

VERCE : Virtual Earthquake and Seismology Research Community in Europe e-science environment



Centre National de la Recherche Scientifique (CNRS-INSU), France
University of Edinburgh (UEDIN), United Kingdom



Royal Netherlands Meteorological Institute (KNMI-ORFEUS), Netherlands



European-Mediterranean Seismological Centre (EMSC), France

Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy

Ludwig-Maximilians-Universität (LMU), Germany

University of Liverpool (ULIV), United Kingdom



Bayerische Akademie der Wissenschaften (BADW-LRZ), Germany

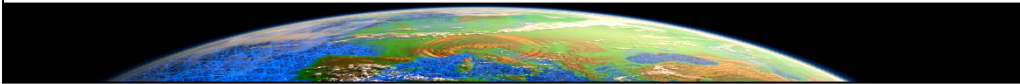


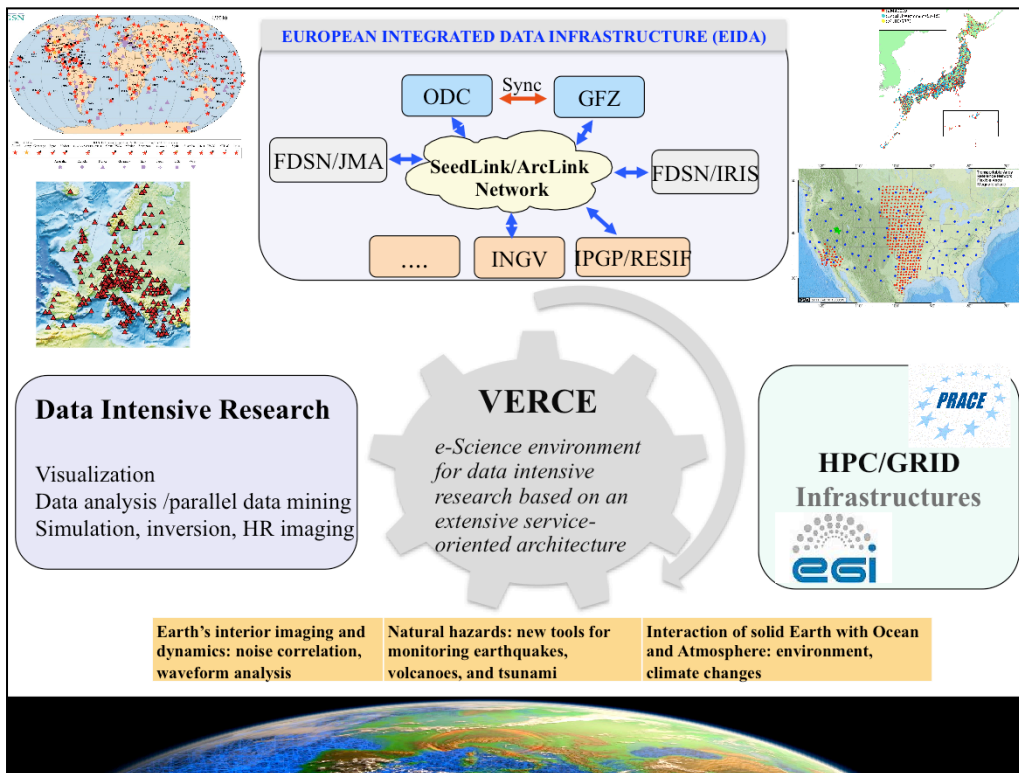
Fraunhofer-Gesellschaft e.V. (SCAI), Germany



Centro di Calcolo Interuniversitario (CINECA), Italy

Jean-Pierre Vilotte (CNRS-IPG Paris), Malcolm Atkinson (UEDIN), Torild van Eyck (ORFEUS-KNMI), Anton Frank (BADW-LRZ)





The earthquake and seismology research is an intrinsically Global undertaking that addresses both fundamental understanding of the Earth's interior structures and evolution, and augmented societal issues concerning natural hazards, energy resources, environmental changes and national security. This community is central in the European Plate Observatory System (EPOS), the ESFRI project of the solid earth community now in the preparatory phase.

Global and regional monitoring networks are continuously transmitting wealth of digital continuous time data from dense seismic arrays to the European Integrated Data Archives (EIDA), part of the International Federation of Digital Seismograph Networks (FDSN) - an international consortium within GEOSS - with well established standards for data exchange formats and data center management systems.

The earthquake and seismology community is facing today a fundamental research paradigm shift: from data and event driven research to data intensive research exploiting the wealth of information provided by dense arrays of continuously recording stations.

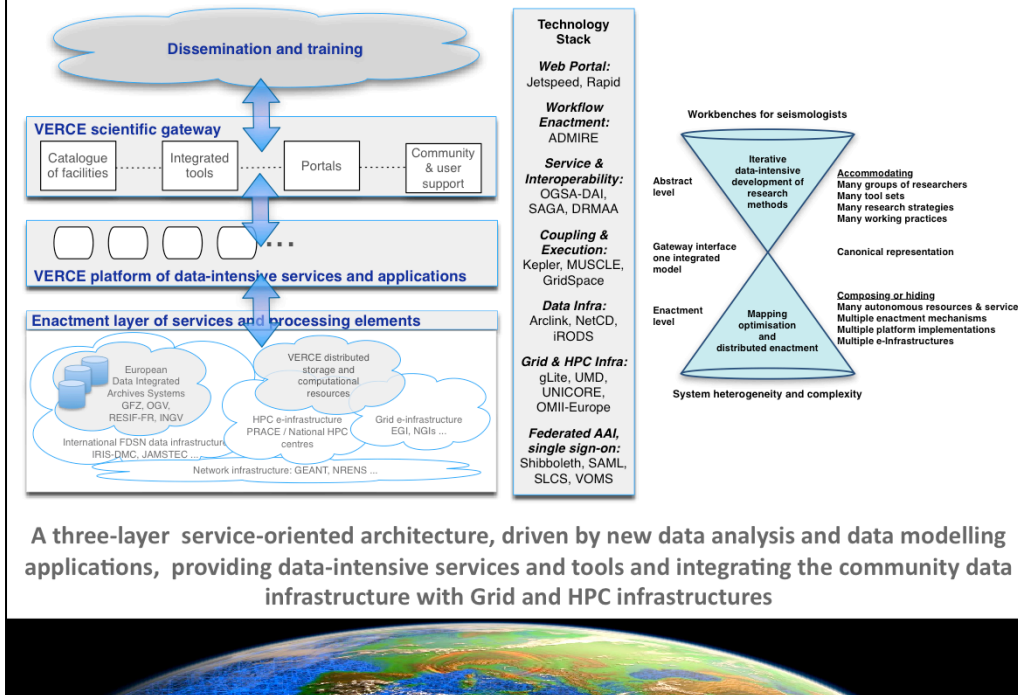
Today, innovative analysis methods extracting spatial and time coherent information using correlation-based and mining methods allow to exploit the whole waveform information even in what was previously disregarded as noise.

These new data analysis and inversion methods require seamless access to large volumes (~100 TB – 1 PB) of distributed multi-sets data and extensive computing and storage capabilities across the Grid and HPC infrastructures. Other critical issues are related to curation, sharing and re-use of processed and simulated data (publishing); collaborative sharing of data and methods, together with orchestrated workflows across multiple service components.

To exploit the full potential of the community data environment and of the emerging European e-infrastructures (Grid and HPC), VERCE aims to build, within the ESFRI-PP project EPOS, a data-intensive e-Science environment based upon a service-oriented architecture and a framework of services and tools integrating Data, HPC and Grid infrastructures.

VERCE will provide to the solid earth community in the EU: a scalable data, mining and modeling infrastructure, a entry point to access data, sharing and knowledge-based dissemination tools in the seismology community and beyond in the Earth Sciences community that will benefit to societal problems.

VERCE: a three-layer service oriented architecture



Technology Stack

Web portal: Jetspeed, Rapid

Workflow enactment: ADMIRE

Coupling and execution environments: Kepler, MUSCLE, GridSpace

Service gateway and interoperability layer: OGSA-DAI, SAGA, DRMAA

Data infrastructures: Arlink, NetCD, iRODS

Grid and HPC infrastructures: gLite, UMD, UNICORE, OMII-Europe

Federated AAI, single sign-on: Shibboleth, SAML, SLCS, VOMS

Coordination and interaction with:

European data integrated archive systems: GFZ, OGV, RESIF-FR, INGV

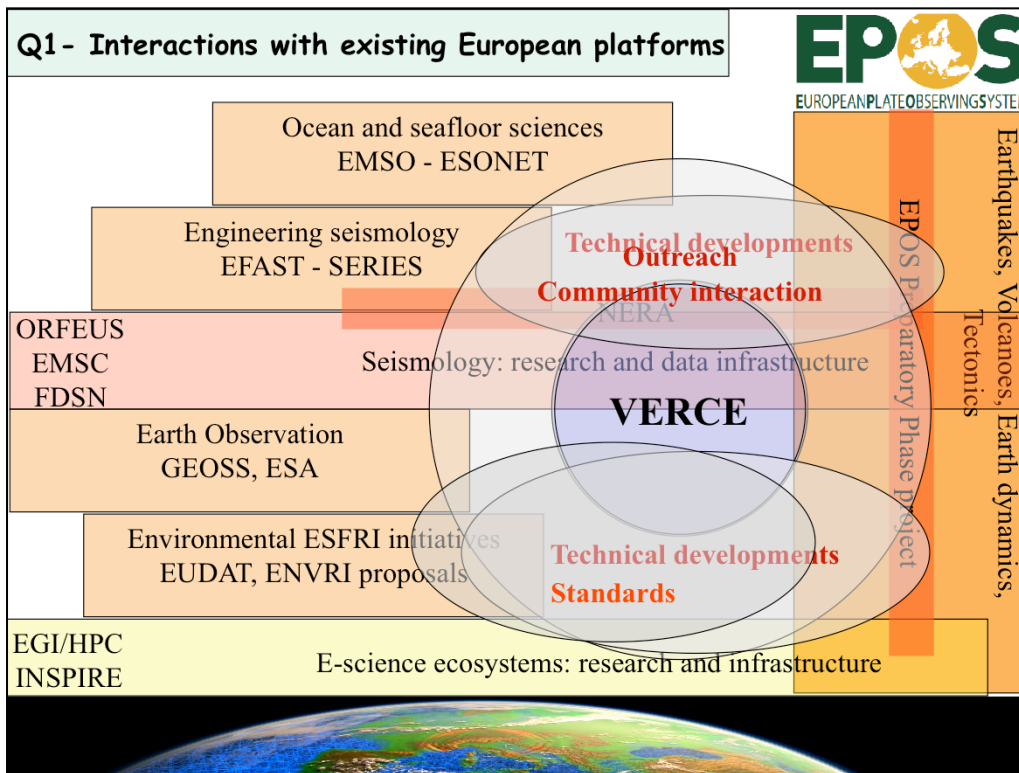
International FDSN data infrastructures: IRIS-DMC, JAMSTEC

Network infrastructures: GEANT, NRENS

HPC infrastructures: PRACE, National HPC centres

Grid infrastructures: EGI, NGIs

VERCE will work in close synergy with European initiatives such as EUDAT and ENVRI, together with EPOS, to insure that the technology stack remains in line with the evolving standards of the community and of the e-Science community.



Q1: How will you interact with existing platforms in the field, such as EMSO and EFAST?

The European Multidisciplinary Seafloor Observatory (EMSO) project is coordinated by INGV and there is close collaboration with EPOS. EMSO is a multidisciplinary project with a wide variety of sensors and physical measurements. The data problematic bares similarities with that of the volcano observatories in EPOS. VERCE will be therefore be in contact with EMSO through INGV and EPOS. A first action will be related with the seismology observations. Discussions between EMSO and EPOS are well advanced. The seismology data of EMSO should be hosted in the EPOS/Seismology data infrastructure. EMSO and EPOS should converge toward a similar data access policy for seismology data. Within this framework, VERCE will interact with EMSO and include the EMSO seismology data access in a first step. Integration of other data will follow the evolution within EPOS, and between EPOS and EMSO. This is in close relation with the Volcano observatories.

EFAST is a European Design Study for Advanced Seismic Testing aiming to create a research infrastructure for shaking table facilities. EFAST is closely linked to Seismic Engineering Research Infrastructures For European Synergies (SERIES) project, collaborating with NERA. Discussions are in progress for developing coordinated portals. VERCE will follow this evolution and possibly interact at some point through SERIES and NERA. The data policy of the Seismic engineering community is more diverse than that of the seismology community. The experimental seismic data of EFAST will probably remain in the facility sites.

VERCE will develop collaborations with ESA under the coordination of EPOS.

Q2- Planned Outreach Strategies: Synergies beyond the consortium

- **European seismology consortia**

ORFEUS (NGO) coordinates waveform seismology archives, European representative in the FDSN (IRIS-DMC, JMA), EPOS representative in ENVRI Initiative; Outreach officer of EPOS, NERA

EMSC (NGO) coordinates earthquake and seismic engineering information
Outreach officer in NERA, SHARE and SERIES

- **IT channels**

NeSC/UEDIN: e-IRG, SSI, ENVRI, EUDAT, EGI, PRACE, ADMIRE, IGE

BADW/LRZ: PRACE and EGI, MAPPER, D-Grid/NGI-DE, IGE

CINECA: PRACE, EUDAT initiative, EMI

SCAI: VRC Earth Sciences (EGI, EGI-Inspire)

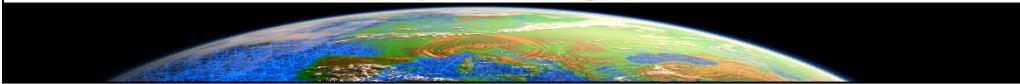
- **Seismology Software and training**

LMU: coordinator of ITN QUEST

CNRS: coordinator of ERC WHISPER, ITN QUEST, NERA, France Grilles/NGI-FR

INGV: coordinator of EPOS and EMSO, EPOS representative in EUDAT, NERA

VERCE outreach strategy will benefit from the extensive and diverse exposure available through the VERCE partners.



Q2 – Please give more details about your planned outreach activities, particularly those aimed at reaching beyond the existing partners.

Q2- Planned Outreach Activities: NA4/NA3

Knowledge-based outreach

Special sessions in large international conferences: AGU, EGU, IASPEI, e-IRG, e-Infrastructures meetings, e-challenges

Coordinated workshops: EPOS (WP8 ORFEUS/EMSC); FDSN (IRIS-DMC, JMA); EC-projects NERA/SHARE/GEM

Scientific gateway: tailored interfaces, data intensive demonstrators, training material

Publications: Int. Journals, EOS/SRL, Nature geosciences

Training and Intellectual ramps

VERCE annual summer camp: community driven training (tutorials, , brain storming ...)

Workshop and training sessions: in synergy with ITN QUEST and ERC WHISPER

Training sessions: in synergy with EGI/NGIs; PRACE/NHPCs

Industrial incentive outreach

Coordinated workshops: with ITN QUEST and ERC WHISPER in direction of the oil and exploration industries (Schlumberger, Total, Shell...) and participation to the EGI and PRACE forums

Public-facing outreach

Website, wiki ... linked to the scientific gateway and ORFEUS/EMSC portals

Early bird initiatives:

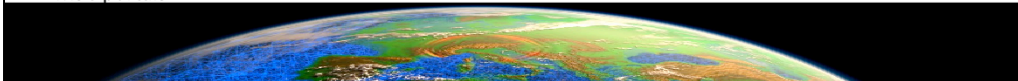
UEDIN/NeSC: e-SI workshops (22-23 November 2010; 7-9 March 2011...)

EPOS: WG7 (e-Science and geosciences) & WG2 (volcanology and supersite initiatives) workshops

NERA: WP9 (web services and portals) technical workshops & WP10 (outreach) user feedback, promotion, interaction

ESC meetings (6-10 September 2010 ...)

EGU meetings (3-8 April 2011...): Earth & Space sciences informatics



Q2 – Please give more details about your planned outreach activities, particularly those aimed at reaching beyond the existing partners.

Q3- Data policies and access rights: some preliminaries

Seismology community

International FDSN open access data policy. In Europe data are available through EIDA.
Standard data formats (miniSEED) and exchange data and metadata formats (SEED)
Standard data and metadata format for earthquake events
Standard data management systems (ArcLink and NetDC)
Enabling service-oriented architecture with Web 2.0 services for data access and retrieval
Central community in EPOS in terms of data infrastructure maturity.

Geodesy community

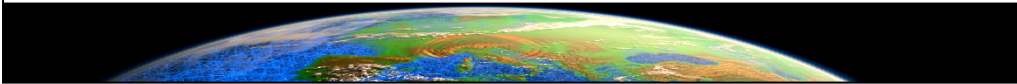
Ground cGPS data: open access data policy for raw data, with data and metadata standards
Standard data management systems and access protocols are progressing rapidly under UNAVCO
Next solid-earth science community to be ready

Other communities

A great variety of data policies for data hinder seamless data access in Europe. Few examples are

Access to ESA/GENESI-DR: policy differs between data sets; access identification done via certificates (Posix ...) but no own root CA: on going access gateways for different authentication methods. VERCE will interact with ESA/GENESI through EPOS and EGI-INSPIRE (VRC Earth Sciences). GENESI/ESA are going toward Shibboleth and SAML in synergy with EGI. VERCE will work on synergy both with EGI and GENESI/ESA.

Access to Earth System Grid (ESG) for climate data: openid and certificates (via myproxy at data centers). Data policies depend on the type of data and today only pragmatic solutions are available. VERCE will be in close contact with this community through EGI-INSPIRE (IPGP/IPSL), and the KNMI. At the KNMI projects of common portals is on progress.

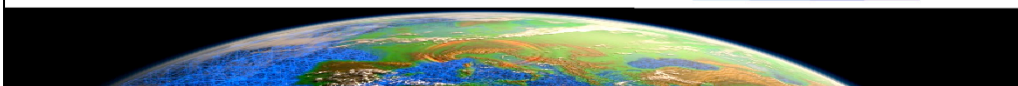
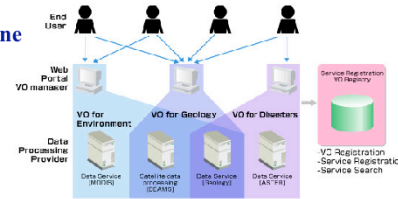


Q3 - How do you propose to manage access rights to the data owned by different communities?

Q3- Managing Data Access - technical strategies for VERCE

Interplay of Socio-Economic and Technical strategies

- **Provenance will be automatically traced and generated**
 - Validate observance of data policies
 - Tracking data ownership
 - Analysis of multi-source data use
- **Support recognized policy rules and data-creator's preferences for their data products**
 - Variations for different data modalities: mobile ground-motion sensors, GPS, Satellite, ...
 - Integrates access constraints in composite data
- **Automated generation of data identifiers for data and originator citation**
 - Increases data-owners' confidence that use of their service will be reported
- **Make it very easy to use data from any source**
 - Simple "press-of-button" on the Science Gateway triggers action across integrated technology
 - Simple access components for workflows hide heterogeneity and handle data-source evolution
 - Prototypes of this pioneering work are already underway in the consortium
- **Heterogeneous platforms use different security models**
 - Automatic mapping of credentials to avoid users having to handle this complexity
 - OGSA-DAI proxy presents appropriate credentials to data owners & implements their rules
 - Make new researchers welcome with light-weight security
- **WS-DAI, SAGA, OGSA-DMI standards and Globus Online**
 - insulate VERCE from e-Infrastructure heterogeneity and evolution



Q3 How do you propose to manage access rights to the data owned by different communities?

There are two aspects to this (1) the socio-technical addressing data-providers and users willingness to use multiple resources, and (2) the technology – e-Infrastructure issues.

The socio-technical strategy is to make things easy for researchers. This means that heterogeneity is hidden by mapping mechanisms and data-access adapters housed in instances of OGSA-DAI. The global agreements on data formats from FDSN, and on protocols from IETF, OASIS and OGF shape these adapters. As the e-Infrastructures (EGI, PRACE, ENVRI & EPOS) agree their standards for services and interfaces the VERCE system will adopt them. Existing investment by the VERCE team will be protected, by adapting the underlying mappings. VERCE will give guarantees to data providers that their access policies will be honored using OGSA-DAI to map credentials and to call security protocols. An example is the GeoGRID. The generation of provenance data, currently being jointly prototyped by KNMI-ORFEUS and UEDIN, supports compliance verification. To make it easy for researchers to make their data available a simple "press-the-button" to validate and submit is provided in the Science Gateway.

The heterogeneity of e-Infrastructures requires mechanisms to manage mappings, e.g. the different security requirements of HPC, Grid and institutions. VOMS (supported by EGI) will maintain a consistent Virtual Organization Membership Service across heterogeneous e-Infrastructures. The OGSA-DMI and the new capabilities of GridFTP (separate certificates for source and destination) will be used for data management and movement. SAGA will make VERCE code transferable between platforms. Where institutions use Shibboleth-based security, OGSA-DAI will provide the mapping, 'Supporting security-oriented inter-disciplinary research: crossing the social, clinical and geospatial domains', Sinnott, et al., e-Social Science, Germany 2009.

Q4- More details on the proposed metrics for the impact of VERCE		
Data intensive research multiplier	New scientific applications and research projects Scientific and technological breakthroughs Cross-interaction among projects and disciplines	# software and tools with number of active users # publications and invited talks referencing VERCE # usage of VERCE by other projects # number of PhD thesis using VERCE
Trust & acceptance	Virtual Organizations Sharing of data and methods Interoperability and efficiency	# reports on VOMS and RBAC use # tailored application interfaces in the scientific gateway and usage statistics # shared repositories of reusable data, models and workflow tools and use reports # orchestrated workflow applications # Monitoring reports
Awareness & agility	Training and tailored demonstrators Ease of data access and resource usage Users and applications lifetime	# participants in training sessions # coordinated meetings with EPOS, QUEST, NERA ... # level of satisfaction of the trainees # new users after training sessions and users retention
European research infrastructures	Contribution to European research infrastructures	# meetings within EPOS # VERCE solutions retains in EPOS e-Science environment # Contributions (use cases, resources) in other EU initiatives (EUDAT, ENVRI ...)
International impact and visibility	In the solid earth community In the e-science community	# joint meetings with FDSN (IRIS-DMC, JMA, UNAVCO ...) and collaborative actions # invited talks in e-Science meetings
Society & Economy	Research and societal cost Industrial and Societal impact	# metrics to be defined in collaboration with ERINA+ and SEQOIA (for SaaS) # Industrial forums (QUEST, WHISPER)

Q4 – Please describe in more details the proposed metrics on impact of the project.

To ensure the earthquake and seismology community can take full advantage of the existing data and computing infrastructures 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.

In the field of earthquake and seismology Virtual Organisations will be established to shape the services of the VERCE platform. The number of VOs using the platform is used as an indicator for the adoption of the services in the scientific community. Alongside some more improvements this will help to increasingly satisfy the needs of the users. This will be measured by assessing certain quality criteria (comfort, efficiency, usability, etc.) from a user viewpoint.

In order to provide the scientists with the most suitable resources it is necessary to combine the use of data intensive and compute intensive methods in a coordinated way. Therefore integration and interoperability of services in the VERCE platform is a crucial factor of the project. Users must be in a position to easily use the services provide by the platform. The time needed for users to really make use of the services will be a key performance indicator for the quality of the platform. Furthermore the platform should adopt as many services from the underlying infrastructures as possible.

For a scientific gateway it is also intended to have a multiplication effect. High quality portals will attract more users and additional scientific communities. This will also result in an increasing number of publications created using the VERCE platform. In addition the number of users that have been trained to use the portal in one of the training sessions provided by VERCE is a key indicator for the adoption of the platform and services in the scientific community addressed here.

In important aspect for the training activity will be the number of science-driven tutorial and demonstrators. These tutorials must be built on simplified version of datascope that appeal to the researchers of the various communities. Through tailored interfaces, the demonstrators must show how the use of the VERCE platform allow data-intensive applications that would not be practical or feasible with local resources.

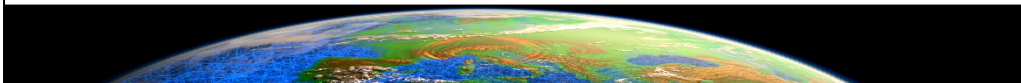
With its close relations to national hazards and risk assessment, energy resources, and environmental change better and faster earthquake and seismology research has a direct impact on society and economy. Improving the research efficiency and thereby reducing the costs through an efficient use of the research infrastructures will help to accelerate the positive effects of this research o the global society.

For the assessment of the impact of the project on society and economy the ERINA+ methodology for socio-economic impact analysis of e-Infrastructure projects will be applied:

- Step 1: Mapping the areas of impact
- Step 2: Defining the baseline and alternatives
- Step 3: Measuring or estimating impact indicators
- Step 4: Exercise final analyses

Q5- Overcoming potential non-technological barriers

- **Build on existing culture**
 - Long-established global sharing of ground-motion data
- **Conform with EPOS, ENVRI, EUDAT policies**
 - Influence and engage in larger community behavior
- **Build Trust**
 - Accurately implement data owners' & researchers' wishes
- **Make it easy to share**
 - "Press-the-button" validate, submit to archive or publish
 - Automatic completion of metadata requirements
 - Support and track different rules for different observation modalities
- **Stimulate sharing by supporting collaboration**
 - Data + methods + workflows + software + opinions + ...
 - Build on Web2.0 methods
 - myExperiment to ADMIRE Registry to Seismic-Community support in the Science Gateway
 - Recognition of contribution; user identifies friends, FOAF recommenders, Citation of contributions
- **Demonstrate benefits of sharing**
 - accelerating and improving research success



Q5: How do you propose to deal with potential non-technological barriers to the sharing of data and other resources?

Established practice of open access to ground-motion sensor data from fixed arrays provides a starting advantage, as does the collaborations within the consortium. Mores and a kernel on which to build. The wider community is a key influence, particularly global agreements in GEOSS and FSDN, and European agreements in EPOS, ENVRI and EUDAT. VERCE will help establish sharing behaviors.

Trust is a necessary precursor to sharing. VERCE will develop this by honoring data providers' and researchers' wishes regarding access to their data.

Sharing is often inhibited because it takes too much effort to prepare data. The provenance tracking and validation will automate as much as possible. This will be conveniently integrated into the Science Gateway using RAPID.

Sharing is helped by providing semantic information in the tools on the Science Gateway, which will build on the ADMIRE Registry.

This uses myExperiment technology. That can used to support much more effective community building and sharing, e.g. based on FOAF recommendations. This will be combined with EU OpenKnowledge project experience.

Towards an e-Science environment for seismology and EPOS

- Provide a data intensive service-oriented e-Science environment to the EPOS community
- Lay the basis for transformative data-intensive research in the solid earth sciences
- Provide solid-earth science contributions to the European data and computing infrastructures
- Put Europe in a leading position in the international seismology and solid earth community
- Build trust and collaborative models for sharing of data , methods and tools
- Engage a new generation of researchers and experts in solid earth data intensive research

A European and International domain context

Integrated European distributed Data Archives (EIDA), part of the international FDSN

A number of coordinated European projects in seismology: NERA, SHARE, GEM, ERC WHISPER, ITN QUEST...

The European Plate Boundary Observation System (EPOS): the ESFRI-PP project

Active collaborations within the FDSN with the US IRIS-DMC, and the Japanese JAMSTEX and NIED

DataNet initiative in the US (IRIS-DMC)

A European e-Science context

Fast evolution of seismology services and applications

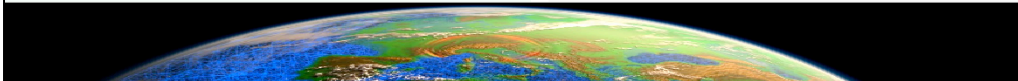
RapidSeis Portal for Accessing & Processing FDSN archives

ADMIRE seismic ambient-noise prototype

European initiatives: EPOS, ENVRI and EUDAT

Converging e-Infrastructure ecosystem: EGI/NGIs, PRACE/NHPCs, GÉANT

Emergent data access and single-sign on protocols



A joint team from KNMI, ULIV and UEDIN, built on the NERIES project, to deliver a convenient portal for accessing and processing data from seismic archives supporting the FDSN standards.

A joint team from KNMI & UEDIN are working under the auspices of the ADMIRE project on a Seismic Noise Data Processing system drawing data from multiple European seismic archives, including the BGS and ORFEUS.

Data Intensive applications

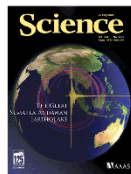
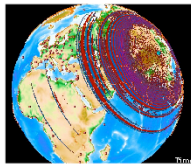
Earthquake and seismology community is facing a fundamental paradigm shift: from data driven to data intensive research:

Large volume data analysis: extracting information from space and time correlations in dense array observations, Data and computing intensive simulation/inversion: 3D wave form information using adjoint methods, stochastic strong motion simulation, Orchestrated workflows across service components.

Seamless access to large volumes of multi-sets data across the Grid and HPC components

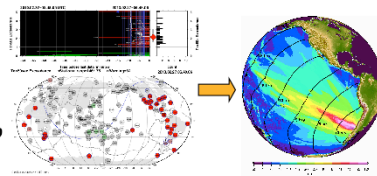
Industrial and societal applications: natural hazards, climate changes and energy resources and national security.

Large earthquake source radiation: Sichuan (Mw 7.9, 2009, China); Sumatra-Andaman (Mw 9.2, 2004, Indonesia)

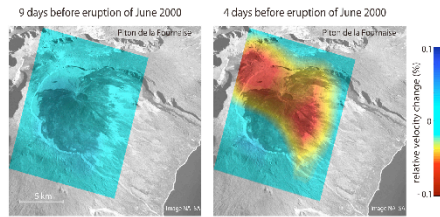


Research groups using SPEC-FEM3D

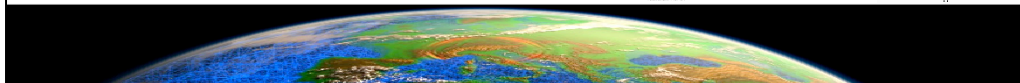
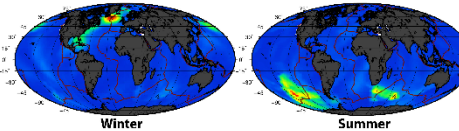
Earthquake detection: tsunami impact maps



Seismic noise correlations: observing precursors to volcanic eruptions



Studying the coupling between the Solid Earth, the Oceans, and the Atmosphere



Thank you!

