NERIES: Seismic Data Gateways and User Composed Datasets Metadata Management

Alessandro Spinuso (1), Luca Trani (1), Linus Kamb (2), and Laurent Frobert (2)

(1) ORFEUS-KNMI, R&D Information and Observation Technology, de Bilt, Netherlands (spinuso@knmi.nl,+31 (0)30 2201 364), (2) European-Mediterranean Seismological, Bruyères le Châtel, FRANCE (kamb@emsc-csem.org,+33 1 69 26 70 00)

One of the NERIES EC project main objectives is to establish and improve the networking of seismic waveform data exchange and access among four main data centers in Europe: INGV, GFZ, ORFEUS and IPGP. Besides the implementation of the data backbone, several investigations and developments have been conducted in order to offer to the users the data available from this network, either programmatically or interactively. One of the challenges is to understand how to enable users' activities such as discovering, aggregating, describing and sharing datasets to obtain a decrease in the replication of similar data queries towards the network, exempting the data centers to guess and create useful pre-packed products. We've started to transfer this task more and more towards the users community, where the users' composed data products could be extensively re-used.

The main link to the data is represented by a centralized webservice (SeismoLink) acting like a single access point to the whole data network. Users can download either waveform data or seismic station inventories directly from their own software routines by connecting to this webservice, which routes the request to the data centers. The provenance of the data is maintained and transferred to the users in the form of URIs, that identify the dataset and implicitly refer to the data provider.

SeismoLink, combined with other webservices (eg EMSC-QuakeML earthquakes catalog service), is used from a community gateway such as the NERIES web portal (http://www.seismicportal.eu). Here the user interacts with a map based portlet which allows the dynamic composition of a data product, binding seismic event's parameters with a set of seismic stations. The requested data is collected by the back-end processes of the portal, preserved and offered to the user in a personal data cart, where metadata can be generated interactively on-demand.

The metadata, expressed in RDF, can also be remotely ingested. They offer rating, provenance and user annotation properties. Once generated they are included into a proprietary taxonomy, used by the overall architecture of the web portal. The metadata are made available through a SPARQL endpoint, thus allowing the datasets to be aggregated and shared among users in a meaningful way, enabling at the same time the development of third party visualization tools beyond the portal infrastructure.

The SEE-GRID-SCI and the JISC-funded RapidSeis projects investigate the usage of this framework to enable the waveform data processing over the Grid.