International Workshop on Spatial Data Quality, February 2018

Towards an Ontology for Geospatial User-centric Metadata

Hasti Ziaimatin[1], Alistair Barros[1], Ivana Ivanova[2] & Tom Spencer[3]

[1] School of Information Systems, Queensland University of Technology, GP Campus, Brisbane, QLD 4000, Australia
hasti.ziaimatin@qut.edu.au, alistair.barros@qut.edu.au
[2] Department of Spatial Sciences, Curtin University, Kent St, Bentley, WA 6102, Australia
ivana.ivanova@curtin.edu.au
[3] PSMA Australia Unit 6, Level 2, 113 Canberra Avenue, Griffith ACT 2603, Australia
Tom.Spencer@psma.com.au

Geospatial data quality has been an active area of research in the GIS community for more than thirty years (Devillers *et al*., 2010). Geospatial data are subject to processes such as generalisation, abstraction and aggregation; consequently, the transformed data can only provide an approximation of the real world, and often suffers from imperfect quality (Goodchild, 1995; Devillers and Jeansoulin, 2006). Thus, data consumers will always be exposed to some level of data uncertainty. With the increased use of geospatial data sources across heterogeneous user groups and domains, assessing fitness for use is emerging as an essential but complicated task (Triglav *et al*., 2011). Users are presented with an increasing choice of data from various data portals, repositories, and clearinghouses. As a result, comparing the quality of different data sources and evaluating a data source's fitness for use, presents major challenges for spatial data users. Consequently, research has focused on the challenge of communicating fitness for use of geospatial data, proposing a more "user-centric approach" to geospatial metadata (Goodchild, 2009).

In this context, the Cooperative Research Centre for Spatial Information (CRCSI) in Australia is conducting research into key informational aspects of spatial data sources that are influential to users for identifying data sources that are fit for their intended uses. This information will be used to create a vocabulary for communicating the fitness for use of spatial data sources to users in the spatial and other domains, in order to enable them to make informed data source selection decisions. As fitness for use of geospatial data depends on the application context, the vocabulary will enable users to describe fitness for use of data sources in the context of their specific application domains. The vocabulary will be predominantly produced by the consumers of spatial data products as they use and gain experience with those products.

The continuous adoption of Semantic Web technologies enables us to transform the vocabulary into a more dynamic and well-grounded formalism; i.e., an ontology. The Geospatial User-centric Metadata (GUM) ontology, represents spatial data users' requirements and implicit knowledge of spatial data sources in the context of specific application domains, using machine-processable concepts defined in widely adopted ontologies. This facilitates the application of reasoning techniques developed by the Semantic Web community. In addition, fitness for use descriptions represented using ontological concepts, can be published and integrated as structured data on the Web, where they can be linked to their corresponding data sources, leading to a seamless aggregation of spatial data and their user-defined metadata. Furthermore, the links between ontological concepts in these metadata and concepts in the Linked Data Cloud (Bizer *et al.*, 2009), can be discovered and used to complement the published data; e.g., user-defined fitness for use descriptions can be integrated with producer-supplied quality metadata. This will provide access to richer, more accurate and more informative quality metadata for spatial data sources.

The GUM ontology will use concepts from standards developed by standardisation bodies, such as ISO/TC 211, OGC, DCMI and FSDF. The ontology can be used not only to identify fitness for use of data sources in the context of specific application domains, but also to identify different use cases and users of data sources, as they describe fitness for use in the context of their application domains. Furthermore, producers can incorporate these user-defined metadata into objective quality measures for their products, allowing providers to meet users' specific requirements. The vocabulary can also be used to complement producer metadata by presenting user defined fitness for use descriptions for various applications of a data source. In addition, the hierarchical structure of ontologies, enables fitness for use to be described at various levels of granularity; e.g., dataset, feature, attribute. Finally, the ontology captures profiles of users that describe fitness for use of spatial data sources. These profiles can be used to assess the fitness for use of spatial data sources, considering attributes such as level of expertise, application domain and roles of users that provide the descriptions.

The ontology will be implemented in prototype software and a usability study will be conducted as an initial "proof of concept" level validation of the design. The project will also apply the ontology to a real industry setting, in order to assess its usefulness in this context.

References

Bizer, C., T. Heath and T. Berners-Lee (2009), "Linked data-the story so far". *International journal on semantic web and information systems*, 5(3): 1-22.

Devillers, R. and R. Jeansoulin (2006), "Spatial Data Quality: Concepts". *In*: R. Devillers and R. Jeansoulin (eds.). *Fundamentals of Spatial Data Quality*. *ISTE Ltd*, London, UK: 31-42.

Devillers, R., Y. Be´dard, P. Fisher, A. Stein, N. Chrisman and W. Shi (2010), "Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities". *Transactions in GIS*, 14(4): 387–400.

Goodchild, M. F. (1995), "Sharing Imperfect Data". *In*: H. J. Onsrud and G. Rushton (eds.). *Sharing Geographic Information*. *Rutgers University Press*, New Brunswick, Canada: 413–425.

Goodchild, M. F. (2009), "Putting Research into Practice". *In:* A. Stein, W. Shi and W. Bijker (eds.). *Quality Aspects of Spatial Data Mining. CRC Press*, Boca Raton, USA: 345–356.

Triglav, J., D. Petrovič and B. Stopar (2011), "Spatio-Temporal Evaluation Matrices for Geospatial Data". *International Journal of Applied Earth Observation and Geoinformation*, 13(1): 100-109.