Geophysical Research Abstracts
Vol. 15, EGU2013-12154, 2013
EGU General Assembly 2013

Making Breakfast with 5 oz of Cinnamon Porridge and 150 gr of Sweet Oatmeal

Matthes Rieke (1), Luis Bermudez (2), Benjamin Pross (3), Aijun Chen (4), and Genong (Eugene) Yu (5)

(1) 52North, Germany (m.rieke@52north.org), (2) Open Geospatial Consortium, United States (lbermudez@opengeospatial.org), (3) 52North, Germany (b.pross@52north.org), (4) George Mason University, United States (aijunchen@gmail.com), (5) George Mason University, United States (gyu@gmu.edu)

Have you ever had two or more data sources for the same region and the data don't seem to fit? One data source has incomplete attributes, features are missing in one source or there seems to be overlapping features representing the same feature? Can we run a process where the differences appear to become lost? Such process is called conflation. Conflation is important to help users see different sources in a clean way, all at once, via one known model.

Conflation is a sophisticated process of unifying two or more separate datasets, which share certain characteristics, into one integrated all-encompassing dataset. In the process, different representations of a feature get folded into one feature. That target feature follows a particular model that is used to capture complementary information.

Conflation was one of the topics advanced in a recent activity within the Open Geospatial Consortium (OGC). OGC is a consortium with about 500 members and serves as a global forum for the collaboration of developers and users of spatial data products and services, and to advance the development of international standards for geospatial interoperability. The OGC Interoperability Program conducts international interoperability testbeds, such as the OGC Web Services Phase 9 (OWS-9), that encourages rapid development, testing, validation, demonstration and adoption of open, consensus based standards and best practices.

The OWS-9 advanced conflation of OGC Web Feature Services (WFS) that use heterogeneous GML application schemas. The conflation was performed in two phases. One performed geometry conflation, while the other performed attribute conflation. The attribute conflation makes use of mappings captured via ontologies in a knowledge base and available via a SPARQL server.

To provide the conflation functionality in an interoperable way, a profile for the OGC Web Processing Service (WPS) was developed. In this service, various conflation rules setups have been integrated to define the behavior of attribute resolving. A simple "add-only" scenario copies those attributes, which are not present in the target dataset. A more advanced "add and update" scenario adjusts the values of present target attributes with values from one of the input datasets. "Qualitative update" methods require the interaction with an expert user who decides on the resulting value of attributes that are present in both the target and input sources.

The work also included capturing of the provenance in a way that will allow a client application to display the information of the resulting conflated dataset at both feature and dataset level using ISO 19115 part I/II Metadata. The captured provenance at dataset level can be registered into an OGC standard compatible catalog service and further retrieved by client applications following interoperable and standard procedures.