

Chapter 1

Semantic Web Standards for Publishing and Integrating Open Data

Axel Polleres

Vienna University of Economics and Business (WU Wien), Austria

Simon Steyskal

Vienna University of Economics and Business (WU Wien), Austria & Vienna University of Technology (TU Wien), Austria

ABSTRACT

The World Wide Web Consortium (W3C) as the main standardization body for Web standards has set a particular focus on publishing and integrating Open Data. In this chapter, the authors explain various standards from the W3C's Semantic Web activity and the–potential–role they play in the context of Open Data: RDF, as a standard data format for publishing and consuming structured information on the Web; the Linked Data principles for interlinking RDF data published across the Web and leveraging a Web of Data; RDFS and OWL to describe vocabularies used in RDF and for describing mappings between such vocabularies. The authors conclude with a review of current deployments of these standards on the Web, particularly within public Open Data initiatives, and discuss potential risks and challenges.

INTRODUCTION

The Semantic Web was founded as an activity of the World Wide Web Consortium (W3C) and comprises a list of standards for leveraging the publication and consumption of structured data on the Web, with the ultimate goal of enabling machines to “understand” Web content, previously only understandable by humans. This list of standards is building up technology “stack,”

that consists of a lightweight interchangeable data model interchange format (RDF), alongside with publishing principles, that enable reusability and combinability of RDF datasets by exploiting basic Web principles of interlinking and dereferenceability and, last but not least, schema languages to describe such published and interlinked data (RDF Schema and OWL), as well as a standard query language for accessing and collecting RDF data (SPARQL). In the following, we will briefly

DOI: 10.4018/978-1-4666-8111-8.ch001

introduce those standards based on a running example revolving around data indicators for cities, which may well be part of data produced by emerging Open Data efforts within communities.

BACKGROUND: SEMANTIC WEB STANDARDS AND OPEN DATA

RDF

The *Resource Description Framework* (RDF) is the basic data model for the Semantic Web. It is built upon one of the simplest structures for representing data: a directed labelled graph. An RDF graph is described by a set of triples of the form *<Subject Predicate Object>*, also called *statements*, which may be viewed as connecting subjects to objects via edges labelled by the predicates. Since any of these can be - dereferenceable - URIs, these edges may be equally viewed as “typed” links between resources and other resources, where resources - unlike in the HTML Web, are no longer restricted to be documents, but arbitrary entities, that can be identified by URIs.

RDF’s flat graph-like representation has the advantage of abstracting away from the data schema, and thus promises to allow for easier integration than customized XML data in different XML dialects: whereas the integration of different XML languages requires the transformation between different tree structures using transformation languages such as XSLT (Kay, 2007) or XQuery (Chamberlin et al., 2007), different RDF graphs can simply be stored and queried alongside, and as soon as they share common URIs, form a joint graph upon a simple merge operation accumulating their respective triples in one joint graph. While the normative syntax to exchange RDF, RDF/XML (Beckett and McBride, 2004), is an XML dialect itself, there are various other serialization formats for RDF, such as RDFa (Adida et al., 2008), a format that allows one to embed RDF within (X) HTML, or non-XML representations such as the

more readable Turtle (Beckett and Berners-Lee, 2008) syntax; likewise RDF stores, that is, special databases for RDF, normally use their own, proprietary internal representations of triples, that do not relate to XML. Various RDF stores that can store and handle RDF data efficiently and at large scale are nowadays available off-the-shelf, both commercial systems and academic ones, such as YARS2 (Harth et al., 2007), Jena TDB (<http://jena.apache.org/documentation/tdb/>), OpenLink Virtuoso (<http://virtuoso.openlinksw.com/>), 4Store (<http://4store.org>), AllegroGraph (<http://www.franz.com/agraph/allegrograph/>), or Sesame (<http://www.openrdf.org/>) to name a few. For an overview of RDF Stores, see also (Haslhofer et al., 2011); a recent article also discusses the use of NoSQL graph databases to store and process RDF (Cudré-Mauroux et al., 2013).

RDF Example

A sample RDF graph consisting of Triples in Turtle syntax that contains information about the city of Vienna stored at DBpedia (Bizer et al., 2009b) - an RDF extract of structured Open Data published in Wikipedia – is shown in Box 1.

This RDF graph consists of four triples, separated by “.”, where the latter three using the Turtle syntax shortcut “;” for grouping together predicate-object pairs of the same subject. Resource URIs are abbreviated using *namespace prefixes*, such as *rdf:;*, or *dpo;*; that is, e.g. the abbreviated URI:Vienna corresponds to <http://dbpedia.org/resource/Vienna>, or, respectively, the predicate URI *dpo:country* corresponds to <http://dbpedia.org/ontology/country>, etc. Particularly, the *rdf:type* predicate plays a special distinct role, in that it allows to model is-a relationships, i.e., class membership; as we will see later, RDF&OWL will enable to define hierarchies of such classes. As the example shows, RDF triples cannot link resources (represented by URIs) to other resources, but also model ordinary predicate-value pairs, by allowing datatyped values in the object position.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/semantic-web-standards-for-publishing-and-integrating-open-data/125283

Related Content

Are Asian Countries Ready to Lead a Global ICT Standardization?

DongBack Seo and Jan Willem Koek (2012). *International Journal of IT Standards and Standardization Research* (pp. 29-44).

www.irma-international.org/article/asian-countries-ready-lead-global/69809

Knowledge Management Policy

Avadh Narayan Yadav (2011). *Handbook of Research on Information Communication Technology Policy: Trends, Issues and Advancements* (pp. 484-496).

www.irma-international.org/chapter/knowledge-management-policy/45402

A Social Bookmarking-Based People Search Service: Building Communities of Practice with Collective Intelligence

Jeff J.S. Huang, Stephen J.H. Yang, Jeng C.C. Chen, Irene Y.S. Li and Indy Y.T. Hsiao (2012). *Information Technology for Intellectual Property Protection: Interdisciplinary Advancements* (pp. 289-305).

www.irma-international.org/chapter/social-bookmarking-based-people-search/60560

Where Are You? Consumers' Associations in Standardization: A Case Study on Switzerland

Christophe Hauert (2010). *International Journal of IT Standards and Standardization Research* (pp. 11-27).

www.irma-international.org/article/you-consumers-associations-standardization/39084

A Data Model for Describing and Exchanging Personal Achieved Learning Outcomes (PALO)

Jad Najjar, Michael Derntl, Tomaž Klobucar, Bernd Simon, Michael Totschnig, Simon Grant and Jan Pawlowski (2010). *International Journal of IT Standards and Standardization Research* (pp. 87-104).

www.irma-international.org/article/data-model-describing-exchanging-personal/46115