Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation

Report

D9.1 EMEPS Initial Implementation and Integration Strategy

Activity:	*European-Mediterranean Earthquake Portal and Services*
Activity number:	*NA9*
Deliverable:	*Initial Implementation and Integration Strategy*
Deliverable number:	*9.1*
Responsible activity leader:	*EMSC*
Responsible participant:	*EMSC*
Author:	*Linus Kamb, ed.*

Seventh Framework Programme
EC project number: 262330

Summary

The NERA NA9 effort was launched with a review-and-planning phase. During this phase we have undertaken both an analysis of the current status of the Earthquake Data Portal and related projects, as well as an initial planning for the future developments and integration of new partners, tools, and services. The Earthquake Data Portal, based on an open standard architecture of JSR-168/286 Java Portlets supported by an underlying layer of web data services, continues to be an operational success, with the numbers of registered users and data requests continuing to increase. The core architectural principles have proven effective both in terms of development and maintenance, and also with regard to the integration of new tools and partners. This architecture has been chosen as the basis for developments in other integrative European projects, including SHARE, VERCE, and EPOS.

As part of the review process, a survey was developed to solicit feedback from the NERA project participants regarding their interests and requirements with respect to the Data Portal and their plans to provide new integrated data access tools within the portal for their datasets. After the review of the response to the internal survey, another survey will be developed to be sent to the registered user-base to solicit detailed feedback concerning usability aspects of the current portal, toolset, and data access services. Responses from these surveys will then provide background and guidance for decisions concerning future developments and enhancements.

The technical developments under NERA NA9 focus on two core activities: the expansion of the Data Portal with new partners, access tools, and data sets, and in particular bringing in the engineering community, and also the definition and development of a standardized data services layer. It is expected that the data access tools developed by the new partners will be developed within the current operational and robust web services and Java Portlet-based architecture. The web services track adopts the OGC (Open Geospatial Consortium) web services stack as the basis for the standardization effort. The Open Geospatial Consortium supports numerous broadly-accepted standards, and the OGC web services stack provide a set of accepted standard interfaces to access a variety of services and information types, including geospatial information, catalogue and inventory information, as well as online service-based processing services. By adhering to accepted standard interfaces, the NERA-developed services will be more broadly accessible and suitable for integration with other projects.

In addition, broad collaboration efforts with other projects are in place to ensure coordinated development efforts and to avoid both duplicative efforts and divergent strategies. The primary targets of these coordination efforts are in the areas of standards development, particularly with respect to service interface definitions, metadata vocabularies, and the overall common services architecture. The definition of a coordinated and coherent resource-oriented uniform identification scheme, along with standard service metadata information and invocation APIs provide the foundation for building the Virtual Data Workbench and the coordinated development of workflows and distributed data and service access.

1. Introduction

NERA NA9 effort, the European-Mediterranean Earthquake Portal and Services, will provide the platform for the integration of data access tools for the data and results of the NERA work packages. The Earthquake Data Portal (EDP) provides interactive tools for the discovery, access, and visualization of available data sets and is the basis for integrative efforts between the seismological, earthquake engineering, and hazard and risk communities. The definition and adoption of standards for service invocation, metadata interfaces, and shared vocabularies, known collectively as the Common Services Architecture, provide the foundations for the development of higher-order applications and scientific gateways.

2. Background Information

NERA NA9 effort, the European-Mediterranean Earthquake Portal and Services, provides the platform for the integration of data access tools for all the NERA work packages. Through the definition and adoption of standard data access interfaces and metadata vocabularies, in conjunction with other leading European projects, the NA9 provides the foundation for cross-domain and inter-project integration efforts. This section provides some background information on the participants, related projects, and other information.

2.1 Participants

2.1.1 EMSC

The European-Mediterranean Seismological Centre (EMSC) leads the NA9 effort. The EMSC was established in 1975 by the European Seismological Commission to provide rapid earthquake information at the Euro-Med and global scales. Hosted by the Laboratoire de Détection Géophysique (LDG) in France, the EMSC provides several core rapid earthquake information and alert services for the public, research communities, and several Emergency Response and Civil Protection services. The main scientific activities include the rapid earthquake information services and the production of the Euro-Med Seismological Bulletin. In addition, the EMSC participates as a partner in numerous national and international research and public safety projects including NERIES, NERA, VERCE, EPOS, SIGMA, RELEMR, and EERWEM. Currently, the EMSC membership includes 84 seismological institutes from 55 countries spanning the entire Euro-Med region.

The rapid information services compile real-time parametric data provided by 64 seismological networks. This compiled information is available on the EMSC web site (http://www.emsc-csem.org). In addition, alert notification services are provided via email, SMS, and fax. Other media are also available, including imode, WAP, and RSS and Twitter feeds. The web site provides lists and map-views of the latest earthquakes, as well as of any selection from the EMSC earthquake catalogue. The provided information includes all the compiled parametric data, as well as additional information including location maps, past regional seismicity, and submitted moment tensor solutions. As part of an ongoing program to increase citizen awareness and involvement, the EMSC also collects eye-witness macro-seismic questionnaires and pictures of earthquake damage and effects. These eye-witness submissions are provided on the earthquake pages if available.

The EMSC also produces automatically-generated earthquake FeltMaps. FeltMaps, developed at the EMSC, provide a visualization of the degree to which an earthquake was felt by the local population. These maps are generated by the real-time analysis and tracking of our web site traffic. When an earthquake has been felt by people in the

Euro-Med region, they often rush to our web site to find information about what they have just felt. This causes a peak in our web site traffic. By statistical analysis and real-time spatial clustering of the location of the IP addresses of the visiting web traffic, we can identify the locations of significant increases in web site traffic. These locations are then plotted on a map of the region. Using advanced analysis techniques, we are able to detect the occurrence of an earthquake in less than two minutes, which is considerably faster than standard automatic earthquake detection algorithms.

The Euro-Med Seismological Bulletin (Godey et al.; 2006) is based on the EMSC's comprehensive collection of bulletins and arrivals provided by participating networks operating within the region. By merging data from 78 contributing networks and by computing locations in a homogeneous way, the EMB provides a reference catalogue of Euro-Med seismicity since 1998. The EMB is currently available through mid-2009 and contains over 250,000 events.

2.1.2 ORFEUS

The Observatories and Research Facilities for EUropean Seismology (ORFEUS), manages the NERA project,and participates among others in NERA-NA9. ORFEUS, founded in 1987, is a non-profit foundation with an international board of directors (14) and more then 60 participants (academic and other reseearc institutes). ORGFEUS objective is to coordinate and promote digital, broadband (BB) seismology in the European-Mediterranean area. The ORFEUS staff is hosted by the seismological division of the Royal Netherlands Meteorological Institute (KNMI) in The Netherlands. An international Executive Committee oversees the day-to-day operations.

Facilitating digital waveform data availability

The ORFEUS Data Center (ODC) retrieves and archives digital waveform data in agreement with observatories in and around Europe within the framework of the Virtual European Broadband Seismograph Networks (VEBSN: www.orfeus-eu.org/Data-info/vebsn.html). Currently more then 55 seismological networks from more then 40 countries are participating in this waveform data exchange pool. The ODC applies a number of Quality Control procedures and thereafter makes this waveform data available to the research community through a number of services (www.orfeus-eu.org/Data-info/available.html) and the European Integrated waveform Data Archive (EIDA). The EIDA is a joint effort of major European seismological data centers; GFZ in Germany, Resif in France, INGV in Italy and the ODC in The Netherlands.

Operating and innovativing data services

Within recent, existing and new project the ODC is working with innovative techniques to improve data quality services and facilitate data access through multiple techniques. Data quality services are being generated (www.orfeus-eu.org/Data-info/dataquality.html) and webservices are implemented to enable users to facilitate data integration in their own programs (www.seismicportal.eu) and section 2.2.1. Developments in NERA-NA9 will further pursue this development is close coordination with the new EC project 'Virtual Earthquake and seismology Research Community in Europe e-science environment (VERCE)' (see section 2.2.2).

Coordinating waveform data exchange in and around Europe

Besides actively participating in the project developments, ORFEUS also either coordinates or manages EC projects or participates in EC projects that aim at improving the European seismological research infrastructure with open data access to digital seismological waveform data of specific concern. Examples of both past and new projects are:

MEREDIAN - www.orfeus-eu.org/Organization/Projects/Meredian/meredian.html (2000 - 2005),
NERIES – www.neries-eu.org (2006 – 2010),

NERA – www.nera-eu.org (2010 – 2014),
EPOS PP – www.epos-eu.org (2010 – 2014), specifically Working Group 1 seismology

The ORFEUS foundation organises further annual observatory coordination and user interaction workshops, facilitate information relevant to research infrastructure users through its web sites of the projects and ORFEUS, and ensures a broad participation of stakeholders beyond the project consortia. Global coordination is facilitated by hosting the secretariat of the International Federation of Digital Seismograph Networks (FDSN – www.fdsn.org) and close coordination of developments with IRIS *) and its Data Managemen Center in the US.

*) IRIS – here Incorporated Research Institutions for Seismology (not be confused with IRIS in section 2.2.6).

2.1.3 SED-ETHZ

The Swiss Seismological Service (SED), attached to the Federal Institute of Technology Zurich (ETHZ), is a combined seismic monitoring facility as well as a research institute for a wide range of topics in seismology.

Responsible for Swiss national seismic monitoring and hazard assessment, the SED runs the Swiss National Broadband Network (SDSnet, currently 36 stations) and the Strong Motion network (SSMnet close to 100 stations, of which an increasing number recording continuously), which are amongst the densest in Europe, but also several smaller networks for reaearch (AlpArray, COGEAR) and monitoring (e.g. geothermal exploration projects, tunnel drilling). SED hosts and provides continuous and event specific seismic waveform data (via EIDA and ArcLink) as well as the national earthquake catalog, and it is the National Data Center of CTBTO.

Earthquake research is carried out in different groups focusing on statistical seismology (mainly probabilistic hazard and risk), earthquake source, engineering seismology, real time ground motion assessment, and computational seismology. SED is involved in university level teaching and hosts a considerable number of doctoral and postdoctoral students.

SED has a strong involvement in European and international infrastructure and research projects. Among others, it is/was a key contributor to the European research programs NERIES, SHARE, MATRIX, GEISER and EPOS (preparatory phase), it hosts the modelling facility of GEM (Global Earthquake Model), the European testing center of CSEP (Collaboratory for the Study of Earthquake Predictability) , and – widely within the framework of NERA – develops EFEHR (the European Facility of Earthquake Hazard and Risk)

2.1.4 VCE

The Vienna Consulting Engineers (VCE) participates among others in NERA - NE9. Within the NE9 VCE contributes to: Initial implementation and integration, Architecture specification for data workbench and User interaction specifications.

The company was founded by Professor K. Wenzel, and has made a major contribution to the technical advancement and aesthetic evolution and refinement of modern bridge design in Austria, and all around the world. And since the formation in 1980 of VCE that contribution has expanded to an international level.

VCE is an independent, high tech oriented consulting firm with its head office in Austria. The company operates in three principal lines of business:

- the transportation sector (including bridges, tunnels and railways)
- the building and industrial sector (as well as general design and management as well as specialized technological expertise such as earthquake engineering)

- and the development sector (from research and development to feasibility and environmental studies, financial engineering, to development aid).

VCE has major operations in Austria, Taiwan, Korea, Eastern Europe, the Middle East and Africa. To date, over 3,500 contracts have been successfully completed in 62 countries world-wide with larger, more exciting projects currently underway. The key personnel at VCE consist of experts in many highly specialized fields with long experience records. The versatility of their skills, professional expertise and on site experience meets the challenges of contemporary engineering and construction around the world. Close cooperation by the firm with major Austrian universities provides additional expertise to the company when required.

The company receives support from the experts at the Technical University of Vienna, Expert Pool and the Austrian Governmental Authorities. With vast experience and wide capabilities VCE is able to provide optimal solutions, resulting in a technical as well as an economic success.

Research & Development plays a key role in the company's strategy. VCE is reliable partner for the development of innovative methods in the field of construction and infrastructure projects. More challenging and interesting projects are underway or under implementation in the 5th Framework Programme. Particular attention is paid to company's coordination skills and the innovative drive in projects. VCE coordinates many research projects such as: MIMOSA, SEISMID, DESTRail, etc. Information about research projects and applied technology is available on the VCE web site (www.vce.at).

2.1.5 JKU / FAW

The Institute for Application Oriented Knowledge Processing (FAW) was initiated by Univ.-Prof. Dr. Roland Wagner as a Research Institute of the Johannes Kepler University (JKU) of Linz. Since February 1992 FAW has been located at the Johannes Kepler University of Linz and at the Softwarepark Hagenberg, 25 km north-east of Linz (Upper Austria).

In 1997 the research institute was integrated into the organization of the university and is now an independent institute of the Johannes Kepler University of Linz.

The research institute's major domains of education, research and development are Advanced Information Systems, Data- and Knowledge Modeling, Data- and Knowledge Centered Systems, Semantics, Web Engineering and Decision Support.

Since its foundation, application orientation is an important attribute of FAW's work. More than hundred projects with research partners and partners from trade and industry have been finished successfully. These cooperations are the basis for an extensive research work at FAW and practice-oriented teaching at the University.

Since 1990 FAW is co-organising the annual international conference event DEXA (Database and Expert Systems Application) which has grown to a world wide well accepted point for publishing and sharing newest research and development result.

2.2 Related Projects

2.2.1 NERIES

Submitted by Alessandro Spinuso, Luca Trani, Torild van Eck, ORFEUS.

The overall vision of NERIES was to integrate the Research and observational infrastructures for earthquake seismology in Europe. And thus maximize the return on the more than 200 MEuro national investments into the seismological infrastructure, and on the 40MEuro annual operating cost. NERIES has created an operational Research Infrastructure concept for the European Plate Observing system (EPOS), including the basic cyber-infrastructure component.

NERIES implemented an advanced comprehensive European integrated Research Infrastructure (RI) for earthquake seismological data that is scalable and sustainable. The project made a significant amount of seismological data openly available, integrated different distributed data archives, implemented and produced many advanced analysis tools and made a large number of relevant advanced software packages and tools available.

A single data portal (www.seismicportal.eu) provides a single access point and overview for many of the seismological data currently (and in the future) available for the earth science research community. A significant additional amount of data access tools and sites have been implemented to meet user and robustness requirements, notably those of the EMSC (www.emsc-csem.org) and ORFEUS (www.orfeus-eu.org).

The datasets compiled in NERIES and available through the portal include among others:

• The Virtual European Broadband Seismic Network (VEBSN) expanded significantly with real-time access to more then 500 stations from > 53 observatories producing a continuous flow of data. This data is continuously monitored, quality controlled and archived in the newly created European Integrated Distributed waveform Archive (EIDA).

• A unique integration of acceleration datasets from seven networks in seven European or associated countries centrally accessible in a homogeneous format, thus forming the core comprehensive European acceleration database. Standardized parameter analysis and actual software are included in the database.

• A Distributed Archive of Historical Earthquake Data (AHEAD) for research purposes, containing among others a comprehensive European Macroseismic Database and Earthquake Catalogue (1000 – 1963, M ≥5.8), including analysis tools.

• Data from 3 one year OBS deployments at three sites, Atlantic, Ionian and Ligurian Sea within the general SEED format, thus creating the core integrated data base for ocean, sea and land based seismological observatories.

NERIES Data Portal

The primary objectives of the NA9 activity were to implement the Earthquake Data Portal and integration strategies as defined in NA7, promote their use, and to collaborate and coordinate with other relevant European and global organizations and projects. The NERIES project Data Portal has been developed and is publicly available, with nearly 1000 registered users and visits from nearly 100 countries around the world. It is available for use at http://www.seismicportal.eu.

An additional focus has been the development of web data services that both support the portal and provide programmatic access to ORFEUS and EMSC data holdings. Several external clients have been written to access these data services.

The development of the portal and supporting services has involved coordination and collaboration with several different organizations, and has fostered on-going discussions with other research projects and global standardization organizations.

Several information dissemination mechanisms are used to provide information and updates about the portal, including maintaining a Blog and publishing updates via Twitter and to an iGoogle Gadget.

A complete report on the portal is made in the NA7 Deliverable 7: Portal Final Report, available on the NERIES project website: http://www.neries-eu.org.

Objectives, methodologies and approaches

One goal of this activity was to solicit and incorporate user requirements and feedback. In order to have feedback, there must exist something about which to comment. This leads to a chicken-and-egg problem: to request feedback, you must build something; but you do not wish to spend considerable resources building something before you have a clear picture of user requirements. To address this, portal development followed an "iterative development" methodology. In this approach preliminary, partial, or prototype versions would be provided for user review and feedback. From the received feedback, changes and new functionality would be considered for inclusion in the next version. This leads to a final product more closely aligned with user needs. The success of the approach in this activity is demonstrated by the highly-positive responses to the final portal.

An additional focus of this activity has been the development of web data services that both support the portal and provide programmatic access to ORFEUS and EMSC data holdings. The foundational architectural principle of the portal is based on Service-Oriented Architecture. This services-based approach allows the re-use and alternative compositions of the core services. The data services provide an accessible interface to outside clients, and several external clients have been written to access these services. This architectural approach, and the availability of core data services, provides the foundations for leveraging the work done within this activity to build future seismological processing applications.

Attempts have been made to adopt and promote community standards. The QuakeML (http://www.quakeml.org) XML format for event parametric data is such a developing community standard. Outside of the NERIES project, which uses it extensively, the US Geological Survey and the Incorporated Research Institutions for Seismology (IRIS) are adopting its use.

The availability and use of the portal and services has been promoted at several conferences, meetings, and through workshops and individual presentations. Posters, presentations, and demonstrations were made at each of the annual European Geophysical Union General Assembly meetings over the duration of the project, as well as at the American Geophysical Union General Assembly meetings in San Francisco, USA. Several targeted workshops were either hosted by the activity or attended by project members to coordinate developments between projects and other NERIES activities, and to present and promote the portal, its architecture, and supporting web services. Through many of these activities, new contacts and avenues for collaboration have been developed.

Several methods have been employed to provide information about the portal and to disseminate updates on available datasets and new earthquakes. The Earthquake Data Portal Blog, at http://neriesdataportalblog.freeflux.net/home/, hosts portal help and background information, as well as publishes latest news updates. The portal blog is also available as an RSS feed. The latest events and shared datasets are published via Twitter (http://twitter.com/EarthquakeDP) and to a downloadable iGoogle Gadget.

Results

The primary result of the activity in NA9 is the development and deployment of the NERIES integrated Data Portal, providing access to several different project activity datasets. The portal, deployed on redundant, high-performance, high-availability hardware, provides a single access point to data and results from the EMSC earthquake catalogue and NERIES activities NA3, NA4, and NA5. In addition, it has been demonstrated as a viable platform on which to integrate access to activities within JRA1 and JRA5.

Another important result is the availability of web data services. These web services support the operation of the portal user interface, and also provide an alternative, programmatic access point to ORFEUS and EMSC data holdings. Publicly available services include:

- QuakeML - Event Service.
  Provides event query interface. Event parametric information returned in QuakeML format.

- SeismoLink - Web service wrapper to access ArcLink / EIDA
  Provides a simplified, web-service access to the ArcLink-based European Integrated Data Archive system.
- SMI-RDF - Seismological Meta Information
  Access to RDF-based meta-information maintained by the portal systems. Provided as SPARQL query endpoint.
- TauP - Travel time calculation
  Theoretical seismic wave travel time calculation. Three velocity models are provided, as well as the option to specify ones own model.

Several client applications to these services have been developed and are available, including:

- JOQUE - Java Orfeus QUake Explorer.
  Java client to the SeismoLink service. Retrieves event-windowed waveforms using CMT or QuakeML catalogues.
- PORSCHE - Perl ORfeus Seed Control and Harvest Engine
  PERL client to the SeismoLink service. Time-delimited waveform retrieval.
- SeismoWeb - Waveform processing toolkit.
  Matlab functions to retrieve waveforms from SeismoLink service.
  Developed by the Research Geophysics Group, TU Vienna.
- getUNID - Retrieve EMSC event Unique Identifier.
  Search for and extract EMSC event Unique Identifier from parametric information in order to perform event data association. Available as both a command-line client from EMSC and as a Matlab function from KOERI.

Other data services have developed in support of the portal, although they are not currently generally available. These include the data cart services. The data cart services support the portal user interfaces and inter-portlet communications. They are also the basis for exporting dataset query results to external services, such as in the RapidSeis system. Further developments will expose these services in a controlled manner to allow linking Data Portal results with other external systems.

More information on the web services and download links for available clients is available at: http://www.seismicportal.eu/jetspeed/portal/web-services.psml.

Impacts

The final revision of the NERIES Data Portal has been publicly available since April, 2010, with previous versions online starting April, 2008. Since the new version has been made available, there have been nearly 1000 new users registered, with visitors from nearly 100 countries around the world. Feedback received from users has been overwhelmingly positive.

The portal has provided improved access to datasets that were previously difficult to impossible to access uniformly. The combination of Event Explorer and Waveform Explorer tools provides an intuitive and coordinated mechanism to access event-oriented waveform data from participating EIDA networks. The Accelerometric Explorer and supporting data system provide the heretofore unavailable ability to request and download available data through a single tool from all the participating accelerometric networks.

The development of the core web data services has enabled the development of external access clients which are being used in operational programs in several locations. These external clients are examples of the next generation of scientific applications which will collect data directly from data sources and orchestrate distributed computation through invocation of web computation services available in the cloud, from high-performance computing centers, or other from other providers. This services-based architecture is proposed as the foundational principle for several proposed and up-coming research and e-Infrastructure projects, including NERA, EPOS, VERCE, and others.

A highly valuable result has been how the portal and web service demonstrations have provided the basis for further discussions with other research and e-Infrastructure projects and international standards organizations. In particular, the OneGeology effort would like to integrate access to the seismic event catalogue in to their online geological maps, and the Open GeoSpatial Consortium (OGC) international GIS standards organization was impressed with the data portal and has expressed interest our input in pursuing on-going discussions to coordinate the data service efforts with developing standards, thereby ensuring a much broader audience for the data services.

2.2.2 VERCE

Submitted by Alessandro Spinuso, Luca Trani, Torild van Eck, ORFEUS.

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.

Community developer 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.

11During the project the developers feed-back will be continuously monitored and analysed to improve and re-orient some of the developments and services.

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.

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.

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.

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.

2.2.3 EPOS

Submitted by Alessandro Spinuso, Luca Trani, Torild van Eck, ORFEUS.

The European Plate Observing System (EPOS – www.epos-eu.org) is a Research Infrastructure initiative within the 2008/2010 roadmap of ESFRI (European Strategic Forum for Research Infrastructures: http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri) of the EC. The EPOS Preparatory Phase project (EPOS-PP; 2010 – 2014) aims at constructing the EPOS business plan. Below we provide a short introduction to the EPOS concept and general plan.

Scientific Landscape

European earth scientists have played a major role in the study of plate tectonics during the past 50 years making great advances, opening new horizons and creating a framework to explain, to first order, plate dynamics and to model deformation processes. European research and monitoring infrastructures have also gathered, largely on national scales, a vast amount of geological and geophysical data. These have been used by research networks to improve our models describing the active deformation processes that generate earthquakes, volcanic eruptions, landslides and tsunamis. Since the 1970s and the 80s (establishment of data centres like ORFEUS and EMSC) there have been European scale initiatives to integrate the existing national research infrastructures in solid earth science to facilitate the exchange of data, information, modelling and monitoring tools. EPOS is timely in integrating these mature initiatives into a single infrastructure enabling earth scientists across Europe to combine, model and interpret multidisciplinary datasets at different scales. The concepts embodied in EPOS are vital to maintain European scientific competitiveness on the international stage.

Meeting the integration challenges

Integrating Earth science infrastructures in Europe presents a double challenge - firstly, a highly heterogeneous geographical pattern of both observational and experimental multidisciplinary data must be integrated to facilitate complex analysis and modelling. Secondly, e-infrastructures and e-science must be developed to support the EPOS construction. The earth science community is well-placed to address these challenges as we have significant experience in coordinating and adopting format standards for data exchange in the global geophysical community, in seismology in particular, and we are well placed in international initiatives. Long term experience in governance of a widely distributed community has been obtained for example in seismology through NGOs like ORFEUS (since 1987) and EMSC (since 1975).

User groups: training the new generation

EPOS will be an open access infrastructure .Several thousand researchers in the earth sciences will benefit from the services provided and this will foster major advances in the understanding of the dynamic processes occurring in the Earth. The EPOS infrastructure will serve as primary source of data for creative young researchers whose scientific discoveries will feed into the development of the infrastructure itself and the service it provides. As a newly integrated RI (Research Infrastructure), EPOS needs to be accompanied by a coherent training programme for the Earth Science user community to ensure that the new generation explores the full potential of the integrated RI and the e-science facilities it incorporates. To this end, EPOS will start a competitive fellowship program dedicated to young researchers. We expect, from current small-scale experiments with e-science inspired laboratories, a revolutionary development in the understanding of solid Earth dynamic processes.

EPOS is aimed at a broad user community including European and Mediterranean countries. Promoting the participation of other countries will further enlarge the initial pan-European composition of EPOS. However, because the development of EPOS is a result of previous and ongoing EC projects (such as NERIES, EXPLORIS or SPICE) with input from European scientific organizations (such as ORFEUS and EMSC), the user community has already strongly contributed and influenced the design of this proposal. Furthermore, access to geophysical and geological data will be attractive for both European and international scientists. It is expected that EPOS will have an impact also on other scientific fields because of its multidisciplinary nature.

e-Science

Solid Earth e-Science

Advances in Information Science Technology (IST) are producing high performance computing networks and advanced web-based computational and visualization tools. EPOS will implement many of the emerging IST developments to create an e-infrastructure for geosciences to facilitate broad use of the Earth science research infrastructure and open up new approaches to scientific research.

A major Earth science data challenge

EPOS faces a major data handling challenge. The amount of primary data and the demand for accessing them is enormous and increasing. On the other hand, the existing data archiving and exchange procedures (and philosophies) appear inadequate to cope with the ever increasing data tsunami. Multidisciplinary scientific research in the solid Earth Sciences is thus hampered by the lack of integration and by the limited access to the large and diverse data volumes needed to make progress (including, for example, geophysical, geological, topographic, and sea level data).

One main EPOS objective is the creation of a comprehensive, easily accessible geo-data volume for the entire European plate at resolutions appropriate to the scale (and to the socio-economic importance) of the target scientific (and non-) issues under investigation. Overall, EPOS will ensure secure storage of geophysical and geological data providing the continued commitment needed for long-term observation of the Earth.

Computational and modelling facilities

Modelling and processing of huge data volumes and visualisation of results will play a central role in testing new scientific hypotheses. EPOS will rely on the world leading European expertise in modelling solid Earth processes developed in national and international projects and in existing European-scale initiatives to promote computational Research Infrastructures including EU-HPC (or PRACE http://www.praceproject.eu) and EGEE (http://www.egi.eu)

Designing the e-infrastructure

EPOS will invest in the development of e-infrastructures to serve a broad audience of Earth scientists, benefiting from the experience gained in other projects and by organizations. The e-infrastructure will include:

- Web portals for data distribution, data mining and archiving, connected to Grid-based computing facilities
- Standardized protocols or protocol services for data exchange
- Web services for integration of new sensors, instruments, observatories and laboratories into the infrastructure
- Repositories for raw observational and experimental data, modelling/simulation data, data products, and meta-data
- Repositories for data assimilation, modelling and visualization tools

Relevance to NERA NA9

Within the EPOS PP project a further integration of the European seismological research infrastructure (all kind of seismological data relevant for the studies of the dynamics of the earth) is discussed and planned within WP6 and WG1. Following this discussion may be relevant for the data integration efforts in NERA-NA9. More important will be the e-Infrastructure discussions within EPOS PP (WG7 within WP6), which will depend for a significant part on the continued developments within NERA and other related projects.

2.2.4 SHARE

Submitted by Philipp Kästli, ETHZ.

SHARE (Seismic Hazard Harmonization in Europe) is an European FP7 project to calculate and communicate a homogenized, state-of-the-art European probabilistic seismic hazard. SHARE has aim to provide information relevant for design and implementation of Eurocode 8, as well as a new baseline and reference for national hazard codes and zonations. It is also the European contribution to the GEM (Global Earthquake Model) initiative.

The work plan comprises:

- A homogenized European Earthquake Catalog for historical and instrumental times.
- A complete catalog of seismogenic sources (faults capable to create earthquakes with a magnitude >= 5.5)
- Strain, and slip rate models, to validate the seismic moment release against observed deformation
- A consensus source zone map, with homogeneously calculated/estimated activity rates.
- A reviewed set of ground motion attenuation relationships for different ground motion parameters and tectonic environments
- A logic tree model over all steps of hazard assessment, allowing for the representation of epistemic as well as aleatory uncertainty

The following results will be presented on a web portal, for different ground motion parameters (acceleration, displacement, spectral acceleration of different frequency bands) and return periods:

- Hazard maps
- Hazard curves
- Hazard spectra
- Disaggregation of the hazard contributions in magnitude and distance

For all types of information, results should be available as medians as well as representing individual branches of the logic tree model, to allow for exploration of the uncertainty and the impact of individual assumptions.

Project Structure

Work package	Lead participant	Contributors
1 – management	SED-ETHZ	
2 – engineering requirements and applications	UPAV	LNEC, AUTH, KOERI, LGIT-UJF, SED-ETHZ, MSO
3 – earthquake sources	INGV	AUTH, NERC-BGS, GFZ, IST, KOERI, NIEP, NKUA, NORSAR-ICG, ROB, CRAAG, MSO, SED-ETHZ
4 – strong ground motion models	LGIT-UJF	METU, BRGM, AUTH, KOERI, SED-ETHZ, MSO
5 – seismic hazard assessment	GFZ	SED-ETHZ, UPAV, INGV, LGIT-UJF
6 – computational infrastructure	SED-ETHZ	BRGM, INGV, UPAV
7 – dissemination	UPAV	INGV, SED-ETHZ

The major IT developments within SHARE are:

- Model building toolkit (to compare seismicity to zoning and fault characterization, and strain rate models, to design the logic tree): A standalone toolkit based on Quantum GIS, Python, and several scientific codes attached is developed at ETHZ. It is intended to be the alpha version for a GEM model building toolkit.
- Hazard engine: SHARE hazard is calculated using the GEM openQuake package. SHARE contributes to openQuake, mostly in specs and in the implementation of attenuation relationships
- Data storage and access: A database and data access web services are developed at ETHZ; core components of the SHARE portal at BRGM. The SHARE portal is based on JSR286 and technically compatible with the NERIES/NERA portal, and intended to be the core of the EFEHR (European Facility of Earthquake Hazard and Risk) data services.

SHARE is planned to run for 3 years, from summer 2009, through summer 2012

2.2.5 MATRIX

Submitted by Philipp Kästli, SED

MATRIX - New Multi-Hazard and Multi-Risk Assessment Methods for Europe, http://matrix.gpi.kit.edu – is an FP7-project looking at the combined impact of whole range of natural hazards including earthquakes, landslides, volcanic eruptions, tsunamis, river floods, winter storms, wildfires and coastal phenomena. Coupling of natural disasters is analyzed on the hazard side, analyzing time-history dependent and cross-correlated incident probability of different natural hazards, as well as on the risk side (analyzing the impact of one disaster on the vulnerability of the infrastructure to others, and on the ability of the society to cope with another disaster). MATRIX is a research project, mostly targeted at providing planners and policy makers with an enhanced knowledge base to design efficient and cost-effective risk-reduction measures.

The matrix workplan is structured as follows:

- **WP2 (single-type risk assessment and comparability)** reviews the state of the art of risk analysis for individual types of natural hazards, and develops a strategy for methodological homogenization of the risk assessment process itself as well as of the description of uncertainty. WP coordinator is **GFZ**, with major contributions of KIT, ISA-CEABN, AMRA, and BRGM.
- **WP3 (Cascade effects in a multi-hazard approach)** reviews current approaches in multi-hazard analysis and develops a basis for MATRIX multi-hazard analysis by providing a universal probabilistic framework, a dictionary of common terms and metrics, and a set of scenarios for the MATRIX test sites. **AMRA**, with ISA-CEABN, GFZ, BRGM, TU-Delft.
- **WP4 (Time-dependent vulnerability)** develops fragility curves – especially conditional fragility taking into account previous damage, and reconstruction. Metrics cover physical and functional vulnerability as well as socio-economic vulnerability. **BRGM**, with ISA-CEABN, AMRA, IIASA.
- **WP5 (Framework for multi-type risk assessment)** is the analogon to WP3 in the risk space, designing a framework for multi-risk (and uncertainty) assessment covering tangible and intangible losses. **NGI**, with ISA-CEABN, GFZ, AMRA, TU-Delft.
- **WP6 (Decision support for mitigation and adaptation in a multi-hazard environment)** connects the risk scenarios to decision-support tools to model the impact of provisions as well as mitigation measures. **IIASA**, with KIT, BRGM, TU-Delft.
- **WP7 ("VirtualCity" and test cases)** implements the outcomes of WP2-6 into IT framework for running multi-risk scenarios based on the MATRIX test sites as well as on synthetic hazard and exposure models. **SED-ETHZ**, with GFZ, BRGM, AMRA.
- WP1 does overall project coordination (**GFZ**, with AMRA, BRGM), while WP8 (**KIT**, with DKKV, AMRA, ISA-CEABN) covers outreach and cooperation of national civil protection authorities in order to translate MATRIX outcome in practice of disaster preparedness and mitigation.

To develop and test its models, MATRIX relies on a series of test regions with a good data coverage and a known exposure to multiple hazards: Naples (mainly earthquakes, volcanism, tsunamis), Cologne (floods, storms, earthquakes), French West Indies (volcanism, storms, earthquakes, landslides). In a second step, a virtual city model should allow to run scenarios on any hazard and exposure profile.

As MATRIX is mainly a research project, public modelling and data dissemination are not a core focus of the project. However, in order to facilitate internal data integration and exchange, MATRIX has a strong focus on implementing best data handling practices, involving the development of data and metadata dictionaries, self-describing exchange formats (xml), and documentation of datasets following the INSPIRE directive.

Matrix started Octobe 1, 2010, and will be terminated by September 30, 2013.

2.2.6 IRIS

Submitted by Erik Sonnleitner, Josef Küng, FAW/JKU.

The IRIS *) project (Integrated European Industrial Risk Reduction System) investigates the possibilities of integrating multiple, different, heterogeneous decision support systems in support of Structural Health Monitoring. In this section we illustrate the functionalities of the resulting approaches and the possibilities towards integration with the NERA project.

*) IRIS project here should not be confused with IRIS – Incorporated Research Institutions for Seismology in the US (see section 2.1.2)

Introduction

During a deep analysis of general behaviour in terms of decision support systems within the area of Structural Health Monitoring (SHM), the Institute for Application-Oriented Knowledge Processing (FAW) has introduced a system which aims to integrate multiple heterogeneous decision support systems, originating from being used within a broad range of different domains in European industry, including bridges and power plants. This system is called PDSIS, Passive Decision Support Integration System and implements a meta-layer above the underlying decision support systems (DSS), which are integrated using ontologies.

Software architecture

In order to implement the collective integration of different DSS, the general software architecture in question uses a three-tier approach:

1. PDSIS component: PDSIS offers a web client, which is responsible for authenticating users, managing sessions, and communicating with the middleware components (the mediator) by requesting operators, operator parameter requirements and retrieving results. The communication between PDSIS and the mediator component is done through SOAP.
2. Mediator component: A middleware, which manages the actual DSS. The mediator is capable of executing data mining-operators natively by directly communicating with the corresponding engines, either through a provided API or via other well-defined service interfaces (e.g. web-services). Moreover, the mediator provides a structured interface to PDSIS, to disclose the operators provided by a certain DSS in a well-defined and unified way. This disclosure includes the kind of input data which is needed to execute, and returns a well-defined result set after the operator execution has finished. By implication, every mediator component is dedicated for handling particularly one specific DSS. Although large portions of mediator-code may be reused during the process of including new DSS, the mediator it self, unlike PDSIS, is not meant to be implemented generically enough to be applied on multiple DSS.
3. Low-level DSS: These are mostly external, third-party engines providing non-unified ways of interaction. PDSIS itself doesn’t care about how to communicate with a certain DSS, since solely the mediator component is aware of and responsible for that.

The mediator component is also responsible for loading the integration ontology, which includes semantic information about all currently integrated operators. The integration ontology is primarily used for identifying applicable data-mining operations for the current process step (e. g. data pre-processing), and dynamically calculate a ranking of these, according to their degree of reasonableness for a certain task. In order to provide a useful classification scheme, the data-mining operators available are meant to be graded manually by the PDSIS users after every operator run. [AnSt10]

Supported DSS

The software architecture has been designed with respect to the fact, that utilization, execution and communication methodologies of certain DSS may differ significantly.

Therefore, one of the main objects was, that integrating a new DSS should be easily possible without considerable restrictions or requirements.

To provide a proof-of-concept implementation, the following engines and systems have already been fully integrated, or implementation is currently in progress:

- Weka (Waikato Environment for Knowledge Analysis), a machine-learning engine by the University of Waikato, NZ (1).
- RapidMiner, an analytical machine-learning, data-mining and text-mining engine, originally developed at the University of Dortmund, and currently maintained by the company Rapid-I (2).
- R, a programming language for statistical computing (3) and part of the GNU project. R aims to be the de-facto standard for statistical computing nowadays.
- Octave, a programming language for numerical computations which aims to be fully compatible with the proprietary MathWorks MATLAB product.
- BRIMOS R system for bridge monitoring and evaluation developed by Vienna Consulting Engineers (VCE).

Conclusion and future work

The meta-DSS approach of the system described in this document or its general architecture may also be of certain interest within the frame of the NERA project. However, PDSIS in general is still under heavy development and not yet ready for production use. Moreover, the current alpha-version only implements one out of originally four proposed workflows. Seamless usage in NERA, without significant adaption and modifications is not realistic at the time of writing.

1. http://www.cs.waikato.ac.nz/~ml/weka/
2. http://rapid-i.com/
3. http://www.r-project.com

[AnSt10] S. Anderlik, R. Stumptner, B. Freudenthaler, M .Fritz: A proposal for ontology-based integration of heterogeneous decision support systems for structural health monitoring. In proceedings of the 12th International Conference on Information Integration and web-based Applications and Services, pp. 166-173, Paris, France, 2010.

2.2.7 ENVRI

ENVRI aims at coordinating where possible standards and common practises within the ESFRI Environment Research Infrastructures. Common Operations of Environmental Research Infrastructures (ENVRI) project involves all ESFRI initiatives that involve distributed systems. For EPOS, the KNMI/ORFEUS participates with a minor role. Within the seismological research infrastructures we are considerably more advanced due to existing global collaborations and long-term existing data exchange routines. NERA NA9 may have some advantages, but will probably mostly provide an example.

2.2.8 ADMIRE

Submitted by Alessandro Spinuso, Luca Trani, Torild van Eck, ORFEUS.

ADMIRE (Advanced Data Mining and Integration Research for Europe) is a project funded by the 7th FP of the European Commission (www.admire-project.eu). Its main goals are to:

- Accelerate access to and increase the benefits from data exploitation
- Deliver consistent and easy to use technology for extracting information and

knowledge

- Cope with complexity, distribution, change and heterogeneity of services, data, and processes, through abstract view of data mining and integration
- Provide power to users and developers of data mining and integration processes

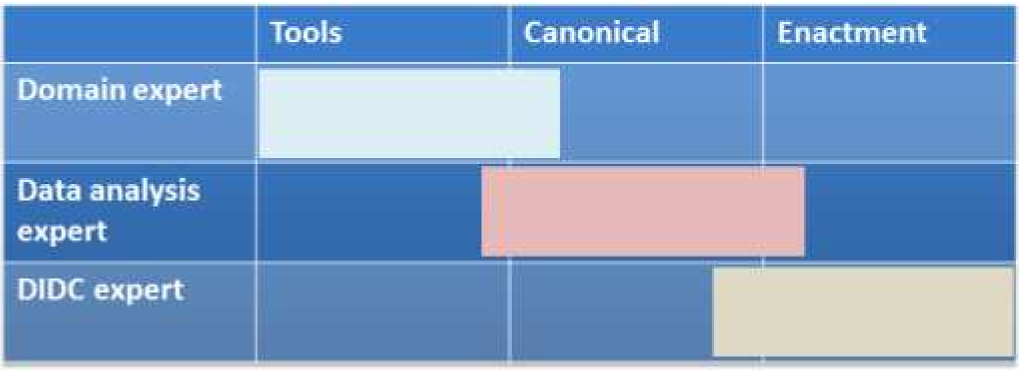
ADMIRE offers a strategy and distributed architecture to assist data analysis experts in making sense of the increasingly large and complex world of digital data.

Separation of concerns

ADMIRE identifies three types of data user:

- **Domain experts** who hold the business perspective for a given data-intensive problem. They need answers to domain-specific questions without having to invest significant effort in development.
- **Data analysis experts** specialise in extracting information from data. They devise the strategies to answer the domain experts' questions.
- **DIDC experts** (data-intensive distributed computing experts) understand the best ways of harnessing computing resources to build data-intensive processing platforms.

ADMIRE separates the concerns of these users, providing familiar *tools* to the domain experts, a standard *canonical approach* to describing data-intensive processes for the data analysis experts, and a comprehensive and efficient *enactment framework* for the DIDC experts.



In this way, ADMIRE divides the challenge of designing and implementing any large-scale data analysis process into three - nearly independent - parts:

1. Design and specification of data-intensive processes in high-level graphical and DISPEL notations;
2. Specification and description of processing elements as user-defined code; and
3. Specification of patterns that compose processing elements and deliver fast ways of generating powerful and frequently used data-intensive processes.

Admire Structure

WP1 DMI Model & Language Research
Development of a new conceptual model for distributed data mining and data integration, supported by rich ontology-driven semantic description and a new data intensive systems process engineering language (**DISPEL**).

WP2 Architecture
Development of a new architecture to support distributed data intensive systems, including semantic registries, DISPEL enactment Gateways and automatic execution optimisation strategies.

WP3 Test, Release & Support
Test and deployment of emerging ADMIRE software for both user tools and a wide-area distributed execution Platform.

WP4 Service Infrastructure

Development of core supporting Gateway services, both generic and DISPEL-specific, as part of the emerging ADMIRE execution Platform.

WP5 Tools

Development of powerful user tools for driving DMI explorations through ADMIRE Gateways. ADMIRE tools will be based on the *eclipse* platform.

WP6 Applications & Evaluation

Demonstration and evaluation of ADMIRE's model, language, architecture and tools through real-world application scenarios.

ADMIRE collaborators

- University of Edinburgh, UK (Coordinator)
- Fujitsu Laboratories of Europe, UK
- University of Vienna, Austria
- Universidad Politécnica de Madrid, Spain
- Institute of Informatics, Slovak Academy of Sciences, Slovakia
- ComArch S.A., Poland

2.3 Required Inputs for Probabilistic Seismic Hazard Assessment (PSHA) within the Context of the IRIS Project

Submitted by Helmut Wenzel, VCE.

After the devastating earthquake in Japan in 2011, a stress test for nuclear facilities in Europe is discussed. The quality of PSHA (Probabilistic Seismic Hazard Assessment) depends on the availability of relevant data and the capability of simulating scenarios that enable risk determination. The seismic information is available with the seismology community and earthquake engineers have developed simulation tools that allow computing large scenarios with detailed elements at risk.

The main innovation required will be a software platform that enables extraction of seismologic information which will be systematically converted into scenarios to find the worst case. The critical issue is the interface between seismology and earthquake engineering.

Risk is a combination of **Hazard** (information provided by the **seismology community**), **Vulnerability** (information provided by the earthquake **engineering community**) and **Consequences** (information provided by the **socio-economic community**). Earthquake Engineers have successfully combined vulnerability and consequences in a software platform called EuroVIS, which is based on the most developed US platform called MaeViz. A module to randomize the seismic input is to be created.

The software platform requires the following input from the respective engineering community:

Seismology:

- Coordinates of the Epicenter
- Magnitude of the Earthquake
- Depths
- Direction of the active fault
- Detailed soil profiles
- VS_{30} profiles
- Ground water table

- A digital terrain module
- Liquefaction susceptibility
- Surface Geology
- etc.

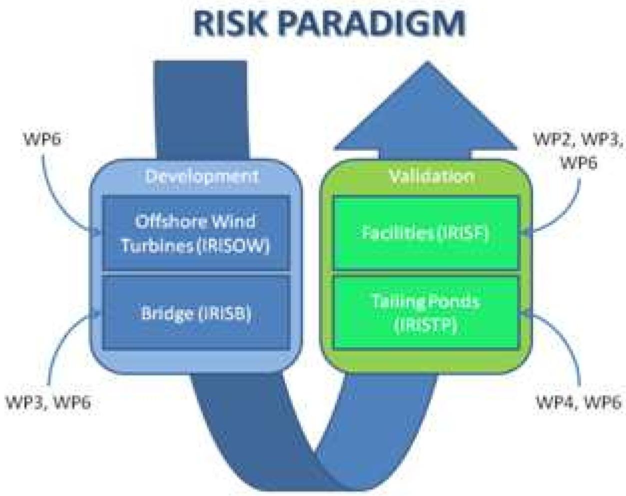
Engineering:

- Detailed fragility curves of elements at risk
- Structural fragility curves
- Fragility curves for systems
- Fragility Mapping
- Repair Cost Estimation
- Content Damage Predictions
- Cost Benefit Module
- Functionality Module
- Interrelation between the systems
- Feedback between the elements of risk
- Quantification of uncertainties
- etc.

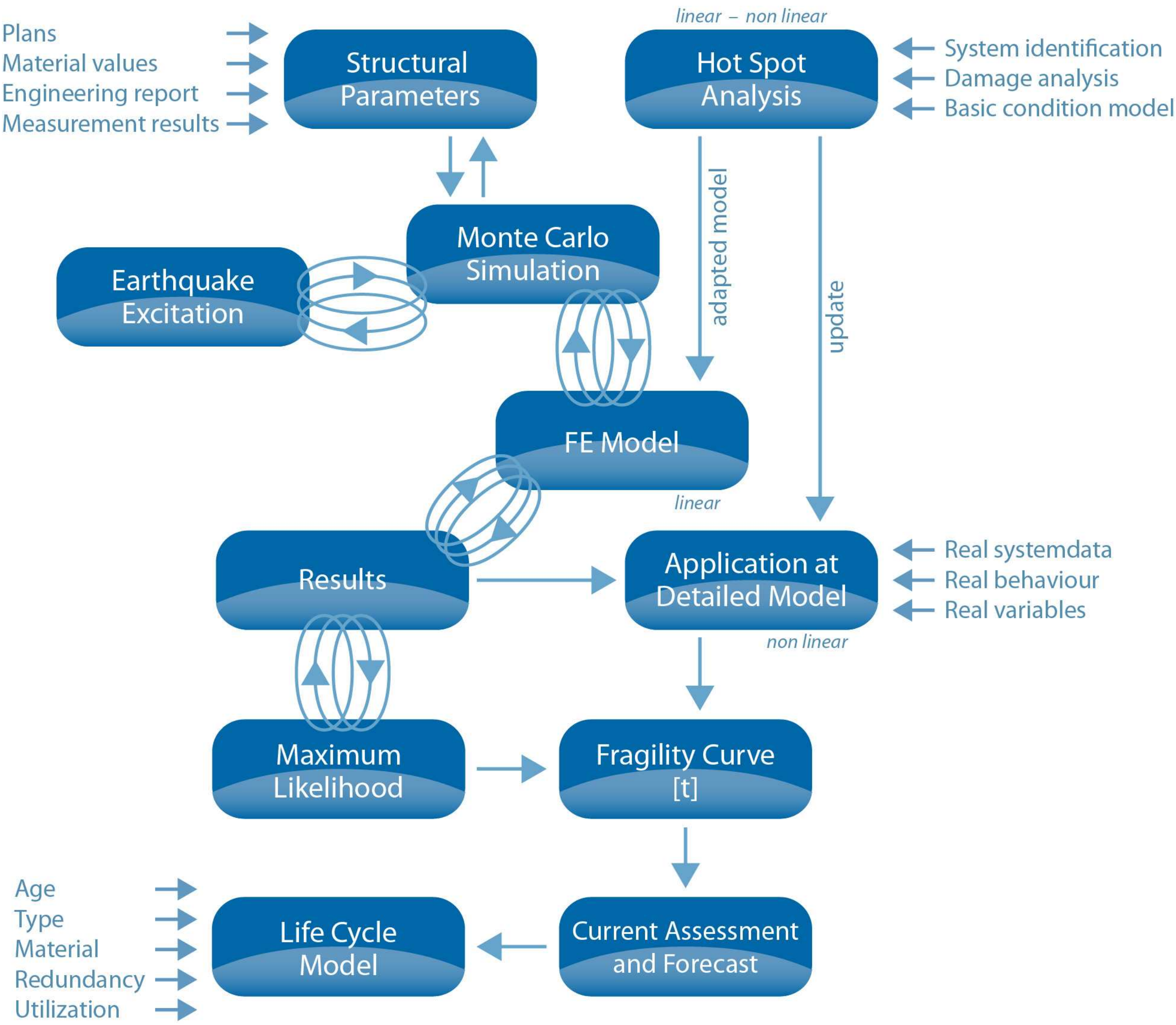
Socio-Economy:

- Number of affected inhabitants
- Death Toll
- Number of people injured
- Number of people seeking shelter
- Number of people displaced for a long time
- Direct property loss
- Indirect losses from failures
- Consequential damages
- Economic Damage
- Business Interruption Loss
- Shelter Supply Requirements
- Temporary Housing
- Business Content Loss
- etc.

The risk paradigm developed by the IRIS project enables the quantification of risk for any scenario developed. The paradigm is described in the subsequent diagram that can be applied for the proposed procedure.



The procedure for the engineering side is given in the following diagram. The interfaces to seismology and socio-economy are to be defined.



2.4 Design Process in Earthquake Engineering

Submitted by Helmut Wenzel, VCE.

Introduction

Engineers are very pragmatic and desire clear rules for the execution of their work. Practice is based on standards which to follow is obligatory. In structural engineering most parameters are directly defined by these standards or guidelines. Earthquake engineering is still developing and new knowledge is generated. It cannot be expected that the engineers working in practice are aware of the state of the art. Strong competition in the community forces practitioners to work efficiently not leaving enough time for extra studies. Therefore it is important to offer the relevant information to the engineering community in a simple and comprehensive way.

The Load Case Earthquakes in the Eurocode

The national annexes to Eurocode 8 provide the engineers with a shake map for the respective country. This map shows a PGA value on regional scale.

For design it is necessary to determine a PGA value for the exact location. The factors offered by EC 8 for the soil classes do not properly represent eventual site effects. Out of these circumstances the following desire exists in the earthquake engineering community:

- Shake maps have to become more detailed and broken down to local scale (optimal resolution 10 meters)
- The respective information on the soil profile used has to be made available because there is direct relevance to the soil parameters used by the engineers. This would also exclude the application of false values in design (the foundation type for example piles, also influences this decision)
- Potential amplification factors from soil conditions or the ground water table have to be specified separately.

To achieve this it would be desirable to have a direct link to the soil parameters used (for example boring logs close to the location of interest). In many cases it would be sufficient to show that there are no additional amplifications to be expected and Eurocode can be applied without further considerations.

It would be nice to get other information like the importance factor of buildings from the same place. Nevertheless this is not a priority.

Design Spectra

For extraordinary structures design spectra are used in conjunction with a modal analysis. The differences in the results can be huge depending on the spectrum used. The spectra proposed by the guidelines are made to cover a wide range of cases and are therefore on the safe side. In engineering we distinguish between our slender and soft structures like bridges and rigid structures like old buildings or palaces. For both, different information is desired.

The best information would be a representative spectrum of a real earthquake of this particular location. As this is never feasible it would be nice to get similarity information. A selection of most probable spectra should be offered out of which the engineer can draw his conclusions.

It has to be mentioned that this work is not done by the average structural engineer but a better educated staff where deeper knowledge can be expected.

2.5 Overview of Earthquake Loss Estimation and Assessment Tools

Submitted by Helmut Wenzel, VCE.

Excerpted from: Open Source Procedure for Assessment of Loss using Global Earthquake Modelling, by James Daniell.

General

Earthquake Loss Assessments (ELAs) are produced in order to detect possible economic, infrastructure and social losses due to an earthquake. In order to produce an effective ELA, four components must be taken into account.

Seismic Loss = Exposure * Vulnerability * Hazard * Damage Loss Conversion

Exposure is defined as the amount of human activity located in the zones of seismic hazard as defined by the stock of infrastructure in that location.

Vulnerability is defined as the susceptibility of the infrastructure stock.

Hazard is defined by risk of a certain ground motion occurring at a location, which can be defined by scenario modelling via stochastic catalogues, PSHA or other methods, and can include different types of earthquake effects.

Damage Loss Conversion can be defined as the mean damage ratio (ration of replacement & demolition to repair and restoration cost, or the social cost.

Constitution of an Earthquake Loss Estimation (ELE) program

First of all there has to be a defined area of interest in which the seismic hazard should be located at every location. The vulnerability of the infrastructure stock exposed to this hazard should be convolved with this hazard and therefore a damage distribution can be established based on various classes of infrastructure damage. From this damage distribution, economic and social losses can be derived. All of these components constitute an ELE. The use of such a tool can be divided in a proactive way (pre-earthquake modelling) or a reactive way (post-earthquake fixed scenario modelling).

Exposure

Exposure is a very critical parameter in an ELE procedure. Unfortunately it is a very difficult component to collect. Data quality can vary from country to country and region to region, so it is very hard to derive an accurate infrastructure stock for any country.

There are currently many different ELE software packages and each software package requires different inputs. The inventory data includes building location, age, use, height, structural type, structural inconsistencies, construction age, number of storeys and population information.

Of course in order to do a detailed ELE, exposure needs to be undertaken on a small area level. On a city or regional level, urban characteristics are required in order to produce a suitable inventory, including location of lifelines and also building stock details. This is generally a large amount of work requiring much money and time for a certain location, but the level of detail that is able to be extracted during this part of the process determines to a great extent the accuracy of the result at the end of an earthquake loss assessment.

One of the most important aspects to be found in an exposure inventory production is that of lifelines within the region being analysed. These lifelines are intrinsically linked to the specific cost section: however, they must be identified for rapid loss estimation process in order to ensure that utility (potable water, waste water, oil, natural gas, electric power and communication) and transport systems (highway, railway, bus, port, and airport) can remain in place during a disaster. Critical Systems must also be identified for rapid response and calculation during the loss estimation module such as medical care facilities, schools and emergency services. In addition, industrial commercial and residential areas should also be identified.

Vulnerability

The principal aim of a vulnerability assessment is to derive the probability of a certain level of damage occurring to a certain infrastructure stock when subjected to a scenario earthquake.

The development of vulnerability curves is not a straightforward process. If it were possible to run a non-linear time history analysis for every single infrastructure item within the infrastructure stock and the information would be freely available, then this would be done in order to calculate the vulnerability. However, this information is not freely available for every infrastructure item and in addition it can take days to produce a single NTHA (Non-linear time history analysis) of a building when given the construction plans as a new model needs to be created for every building. Although it can be taken into account, that single (important) buildings are being computed that way.

Empirical Methods are vulnerability assessment methods on observed damage data. These methods have been employed to define the vulnerability of infrastructure stock from the 1970s. Many of these methods only use macro-seismic intensity of PGA (Peak Ground Acceleration) rather than spectral ordinates which created large scatter of results. However, these were initially the only possible methods for large scale seismic risk analysis.

> Damage Probability matrices (DPM) are methods to determine damage due to strong motion which are simply the conditional probability of obtaining a certain damage level (j) due to a certain ground motion intensity (i).
>
> For a given building class, subjected to a given intensity, there will be a certain percentage of buildings associated with the combination which corresponds directly to a given damage ratio, as classed within the damage state index used. Thus for this percentage of buildings, the ratio of repair to replacement cost corresponds to the values given in the DPM.

The disadvantage of using macro-seismic methods for observed damage of building stock is that the vulnerability and ground motion input are both based on observed damage due to earthquakes which is not correct. If such observed damage values are going to be used, there are also not many recordings of earthquakes with large intensities which occur close to cities. Therefore there is a lack of data in the high damage and ground motion section of the vulnerability matrix and so the statistical certainty is less towards the higher end of the spectrum. Another issue results from the fact that PGA and spectral ordinates are generally used for seismic hazard maps and these are not directly related to intensity scales which are slightly subjective in nature. PGA when derived for empirical vulnerability does not take onto account the relationship of vibration frequency content of buildings versus that of the ground motion and this is why spectral ordinates are more desirable.

Empirical Vulnerability Index Methods

These Methods are usually based on much survey data after an earthquake in order to gain information as to relationships between damage and intensity based on parameters influencing vulnerability. These methods have been used extensively throughout Italy previously using these parameters, and thus expert judgement can be used in order to calculate the vulnerability index and then produce an indirect relationship with a damage factor for a given PGA/macro-seismic intensity.

These methods have been also used as part of the RISK-UE project.

This methodology is easily adaptable to large-scale assessment of groups of buildings but still requires expert opinion. It also requires the use of extensive field surveying which not available in many regions. It is somewhat subjective and therefore is not exact, and thus surveyors may have different ideas as to define building characteristics without strict guidelines. Therefore this gives discrepancies.

Continuous Vulnerability Curves

Those methods are another empirical method which has been used for vulnerability assessment by directly utilising the probability of the damage of buildings to earthquakes. These have been produced using ISTAT data for the production of seismic risk by taking DPMs to produce vulnerability curves in terms of spectral displacement at the period of vibration.

This method was undertaken in order to overcome the inaccuracies or continuous curves based on PGA or macro-seismic intensity converted from PSI (parameterless scale of intensity).

Analytical and Hybrid Methods of Vulnerability Analysis

Analytical methods are based on structural mechanics principles and are fast becoming the preferred method of large-scale vulnerability assessment due to their proactive capacity, direct correlation with damage and non-reliance an observed damage data. Analytical methods are mainly based on non-linear analysis as this allows for stiffness degradation of existing buildings to be taken into account.

Analytically derived vulnerability and fragility curves and DPMs (Damage Probability Matrix)

Those curves can be produced by computational intensive analyses rather than observed damage data to obtain the structural performance via a given intensity measure. (Europe, Rossetto and Elnashai (2005), Dumova-Jovanovska (2004) and Masi (2004)).

Singhal and Kiremidjian (1996) derived fragility curves and DPMs from Monte Carlo simulation (random variation of material properties) for reinforced concrete frame structures using a variety of ground motions with non-linear dynamic analysis (NTHA) in order to produce the structural damage probabilities. The results for each of the nonlinear analyses was done based on the Park and Ang (1985) damage index and the statistical analysis used in order to produce the DPMs and fragility curves. The vulnerability curves were then updated with the Northridge data with an additional weighting system.

The most used damage scale for analytical fragility functions is the Park and Ang (1985) damage index.

Masi (2004) derived vulnerability curves using the EMS-98 scale, and NTHA with synthetic and real accelerograms, utilising design code and handbook derived structural models from Italian buildings from the 1970s onwards.

Rosetto and Elnashai (2005) used the damage scale derived from their 2003 paper and produced adaptive pushover curves and thus via the capacity spectrum method the performance point was defined to a damage state.

The computational time needed for analytical methods impacts upon their usefulness for countries where there are many different construction types and characteristics, and thus although not useful on their own, can be used to support empirical DPMs, fragility and vulnerability curves via their use in hybrid DPM and vulnerability curve methods, which use analytical methods to fill in the gaps and data within the damage band for certain intensity levels where there is no empirical data in that location. This is therefore faster than analytical methods.

Capacity Spectrum Method

This method is widely used within loss assessment models due to its ability to relate the crossover point of capacity via a pushover curve and demand via an ADRS (Acceleration-Displacement Response Spectrum) to a given damage state.

Essentially, the capacity spectrum method relies on an iteration method from the initial ADRS (usually set at 5%) in order to relate it directly to the pushover curve to achieve the performance point which defines the damage state taking into account both the equivalent non-linear damping and ductility (representing the horizontal displacement of the structure under increasing horizontal loading). The iteration from FEMA-440 shows the ratio beneath the capacity curve (i.e. maximum strain energy) from the performance point, to the total hysteresis loop are which is the energy dissipated by damping.

As the ground motion increases (i.e. higher ground shaking), the amount of inelastic deformation increases (i.e. larger displacements for a certain acceleration), and period lengthening occurs for the structure. More ductile structures will have larger displacement ductility associated with their capacity and stronger structures will be able to resist greater accelerations for a certain displacement.

HAZUS is a very simple hence useful and adaptive procedure. It has many simplifications as it assumes the same capacity curve for a certain location. It is also difficult to adapt the capacity curves to other locations in the world as the building classes have been derived for limited height buildings in the U.S. These buildings are also put into binned height classes and therefore the building capacity curves and vulnerability functions are approximate. Thus, in order to adapt these buildings to other locations in the world, a large amount of building information is required in order to carry out a reliable non-linear static procedure.

Giovinazzi (2005) presents a displacement-based mechanical procedure to assess masonry and RC (Reinforced concrete) frames by using s capacity cure which has been converted to a Sa-Sd plot.

Molina and Lindholm (2005) as part of their SELENA ELE software incorporate a logic tree approach within their capacity spectrum methodology in order to reduce epistemic uncertainty.

Displacement-based methods have been produced recently due to their ability to relate to damage states better than original force-based methods and proposed through Priestley (2003) as damage is strain dependent. Thus, through strains, curvatures can be derived and subsequently rotations and displacements.

<u>Hazard</u>

Identification of possible hazards in addition to ground shaking has been undertaken by Bird and Bommer (2004) and the impact to building damage of these potential hazards has been found to be much less than the ground shaking due to earthquakes (ground failure effects such as liquefaction, fault rupture landslide, slope stability and bearing capacity; tsunamis and seiche). However, these secondary effects cause a lot of damage to lifelines (Bommer et al, 2006).

Ground Failure

Liquefaction involves the changing of soils from slid to liquid state and is usually caused by induced cyclic shear. It is thus energy related and so is extremely complicated to incorporate into loss models. There are many uncertainties over the area and extent of liquefaction but by using simplified methods such as relationships between PGA and susceptibility liquefaction can be applied into loss models.

Bearing Capacity failures using the methods of Richards et al (1993) or Kumar et al. (2003) can be used in order to determine the damage of structures due to loss of bearing capacity, but currently there are no loss estimation models that take this into account.

Fault Rupture causes localised effects and thus is usually not incorporated into the ELE assessment.

Landslides and Slope Stability are more difficult parameters to constrain because of the need to determine the rainfall that has occurs in the area before the earthquake to have an ideas of possible landslides as they can also be rainfall-induced. Most of the analysis methods include a simple ratio between PGA and the factor of safety (FS) is based on a critical acceleration for the slide mass. Many methods have been established including a probabilistic framework by Del Gausio et al. (2003) producing damage functions for structures based on the movements and probability of slope failure. GIS evaluation models can simply identify susceptible areas to landslides, and this combined with an intensity measure approach (Wilson, 1993) may be the best method for application.

Tsunamis and Seiche (standing wave induced phenomena) have increased in importance since the Boxing Day Tsunami of 2004. Although important, this type of secondary effect should be generally applied at a rapid response level, due to the unknown nature of sea-floor bathymetry with undersea quakes and the relative lack of knowledge in the area worldwide, given the depth and uncertainty of such phenomena.

Methods of Seismic Hazard Assessment

There are two main methods of seismic hazard assessment: ones which are deterministic (DSHA) and that include a single scenario earthquake (historical, MCE or user-defined); or a probabilistic combination of earthquake scenarios in order to determine the hazard for the given area (PSHA).

A seismic hazard assessment consists of 3 components: recurrence relations (magnitude function), source zones and earthquake catalogues (historic and stochastic). The recurrence relationship comes about as a probabilistic result of the minimum and maximum earthquake possible from an earthquake catalogue for the given source to produce a probability density function giving the ARE (Annual Rate of Exceedance) of different magnitudes. Earthquake catalogues are extremely important in hazard assessment and detail the magnitude and spatial position of previous recorded earthquakes.

Source zones are the spatial regions where the future earthquakes are expected to occur, defined by tectonics, geology and and observed seismicity. Source, path and site effect calculated via GMPEs (Ground Motion Prediction Equation) and local site conditions define the ground motion field away from the sources.

A **deterministic seismic hazard assessment** (DSHA) consists of 3 main steps and has been carried out for many locations where a complete worst case scenario or historical repetitive earthquake is waiting to be modelled:

1) Define all possible sources to cause significant hazard at a site from historic tectonic, geologic or geotechnical data.
2) Choose a fixed distance, fixed magnitude earthquake and place it on the closest position to the site on each source, defined via empirical equations on the basis of geological evidence using Wells and Coppersmith (1994) of by just adding 0.5 magnitude units to the largest historical earthquake
3) Estimate ground motions via GMPEs to determine the ground motions at the site in terms of spectral ordinates. Variability can be modelled for the ground motions within a DSHA; however a common way is to use motions which are one logarithmic standard deviation above logarithmic mean. Each of these DSHA is very useful for lifeline and critical facility locations and is increasingly being used to supplement a PSHA

A **probabilistic seismic hazard assessment** (PSHA) considers all M-D-e combinations taking into account all probabilities and scenarios possible for magnitude and distance to calculate the hazard. The steps involves are adapted from Akkar and Boore

1) Define a probability of potential rupture locations for each source.
2) Determination of the temporal distribution via recurrence relationships. The Guttenberg- Richter relationship is commonly used where Nm is the mean annual rate of exceedance of magnitude M, b is the activity parameter expressing likelihood of large and small earthquakes and , a , describes the yearly rate in logarithmic space of earthquake

$$\log(N_m) = a - bM$$

   But other relationships adapt this to calculate a characteristic magnitude and thus describe truncated normal and lognormal, exponential, uniform and Qouns and Coppersmith Characteristic Equation and Delta magnitude recurrence relations.
3) GMPEs are used for the range of distances for each magnitude to produce spectral ordinates dependent on the tectonic regime with aleatory variability, σ (interevent and intraevent) of each relationship taken into account as well as the applied variability, e
4) The hazard must then be integrated by combining the effects of different size, locations, source zones and occurrence probability earthquakes in order to calculate the expected number of exceedance of ground motions due to the PDF (Probability Density Function) of magnitude, distance between source and site and also the probability calculation for spectral ordinate values away from the mean value. From this, annual rates of occurrence are derived giving a hazard curve
5) A PSHA assumption common made is the Poissonian model that takes the annual frequency of exceedance from this analysis and assumes that each earthquake is independent of other earthquake, where q(z)= is the probability of exceedance of a user-defined ground motion level for a given time, t in years, where λ(z) is the annual rate of exceedance is thus';

$$q(z) = 1 - e^{-\lambda t}$$

From this a 10% probability in 50 years gives the 1 in 475 year return period for this earthquake.

Damage Loss Conversion, Economic and Social Costs

By convolving the impacts of hazard, vulnerability and exposure, the conversion into a damage loss and specific cost in terms of economic and social cost can be applied.

There are two different scenario components for the socio-economic module that should be addressed:

a) Direct conversion of damage to fatalities, injuries, homeless and economic impact
b) Direct and Indirect socio-economic impacts and complex indicators including social and economic vulnerability

There are many different indices and databases which can be used for the evaluation of socio-economic consequence functions. Most of the historic social and economic functions have been damage-based i.e. building class related, however there is the need for pure economic and social functions which are unrelated to damage, and may be intensity-based, or based on some other parameter. Indirect social and economic functions should also be accounted for.

Social and Economic Vulnerability

Increased social vulnerability can be looked at in advance, once danger has been identified such as currently in Adelaide, Australia for a fictitious case. By using an in-depth assessment technique such as DBELA and then applying it to Adelaide, to find out which types of houses are most susceptible, people can be warned prepared and educated. In some cases the government will undertake screening methods and the retrofitting methods. Thos has been found to work in advanced nations, but in developing nations this can be a problem and so the results of ELE software need to work out some sorts of social vulnerability function based on the nation in term of building practice and not directly to damage.

It can be generally assumed that within developed countries, the seismic codes employed will be more up to date than those of developing countries. Therefore, it can be assumed that less damage will occur in these countries for an earthquake striking a high population area.

In most cases, if an earthquake strikes a location of high population density such as in the city of Kobe in Japan in 1995, it can be expected that the economic loss will be high, with a large number of death and with this value of deaths increasing with decreasing development.

It can be seen that by correlating the population growth to the locations of population density and also those of the development index, the increasing exposure of a nation to earthquake can be observed. Assuming the current population level increase, people will have to build in locations of higher risk or with increasing speed and this will cause problems in the future.

There are a number of socio-economic indices that have been established. Risk analysis when applying indicators for the socio-economic assessment on any geographical level can be undertaken using a variety of Socio-Economic Indices.

- Urban Seismic Risk Index (USRi) – Carreno et al. (2006)
- Social Vulnerability Index – Cutter et al. (2003)
- Disaster Deficit Index (DDI) – IADB
- Seismic Safety Indices
- Disaster Risk Management Index (DRMI) – IDB IDEA
- Earthquake Disaster Risk Index – Davidson (1997)

- Cities Project – Granger et al. (1999)
- Hotspots Project by the World Bank

2.6 Ontologies

Submitted by Erik Sonnleitner, Josef Küng, FAW/JKU.

Introduction

Ontologies constitute a central component in semantic technologies. Originating from the scientific field of artificial intelligence in the 1970s, an ontology represents items of knowledge in a way that defines the relationships and classifications of concepts within a specified knowledge domain [Jep09].

An ontology basically consists of classes and instances of classes and, especially, the relationships among them in order to establish an interconnected view of the domain in question. To achieve this kind of interconnection, classes define slots (also called roles or properties) and facets (also called role restrictions). A slot is is a certain property of a class which every instance of a class all all its sub-classes are expected to provide, while a facet defines the value type, allowed values and their cardinality.

Moreover, since an ontology is strictly formally defined, it's possible to perform software-supported automatic semantic reasoning. Using so called reasoners, it's easily possible to determine logical consequences.

Modelling example

Let's illustrate these rather theoretical terms in a more comprehensible real world example by modelling automotive manufacturers. For accomplishing that, we create a class Marque. Instances of this newly created class could possibly be BMW, Renault and Mazda. If we want to model where these marques have their origins, we first have to create an additional class Country and their respective instances Germany, France and Japan.

Until now, no relationship between these two classes has yet been defined, for this requires the definition of a slot in order to interconnect Marque and Country. So let's add a slot originatesFrom to class Marque.

The final step is to declare a facet to our new slot to restrict the values which may be assigned to it. Facets can be regular value-types (strings, numbers, enumerations, etc.) on the one hand, and so called instance-types on the other hand. In our case, the facet would be an instance-type of class Country.

As a matter of fact, ontologies are always subject to a specific knowledge domain, and are rarely ever considered to be complete [Noy01]. For instance, we may add additional subclasses to Marque like Model having properties name and engine. Alternatively, we could also introduce a new class Engine and define corresponding slots like cubicSize and power.

Languages and modelling tools

Since ontologies emerged, a range of different modelling languages and formats emerged. Nowadays, the probably most common ontology language is the Web Ontology Language (OWL) standardized by W3C.

Alongside these languages, different tools have been created in order to support ontology modelling, like Protégé (1), developed by the University of Stanford, and Chimaera (2). These tools not only support the creation of new ontologies, but also offer advanced features like reasoning and automated graphical ontology rendering.

1) http://protege.stanford.edu/

2) http://ksl.stanford.edu/software/chimaera/

References

[Jep09] Thomas C. Jepsen, 2009, Just what is an ontology, anyway? IEEE IT Professional, 11:22-27

[Noy01] Natalya F. Noy and Deborah L. McGuinnes, 2001, Ontology development 101: A guide to creating your first ontology, Stanford University

3. Current Status

3.1 Earthquake Data Portal

The Earthquake Data Portal, originally developed under the NERIES project and discussed at length above in section 2.2.1, continues to operate and serve the community with a growing user-base of over 1000 registered users to date and with visitors coming from 135 countries from around the world. The Data Portal and supporting web services provide the basis for the integration of new developments and tools provided by the NERA project.

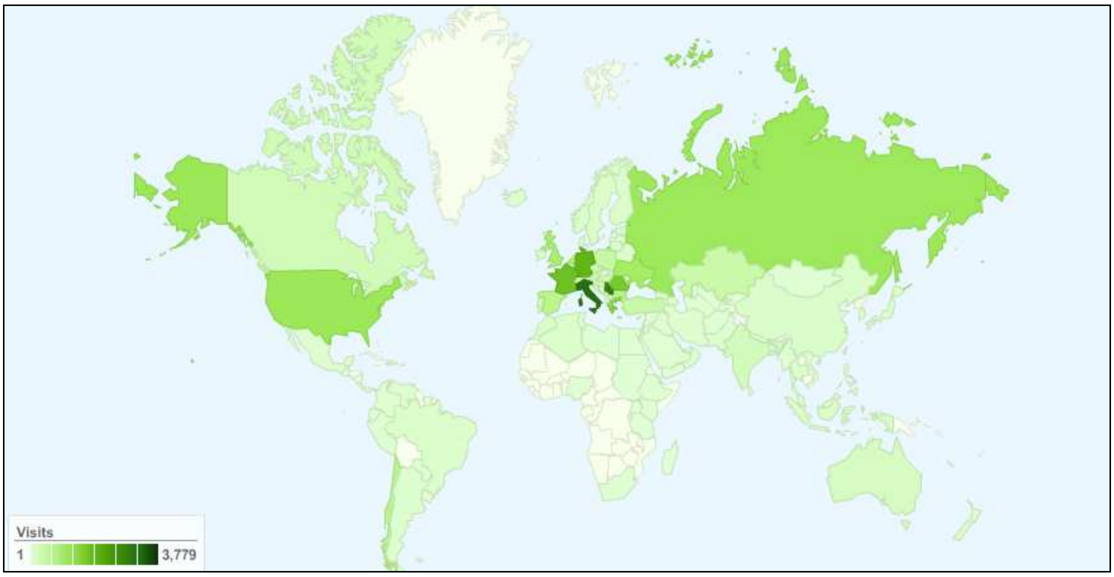


Figure 1 Geographic origin of users of the Earthquake Data Portal

The portal aggregates data access tools from several data centres and providers into a single unified web interface, including the event, broadband and accelerometric data explorers, request management "data carts," the historical earthquake catalogue browser, and a visualization tool for the European seismological reference model. These tools provide access to a range of data sets from different data centres and broad array of data providers.

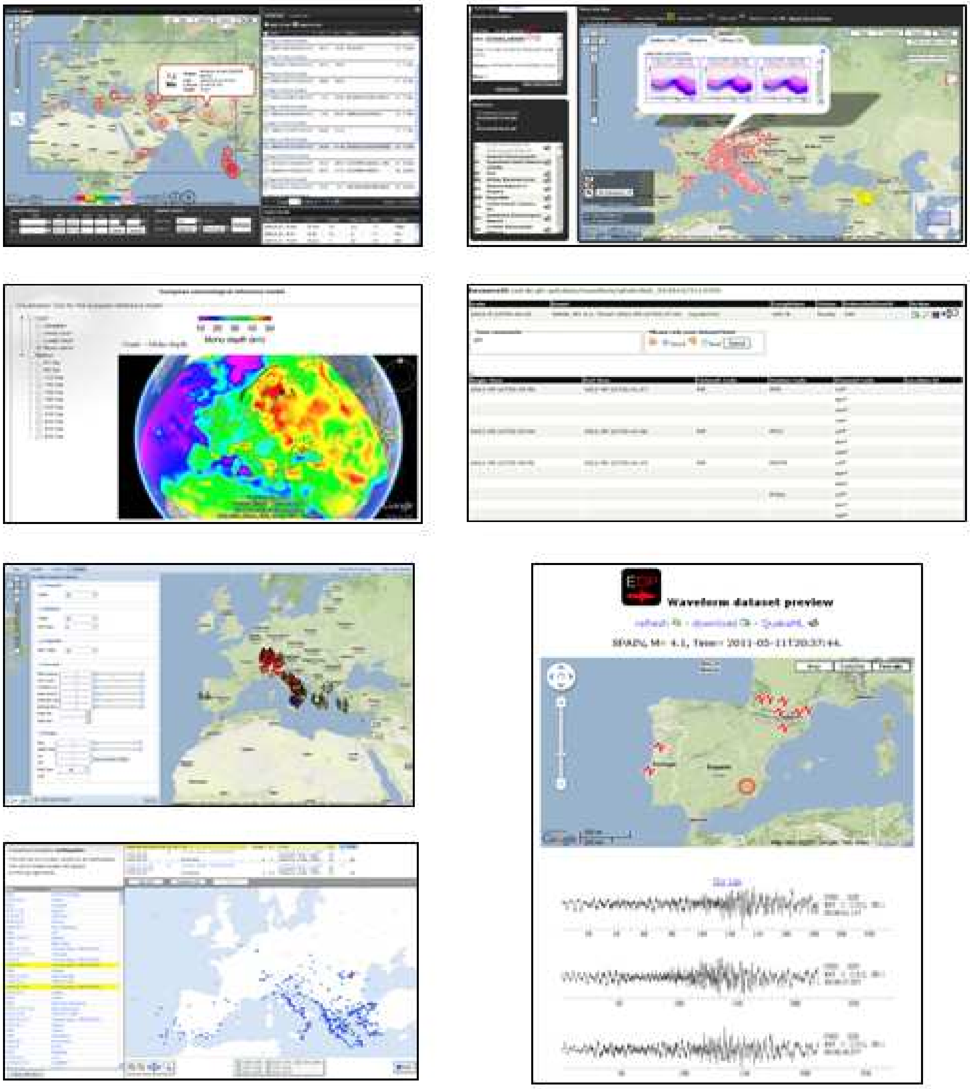


Figure 2 Tools available within the Earthquake Data Portal

3.2 Portal Architecture and Web Data Services

The portal architecture is based on the Java Portlet standards (JSR-168/286). Individual Java portlets are aggregated within the Portal, providing a single access point to the different data centres and access tools. The portal is hosted by and runs at the EMSC, while the individual Java portlets run at the distributed data providers. The portlets in turn access an underlying Service-Oriented Architecture of data web services.

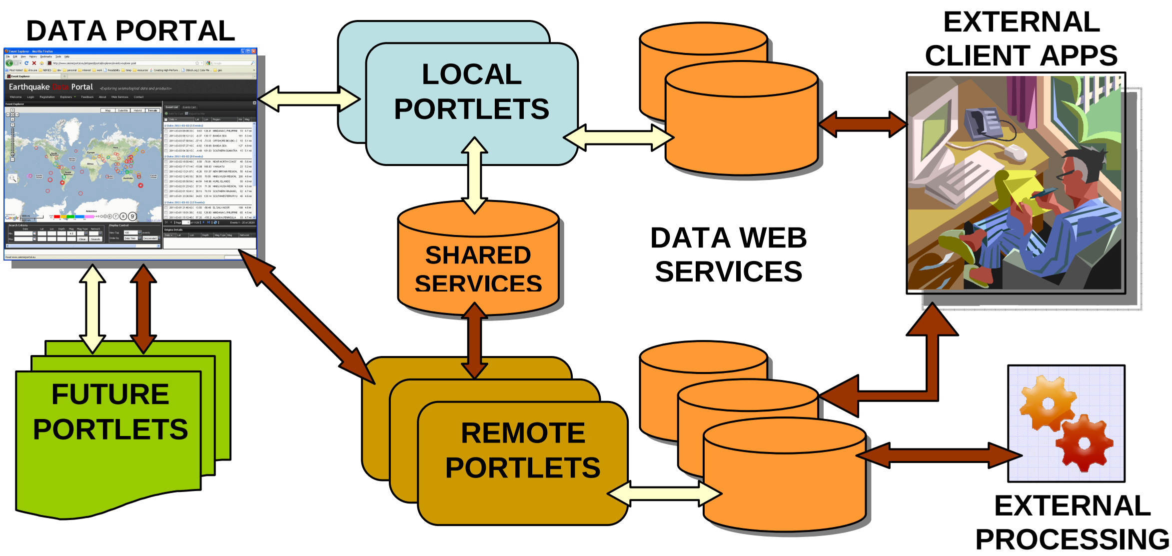


Figure 3 Portal architecture based on JSR-168 Portlets and Web Services

Some of the inter-portlet communications are managed through shared data services. For instance, the Broadband Waveform Explorer accesses the user's Event Data Cart to build event-oriented waveform requests for selected events.

The data services are also available for access by external client applications, supporting the development of individual and community user applications. There are currently several external sites regularly accessing the web services directly.

In addition, the data services are the basis for extending the architecture with external processing services. In response to invocation requests, these external services access data sets by reference from the web data services.

Currently, the data services do not conform to any standard or common API. There is some adherence to standard data formats, eg QuakeML, but the application interfaces are specific to each service and data type. It is therefore necessary to write custom code to access each service, thereby limiting both scalability and integration of new data sources and services.

3.2.1 OGC Services

Implementation has begun on the development of Open Geospatial Consortium (OGC) standards-compliant web services within NA9. The EMSC is working on a service providing earthquake location information through OGC Web Feature Service (WFS) and Web Map Service (WMS) services. These services return earthquake location information in a standard format for consumption and/or display in OGC-compliant GIS systems and other client applications. A collection of example clients in various languages and platforms are also provided. The preliminary access page is available at http://www.seismicportal.eu/ogc-beta.

3.3 Feedback Surveys

As part of the review and planning process, a survey was developed to solicit feedback from the NERA project participants regarding their interests and requirements with respect to the Data Portal and their plans to provide new integrated data access tools within the portal for their datasets. After the review of the response to the internal survey, another survey will be developed to be sent to the registered user-base to solicit detailed feedback concerning the usability aspects of the current portal implementation, toolset, and data access services. Responses from these surveys will then provide background and guidance for decisions concerning future developments and enhancements.

The first part of the internal survey gauged the respondents' familiarity with the EDP and other data portals, and the frequency of use in their regular work processes. The next part solicited feedback on general categories of services that should be implemented within the portal, such as emphasis on data visualization services, data processing services, or subscription-style notifications of new data availability. Finally, a series of questions were presented concerning the work package plans and expectations for providing access to their data and results through the portal.

Following the internal survey, a broader survey directed at portal users is planned. This survey will be sent to all the portal registered users. Like the internal survey, it will gauge the users' familiarity with the EDP and other data portals. It will also enquire about their usage patterns on the EDP and other data portals. Finally, it will solicit feedback on specific possible features and functionality, as well as provide an opportunity for open commentary and suggestions.

4. Development Strategy

The Earthquake Data Portal provides an interactive toolset for discovery and access to some of the varied data and products developed and delivered under the NERIES project, with development and expansion continuing into the NERA effort. This interactive approach is appropriate for discovery and browsing, allowing the user to find and visualize what data sets are available. It also works adequately for occasional one-off, limited, or specific data accesses. However, modern approaches to science often require access to large data sets, or scripted access to broad swaths of data matching a certain set of constraints or in response to external stimuli, such as the occurrence of an earthquake matching defined criteria. While development, improvements, and expansion of the available tools and data sets will proceed under the NERA NA9 effort, a parallel core focus of the work package will be the definition and development of common standard services architecture.

Both of these development avenues are important, with the interactive activities typically preceding, encouraging, and supporting programmatic access by providing data discovery and visualization, informational pages, and help and access tutorials, etc. As new data providers and data sets become available, it is important to provide an intuitive interactive graphical interface to explore the available data and to provide informational pages. As familiarity with the data grows, and for data sets that regularly change or grow, users will often want to build applications to automatically retrieve new data. The existence of, and adherence to, a published uniform data access standard makes this task tractable in the face of multiple heterogeneous data sets. At the same time, this heterogeneous nature of the data providers and data sets is the very reason the definition and implementation of common interface definitions suitable across all participating domains and data types is an ongoing challenge. To that end, an emphasis is placed on the definition, adoption, and development of a coherent standardized web-based services architecture.

4.1 Expand Current Portal Tool Set

As new data providers and data sets become available, it is important to provide an intuitive interactive graphical interface to explore the available data and to provide informational pages. Within the context of the NERA Data Portal, it is important that there be a fairly consistent and uniform look-and-feel and operational mode across the deployed tools, to the extent possible between the different data sets. This makes it easier for the user to switch between tools and to have a common intuitive understanding on how to proceed with each tool.

The architecture and implementation of the Data Portal and supporting services have been shown to be viable for both the development and integration of interactive tools, as well as for the support of external client applications. The Java Portlet-based portal architecture allows the independent development and management of the data access tools, while still presenting a single, unified access point to the project datasets. In addition, this architecture has been adopted by other European projects, enabling collaboration and integration efforts between projects.

We intend to leverage the mature and stable portal with its existing tools and services, not to mention core developer knowledge, by developing new tools within this established architecture. It is anticipated that the tools provided by new partners and from the other work packages will be built and delivered within this architecture to promote use of common technologies to facilitate their integration into the Portal. Existing tools will also see improvements and new features added within the existing architecture as identified from the user feedback surveys.

4.2 Extend and Standardize Web Services

An important effort within NA9 is the specification of a standard and consistent common services architecture for federated waveform and earthquake data services. The design process of this services architecture will be coordinated with similar efforts in other EU projects and international organization such as IRIS.

A tight collaboration among NERA and VERCE will evaluate the possibility to integrate such services in the data access components that will be described by a metadata registry, providing semantic and syntactic descriptive information about the service and its inputs and outputs. This registry provides standard descriptive metadata such as the service name, description, provider, and similar kinds of information. It may also provide more detailed information such as the algorithm or quality constraints, where appropriate. It must also provide detailed information about the required and optional input parameters and output results. This information includes the names, descriptions, and – most importantly – accepted data types of inputs, and descriptions and available formats for outputs.

4.2.1 Resource Orientation

Resource Orientation is an approach to developing a services architecture that puts a strong focus on the identification of all elements within the architecture, from service endpoints to data result sets to metadata catalogues, as identified resources. Essentially anything that can be identified is considered to be a resource. All resources are identified with a Universal Resource Identifier, or URI. HTTP URLs are one form of URI. As an example, a web page is a resource identified by its HTTP URL. Another example is detailed event parametric information available from the EMSC QuakeML service, such as: http://www.seismicportal.eu/services/event/detail/19990817_0000001. Note that the service itself, with URI http://www.seismicportal.eu/services/event/detail, is also an identified resource.

Not all resources are necessarily static, which is to say that, although uniquely identified, they may not always have the same content. Take for example a query for all the events in the region bounded by 26.1 and 45.0 W longitude and 36.3 and 42.0 N latitude. This resource can be identified by the URL http://www.seismicportal.eu/services/event/search?lonMin=26.1&lonMax=45&latMin=36.3&latMax=42. This resource will often return a different result set as new events occur in the region. However, even if the representation of this resource may change over time, it is still considered an identifiable resource. As such, it can be tracked and managed, and additional information – metadata – may be associated with it, as with all resources.

One important bit of metadata is the format of the returned representation, that is to say the output data format. In this case, it is QuakeML event parameters. By tracking this output format and associating it with the resource, in this case a data service query, it becomes possible to then pass the output of this resource to another resource that can accept that format.

The development of a coherent and consistent URI-based identification scheme for all resources with the system is the foundation of Resource Orientation. From this foundation, we can build a Resource Oriented Architecture, in which all components of the system, which are themselves resources, operate on or with other resources.

4.2.2 OGC Services

The Open Geospatial Consortium (OGC) defines a broad set of standards for the access, management, manipulation, transfer, and encoding of geo-located information. These standards are broadly accepted, with over 640 implementing products registered with the OGC and innumerable academic and other research projects based on or investigating aspects of the OGC standards stack.

Given the broad acceptance of the OGC standard interfaces, and thus opportunities for direct integration with other projects supporting the same standards such as the OneGeology project, we will investigate the feasibility of using the OGC standard interfaces as the basis for the NERA standard services architecture, both for data access and data processing services. Of course, for this to be feasible, it must work for all data sets and access tools.

4.3 Virtual Data Workbench

The current portal architecture includes Data Explorer tools to discover and submit requests for available data. These requests are subsequently managed within so-called Data Carts, through which request status can be checked, data can be downloaded, and certain other operations may be available to be performed on the requests or the result data sets, depending on the data type. For instance, within the Broadband Waveform Data Cart, completed requests (*ie*, requests for which all available data has been gathered and packaged) maybe be previewed or displayed using an available data visualization tool. Each Data Cart tool is supported by a data cart web service. Some of the explorer tools are also able access the data carts of other explorers. This mechanism provides the inter-tool communication within the portal. For example, the Broadband Explorer accesses a user's Event Data Cart in order to provide event-oriented data requests.

The idea of the Workbench consist in an extension of the current user carts with new tools allowing, for example, the subscription to the retrieval of heterogeneous new datasets, triggered by the occurrence of user defined conditions. After the datasets generation is triggered, the user data carts will be populated consistently. Moreover, in collaboration with the VERCE infrastructure developments, new data customisation tools will be available in order to apply core and fundamental transformations on the collected data, on demand.

4.4 Requirements-Based Development

It is essential for any successful product to base their developments on clearly articulated user requirements. However, it is often difficult to get accurate user feedback on *proposed* developments and tools, as users may have varying interpretations of the proposal or may fail to grasp important aspects of the design. The development of the portal under the NERIES project followed an iterative development approach to address this "chicken-and-egg" problem, providing early preliminary versions for user review and feedback and then adjusting the trajectory accordingly. The NERA project benefits from the existence of the final version of the NERIES

Earthquake Data Portal and a large user-base from which to solicit comments and feedback.

4.4.1 Internal Feedback

As the NA9 portal is expected to provide the framework for the access and distribution of NERA project data sets and products, a usage and requirements survey has been sent to all of the project participants requesting feedback on their current usage patterns, expectations, and foreseen requirements necessary for producing access tools for their work package data and results. These results will help advise structural and technical discussions on how best to accommodate the varied data providers and data sets produced within the NERA project.

4.4.2 User Feedback

Following the evaluation of the responses to the internal solicitation, a new survey will be produced and sent to all of the registered users of the portal, of which there are currently over 1000. The intent of this survey will be to better characterize the user experience with and reaction to the current toolset, portal design, and web services, and also to solicit, in a controlled manner, suggestions and opinions on proposed new features and functionality. A road map can then be developed to plan the evolution of the portal tool set and features that takes in to consideration user requirements and requested features, along with project aims and inter-project coordination efforts.

4.5 Coordination

In order to avoid both duplicated effort as well as divergent strategies and implementations, it is imperative that related projects within the European and global seismology and eScience IT efforts coordinate their work. We are fortunate within this project to have close associations with several related efforts.

4.5.1 Internal Coordination

Within the NERA project, it is expected that project-generated data, products, and results will be made available through access tools provided on the portal. Several working groups have already expressed their intentions in this regard. There is already a strong coordination effort in place between the services interface development in NA9 and the development of processing tools in JRA2. In addition, the solicitation of feedback from the NERA participants will aid in both the understanding of the other working packages' requirements, as well as increasing their awareness about the portal.

One of the primary goals of the NERA project is to bring together the seismological research and earthquake engineering communities in order to foster increased awareness and cooperation, leverage each community's expertise, avoid duplication of effort, and advance the science in each domain. While these communities have significant overlap in areas of interest, they have generally progressed largely independent of each other. The portal therefore aims to provide the framework to integrate multi-disciplinary data sets and data providers, including earthquake engineering data sets, seismic hazard, and seismological data sets. Open issues remain, however, in the identification of the data and content that will be furnished by the engineering community within the portal, as well as identification of what seismological tools and data can be provided to be used by the engineering communities. The participants are engaged in a discussions and the exchange of background and informational documents in order to explore the possible avenues of integration.

4.5.2 External coordination

4.5.2.1. SHARE

The SHARE project is developing REST-based data access services as well as portlet-based viewers for their hazard products (hazard curves, hazard maps, hazard spectra, and disaggregation data. Before having time dependent hazard assessment routines in place, the SHARE portlets and services are the data backbone of EFEHR. However, as soon as authoritative SHARE datasets are available, SHARE/EFEHR portlets will also be integrated into the EDP, and data services made available to the Workbench. In addition, coordination between the SHARE project and the engineering community to discuss engineering requirements for SHARE products is ongoing.

4.5.2.2. VERCE

The recently-funded FP7 Virtual Earthquake Research Communities for Europe (VERCE) project aims to provide a comprehensive architecture and framework adapted to the scale and the diversity of data-intensive applications in data mining and modelling, and to integrate 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.

The NERA NA9 effort will work closely with the VERCE project to coordinate the planning and design of the Common Services Architecture in order to advance a uniform standard within the community. The resource-orientation and enhanced metadata facilities of the services architecture, which in turn enables the Virtual Data Workbench, will all be building blocks for the development of the workflow tools and data sharing mechanisms on which VERCE will depend.

4.5.2.3. EPOS

The European Plate Observing System (EPOS) project is currently in the 5-year Preparatory Phase, during which time it will develop its legal and administrative bodies, work towards the effective integration of national Research Infrastructures, and develop a prototype e-science platform supporting the scientific and integrative goals of the project.

4.5.2.4. Engineering community

The European earthquake and geotechnical engineering communities are represented within NERA by several organizations and institutes. Through contacts with these organizations, we will foster collaborative developments.

4.5.2.5. IRIS-DMC

The Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC) is an important data center for seismological waveform data and products. Currently they provide several interactive web access tools and a programmatic (CORBA) interface into their data holdings. They recently launched a broad set of web service-based access interfaces. ORFEUS and IRIS-DMC are presently working on synchronising where possible each others web services in order to provide a homogeneous service towards users. Also the technical implications for the integration of the IRIS-DMC within the EIDA (see section 2.1.2) are being worked out. A joint project initiative is foreseen in the near future to speed up this service synchronisation. As part of the services development within NA9, we will work to synchronize service interfaces with the developments at the IRIS-DMC.

4.6 Planning

The design and development of the Portal and supporting services architecture will pose both technical and organizational challenges. From the technical perspective, one of the challenges is the definition of common standards across the heterogeneous data sets, archive structures, and typical access patterns and is a primary concern. It has been shown in numerous other projects that one-size-fits-all approaches are at best difficult to achieve, and usually with significant sacrifice in the ultimate functionality. From an organizational point-of-view, the independent and distributed nature of the project teams poses significant communication challenges. Open communication channels with frequent meetings, both virtual and in-person, are necessary to develop and maintain a coherent view of the portal and services plan as well as an awareness of the progress and activities of each partner.

To address the issue of different data types and access modes, we are considering the entire suite of Open Geospatial Consortium service interface standards. While at the lowest level, these service standards are data-type agnostic, they do provide different yet related standards that support different classes of data and access modes. For instance, the mapping services provide both feature-based and coverage-based interfaces. In addition, the OGC also provides a set of standard service interfaces for access to sensor data. These services may be suitable for accessing waveform data. It will be necessary to investigate the SensorML and Sensor Web Enablement (SWE) suite of standards in regards to their applicability to windowed and continuous data. The OGC has also defined the Web Processing Service interface which, at first view, appears to provide a suitable interface for web-available processing services.

The Virtual Data Workbench depends on the definition and implementation of the Common Services Architecture and an associated solid Resource Identification Scheme. This will require some thought to ensure a logical, coherent, and stable naming scheme. The naming scheme needs to be adequately powerful and flexible to support all the foreseeable resource types, while remaining simple enough to be implementable within the constraints of the project.

As the Portal evolves towards a Scientific Gateway, it will important that participating support and processing services are self-describing. Each service must support a common standard interface that may be queried for the service description and metadata information. The format and content of the returned metadata is to be determined, but it needs to provide identifying and service description information, including the format and nature of any input parameters and controls, and the available output formats. In this way, any new service can be introduced within the Portal and can be queried to provide the necessary descriptive information in order to include the service into processing chains. This interface is called the Service Description Interface (SDI) and is a facet or sub-interface within the Common Services Architecture, and is similar to the getCapabilities operation common to many of the OGC standard service interfaces.

Portal Vision

The Earthquake Data Portal rests in a unique position within the seismological community. It provides a single point of access to a number of data access tools for distributed and diverse data sets. The flexibility and power of the architecture allow these tools to be independently modified and updated. This makes the Portal an ideal jump-off point for the discovery of available data sets. This position can be leveraged and expanded with new tools, and through improvements to existing tools. Over the course of the NERA project and collaborations with other projects, notably VERCE and EPOS, the Portal will develop into the primary access point for the varied data sets and associated processing resources. It will also provide informational resources, including tutorials, sample applications, and client libraries, for accessing the underlying data services layer.

The Discovery and Access tools within the Portal can be used to interactively build data requests which can then subsequently be saved, manipulated, and replayed, or exported for use by the data service clients. These saved queries are themselves resources that will be managed within the Workbench.

5. Conclusion

The European-Mediterranean Earthquake Portal and Services (NA9) effort will provide the foundations and tools necessary for the discovery and use of NERA project data and results. The Data Portal provides intuitive interactive tools for data discovery and preview. The data services, as programmatic interfaces defined within the structure of the Common Services Architecture, provide the building blocks with which to develop external access and higher-order scientific applications and workflows. In coordination with other major European projects, the approach and architecture adopted within the EMEPS will provide the basis for the development of Scientific Gateways and the coordinated services architecture necessary to build an operational and sustainable e-Science Infrastructure.

Building these foundations involves coordinated yet independent developments between several organizations and teams, as well as across projects. Having completed the NERIES project, the NERA project already has several developments in place, which include the experience and relationships developed during their implementation. This experience and existing infrastructure will be leveraged in the development of the new services. We also intend that it will be leveraged in the collaboration with other projects. Other projects, notably VERCE and EPOS, will base their developments on the standards and advances achieved within NA9.

The EMEPS development will proceed along two related tracks: the expansion and improvement of the Earthquake Data Portal, and the development and standardization of the supporting data services under the Common Services Architecture. The Portal will be extended with new data access and manipulation tools from the NA9 partners and from other NERA work packages. As the new tools are added and existing tools are updated, the Portal will migrate towards a full Scientific Gateway, providing the tools, informational materials, and management utilities for researchers to orchestrate the data access and processing. The Common Services Architecture will provide a Resource-Oriented set of standardized service interfaces implemented by all participants, supporting the development of both Portal tools, including the Virtual Data Workbench, and external applications. The Open Geospatial Consortium (OGC) suite of standards is investigated as the possible basis for the Common Services Architecture. One of the first development targets is an investigation of the suitability of the OGC suite of service standards for all of the relevant data types.