Linked Open Data for Public Sector Information

Diploma Thesis Review

Diploma Thesis Author: Jindřich Mynarz

Reviewed by: Stefan Urbanek <stefan.urbanek@gmail.com>

The domain of Open Data can be considered quite new within the European Union – the geographic area of thesis' focus. As the domain is still not common knowledge, the author commits majority of the work to description of Public Sector Information and Open Data itself where author shows very extensive understanding of the researched domain.

The work contains excellent summary of modern Open Data knowledge and related topics. It would be very useful, if parts of the work were published for educational purposes. The slight data literacy schooling might be required as well, however it is out-of-scope of the thesis topic.

The thesis can be split into two parts that might have been presented independently: general research of current open-data state and description of RDF based Linked Data. Based on the thesis title "Linked Open Data for Public Sector Information" one would expect more focus on the "Linked Data" together with concrete explained examples and use-cases. Author should have dropped the linked data part and focus more on the open-data with examples or should have focused less on the open-data and dive into the linked data in more details.

On the other hand, since the topic is very new, it would be really difficult to achieve the original goal of the thesis' topic without detailed description of the context. Moreover, adding more to the core topic would make the thesis go over required limit. Therefore the little disconnection should not be considered as defect.

The author has done an outstanding work.

Comments and Questions

1. **General:** Author focuses on the Resource Description Framework (RDF), which can be considered as highest and most universal abstraction of data representation. Thesis is missing description or at least mention of existence of alternative ways of achieving "linked data".
2. **Question:** Give an example of achieving linked data not based on RDF. Does not have to be globally linkable.
3. **p.30: Machine Readability:** it should be noted, that even structured formats can be abused (for example: XML with three fields: id, date and "content" with original document data.)
4. **p.53: 4.3.2.1. Completeness:** Data quality dimension *completeness* refers rather to existence of expected data attributes. It should be highlighted that using some means of materialization requires existence of a broader context description that help to increase

the completeness. Just by having a data set described in RDF can hardly fill in the originally missing values. That is true of any technology and not related to RDF exclusively, though.

5. **p. 54: 4.3.2.2. Usability:** "Usability may be perceived as the weakest point of linked data." - Author probably meant "Linked Data" (capitalized) based on RDF.
6. **Question:** Give a use-case example, where other data description that RDF would be more appropriate (for example tabular). Explain why.

Summary

Decision: Accepted

Grade: A

22. May 2012, Bratislava

Ing. Stefan Urbanek