

2nd Special Issue on Linked Dataset Descriptions

Editorial

Jens Lehmann^a, Oscar Corcho^b

^a *University of Leipzig, Institute of Computer Science, AKSW Group, Leipzig, Germany*
E-mail: lehmann@informatik.uni-leipzig.de

^b *Ontology Engineering Group, Universidad Politécnica de Madrid. Boadilla del Monte, Spain*
E-mail: ocorcho@fi.upm.es

Abstract. The Linked Dataset special issue gathers descriptions of mature and useful datasets thereby awarding researchers and practitioners who create and maintain those datasets with academic recognition. In this editorial, we briefly summarise the contributions to the 2nd special issue on Linked Dataset Descriptions at the Semantic Web Journal.

Keywords: RDF, Linked Data

In 2012, the Semantic Web journal released the first call for Linked Dataset description papers¹. As a result of this call, a special issue with 15 Linked Dataset description papers was published [1]. More importantly, the response was so well received by the Semantic Web community that dataset descriptions were considered since then as a new standing paper type for the journal².

Despite the amount of high-quality work usually required to publish high-impact Linked Datasets, authors were finding it difficult to get appropriate academic credit for such work, and those using the datasets were usually not able to cite them appropriately. Hence, Linked Dataset description papers provide a publication outlet for these contributions, making them more visible to a broader research community, archiving them as notable contributions to development of the Semantic Web, and allowing for peer-review based quality control with a transparent set of review criteria.

In this second special issue³ we have accepted 7 papers for publication. Together with the papers accepted for the first special issue plus other papers that have been accepted regularly, there are now approximately 30 Linked Dataset description papers accepted in the journal, which suggests that the decision on allowing this type of work to be published was well-received in the community.

These papers are evaluated according to a different set of criteria than the ones used for other types of papers. This includes an analysis of the quality and maturity of the dataset, of its current or potential usefulness, and of the clarity and completeness of the descriptions provided in the paper. Particular emphasis was put during the evaluation on making sure that sustainability strategies for the dataset are provided, in particular to handle future updates as well as hosting of the dataset.

The accepted papers cover several domain areas: statistical data, bibliographic data, urban data,

¹<http://www.semantic-web-journal.net/blog/semantic-web-journal-special-call-linked-dataset-descriptions>

²<http://www.semantic-web-journal.net/authors#types>

³<http://www.semantic-web-journal.net/blog/semantic-web-journal-call-2nd-special-issue-linked-dataset-descriptions>

energy and building data, biomedical institution data, protein data and Open Government Data. This shows that Linked Data principles can be applied to a wide range of domains. The sizes of datasets are mostly in the range of several million triples by the time of publishing the papers, although there are a few smaller datasets that have been selected for their novelty in their respective domains. Furthermore, licenses vary as well, with Creative Commons Attribution being the most used one. In all cases they are open data licenses.

- Linked SDMX Data: 600M triples; CC0 1.0
- AGRIS: 130M triples; mixed licenses
- Urbanopoly: 280K triples; Open Data Commons Open Database License (ODbL)
- RÉPENER: 150K triples; CC BY 3.0
- eagle-i: 1.3M triples; CC BY 3.0
- nextProt: 97M triples; CC BY-SA 3.0
- Public Spending: 63M triples: CC BY 3.0

We will now briefly summarise the articles which appeared in this special issue.

Linked SDMX Data by Capadisi et. al describes an ongoing effort to automate the transformation of statistical data from several international organisations (e.g. OECD, ECB, FAO, IMF) into Linked Data according to the W3C RDF Data Cube. This transformation work may be applied to any organisation that provides their data according to SDMX, and in fact the work has continued also with other datasets that were not originally described in this paper.

Migrating Bibliographic Datasets to the Semantic Web: the AGRIS case by Anibaldi et. al describes the dataset resulting from the transformation of a more than 5 million bibliographic references collected over several years by the Food and Agriculture Organisation of the United Nations (FAO). It was originally modeled using a qualified Dublin Core metadata format.

Geospatial Dataset Curation through a Location-based Game by Celino describes the datasets that result from a data curation campaign on available geospatial open datasets like OpenStreetMap, conducted through a location-based Game with a Purpose inspired by the monopoly board game.

The RÉPENER Linked Dataset by Sicilia describes a dataset that contains integrated informa-

tion of the Spanish territory regarding energy certification, building monitoring, and geographical data.

eagle-i: biomedical research resource datasets by Torniai describes a set of linked datasets that result from the transformation of biomedical resources available at 25 institutions. The datasets cover various areas of interest to these institutions, such as biological specimens, documents, human studies, instruments, organisms, reagents, research funding opportunities, etc.

Converting neXtProt into Linked Data and nanopublications by Chichester et. al describes how the RDF/XML annotations specific to neXtProt, a comprehensive knowledgebase on human proteins, have been dealt with according to the nanopublication model, so as to give appropriate attribution and provenance to all data.

Public spending as LOD: the case of Greece by Vafopoulos et. al describes a dataset that results from the generation, curation, interlinking and publishing of daily updated economic data on public expenditure, which is gathered by the Greek Transparency Initiative.

The past decade has provided evidence that the actual availability of structured data can fuel advances in research by providing a testbed for new algorithms and technology. In this spirit further Linked Dataset special issues in the Semantic Web Journal are planned and will continue to positively influence Semantic Web research.

Acknowledgements

This work was supported by a grant from the European Unions 7th Framework Programme provided for the project GeoKnow (GA no. 318159). Finally, we would like to thank all referees for helping us shape this special issue and providing very constructive comments to all papers.

References

- [1] P. Hitzler and K. Janowicz. Linked data, big data, and the 4th paradigm. *Semantic Web*, 4(3):233–235, 2013.