Editorial: Special issue on quality management of Semantic Web assets (data, services and systems)

Amrapali Zaveri [a,*], Dimitris Kontokostas [b], Sebastian Hellmann [b] and Jürgen Umbrich [c]

[a] *Institute of Data Science, Maastricht University, 6229ER, Netherlands*
E-mail: amrapali.zaveri@maastrichtuniversity.nl
[b] *AKSW group, University of Leipzig, Leipzig, Germany*
E-mails: kontokostas@informatik.uni-leipzig.de, hellmann@informatik.uni-leipzig.de
[c] *Vienna University of Economics and Business, Vienna, Austria*
E-mail: juergen.umbrich@wu.ac.at

Abstract. This editorial summarizes the content of the Special Issue on Quality Management of Semantic Web Assets (Data, Services and Systems) part of the Semantic Web Journal.

Keywords: Data quality, linked data, data quality management, Semantic Web

The standardization and adoption of Semantic Web technologies has resulted in a variety of assets, including an unprecedented volume of data being semantically enriched and systems and services, which consume or publish this data. Although gathering, processing and publishing data is a step towards further adoption of Semantic Web, quality does not yet play a central role in these assets (e.g., data lifecycle, system/service development) [2].

Quality management of Semantic Web Assets (data, services and systems), in particular, presents new challenges that were not handled before in other research areas. Thus, adopting existing approaches for data quality management is not a straightforward solution. These challenges are related to the openness of the Semantic Web, the diversity of the information and the unbounded, dynamic set of autonomous data sources, publishers and consumers (legal and software agents). Additionally, detecting the quality of available data sources and making the information explicit is yet another challenge. Moreover, noise in one data set, or missing links between different data sets, propagates throughout the Web of Data, and imposes great challenges on the data value chain The potential heterogeneity and incompatibility of different implementations (potentially only partially adhering to standards) additionally poses several challenges for the quality assessments in and for such systems and services [1].

This Special Issue was addressed to those members of the community interested in providing novel methodologies or frameworks in managing, assessing, monitoring, maintaining and improving the quality of the Semantic Web data, services and systems and also introduce tools and user interfaces which can effectively assist in this management.

Overall, we received 9 submissions of which the following 5 papers were accepted:

In "A Comprehensive Quality Model for Linked Data", the authors present a quality model for Linked Data as an extension of the ISO 25012 data quality model altered and include aspects which are unique to linked data quality. In addition, the authors extended the W3C Data Quality Vocabulary (DQV) in order to serve the proposed quality model. They provide an im-

---

*Corresponding author.
E-mail: amrapali.zaveri@maastrichtuniversity.nl.

plementation and a use case in which the benefits of the quality model proposed in this paper are presented in the tool for Linked Data evaluation.

In "RODI: Benchmarking Relational-to-Ontology Mapping Generation Quality", the authors present a benchmarking framework called RODI for relational data-to-ontology mappers. The framework consists of several evaluation scenarios covering three application domains. RODI includes test scenarios from the domains of scientific conferences, geographical data, and oil and gas exploration. Systems that compute relational-to-ontology mappings can be evaluated using RODI by checking how well they can handle various features of relational schemas and ontologies, and how well the computed mappings work for query answering. Using RODI, the authors have conducted a comprehensive evaluation of seven systems.

In "Linked data schemata: fixing unsound foundations", the authors describe tools and method for the evaluation of the practical and logical implications of combining common linked data vocabularies into a single local logical model for the purpose of reasoning or performing quality evaluations. They found that strong interdependencies between vocabularies are common and that a significant number of logical and practical problems make this model unification inconsistent. In addition to identifying problems, this paper suggests a set of recommendations for linked data ontology design best practice. Finally they make suggestions for improving OWL's support for distributed authoring and ontology reuse.

In "Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO", the authors focus on the quality of knowledge graphs and measure the quality of five famous knowledge graphs based on their defined and collected quality criteria, which are outlined in great detail in the manuscript. In addition, the authors proposes a framework to recommend the most suitable knowledge graph for a given setting. Especially the comparisons of the different knowledge graphs provides a better understanding of the differences in knowledge graphs.

In "Literally Better: Analyzing and Improving the Quality of Literals", the authors focus on analyzing and improving the quality of literals since literals form a substantial (one in seven statements) and crucial part of the Semantic Web. They present a tool chain that builds on the LOD Laundromat data cleaning and republishing infrastructure and that allows them to analyze the quality of literals on a very large scale, using a collection of quality criteria they specify in a systematic way. The authors illustrate the viability of their approach by focusing on two particular aspects in which the current LOD Cloud can be immediately improved by automated means: value canonization and language tagging. Additionally, they give an overview of other problems that can be used to guide future endeavors in tooling, training, and best practice formulation.

Last but not the least, we would like to thank all the authors for their contributions to this special issue and all reviewers (Mathieu d'Aquin, John McCrae, Tomas Knap, Peter Patel Schneider, Jeremy Debattista, Gavin Mendel-Gleason, Nandana Mihindukulasooriya, Heiko Paulheim, Volha Bryl, Christoph Lange, Patrick Westphal, Anastasia Dimou, Zhigang Wang, Sebastian Mellorm, Riccardo Albertoni, Bojan Bozic Peter, Ioannis Chrysakis, Heiko Paulheim and Monika Solanki) for their valuable and careful review work that made the publication of this special issue a success.

References

[1] A. Zaveri, A. Maurino and L.-B. Equille, Web data quality: Current state and new challenges, *Int. J. Semant. Web Inf. Syst.* **10**(2) (2014), 1–6. doi:10.4018/ijswis.2014040101.

[2] A. Zaveri, A. Rula, A. Maurino, R. Pietrobon, J. Lehmann and S. Auer, Quality assessment for Linked Data: A survey, *Semantic Web Journal* **7** (2016), 63–93. Available at http://www.semantic-web-journal.net/content/quality-assessment-linked-data-survey. doi:10.3233/SW-150175.