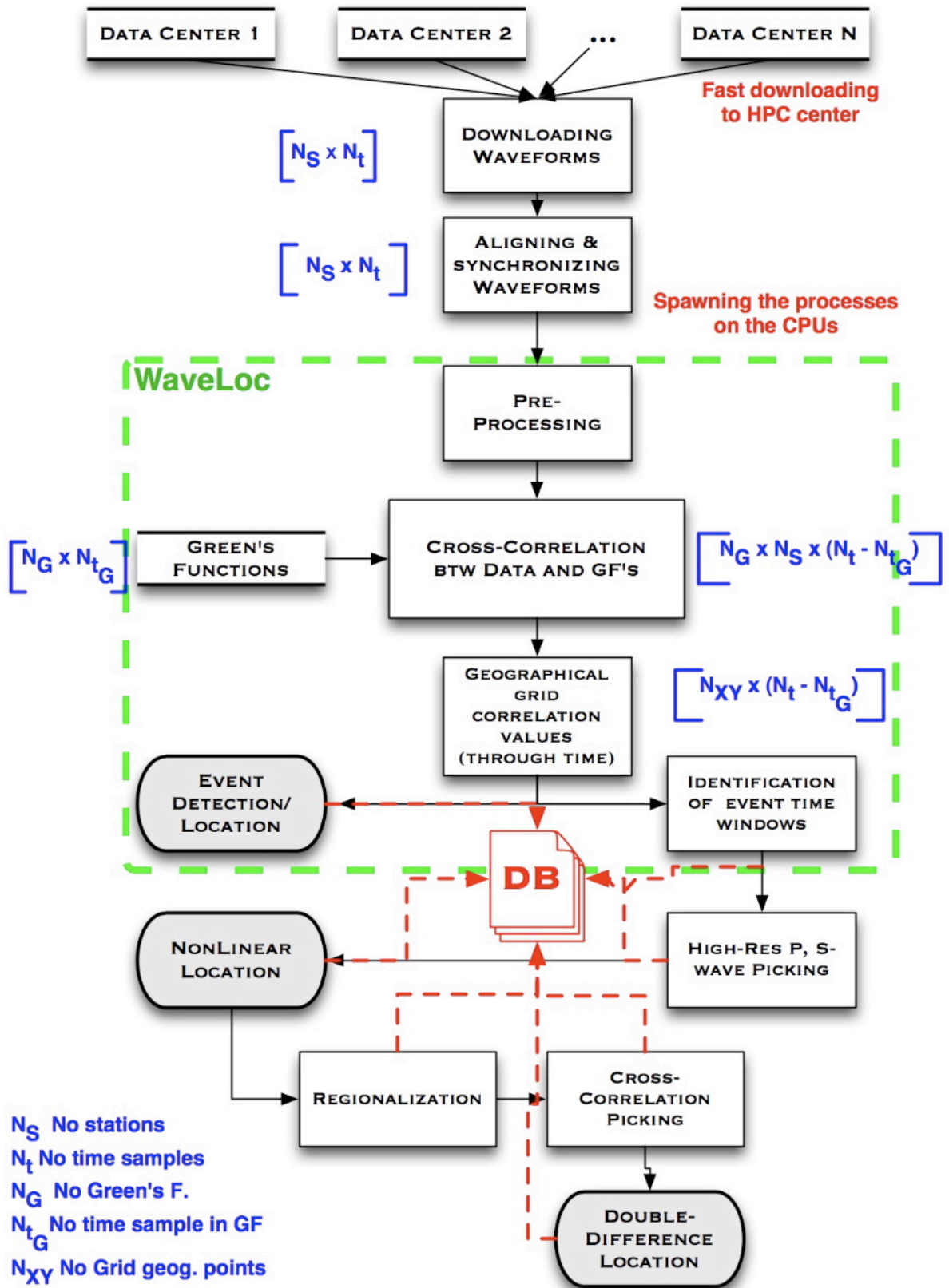**A.2.3 Waveform High Resolution Locations (WaveHRL)**,

Continuous high resolution seismic source location of large volumes of data at regional scale (e.g., Italy and neighboring areas) is a very important but daunting problem in seismology. Accurate and precise earthquake locations, using 3D velocity models and applications of the double-difference technique, have shown that earthquakes do tend to align along preferential planes in 3D, evidencing active faults. This is not, however, the practice used routinely by most seismic monitoring centers, who, in the preparation of their earthquake bulletins, rely on locating earthquakes singularly, suffer from approximate knowledge of the velocity structure (mostly only 1D), use diverse sets of data (i.e., variable number of phases for earthquakes in the same area) and sometimes inconsistent P- and S-wave onset time readings. These factors all contribute to smearing possible alignments, and impeding the recognition of activated faults. We note that although the effort expended to compile earthquake bulletins containing tens of thousands of earthquakes is immense, the results obtained using standard techniques based on taking earthquakes singularly are generally qualitatively poor.

WaveHRL is an innovative modular procedure, developed in collaboration between A. Maggi (EOST, University of Strasbourg, France), A. Michellini (INGV, Rome, Italy) and A. Lomax (Mouans-Sartoux, France), that performs sequentially and automatically the following analyses on very large volumes of continuously recorded data.

- Detection and initial location using WaveLoc – a full waveform detection/location algorithm capable of analysing seamlessly large volumes of waveform data to detect seismic sources and provide seismic event locations
- Consistent waveform picking on detected events
    - ➢ using state-of-the-art 3C wave onset detectors
    - ➢ using cross-correlation picking on detected clusters
- Earthquake location
    - ➢ using non-linear, global search methodologies and/or
    - ➢ global double difference locations using sets of reference events for individual areas of the target region.

This modular processing is expected to provide a very big improvement in the analysis of large volumes of data in terms of both reduced need of human resources (i.e., seismic analysts time) and much improved earthquake location quality (lower magnitude detection level and much higher relative and absolute locations).

DATA CENTER 1 DATA CENTER 2 ... DATA CENTER N

Fast downloading
to HPC center

$\begin{bmatrix} N_S \times N_t \end{bmatrix}$

DOWNLOADING WAVEFORMS

$\begin{bmatrix} N_S \times N_t \end{bmatrix}$

ALIGNING & SYNCHRONIZING WAVEFORMS

Spawning the processes
on the CPUs

WaveLoc

PRE-PROCESSING

$\begin{bmatrix} N_G \times N_{t_G} \end{bmatrix}$ GREEN'S FUNCTIONS

CROSS-CORRELATION BTW DATA AND GF'S

$\begin{bmatrix} N_G \times N_S \times (N_t - N_{t_G}) \end{bmatrix}$

GEOGRAPHICAL GRID CORRELATION VALUES (THROUGH TIME)

$\begin{bmatrix} N_{XY} \times (N_t - N_{t_G}) \end{bmatrix}$

EVENT DETECTION/ LOCATION

IDENTIFICATION OF EVENT TIME WINDOWS

DB

NONLINEAR LOCATION

HIGH-RES P, S-WAVE PICKING

REGIONALIZATION

CROSS-CORRELATION PICKING

DOUBLE-DIFFERENCE LOCATION

N_S No stations
N_t No time samples
N_G No Green's F.
N_{t_G} No time sample in GF
N_{XY} No Grid geog. points

114

A.2.4 : TSUnami impact MAPS in near real-time (TSUMAPS)

<u>Application description</u>
As confirmed by the very last tsunami events (e.g., Maule, Chile, M8.8, February 2010; Mentawai, Indonesia, M7.7, October 2010), rapid assessment of the tsunami impact after a submarine earthquake is still a fundamental issue for tsunami warning centers. This is especially crucial for those communities that might be "attacked" by tsunami waves in the very first minutes following the mainshock.
On the other side, current HPC facilities meet the requirements for real-time tsunami modeling to the purpose of early warning.
TSUMAPS is a modular procedure that, starting from real-time seismic data recorded during an earthquake, is able to provide i.) fast and accurate estimates of maximum water elevation along coastlines, ii.) tsunami travel times and iii.) predicted waveforms at selected coastal locations. The forecasts can be updated based on seismic and tsunami information continuously fed into the system. This application exploits an open source tsunami propagation code (COMCOT), extensively used at INGV, to forecast in near real-time (<5 min) tsunami heights after a submarine earthquake.

TSUMAPS consists of 3 modules:
Module 1, Rapid Detection and Characterization of the Earthquake;
Module 2, Source Setup;
Module 3, Fast Tsunami Modeling/Forecasting.

When an earthquake occurs, seismic waveforms are downloaded to HPC centers and used by Module 1 to calculate origin time, hypocenter and magnitude of the seismic source. Furthermore, this module computes the apparent rupture duration and the dominant period of the recorded P-waveforms, which have been demonstrated to give a first order estimation of the tsunamigenic potential of an earthquake (Lomax and Michelini, 2009). Module 1 is already operational at the INGV seismic center (see http://s3.rm.ingv.it/D-E.php).
Module 2 combines the above parameters with information on tectonic setting and historic moment tensors of the source zone through classical earthquake scaling laws, in order to set-up a finite fault model of the seismic source.
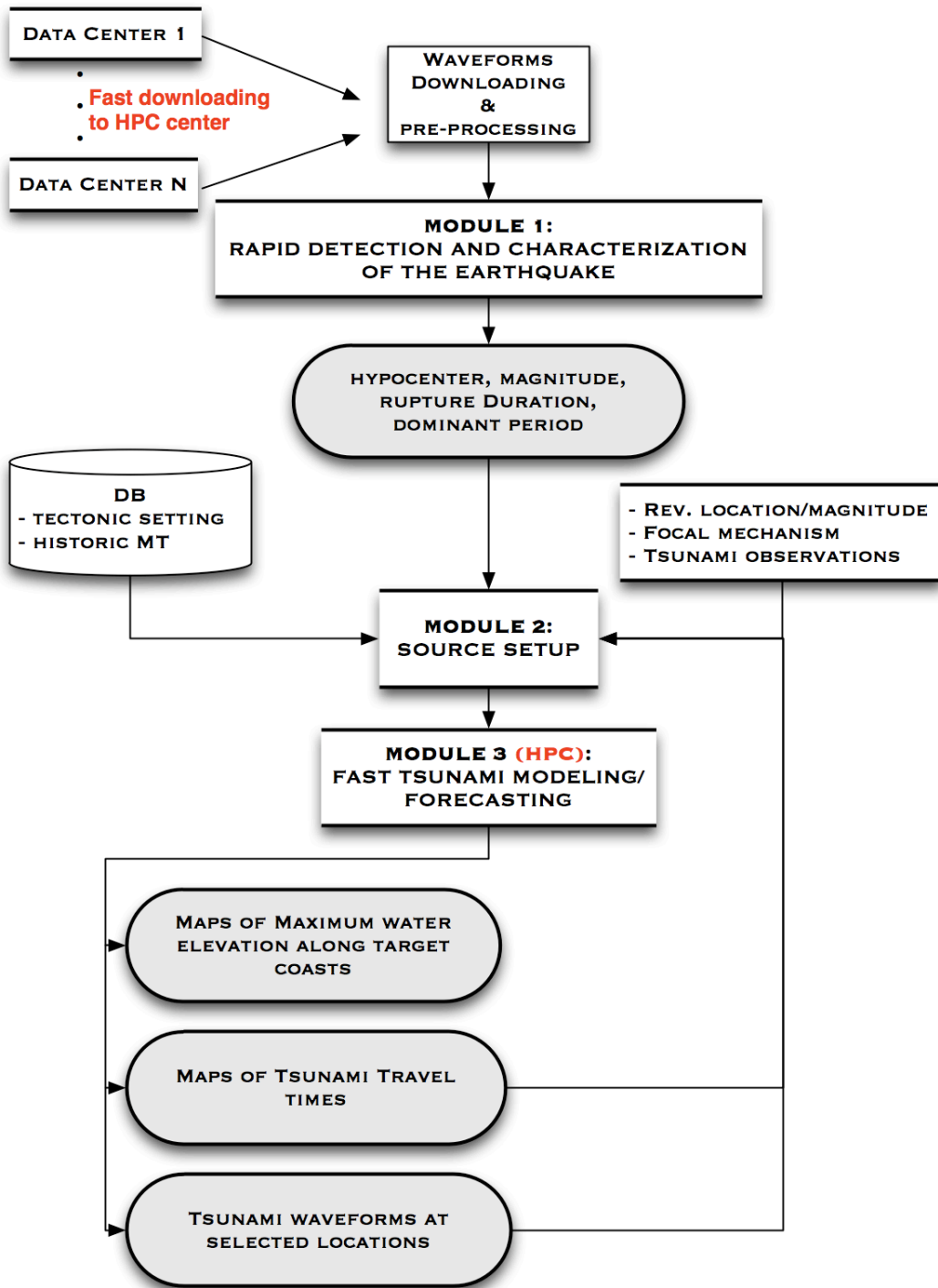Module 3 initializes the calculation of explicit tsunami propagation (from the source to the target coastlines) on a HPC resource.
Outputs of Module 3 are the maximum wave heights forecasted in front of the target coastlines, coded in terms of threat level threshold exceedance. Tsunami travel time maps will complement wave height information and, at specific sites, sea-level waveforms are provided as well.
The calculations can be replicated to form a set of scenarios, for the treatment of the uncertainties with a Bayesian approach. When more information on the seismic source (focal mechanism, finite fault parameters) and/or tsunami observations (tide-gauges, buoys) are available, these can be assimilated as well into the forecast.
The forecasted TSUMAPS may be quickly disseminated to the centers that are in charge for tsunami warning. This information should be used by warning centers (e.g. NEAMTWS) based on an appropriate decision matrix.

TSUMAPS FLOWCHART

DATA CENTER 1

Fast downloading to HPC center

DATA CENTER N

WAVEFORMS DOWNLOADING & PRE-PROCESSING

MODULE 1: RAPID DETECTION AND CHARACTERIZATION OF THE EARTHQUAKE

HYPOCENTER, MAGNITUDE, RUPTURE DURATION, DOMINANT PERIOD

DB
- TECTONIC SETTING
- HISTORIC MT

- REV. LOCATION/MAGNITUDE
- FOCAL MECHANISM
- TSUNAMI OBSERVATIONS

MODULE 2: SOURCE SETUP

MODULE 3 (HPC): FAST TSUNAMI MODELING/ FORECASTING

MAPS OF MAXIMUM WATER ELEVATION ALONG TARGET COASTS

MAPS OF TSUNAMI TRAVEL TIMES

TSUNAMI WAVEFORMS AT SELECTED LOCATIONS

116

A.2.5 **Non-linear Inversion for Kinematic Earthquake Source (NIKES)**

NIKES (Non-linear Inversion for Kinematic Earthquake Source) is a procedure that provides the complete spatio-temporal description of the earthquake source, by jointly inverting different kind of ground motion data (strong motion records, static and dynamic GPS, SAR interferograms). This procedure has been recently developed by Piatanesi et al. at INGV and is based on a two-stage nonlinear technique. During the first stage of the inversion, a "heat-bath" simulated-annealing algorithm explores the model space to generate an ensemble of models that efficiently sample the good data-fitting regions. In the second stage (appraisal), the algorithm performs a statistical analysis of the model ensemble and computes an average model and its standard deviation as well as the correlation between the model parameters.

Implementation and workflow

Problem set-up
NIKES is conceived to start when some parameters (namely the hypocenter, magnitude and focal mechanism) of the causative source are already estimated. On the basis of these information NIKES sets up the problem geometry by defining the position, strike and dip of the finite-fault and calculates the traction Green's functions on the fault for various observer (station) locations. Then, the extended fault is divided into subfaults with model parameters assigned at the corners; the value of every parameter is not constant inside the subfault but it spatially varies through a bilinear interpolation of the nodal values. At each point on the fault the rupture model is described by four model parameters: rise time, rupture time, peak slip velocity and rake angle. Each point on the fault can slip only once (single window approach) and the source time function can be selected among different analytical forms (e.g. box-car, triangular, exponential, regularized Yoffe).
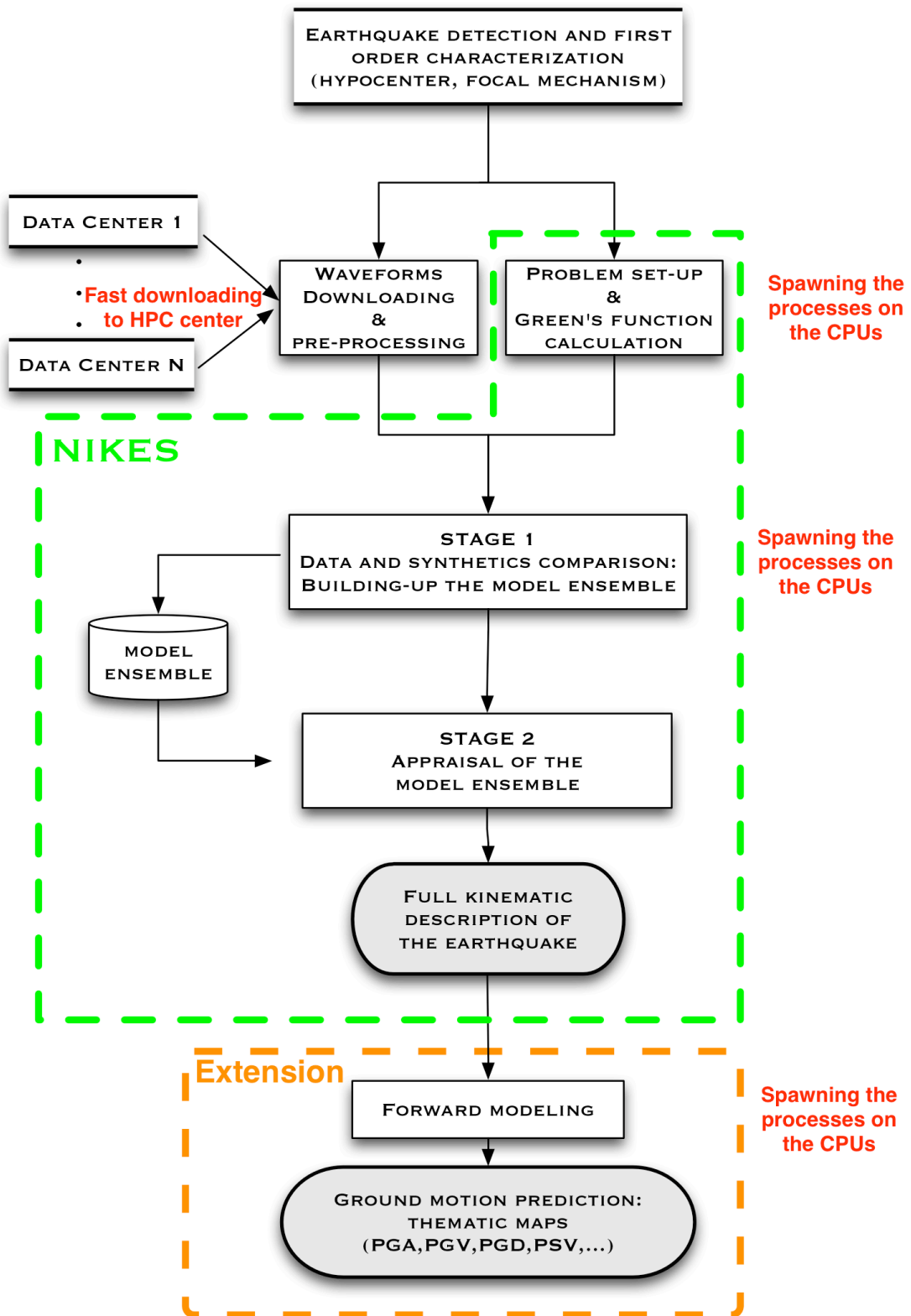
Data gathering and processing
To perform the inversion NIKES requires the data (seismograms) in the Fourier domain. The gathered data have to be converted in SAC format and processed by a SAC script that performs all the operations needed. These are mainly: time alignment to the origin time of the earthquake, cutting, band-pass filtering, eventually integration to velocity, rotation and Fourier transforming.

Stage 1: building up the model ensemble
In this stage a heat-bath simulated annealing algorithm builds up the model ensemble. The algorithm starts its search by a random model and then it perturbs the model parameters one by one. Then, for each possible configuration, the forward modeling is performed with a Discrete Wave-Number Integration technique (*Spudich and Xu*, 2003), whose Green's function includes the complete response of the 1-D Earth structure. Observed and predicted data are compared in the frequency domain. For strong motion data an objective cost function that is an hybrid representation between L1 and L2 norms is used, while the cost function related to the GPS measurements is a sum-squared of the residuals between synthetic and recorded static displacements normalized to the observed data. The total cost function is computed from the summation of the weighted cost functions of the two datasets.

Stage 2: appraising the model ensemble
During the first stage, all models and their cost function values are saved to build up the model ensemble. In the second stage the algorithm performs a statistical analysis of the ensemble providing the best-fitting model, the average model, the associated standard deviation computed by weighting all models of the ensemble by the inverse of their cost function values and the correlation among the model parameters. These estimates represent the ensemble properties and are the actual solution of the nonlinear inverse problem. This approach allows extracting the most stable features of the rupture process that are consistent with the data as well as to assess model uncertainties.

EARTHQUAKE DETECTION AND FIRST ORDER CHARACTERIZATION (HYPOCENTER, FOCAL MECHANISM)

DATA CENTER 1

·Fast downloading to HPC center

DATA CENTER N

WAVEFORMS DOWNLOADING & PRE-PROCESSING

PROBLEM SET-UP & GREEN'S FUNCTION CALCULATION

Spawning the processes on the CPUs

NIKES

STAGE 1
DATA AND SYNTHETICS COMPARISON: BUILDING-UP THE MODEL ENSEMBLE

Spawning the processes on the CPUs

MODEL ENSEMBLE

STAGE 2
APPRAISAL OF THE MODEL ENSEMBLE

FULL KINEMATIC DESCRIPTION OF THE EARTHQUAKE

Extension

FORWARD MODELING

Spawning the processes on the CPUs

GROUND MOTION PREDICTION: THEMATIC MAPS (PGA, PGV, PGD, PSV, ...)

EP**O**S
EUROPEANPLATEOBSERVINGSYSTEM

Research Infrastructure and E-science for Data and Observatories on
Earthquakes, Volcanoes, Surface Dynamics And Tectonics

Letter of Intent

Rome 16 November 2010

To: Jean-Pierre Vilotte
       Equipe de Sismologie
       Institut de Physique du Globe de Paris
       (CNRS-UMR 7154)
       4 Place Jussieu, 75252
       Paris Cedex 05

As coordinator of the ESFRI infrastructure "*European Plate Observing System* (**EPOS**)"
I hereby state that we have identified common interests with the I3 project **VERCE**
(*Virtual Earthquake and Seismology Research Community in Europe*) coordinated by
your institution, which is going to be submitted to the INFSO Call INFRA-2011-1.2.1.

In particular, the main objectives of the VERCE proposal, namely supporting the
development and deployment of e-Science environments, integrating the European
seismological Data infrastricture with the European Grid and HPC e-infrastructures, and
the implementation of dedicated, high-level services for the users, are of relevance for
EPOS now entering in its preparatory phase. VERCE will have a strong impact to our e-
infrastructure implementation plan, one of the key issues of the EPOS project.

In this context, VERCE can provide an ideal architecture and platform to define a Data,
Grid and HPC support strategy to fit the growing needs of the EPOS community for the
analysis of large data volumes and for state-of-the-art modelling and inversion of the
seismic wavefield in complex structures.

We have included in the EPOS-Preparatory Phase project a dedicated task aiming at
formalizing the cooperation with other projects, not limited to the ESFRI roadmap. We
foresee that in the frame of such cooperation we can optimize both investments for a
better study of solid Earth and to face modern challenges in geophysical sciences.

Yours sincerely

Massimo Cocco
EPOS Coordinator

CN/RS CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

UNIVERSITE
JOSEPH FOURIER
SCIENCES TECHNOLOGIE MEDECINE

Observatoire de Grenoble   **LGIT** *Grenoble*   *Laboratoire de Géophysique Interne et Tectonophysique*

Grenoble, 20 november 2010

Letter of support for VERCE

As the scientific leader of the ERC Avdanced Grant project WHISPER, I am willing to express my support to the VERCE proposal. The goal of WHISPER is to develop the methodology towards a continuous monitoring of the continuously changing Earth. Our project focuses on the use of the seismic ambient noise to monitor slight changes of properties in the solid Earth. The implications are the detection of change of strain at depth with applications in different contexts. A major field of application is the monitoring of potentially dangerous structures like volcanoes or active fault zones. The project includes new methodological developments, massive processing of existing data and field experiments. Applications in regions where changes are induced by industrial activity are important both for the quantitative refinement of the method and for the important economic and social implications of these problems. Developing the WHISPER project needs to use huge quantity of continuous recordings of the ground motion, i.e. years of records by modern networks. As seismologists we are able to develop new methodologies for the processing and to work out the necessary theoretical background, as expressed in our proposal. On the other hand we realize that to shape the future techniques, we have to envision the data mining at scales much larger than what we are currently doing. This means to make use of new e-infrastructure technologies to better exploit the increasing wealth of data from the networks in developments with hundreds or thousands of sensors. This implies an application driven e-infrastructure architecture and framework integrating data and distributed computing infrastructures, as well as new workflow and data streaming tools. These tasks require a synergy with data-aware distributed computing, data mining and data integration experts for an optimal design of the architecture and the adaptation/optimization of our software, with up-to-date concepts in these fields. WHISPER can be a starting base for these developments in the field of monitoring. Two engineers are presently working at the optimization of our procedures for a local grid. We will be glad to share our expertise and equipments in the framework of the VERCE project and to participate in a European effort to handle the challenge of massive seismic data analysis. I therefore consider VERCE as an exciting initiative that corresponds to the needs of the emerging science challenges in seismology. I give it my entire support.

Michel Campillo
Professeur à l'Université Joseph Fourier et à l'Institut Universitaire de France

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

GEOPHYSIK/SEISMOLOGIE
DEPARTMENT FÜR GEO- UND UMWELTWISSENSCHAFTEN

LMU • Geophysik • Theresienstraße 41/IV • 80333 München

Prof. Dr. Heiner Igel

Telefon +49 (0)89 2180-4204
Telefax +49 (0)89 2180-4205

Heiner.Igel@lmu.de
www.geophysik.uni-
muenchen.de
Department für Geo-und
Umweltwissenschaften
Geophysik
Theresienstraße 41/IV
80333 München

Prof. Jean-Pierre Vilotte
Equipe de Sismologie
Institut de Physique du Globe de Paris
(CNRS-UMR 7154)
4 Place Jussieu, 75252 - Paris Cedex 05
France

Ihr Zeichen, Ihre Nachricht vom   Unser Zeichen

München, 12.11.10

Re: Integrated Infrastructure Initiative projects (I3) Call topics: INFRA-2011-1.2.1

Dear Jean-Pierre,

in my function as coordinator of QUEST, the Geophysics Research Group, Ludwig-Maximilians-University, and the QUEST Consortium I hereby express strongest support for the VERCE (Virtual Earthquake and seismology Research Community in Europe: a data-intensive e-science environment) initiative. The EU funded QUEST project links together leading European scientists in the field of seismology, seismic hazard, earthquake physics, physical inverse problems, applied mathematics and high-performance computing from Departments of Mathematics, Physics, Earth Sciences, and Oceanography. Such a multidisciplinary approach has previously proven very successful in combining leading scientific collaborations with strong industrial links and will provide the "market" for the applications to be engineered in VERCE.

QUEST (quest-itn.org) is the second, large, training network co-ordinated by LMU Geophysics (the SPICE project was completed in 2007). Both networks specifically aim at training early stage researchers (PhD and early postdoctoral fellows) in the field of computational seismology, with particular emphasis on providing the participants with transferable skills and the necessary tools to succeed in either academia or industry. In the case of QUEST this is reflected in the inclusion of 4 industrial partners; 2 full-partners and 2 associated-partners.

In the light of this VERCE would provide a unique complement to QUEST (and other initiatives like NERA) as a provider of professional services for the development, maintenance, and implementation of scientific software and access to the European GRID and supercomputer infrastructure as well as joint training of a new generation of scientists. VERCE would have a tremendous impact on the seismological community and would lead to a new paradigm how to deal with the increasing data volumes.

Yours sincerely,

Leibniz-Rechenzentrum
der Bayerischen Akademie der Wissenschaften

Jean-Pierre Vilotte
Seismology, Office 329
Institut de Physique du Globe de Paris
(CNRS-0UMR 7154)
1 rue Jussieu, 75238 – Paris Cedex 05
France

Garching, 19/11/10
Telefon: +49-89-35831-8823
Telefax: +49-89-35831-8623
E-Mail: helmut.heller@lrz.de

Subject: Letter of Support for VERCE (Integrated Research Environments for Earth Science)

Dear Dr. Vilotte,

The Initiative for Globus in Europe (IGE) coordinates European Globus activities, drives forward Globus developments according to the requirements of European users, and strengthens the influence of European developers in the Globus Alliance. IGE serves as a comprehensive service provider for the European e-infrastructures regarding the development, customization, training, support, and maintenance of components of the Globus Toolkit. If VERCE receives funding, we will also offer these services to VERCE.

The IGE EU FP7 project is coordinated by the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities in Garching near Munich (BADW-LRZ).

On behalf of the IGE consortium, I would like to offer my strong support for the project

VERCE (Integrated Research Environments for Earth Science)

in response to the EU FP7 Call – Topic: INFRA-2011-1.2.1: e-Science environments, Call-Identifier: FP7-INFRASTRUCTURES-2011-2.

Sincerely,

Dr. Helmut Heller
(IGE Project Coordinator)

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR INFORMATIK
LEHR- UND FORSCHUNGSEINHEIT FÜR KOMMUNIKATIONS-
SYSTEME UND SYSTEMPROGRAMMIERUNG

Prof. Dr. Dieter Kranzlmüller · Institut für Informatik
Universität München · Oettingenstr. 67 · 80538 München

Jean-Pierre Vilotte
Seismology, Office 329
Institut de Physique du Globe de Paris
(CNRS-UMR 7154)
1 rue Jussieu, 75238 - Paris Cedex 05
France

Prof. Dr. Dieter Kranzlmüller
Oettingenstr. 67, 80538 München
Tel: 0 89 / 21 80- 9146
Fax: 0 89 / 21 80- 9147

kranzlmueller@ifi.lmu.de
www.nm.ifi.lmu.de

München, den 19.11.2010

Subject: Letter of Support for **VERCE (Integrated Research Environments for Earth Science)**

Dear Dr. Vilotte,

The German NGI-DE (National Grid Initiative Deutschland) is represented by the Gauß-Allianz e.V. (GA), an association founded by the German High Performance Computing (HPC) sites to support the scientific community in Germany for sustainable and efficient usage of HPC resources. The activities of NGI-DE are performed within the GA by a consortium of Karlsruhe Institute of Technology (KIT), Leibniz Supercomputing Centre in Garching near Munich (LRZ), Research Center Juelich (FZJ), and German Electron Synchroton Hamburg (DESY), which interact with partners on both, national and international level.

On behalf of German NGI-DE and in my role as representative of NGI-DE on the EGI Council, I would like to offer my strong support for the project

VERCE (Integrated Research Environments for Earth Science)

in response to the EU FP7 Call – Topic: INFRA-2011-1.2.1: e-Science environments, Call-Identifier: FP7-INFRASTRUCTURES-2011-2.

Sincerely,

Prof. Dr. Dieter Kranzlmüller
(NGI-DE Representative)

Software Sustainability Institute

EPCC
JCMB, Mayfield Road
Edinburgh, EH9 3JZ
United Kingdom

Prof. Jean-Pierre Vilotte   21 November 2010
Institut de physique du globe de Paris
4, place Jussieu - Case 89
75252 Paris
France

Dear Jean-Pierre,

On behalf of the Software Sustainability Institute, I am writing in strong support of your Virtual Earthquake and Seismology Research Community in Europe (VERCE) proposal to the EC to provide tools for the research community to better exploit the under-utilised seismological data which is available but still distributed across many different organisations and countries.

The Software Sustainability Institute, centred at the University of Edinburgh, works in partnership with research communities to identify key software and helps them make it sustainable. Our vision is to make software useful for future generations of researchers by improving usability, maintainability and quality. This increases research productivity by enabling more people to use software, delivers added value by leveraging our expertise and connections, and ensures researchers can continue to use software as a cornerstone of their research.

The commitment of VERCE to work closely with existing infrastructure providers and other related projects is important to ensure sustainability of its outputs, and follows the best examples we have seen at the SSI. The consortium is led by some of the leading institutions in seismology and earthquake research communities which in our experience will lead to a far greater, and more sustainable, impact from the project than if it were purely technology driven.

The SSI will be glad to offer guidance and expertise in community development and engagement; suggest processes and infrastructure to aid long term-sustainability of the software; and promote and disseminate the work and successes of the project through our networks. We work closely with the EGI, UK NGS/NGI and many middleware providers, particularly to help define the unified middleware distribution that will be taken up across the European infrastructural providers and would also be able to provide expert knowledge in this area.

We look forward to working with your team in the future.

Yours sincerely,

Neil Chue Hong
Director, Software Sustainability Institute

……

22 November 2010

To whom it may concern

Dear Sir/Madam,

Re: VERCE proposal

The purpose of this letter is to provide written support for the above proposal to the EU for a Virtual Earthquake and Seismology Research Community in Europe. VERCE will enable the earthquake research and Seismology community to significantly advance its understanding of the earth's structure and processes, thereby aiding society in the management of natural hazards, energy resources, environmental changes, and national security concerns.

I have been aware of Jean-Pierre Villotte's work in seismology in the public literature and personally since visiting Paris on sabbatical in 1999, and Michael Atkinson's through my current Royal Society of Edinburgh fellowship on data mining in hydrocarbon reservoir dynamics and induced seismicity based in the School of Informatics. I was also involved in the EU NERIES project that preceded the current proposal, as project leader in aspects to do with uncertainty analysis in seismic hazard analysis.

I am writing to confirm that Geoscientists from UoE will be actively engaged as contributors and users of the VERCE platform, both in the direct application of the seismic data base to problems in inversion of continuous monitoring data (e.g. for time-lapse tomography based on analysis of cross-correlated seismic 'noise') and it time to explore applications in wider geophysical problems such as meteorology, climate change, environmental pollution and impact on society. A recent meeting on GeoInformatics at Edinburgh attracted some 30 attendees from the two Schools testament to the time being right to harvest the ongoing 'data bonanza' in combined Geospatial and GeoTemporal data for a variety of applications.

The seismology group at Edinburgh has good contacts with the British Geological Survey, formally the primary UK national providers of broadband seismic data to organisations such as ORPHEUS, and who also have their main seismology group on the Edinburgh University Science and Engineering campus. This will provide a useful sounding board of data providers and users to give advice and support to the overall project, and ensure products are developed in such a way as to be compatible with a variety of user and provider-led interests and constraints.

Yours sincerely,

Ian Main
Professor of Seismology and Rock Physics
Scottish Government/Royal Society of Edinburgh Research Fellow

British
Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Edinburgh

Murchison House
West Mains Road
Edinburgh
United Kingdom
EH9 3LA

Telephone +44(0)131 6671000
Direct Line +44(0)131 6500222
Main Fax +44(0)131 6681535
Direct Fax +44(0)131 6671877
E-mail bbap@bgs.ac.uk

Professor Malcolm Atkinson
Director
National e-Science Centre
School of Informatics
University of Edinburgh
UK

Dear Professor Atkinson,

As an active member of the seismology community BGS supports the aims of the VERCE proposal to improve and better integrate resources for seismic data. Scientists at the BGS would be happy to be actively engaged as contributors and users of the proposed European continuous monitoring database outlined in the proposal. In addition, I believe that improvements in infrastructure in data access, transfer and exploitations are essential for advancing research and future mitigation of earthquake hazard.

Yours sincerely

Brian Baptie

INVESTOR IN PEOPLE