RDF Data Cube Vocabulary: Implementation report

Summary: This document gives an implementation report for the W3C RDF Data Cube Vocabulary.

Editors: Sebastian Bayerl, Michael Granitzer, Kai Schlegel, Florian Stegmaier

Affiliation: University of Passau

Innstrasse 43

94032 Passau

Germany

CODE Consortium:

1 Abstract

A crucial task in a researchers' daily work is the analysis of primary research data to estimate the evolution of certain fields or technologies, e.g. tables in publications or tabular benchmark results. Due to a lack of comparability and reliability of published primary research data, this becomes more and more time-consuming leading to contradicting facts, as has been shown for ad-hoc retrieval [1]. The CODE project[1] [2] aims at contributing to a Linked Science Data Cloud by integrating unstructured research information with semantically represented research data. Through crowdsourcing techniques, data centric tasks like data extraction; integration and analysis in combination with sustainable data marketplace concepts will establish a sustainable, high-impact ecosystem.

The RDF Data Cube Vocabulary is utilized within the whole project to reach a homogeneous data integration of primary research data as well as to generate an OLAP-aware storage. Besides, this standardized data model also fosters the interaction with consuming peers, such as the envisioned visual analytics component.

2 Details on the integrity constraints

The details of the evaluation can be found in Table 1. If a integrity constraint has failed, an explanation has been added.

Table 1: Details on the integrity constraints test

Test case	Result	Description
IC-1. Unique DataSet	Fail	The observation resource uses a control variable. Identity is guaranteed by the named graphs.
IC-2. Unique DSD	Pass	
IC-3. DSD includes measure	Fail	:Water_level a rdf:Property , qb:MeasureProperty ;
IC-4. Dimensions have range	Fail	Range not yet defined in the prototype.
IC-5. Concept dimensions have code lists	Pass	Code lists are not used.
IC-6. Only attributes may be optional	Pass	qb:AttributeProperty are not yet used.
IC-7. Slice Keys must be declared	Pass	Slices are not used.
IC-8. Slice Keys consistent with DSD	Pass	See IC-7
IC-9. Unique slice structure	Pass	See IC-7
IC-10. Slice dimensions complete	Pass	See IC-7
IC-11. All dimensions required	Pass	

[1] http://code-research.eu/

IC-12. No duplicate observation	Pass	
IC-13. Required attributes	Pass	
IC-14. All measures present	Pass	
IC-15. Measure dimension consistent	Pass	
IC-16. Single measure on measure dimension observation	Pass	
IC-17. All measures present in measures dimension cube	Pass	
IC-18. Consistent data set links	Pass	See IC-7
IC-19. Codes from code list	Pass	See IC-5
IC-20. Codes from hierarchy	Pass	See IC-5
IC-21. Codes from hierarchy (inverse)	Pass	See IC-5

3 SPARQL Endpoint

The data cubes generated by prototypes of the CODE project are hosted at the following link:

http://zaire.dimis.fim.uni-passau.de:8080/bigdata/sparql

Please be aware that this is on-going research. Changes and updates to the endpoint as well as data cubes may be applied.

References

[1] T. G. Armstrong, A. Moffat, W. Webber, and J. Zobel, “Improvements that don’t add up: ad-hoc retrieval results since 1998.,” in Proceedings of the Conference on Information and Knowledge Management, pp. 601–610, 2009.

[2] F. Stegmaier, C. Seifert, R. Kern, P. Höfler, S. Bayerl, M. Granitzer, H. Kosch, S. Lindstaedt, B. Mutlu, V. Sabol, K. Schlegel, and S. Zwicklbauer, “Unleashing semantics of research data,” in Proceedings of the 2nd Workshop on Big Data Bench- marking, 2012.