Formal Representation and Harmonization of Open Government Data with Semantic Web Technologies

Markus UNGERSBOECK

University of Applied Sciences Wiener Neustadt, Email: markus.ungersboeck@hotmail.com

Abstract

The Initiative Open Government Data targets the publication of non-personal data of the public administration for private and commercial use. Open Government Data is published autonomous and unstructured by multiple agencies of the Austrian administration.

This results in data sets of different spatial, temporal and thematic granularity, which strongly limits the usability of the data. To achieve increased benefit and usability, the harmonization of data which is provided in accordance with the initiative, is desirable.

The purpose of this paper is to introduce a data model using semantic web technologies, like the Resource Description Framework (RDF), to formally represent Open Government Data. The conceptualization of multi-dimensional data is accomplished by the use of the Data Cube Vocabulary. The temporal dimension is modeled using OWL-Time. Domain specific concepts are defined with the Simple Knowledge Organisation System (SKOS).

Aggregation concepts are built upon this data model to harmonize the geographic reference of Open Government Data. An administrative hierarchy of geographical objects for Austria, provided by GeoNames, is utilized for these aggregation concepts. Open Government Data, which can be aggregated to a spatial object of lower granularity, is determined by hierarchical completeness. This is accomplished by extending the formal model and making use of the Web Ontology Language (OWL). The extension manages the Open World Assumption as well as the Non Unique Naming Assumption, which are both ubiquitous in the Semantic Web. The actual aggregation is realized using SPARQL, an query language for RDF.

This paper only recognizes statistical data for the representation and aggregation of Open Government Data.