SECOND EUROPEAN CONFERENCE ON EARTHQUAKE ENGINEERING
AND SEISMOLOGY, ISTANBUL AUG. 25-29, 2014

TOWARDS QUAKEML 2.0: NEW PACKAGES, TECHNIQUES, TOOLS

Fabian EUCHNER[1] and Philipp KÄSTLI[2]

During recent years, QuakeML has made great progress to be recognized as the most modern and practical data format for the exchange of seismic event parameter information. Its current version 1.2 was released in February 2013 and has been established since then for data dissemination at major seismological data centers throughout the world, like USGS, IRIS DMC, and GeoNet New Zealand.

QuakeML 1.2 features a modular architecture, with an umbrella XML Schema, and imported sub-schemas per thematic package . However, with v1.2 only one package, Basic Event Description (the "original" QuakeML) was published. Meanwhile, data models for several other thematic areas have advanced sufficiently, so that they could be included in the working model of the next generation of QuakeML at ETH, and relational databases based on the models are in productive use (peak ground motion, site and station characterization, hydraulic parameters of borehole injection processes). The inclusion of new packages lead to further partitioning of the schemas, creating separate packages for common, frequently reused concepts, such as identifiers, error models for numeric values, or representation of persons and bibliographic references. The new package layout introduces additional namespaces which will require some adaption when transferring old instance documents. The novel packages drafted at ETH can be considered sufficiently mature to be exposed to a public Request for Comments process.

QuakeML data model development at ETH uses a UML class diagram as the basic representation. Distributed model development is managed by keeping XMI serializations of the packages in a Subversion code repository. Other model representations are created automatically by a code generator written in Python, and include XML Schema, Relax NG, SQL Schema, SKOS, and Python class stubs.

In QuakeML 1.2 (and older versions), controlled vocabularies (like the one used for seismic event type) are simple flat enumerations, i.e., the hierarchy of terms is only implicit and cannot be evaluated programmatically. In QuakeML 2.0, we propose to express hierarchy explicitly by using knowledge representation categories of broader, narrower, and related. Also, support for multilingual terms and synonyms will be provided. Vocabularies will be exported into an SKOS definition per package, and thus can be used to supplement schema information in users' program codes.

Taking advantage of its improved modularization and knowledge representation features, QuakeML 2.0 will be enabled to make the step from a successful exchange format for a single parametric data application (earthquake source parameters) to a data modelling paradigm for a broad field of parametric data use cases in seismology, and possibly beyond. We hope that consistent data modelling and best practices in format definition will facilitate the development of an integrated data infrastructure for seismological research and monitoring in Europe and world-wide.

[1]Institute of Geophysics, Federal Institute of Technology, Zurich, Switzerland, fabian@sed.ethz.ch
[2]Swiss Seismological Service, Federal Institute of Technology, Zurich, Switzerland, kaestli@sed.ethz.ch