Geophysical Research Abstracts
Vol. 19, EGU2017-14781-2, 2017
EGU General Assembly 2017

Datacube Interoperability, Encoding Independence, and Analytics

Peter Baumann (1), Eric Hirschorn (2), and Joan Maso (3)
(1) Jacobs University, Bremen, Germany (p.baumann@jacobs-university.de), (2) KEYW Corp., (3) CREAF

Datacubes are commonly accepted as an enabling paradigm which provides a handy abstraction for accessing and analyzing the zillions of image files delivered by the manifold satellite instruments and climate simulations, among others. Additionally, datacubes are the classic model for statistical and OLAP datacubes, so a further information category can be integrated.

From a standards perspective, spatio-temporal datacubes naturally are included in the concept of coverages which encompass regular and irregular grids, point clouds, and general meshes – or, more abstractly, digital representations of spatio-temporally varying phenomena. ISO 19123, which is identical to OGC Abstract Topic 6, gives a high-level abstract definition which is complemented by the OGC Coverage Implementation Schema (CIS) which is an interoperable, yet format independent concretization of the abstract model. Currently, ISO is working on adopting OGC CIS as ISO 19123-2; the existing ISO 19123 standard is under revision by one of the abstract authors and will become ISO 19123-1. The roadmap agreed by ISO further foresees adoption of the OGC Web Coverage Service (WCS) as an ISO standard so that a complete data and service model will exist. In 2016, INSPIRE has adopted WCS as Coverage Download Service, including the datacube analytics language Web Coverage Processing Service (WCPS). The rasdaman technology (www.rasdaman.org) is both OGC and INSPIRE Reference Implementation. In the global EarthServer initiative rasdaman database sizes are exceeding 250 TB today, heading for the Petabyte frontier well in 2017.

Technically, CIS defines a compact, efficient model for representing multi-dimensional datacubes in several ways. The classical coverage cube defines a domain set (where are values?), a range set (what are these values?), and range type (what do the values mean?), as well as a "bag" for arbitrary metadata. With CIS 1.1, coordinate/value pair sequences have been added, as well as tiled representations. Further, CIS 1.1 offers a unified model for any kind of regular and irregular grids, also allowing sensor models as per SensorML. Encodings include ASCII formats like GML, JSON, RDF as well as binary formats like GeoTIFF, NetCDF, JPEG2000, and GRIB2; further, a container concept allows mixed representations within one coverage file utilizing zip or other convenient package formats. Through the tight integration with the Sensor Web Enablement (SWE), a lossless "transport" from sensor into coverage world is ensured. The corresponding service model of WCS supports datacube operations ranging from simple data extraction to complex ad-hoc analytics with WPCS.

Notably, W3C is working has set out on a coverage model as well; it has been designed relatively independently from the abovementioned standards, but there is informal agreement to link it into the CIS universe (which allows for different, yet interchangeable representations). Particularly interesting in the W3C proposal is the detailed semantic modeling of metadata; as CIS 1.1 supports RDF, a tight coupling seems feasible.