

Extended abstract

Making sense of a flat list of terms into Linked Open Data SKOS vocabularies

Helder Noel Firmino and Ana Alice Baptista

CentroAlgoritmi, University of Minho, Guimarães, Portugal

E-mails: pg19785@alunos.uminho.pt, analice@dsi.uminho.pt

1. Introduction

The emergence of the Web in 1991, within the CERN (European Organisation for Nuclear Research) in Genève, Switzerland, was a milestone in the history of humanity, changing the way communication, democratizing access to information and knowledge. With the Web, humanity passes into a stage of collective intelligence [3], where everyone, except the info-excluded, has the possibility to share information. The Web was invented by the visionary British scientist, Tim Berners-Lee. He suggested the so-called global hypertext system.

Tim Berners-Lee and the W3C (World Wide Web Consortium) are now involved in another breakthrough project: the Web of data. “The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network”. Tim Berners-Lee, suggested a 5 star deployment scheme for Open Data – see Table 1. According to Tim Berners-Lee, the Semantic Web is not just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. Linked data technologies enable the description and explicit relation of resources through URIs (Uniform Resource Identifiers), meaning that implicit relations with other resources can be found or inferred. When the linked data is open, the expression Open Lined Data is used. “Linked Open Data (LOD) is Linked data which is released under an open licence, which does not impede its reuse for free” [1], which means that in a LOD scenario, the five star level data is open.

Table 1
The five star scheme [1]

*	Available on the web (whatever format) <i>but with an open licence, to be Open Data</i>
**	Available as machine-readable structured data (e.g. excel instead of image scan of a table)
***	Use non-proprietary formats (e.g. CSV instead of excel)
****	All the above plus use open standards from W3C (RDF and SPARQL) to identify things (URIs), so that people can point at your stuff
*****	All the above plus link your data to other people’s data to provide context

“The term ‘Semantic Web’ refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL and SKOS” [7].

SKOS, which became a W3C recommendation in 2009, provides a model for expressing the basic structure and content of concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies, folksonomies, and other similar types of controlled vocabulary. As an application of the Resource Description Framework (RDF), SKOS allows concepts to be composed and published on the World Wide Web, linked with data on the Web and integrated into other concept schemes [2].

2. Problem

ANACOM (Autoridade Nacional de Comunicações – Portuguese National Authority on Communications) is the regulator, supervisor and representative of the communications (postal and electronic communications) sector in Portugal. In the scope of the first phase of the SWAN (Semantic Web at ANACOM) project we analysed the requirements for a Semantic Web presence of ANACOM. In fact, ANACOM already has all its Web Site resources described using Dublin Core (DC), encoded in HTML, which corresponds to the level one of interoperability for DC metadata (see [4]) and in level 3 of the Tim Berners-Lee’ five star scheme.

Through the years, ANACOM has been developing a kind of a flat controlled vocabulary: a list of controlled terms, without any structure and without specific a priori relations to the DC properties in use. One of the requirements of the project is to organize that list into one or more controlled vocabularies, relate them with other LOD vocabularies, encode them in SKOS and make them open in at least two languages (Portuguese and English).

3. Objectives

- Analyse and organize the terms being used to describe resources at ANACOM’s website;
- Relate the resulting controlled vocabularies with other existing vocabularies;
- Encode the resulting controlled vocabularies in SKOS in, at least, two languages, and under a five star LOD perspective.

4. Methodology

We will adopt a hybrid approach, combining the best features of each methodology. Specifically with regard to the construction of vocabularies, we will combine Ontology Development 101 [5] and the methodology proposed by the European Programme ISA [6], one of the most recent outputs in this area.

We will use also qualitative research techniques, such as interviews and focus groups to analyse and validate the relations between the terms.

5. Expected results

This article will present a work in progress comprised of the whole process starting on the organization of the terms and ending in the first version of the controlled vocabularies encoded in SKOS.

Acknowledgements

This article is funded by FEDER funds by the Programa Operacional Fatores de Competitividade – COMPETE and by National Funds through FCT – Foundation for Science and Technology under the Project: FCOMP-01-0124-FEDER-022674.

References

- [1] T. Berners-Lee, Linked data – Design issues, 2009, available at: <http://www.w3.org/DesignIssues/LinkedData.html>; cited 2013, January 21.
- [2] A. Isaac and E. Summers, SKOS simple knowledge organization system primer, 2009, available at: <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>; cited 2012, November 5.
- [3] P. Lévy, *Cibercultura*, Editora 34, São Paulo, 1999.
- [4] M. Nilsson, T. Baker and P. Johnston, Interoperability levels for Dublin core metadata, 2009, available at: <http://dublincore.org/documents/interoperability-levels/>; cited 2010, June 9.
- [5] N. Noy and D. McGuinness, Ontology development 101: A guide to creating your first ontology, Stanford University, 2001, available at: <http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>; cited 2012, September 20.
- [6] PwC EU Services EESV, D3.1 – Process and methodology for core vocabularies, European Commission, Brussels, 2012, p. 59, available at: http://joinup.ec.europa.eu/sites/default/files/D3.1-Process%20and%20Methodology%20for%20Core%20Vocabularies_v1.01.pdf.
- [7] W3C, Semantic Web, 2013, available at: <http://www.w3.org/standards/semanticweb/>; cited 2013, January 21.