Grant Agreement no. 226967

Seismic Hazard Harmonization in Europe

Project Acronym: SHARE

SP 1-Cooperation

Collaborative project: Small or medium-scale focused research project

THEME 6: Environment

Call: ENV.2008.1.3.1.1 Development of a common methodology and tools to evaluate earthquake hazard in Europe

D 6.3 – Web service-oriented architecture providing access to all databases

Due date of deliverable: 31.05.2010
Actual submission date: 30.11.2010

Start date of project: 2009-06-01 Duration: 36

Swiss Seismological Service, Eidgenössische Technische Hochschule (SED-ETHZ)

P. Kästli, R. Krishnamurty, R. Siegert, L. Danciu, J. Woessner

Bureau de recherches Géologiques et Minières (BRGM)

A. Tellez-Arenas, F. Husson

Revision: 1

Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	X
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Table of Contents

1. Introduction

1.1 The Framework

The main objective of SHARE is to provide a community-based living seismic hazard model for the Euro-Mediterranean region and to provide a sustainable mechanism for future updates as new scientific sound methods and data are available. The project aims to establish new standards in Probabilistic Seismic Hazard Assessment (PSHA) practice by a close cooperation of leading European seismologists and engineers. In particular, the project employs state-of-the art Information Technology (IT) to disseminate the data and results to expert and interested non-expert users such as the media and general public.

Access to data and results will be granted to the public and specialist through a single entry point, the SHARE Portal via the project web-page at: www.share-eu.org. SHARE ensures full technical compatibility with the portal technology adopted for other portals in the seismological community. In cooperation with the European-Mediterranean Seismologcial Centre (EMSC) and the foundation for Observatories and Research Facilities for European Seismology (ORFEUS), we ensure not only compatibility in terms of portlets consumable in the different portals, but also in terms of the data services it is based on. SHARE designs its service infrastructure as a comprehensive and homogeneous access point to European seismic hazard data that will be further enhanced within the FP7-funded project "Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation " (NERA) to serve the expert seismological, engineering seismology and engineering community as well as non-expert communities. SHARE started to design a facility for hazard and risk assessment as the European component of the Global Earthquake Model Foundation (GEM) and will in cooperation with NERA cooperatively ensure the implementation of the European Facility for Hazard and Risk (EFEHR) as primary access to hazard and risk data and results in Europe.

This document describes the design of the architecture that works behind the scenes of the portal GUI and allows interactive as well as script-based data access: the web services. We outline the technical and non-technical requirements, the implementation strategy and defines remaining tasks.

1.2 Why using web services

While in the first half of the last decade, it was still common to distribute scientific data along with viewer applications and specific analysis tools on offline media, the availability of stable, fast, and ubiquitous internet connections nowadays allow for centralized data storage and on-the-fly access in each visualisation or analysis step executed or requested by a remote user. The advantages are obvious:

1. Centralized data maintenance, update, and backup by the data owner
2. Media-free data distribution: cheaper and more tailored to the user's need
3. Access control and behaviour-monitoring of data users allow the design of data products and services more adequately tailored to the needs of the user
4. Depending on data and performance requirements, calculation services can be provided by the data providers themselves, executed client-side or outsourced to high-performance computing centers.

Web services allow access to data as well as to functionality using the most general and commonly available type of networking (TCP/IP networks including public internet) at low entry costs (no specific client required, port 80 connectivity sufficient).This flexibility overwhelms techniques with binary data transfer, but the requirement of custom thick clients.

1.3 Development history

With the initial implementation phase of the Global Earthquake Model (GEM1, see http://www.globalquakemodel.org/gem1) and the first year of SHARE implementation, a uniform framework for data storage (in postgres databases), calculation (java engines), frontend (JSR 286 portal / portlets), and interlinks (plain data in community based XML formats over SOAP web services) was envisioned. Web service providers and consumers, for earthquake hazard and risk calculation as well as for data discovery and visualisation, were implemented using the same software stack based on java classes with JAXB bindings 1.

After one year of development and with the end of the GEM1 pilot project, this IT framework was reviewed by an international panel of experts. This review resulted in the following criticism:

1. If the framework should allow everybody to do his own earthquake hazard and risk calculations, then it may not be scalable enough
2. For some data types, alternatives to a relational database as data backend should be considered
3. A web services API may be heavy in places where it is not used publicly anyway, e.g. between hazard calculation and data storage
4. Entry costs for RESTful web services may be lower than for SOAP based web services
5. Portal/portlet framework may be heavy, expensive to develop, and old-fashioned (not a-priori integrating with social APIs of facebook, google etc.)

After this review, in July 2010, the development teams and strategies of GEM and SHARE were split, with GEM focussing on the development of a multirisk calculator infrastructure (without frontend for the time being), while SHARE working at the presentation of earthquake hazard data to a scientific public.

The feedback of the GEM1 IT review was used as follows:

1. The scalability issue does not affect the SHARE project, as its scope is to provide state-of-the-art hazard calculations for Europe rather than a framework that allows others to do it.
2. Although all data is primarily stored in a relational database, data usually delivered in big junks (especially map data for display) is cached in binary grid files as a secondary data storage, for more quick delivery.

[1] For a detailed description of this stage of the OpenGEM IT design, see Krishnamurthy R., Euchner F., Mömke A., Siegert R., Kästli P. (2010) OpenGEM System Design Document, GEM Technical Report 2010-6, GEM Foundation, Pavia, Italy. (http://www.globalquakemodel.org/system/files/doc/GEM-TechnicalReport_2010-6.pdf)

3. No developments are currently carried out at the web services interfaces between data storage and hazard calculators; later on, results may be written to storage by code accessing the DB directly.
4. Web service development is moved from SOAP to RESTful style; existing services are redone as soon as they need functional modifications, or new clients are developed. While initially SOAP was adopted anticipating complex session-based applications with server-side state control, these requirements did not materialize, and REST provides uniquely simple service access using HTTP requests issued in the simplest case even from wget or a web browser.
5. The dedication to the portal/portlet framework is maintained, for the following reasons:
    a. SHARE is to integrate with other European Framework projects, such as Network of Research Infrastructures for European Seismology (NERIES, www.neries-eu.org), Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation (NERA), Research Infrastructure and E-Science for Data and Observatories on Earthquakes, Volcanoes, Surface Dynamics and Tectonics. (EPOS, www.epos-eu.org), which use the same techniques;
    b. For long-term sustainability, standards-based development is unavoidable – parallel and sequential project integration over JSR 286 is a heavier argument at long term than pure development speed;
    c. In contrary to GEM, the integration in social networks is not a primary goal of SHARE.

2. Web services in the framework of the overall SHARE-IT design

2.1 Types of services

In the implementation of the SHARE web services framework, four different types of services are distinguished:

Data discovery services: Services for looking up, and filtering, indexes of available data products (e.g. list all hazard models for a specific area expressing hazard using a specific ground motion parameter). Data is delivered in XML, following simple ad-hoc formats.

Data access services: services for looking up data products referenced by an index (e.g. provide the mean standard model hazard curve for a return period of 475 years). Data is delivered in XML, following community-based standards (pre-existing ones, as e.g. quakeML for earthquake catalogs, or data standards developed in the framework of SHARE, GEM, and related projects (as e.g. SHAML [Seismic Hazard Assessment Markup Language, Pagani et al. 2009] / NRML [Natural Risk Markup Language, current development project within the GEM framework]).

Map display services: Services for screen display of hazard input data (e.g. colored screen views of earthquake catalogs, faults), and results (e.g. hazard maps). Data is delivered as georeferenced png24-images.

Map data services: Delivery of parametric data of preliminary spatial nature, e.g. the grid of ground motion levels with the same probability of exceedance within a given time period for central Europe, given specific model assumptions.

2.2 Implementation strategy

For the time being, all services developed as needed for the functionality of the European Hazard Portal (see deliverable 6.5). However, all services that are of general interest, implemented following the latest development paradigms (especially ported to rest style), and stable, are published at **http://appsrvr.share-eu.org:8080/share/**, for direct usage by 3rd party clients.

2.3 Implementation of data discovery services

Services are implemented in JAVA using the standard http-request and response techniques without overhead. Data is accessed based on a object-relational mapping based on a Hibernate layer initially generated programmatically, however, manually adapted for the support of spatial data types: PostGIS in the database, Hibernate spatial / Java Topology Suite in Java: simple <latitude> / <longitude> attributes or GML objects in the delivered XML.

The REST paradigm of one URL corresponding to one resource is applied trying to replicate the natural ontology of the objects referenced. However, for the ease of understanding, where data restrictions do not translate directly into subordinated resources, query parameters are used. This may be explained using the following example: A hazard map may be represented as a sub-ressource of the hazard model it was generated on:

http://serviceprovider/hazardmodels/[mymodel]/maps/[mymap]

However, a hazard value at a specific lat/lon coordinate pair may not easily be described as a subressource, when referencing it with latitude and longitude (as a latitude value is not a subressource of a longitude value, nor vice versa). For the ease of understanding, instead of introducing a "geoaddress"-resource integrating latitude and longitude properties, we work with selection parameters:

http://serviceprovider/hazardmodels/[mymodel]/maps/[mymap]/hazardvalues?latitude=lat&longitude=lon

rather than:

http://serviceprovider/hazardmodels/[mymodel]/maps/[mymap]/hazardvalues/[myGeoAddressAsFunctionOfLatLon]

2.3.1 <u>Example: Hazard Map</u>

The character of the services currently implemented may be illustrated with examples of those implemented as a selection of a hazard map on the portal at http://portal.share-eu.org:8080/jetspeed/portal/HazardMaps.psml :

Task:	provide an index of all hazard models that are defined for a specific point of interest
Service URL:	http://appsrvr.share-eu.org:8080/share/model?latitude=47.0&longitude=6.0
Server response:	<models> <model> <id>**1**</id> <name>**GFZ Europe Model**</name> </model> <model> <id>**11**</id> <name>**GEM1 Hi-Res Global Model**</name> </model> <model> <id>**7**</id> <name>**GEM1 Global Model**</name> </model> <model> <id>**1**</id> <name>**GFZ Europe Model**</name>

	</model> </models>

Task:	Provide all intensity measurement types for which model 1 (in our case the preliminary GFZ Europe model) provides hazard information
Service URL:	http://appsrvr.share-eu.org:8080/share/model?id=1
Server response:	<imtcodes> <imtcode> <code>**PGA**</code> </imtcode> </imtcodes>

Task:	Provide all probabilities of exceedances of peak ground aceleration, and time periods for which model 1 provides hazard maps
Service URL:	http://appsrvr.share-eu.org:8080/share/model?id=1&imt=PGA
Server response:	<exceedances> <exceedance> <hmapexceedprob>**0.02**</hmapexceedprob> <hmapexceedyears>**50**</hmapexceedyears> </exceedance> <exceedance> <hmapexceedprob>**0.05**</hmapexceedprob> <hmapexceedyears>**50**</hmapexceedyears> </exceedance> <exceedance> <hmapexceedprob>**0.1**</hmapexceedprob> <hmapexceedyears>**50**</hmapexceedyears> </exceedance> </exceedances>

Task:	Provide all soil types, for which model 1 provides Peak ground acceleration information for a probability of exceedance of 20% within 50 years
Service URL:	http://appsrvr.share-eu.org:8080/share/model?id=1&imt=PGA&hmapexceedprob=20&hmapexceedyears=50
Server response:	<soiltype> <type>**rock**</type> </soiltype>

Task:	Provide all soil types, for which model 1 provides Peak ground acceleration information for a probability of exceedance of 20% within 50 years
Service URL:	http://appsrvr.share-eu.org:8080/share/model?id=1&imt=PGA&hmapexceedprob=20&hmapexceedyears=50
Server response:	<soiltype> <type>**rock**</type> </soiltype>

Task:	Provide the map identity, and the layer reference for the web map service, of the hazard map for PGA with a probability of 20% within 50 years, from model 1, calculated on rock
Service URL:	http://appsrvr.share-eu.org:8080/share/model?id=1&imt=PGA&hmapexceedprob=0.02&hmapexceedyears=50&soiltype=rock
Server response:	<hazardmaplocation> <hmapid>**26**</hmapid> <hmapwms>**hmap26**</hmapwms> </hazardmaplocation>

Note: All index choices are dependent on previous selections; thus the querying sequence matters. However, the querying sequence does not represent a selection from a general resource to a special subressource, rather a data mining workflow.

2.4 Implementation of data delivery services

The classical example of this service type is the hazard curve service, providing the hazard curve information (pairs of ground motion and probability of exceedance) for the hazard curve portlet (http://portal.share-eu.org:8080/jetspeed/portal/HazardCurves.psml)

In contrary of the responses of the index services, which require the client to track the context of a response), data products should be full-featured, i.e. contain all meta-information of a model, soil type, logic tree branch etc. along with the hazard curve data, in order to be useful also after being stored out of context. The respective services have been implemented on the common development platform with GEM (using draft SHAML format, see example in appendix 1, and SOAP-based services). In the meantime, SHAML is re-engineered by GEM. Thus SHARE is currently re-evaluating the data format question for the reimplementation of the services in REST style.

2.5 Implementation of map display, and data services

Both service types are provided using the OGIS service standards Web Map Service (WMS) and Web Feature Service (WFS), provided by a UMN Mapserver installation. While WMS returns georeferenced imaginery for screen display, WFS provides GML-(geographic markup language) encoded spatial information with hazard attributes, and is readable by GIS clients such as Quantum GIS or ESRI ArcGIS. Server-side, WMS are served based on database extracts to binary (GeoTIFF) files, which are resampled, tiled, and enhanced with pyramids for optimal graphical results. This enhances the return speed of the service, while it requires manual (or script-based) consistency control with and update from the primary database. WMS services are behind all hazard map displays of the current preliminary portal implementation at http://portal.share-eu.org:8080/jetspeed/portal/HazardMaps.psml; they are currently not published explicitely (e.g. to WMS get-capabilities requests), in order to avoid abuse and misinterpretation, as all hazard maps currently available are only preliminary test datasets.

WFS services for attributed map data are served by the same mapserver installation, however based on data coming directly from the primary postgreSQL / postGIS data (without the performance enhancements and consistency limitations of external binary file dumps). They are used internally for testing, however, they do not yet have any publicly accessible application within the portal.

3 Outlook

The short-term scheduling of further web service developments are timed in a way to optimally support parallel scientific tasks with portal-based functionality:

December 2010	• Port hazard curve data services from SOAP to REST and from SHAML format to NRML (Natural Risk Markup Language) drafted within the GEM OpenQuake software development project • Allow efficient multi-site hazard curve requests Both service development steps are required for finalizing the hazard curve display portlet for single-branch and aggregated models. They can be started without any further preconditions.
January 2011	• Provide a data submission service (initially for internal usage), for easy submission of hazard calculation results (representing curves or maps) into the SHARE databases. This step is required in order to use the portal as a tool for an initial model comparison workshop end of January 2010. This development can be started without any further preconditions.
February 2011	• Provide data services for discovering and retrieving datasets belonging to individual branches of logic tree models.

	• Provide mapping services for hazard input data (earthquake catalogs, source zones, faults) These steps are required in order to enable the hazard map and hazard curve portlets with displaying capabilities on logic tree branch level. Pending requirements are some clarifications on the possible structure of the logic trees, expected by December 2010 • Development of data discovery and access services for hazard spectra data. This step is required in order to back the hazard spectra viewer. It requires some further specification of the data backend as well as for the required discovery functionality, both scheduled to be provided by a workshop end of January 2011.
March 2011	• Development of data discovery and access services for information on hazard disaggregation. Depending on the axes of disaggregation, these services will have to look up data from storage, or calculate it on the fly. Preliminary decisions on the nature of disaggregation have been taken in September 2010, but they need confirmation when the logic tree structures are finalized.
Late Spring 2010	Model metadata infrastructure (data service backend for the model explorer – details still need discussion)

In mid term, the SHARE service developments seamlessly feed into the setup of EFEHR, the new European Facility of Earthquake Hazard and Risk hosted at ETHZ. By June 2011, this facility will get its first project-independent strategy planning, which will have a considerable impact on the scaling and priority of data services related developments.

In parallel to this, the NERA project started in November 2010 has a strong initiative to leverage data services on earthquake source parameters, waveforms, and, hazard information into an integrated seismic data analysis workbench environment, serving as a case study for the multi-geosciences European Plate Observatory which will take off in 2014. NERA will develop its technology blueprints by May 2011, and have a strong impact on the further technical implementation of the SHARE data service framework.

3. Appendix 1: Hazard curve in SHAML format (Version Spring 2010)


```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
   <shaml:HazardResultList xmlns:shaml="http://shaml.org/xmlns/shaml/0.1"
xmlns:gml="http://www.opengis.net/gml" xmlns:ns4="http://www.w3.org/2001/SMIL20/"
xmlns:ns5="http://www.w3.org/1999/xlink"
xmlns:ns6="http://www.w3.org/2001/SMIL20/Language">
      <shaml:Result timeSpanDuration="50.0" IDmodel="Model_Id" IMT="PGA">
         <shaml:Descriptor>
            <shaml:endBranchLabel>2_1_3</shaml:endBranchLabel>
         </shaml:Descriptor>
         <shaml:Values>
            <shaml:HazardCurveList>
               <shaml:IML>-5.298317366548036 -4.961845129926823 -4.625372893305611 -
4.290359446148058 -3.952844999948401 -3.6156289923743437 -3.280751228586288 -
2.9431398234348203 -2.6063965473757102 -2.2730262907525014 -1.9310215365615626 -
1.5945492999403497 -1.258781040820931 -0.9238189982949466 -0.5869869847315545 -
0.2510287548037454 0.0861776962410524 0.41871033485818504
0.7561219797213337</shaml:IML>
               <shaml:List>
                  <shaml:Curve maxProb="0.5479383429976421" minProb=
"0.0016018315252598336">
                     <shaml:Site>
                        <shaml:Site>
                           <gml:pos>-119.8 39.0</gml:pos>
                        </shaml:Site>
                        <shaml:Values>0.5479383429976421 0.5142291945445038
0.47046288878267073 0.4239089953599543 0.381035575668243 0.34397617314291096
0.30947590290782523 0.2718765686623872 0.22873125860109678 0.18348078588762617
0.14299504836070553 0.11483651754715807 0.0958847708686058 0.07898698751263844
0.05905857724899444 0.036959331482577795 0.017938297828110406 0.0064972365153485034
0.0016018315252598336</shaml:Values>
                        <shaml:vs30>760.0</shaml:vs30>
                     </shaml:Site>
                  </shaml:Curve>
               </shaml:List>
            </shaml:HazardCurveList>
         </shaml:Values>
      </shaml:Result>
```