Geophysical Research Abstracts
Vol. 13, EGU2011-13462, 2011
EGU General Assembly 2011

The development of a comprehensive format for the representation of seismic tomography models

Peter Danecek, Luca Postpischl, and Andrea Morelli
Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy (Peter.Danecek@bo.ingv.it)

The need to establish open data standards is widely acknowledged, in most scientific disciplines, not only in computer science. Such open standards facilitate not only exchange and immediate reuse of scientific data or results, they also are a precondition for the bibliographic conservation of scientific results and the foundation for any effective and accessible data repositories.

In seismology, the most successful branch to establish standard data formats and exchanges protocols is probably observational seismology, that thereby facilitates exchange of seismograms (SEED), seismic events and event characterization (QuakeML). Modellers in seismology usually rely on various formats, which are given by their specific methods, parameterisation or applications, even if in some cases they share common code through software repositories such as SPICE (http://www.spice-rtn.org) or CIG (http://www.geodynamics.org). In particular, it is laborious and time-consuming to move the huge variety of 2D and 3D of tomographic models between applications. GIS related data benefit from well established data exchange standards, but focus on the representation of Earth surface features and georeferenced information, which are often unrelated to physical modelling. Other Earth science communities, notably meteorology and climatology, use standardised metadata conventions (COARDS, CF), which also serve modellers and find applications beyond this specific community, including seismology. However, these approaches fail to accommodate all requirements of seismological 3D modelling. They are targeted specifically to applications in their original field and provide only limited support for seismological data and tomographic models.

First approaches sample the model on a regular latitude-longitude grid, which is justified, if focused on visualisation and comparison of tomographic models. The utility of the resampled model representations beyond visualization is very limited, because the reparameterisation introduces uncontrolled approximations and inaccuracies, usually inacceptable for reuse in subsequent modelling. A comprehensive standard format for the representation of the volumetric dataset of a tomography model, would ideally preserve the original parameterisation and geometry of the model. This involves the self-description of the represented model, a detailed and complete description of the model geometry and parameterisation in a formal, consistent way and the representation of the model's volumetric field data in its original parameterisation.

We discuss the design goals in detail and present an approach which combines basic objects in order to "construct" the parameterisation and which describe geometry in the original coordinates. We propose a format draft for the related data structures and attributes, which are required to accommodate the necessary information. Due to the intended use with semantic web technologies and for web services, we represent the data structures using JSON – a hierarchical lightweight data-interchange format, which is human-readable and supported by most high-level programming language – and discuss examples in this TomoJSON format. Further, we discuss the limitations of a text-based, sequential format and outline strategies to eliminate these problems by translating the defined structures into equivalent binary formats.

We are convinced that the proposed format could be equally useful to other datasets, where georeferences volumetric data need to be represented. As this is an effort to establish a standard format for the representation of tomographic models, we very much appreciate feedback and collaboration from the wider community.