SPATIAL DATA SUPPLY CHAINS

David A. McMeekin, Geoff West, Premalatha Varadharajulu, Muhammad Azeem Sadiq, Feiyan Yu, Lesley Arnold, Simon Moncrieff

Department of Spatial Sciences, Curtin University, GPO Box U1987, Perth 6845, Western Australia, Australia,
Cooperative Research Centre for Spatial Information, Australia.
d.mcmeekin@curtin.edu.au

Commission VI, WG VI/5

KEY WORDS: Spatial Data Supply Chains, Provenance, Conflation, Semantic Web, Ontologies, Artificial Intelligence

ABSTRACT:

This abstract describes current research into the supply of spatial data to the end user in as close to real time as possible via the World Wide Web. The Spatial Data Infrastructure paradigm has been discussed since the early 1990s. The concept has evolved significantly since then but has almost always examined data from the perspective of the supplier. It has been a supplier driven focus rather than a user driven focus. The current research being conducted is making a paradigm shift and looking at the supply of spatial data as a supply chain, similar to a manufacturing supply chain in which users play a significant part.

A comprehensive consultation process has taken place within Australia incorporating a large number of stakeholders. Three research projects that have arisen from the consultation process are discussed within this abstract.

The next generation spatial infrastructures must address multiple contemporaneous issues within the spatial data supply chains (SDSC). A SDSCs consists of numerous value add processes along the chain. At each value add point in the chain there may be heterogeneous geo-processes, methods, models and workflows combining to generate, modify and consume spatial data. The value add processes occurring in integrating and processing multiple datasets raises questions about data trust, quality, its fitness for purpose, currency and authoritative level because these datasets originated from different sources having had different geo processes executed upon them to arrive at this final product. Knowing how data is collected, what level of accuracy was used gives understanding to what purpose the data can be used for.

Geospatial data sharing is extremely important in Spatial Infrastructures (SI) as huge amounts of data are supplied by a variety of organizations, stored in different formats, and managed at different user levels. In Australia, the increased dependency on timely spatial data has led to an identified need to consider a supply chains model for spatial data from local the local government authorities all the way through to the Commonwealth government. Much of the Australian spatial data is acquired at the local government level, combined to form the State or Territory level datasets and then used to create Commonwealth level datasets. Many processes used in spatial data generation are manual and undocumented as well as implicitly requiring human intervention. There is a lack of or no linking mechanisms at all between datasets as well as multiple versions of data sets being used which may lead to an inaccurate or out of date dataset being used. There are dependencies between the different data at different levels including differing formats and human intervention. These factors complicate the integration at different levels. This research is examining technical solutions to the spatial industry supply chain problems through the application of semantic web and linked data technologies to address the boundaries and gaps identified that prevent seamless supply chain integration and operation. The research is concentrated on determining a universal approach that can be framed to deal with SI supply chain issues that allow for the understanding and automation of the supply chain process that incorporates multiple data sources, even crowd sourced data.

Semantic web technologies provide new methods for spatial data exploration and understanding. Data provenance as it is disseminated to users is an area that has not been examined. This research is addressing provenance in spatial data using semantic web technologies and artificial intelligence to resolve gaps in data provenance understanding along the SDSC. The W3C and OGC have recently proposed models for provenance that have been explored. A geospatial provenance model is proposed using semantic technologies as one possible solution to increase the "trustability" of both spatial datasets and processes that builds on the work of the W3C, OGC and others.

In Australia many organisations at local government level, state government departments in the different jurisdiction as well as commonwealth agencies, acquire spatial data for specific areas or points of interest. This leads to data duplication at multiple points on the SDSC. Lack of awareness or simply because no single dataset suits multiple agencies' needs leads to this duplication. To improve the SDSC process, this research is examining data conflation as a means to reduce or even remove the data duplication within the SDSC. Through combining multiple, overlapping data sources into a single point of truth dataset while retaining accuracy, reducing redundancy, reconciling data conflicts and obtaining richer attributes is the aim of this research. This research is applying semantic web technologies to automate this conflation process. The focus is on creating ontologies using OWL-2 for spatial datasets and coding relevant geometry, topology, and policy rules that can either be mined from the data or supporting documents. The satisfying of these rules (Description Logic or DL) through computer reasoning, relevant datasets may be intelligently linked and integrated. The conflated dataset then becomes a single authoritative, trusted data source, fit for multiple purposes. The data can also

be co-maintained, providing agencies etc. with an understanding that they maintain their own data and yet be used by multiple organisations eliminating the need for siloed systems.

Changes occurring along the supply chain occur because of new property developments, new roads and other changes to the environment. Many of these changes have to abide by laws and policies enacted by governments and typically are manually applied to determine whether changes are allowed or not. Creating new housing estates and the consequent new roads is one case being considered. Policies are being encoded as rules in OWL-2 and used with ontologies to provide decision support for the changes. OWL-2 allows an ontology describing knowledge about properties and roads along with axioms and constraints such as a new house has to have an entrance to or frontage onto a proposed road. Other rules implemented in Description Logic process the new data, axioms and constraints and advise the user as to whether the development is allowed or not.

Importantly the use of semantic web technologies for these three issues enables transparency because the data, rules, constraints etc. are publishable in W3C standard formats.