

Eliciting User and Producer Views on Geospatial Data Quality

Hasti Ziainatin¹, Alireza Nili¹, Alistair Barros¹, Ivana Ivanova² & Tom Spencer³

¹ School of Information Systems, Queensland University of Technology, GP Campus, Brisbane, QLD 4000, Australia
hasti.ziainatin@qut.edu.au, a.nili@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

Spatial data supply chains today push data to the users via spatial Web portals or Web Services. The value of this information depends on the ability to anticipate users' needs and quality requirements. This approach is proving problematic given the unforeseeable and diverse nature of user requirements in the context of various application domains. The value of spatial data products is realised when the delivered knowledge enables users to achieve their intended purposes. In other words, the value of spatial data to consumers is not the actual data itself, but rather the fitness for use of the knowledge that is extracted from the data. Therefore, the value proposition is not so much about how consumers use information; but rather why they are using it (Arnold, 2016). Despite detailed recommendations of standardisation bodies such as ISO/TC 211, INSPIRE, OGC, DCMI and the existence of formal metadata standards, often data quality information is not communicated to users in a consistent and standardised way (Boin and Hunter, 2006). Furthermore, metadata standards are mostly focused on data production rather than potential data use and application; therefore, a typical metadata document is not sufficient to effectively communicate "fitness for use" to consumers from a variety of domains and expertise levels (Goodchild, 2009).

There is currently a shortage of empirical research relating to how people interpret and use quality metadata to make informed data source selection decisions in a real-world environment. The Cooperative Research Centre for Spatial Information (CRCSI) in Australia is conducting research into how users and producers of geospatial data across Australia and New Zealand evaluate the quality of data sources. Towards this goal, the project is in the process of conducting semi-structured interviews with spatial data users and producers from a variety of domains and applications. Interviews with geospatial data users, aim to identify key informational aspects of geospatial data sources that are influential in evaluating quality and fitness for use, i.e., external quality (Devillers and Jeansoulin, 2006). Interviews conducted with geospatial data producers, aim to identify producers' perceptions of geospatial data quality, i.e., internal quality (Devillers and Jeansoulin, 2006) and objective quality measures and elements of geospatial data that are used to describe the internal quality of data sources.

We have performed a systematic review of studies and research conducted into the assessment and evaluation of geospatial data quality and data source selection. The findings of this review will be used to develop an a-priori list of data quality elements; e.g., positional accuracy, attribute accuracy, temporal accuracy; and data quality sub elements; e.g., conceptual consistency, temporal consistency (Devillers and Jeansoulin, 2006). The semi-structured interviews with the producers and users of geospatial data will be analysed inductively (Miles *et al.*, 2014; Denzin and Lincoln, 1994) to capture generic and specific quality indicators identified by users as being influential in assessing the fitness for use of spatial data sources. Furthermore, an operational definition will be provided for each of the identified quality indicators. Data collection will continue until theoretical saturation is achieved; i.e., until no new information or new quality themes emerge from the interviews (Bowen, 2008). The results of the interviews will be used to refine the a-priori list, followed by an inter-coder reliability check to ensure the reliability of the qualitative coding of data. We have chosen Krippendorff's alpha (Krippendorff, 2004) as our method of inter-coder reliability check. The findings for each spatial data user will then be sent to them for verification. Once the findings are verified, they will be used to profile spatial data users according to the data quality elements identified as influential to them, for assessing fitness for use of data sources, in the context of their specific use cases. Producers will also be profiled, based on the data quality elements that they provide for the intended uses of their data sources.

This analysis and comparison of producer and user profiles will be used to identify the gaps that exist between internal quality (producer supplied) and external quality (consumer described). The information will inform further phases of the project, which aim to design a vocabulary that communicates the fitness for use of geospatial data sources to spatial data users in the context of specific applications and domains, in order to enable them to make informed data source selection decisions.

References

- Arnold, L. (2016), "Spatial Data Supply Chain and End User Frameworks: Towards an Ontology for Value Creation". *GeoValue Workshop*, Paris, France.
- Boin, A. T. and G. J. Hunter (2006), "Do Spatial Data Consumers Really Understand Data Quality Information?" *In: M. Caetano and M. Painho (eds.). 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Lisboa, Portugal.
- Bowen, G. A. (2008), "Naturalistic inquiry and the saturation concept: a research note". *Qualitative Research*, 8(1): 137-152.
- Denzin, N. and Y. Lincoln (1994), "Introduction – Entering the field of qualitative research". *In: N. Denzin and Y. Lincoln (eds.). Handbook of qualitative research*. Sage, Thousand Oaks, CA, USA: 1-17.
- Devilleers, 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.
- 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.
- Krippendorff, K. (2004), *Content analysis: An introduction to its methodology* (2nd ed.), Sage, Thousand Oaks, CA: USA.
- Miles, M.B., A. M. Huberman and J. Saldaña (2014), *Qualitative data analysis, A methods sourcebook*, Sage, Thousand Oaks, CA, USA.