Distributed system for strong motion data retrieval and archiving : metadata, databases and data exchange within the NA5 framework

C. Pequegnat (1,3), P. Gueguen (2,3), and R. Jacquot (3)

(1) LGIT, Centre National de la Recherche Scientifique, (2) LGIT, Laboratoire Central des Ponts et Chaussée, (3) Grenoble University

The goal of the NERIES NA5 activity (http://www.neries-eu.org, Improving Accelerometric Data Access) is the development of common access to equally formatted event based accelerometric data and to the corresponding sheet of strong motion parameters. The core of the NA5 is made of 5 European institutes and the final protocol should permit other European institutes to integrate the NA5 portal. More precisely, the aim of the NA5 distributed data system is (1) to make available the data in a in specific format for the engineering community (i.e., ASCII) and in standard format for the seismological community (i.e. full SEED, SAC) and (2) to retrieve data at an unique portal on seismological and - accelerometric criteria, using relations between seismic sources and recordings and using specific parameters for the engineering community, i.e. site conditions and parameter thresholds (e.g., PGA, Ia, Duration, Sa(T), Sv(T)...). Parametric data as well as the procedures to compute them have been defined, implemented and make avalaible for all the NA5 partners.

The final product will be a system based on a distributed 3̃ tiersárchitecture, the three main nodes of which are : (1) the primary data servers of NA5 data providers, who make available waveforms (in ascii format) and the associated parameters and events-records tables, via ftp or http protocols (2) the NA5 portal, which supports metadata databases (events and stations metadata) and the associated user interfaces and webservices (3) the NA5 dataserver, the main function of which is the evaluation of the end-users requests, involving data retrieval, data conversion (sac, ascii and miniseed) and metadata formatting (sac, ascii and seed headers). Both NA5 portal and NA5 dataserver are presently under development, the former at EMSC, the latter at LGIT.

Our presentation will point out the main features and resources of the NA5 dataserver :

- a database of the instrument response files for the accelerometric channels (coefficient, poles and zeros) called PZ database has been set up, as well as some database maintenance tools. The PZ database is hierarchically structured into three directories : sensor, ana_filter and digitizer.

- user requests are materialized by a specific xml structure generated at the portal and pushed to the dataserver via QWID/EIDS (Earthquake Information Distribution System). A NA5 xml request contains all the information (metadata) needed for its evaluation : requested data files matching the user search criteria are described by their URL, their network, site of installation and acquisition channel, the events and hypocentral informations they are linked to, as well as values of the parametric data.

- the NA5 server notifies a xml parser, which checks and compiles the structure and store the information into an internal permanent relational metadata database. This database and the PZ directory are linked together by the way of a dedicated table, which contains directories and files names. Thus, the NA5 dataserver database will progressively be filled up by values carried on by NA5 requests.

- once a request has been loaded into the database, the NA5 archiver process will then fork into four different sub-processes, each of them calling the internal database to get value(s) for its parameters :

(a) get the ascii native waveforms files by downloading their URL if needed, i-e if the data are not already stored in the local cache, convert them into sac binary and miniseed (if needed, ie if miniseed files have not already been generated for this native waveforms files).

(b) build the linked seed dataless, if the PZ files for these channels are available.

(c) build the headers files (such headers will contain some more specific information about event, site condition and soil classification, max accaleration, aso.)

(d) pack the parametric data before archiving and compressing the result.

The archiver process will end by filling up and re-writing the xml request file with updated information (archive

URL for download, waveforms found and packed, size, etc.), before pushing it to the NA5 portal, for user notification.