



A SKOS vocabulary for Linked Land Administration: Cadastre and Land Administration Thesaurus



Volkan Çağdaş^{a,*}, Erik Stubkjær^b

^a Yildiz Technical University, Department of Surveying Engineering, 34220 Istanbul, Turkey

^b Aalborg University, Department of Development and Planning, 9220 Aalborg, Denmark

ARTICLE INFO

Keywords:

Cadastre
Land administration
Linked Data
Cadastre and Land Administration
Thesaurus
Simple Knowledge Organization System
Linked Land Administration

ABSTRACT

Intensification of international communication calls for multilingual terminology databases and other linguistic tools. The World Wide Web Consortium framed the further specification of the Simple Knowledge Organization System (SKOS), which is a formal language designed for standardized representation of structured vocabularies, as well as the Linked Data methodology of publishing structured data in a machine-readable and interlinked way, that they become more useful. Applying the Linked Data approach, the European Commission recently established e-Government Core Vocabularies. This paper describes the development of a Knowledge Organization System (KOS) in terms of a thesaurus for the domain of cadastre and land administration. The main purpose is to contribute towards the development of 'Linked Land Administration' that adopts Linked Data technologies for semantic management of datasets kept in public registries, and scholarly and legislative resources kept in libraries. The proposed controlled vocabulary in SKOS format may be used for specifying metadata records of scholarly and legislative resources, as well as enrichment of these resources through semantic annotations. It also provides a base for further ontology development initiatives. The thesaurus is mainly derived from terms of the standard ISO 19152:2012 Land Administration Domain Model, which represents the static aspect of the domain. Reports of the project Property Formation in the Nordic countries (Kort og Matrikelstyrelsen, 2006) provide the basis for the temporal or activity aspect of the domain, while the ANSI/NISO, 2005 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies provided methodological advice. The thesaurus consists of 143 terms, the relations of which are recorded according to the mentioned SKOS standard. About one fourth of the terms are adopted from existing thesauri, including the AGROVOC multilingual agricultural vocabulary, the GEMET Thesaurus with INSPIRE Spatial Data Themes, and the STW Thesaurus for Economics.

© 2015 Elsevier Ltd. All rights reserved.

Introduction

A Knowledge Organization System (KOS) (see Hodge, 2000) in terms of a thesaurus is developed for the domain of cadastre and land administration. It is mainly derived from terms of the standard ISO 19152:2012 Land Administration Domain Model (LADM), which represents the static aspect of the domain; and reports of the project Property Formation in the Nordic countries (Kort og Matrikelstyrelsen, 2006), which represents the temporal or process aspect of the domain. The designed terminological

artefact is represented through the Simple Knowledge Organization System² (SKOS) specifications developed by the World Wide Web Consortium³ (W3C), and made available at the web site <http://cadastralvocabulary.org>. The purpose is to contribute towards the idea of 'Linked Land Administration' through a domain thesaurus which could assist 'semantic publication' (see Shotton et al., 2009) of cadastre and land administration related scholarly and legislative resources in a Linked Data environment.

Linked Data refers to data published on the Web in such a way that it is machine-readable, its meaning is explicitly defined, it is linked to other external data sets, and can in turn be linked to from external data sets (Bizer et al., 2009, p. 2). More

* Corresponding author at: Yildiz Teknik Universitesi, Harita Muhendisligi Bolümü, Davutpasa Kampusu, Esenler, 34220 Istanbul, Turkey. Tel.: +90 212 383 5313.

E-mail addresses: volkan@yildiz.edu.tr (V. Çağdaş), est@land.aau.dk (E. Stubkjær).

² <http://www.w3.org/2004/02/skos/>.

³ <http://www.w3.org/>.

specifically, a resource is identified and accessed by Uniform Resource Identifiers⁴ (URI) over the Web, described by Resource Description Framework⁵ (RDF) based on standardized RDF vocabularies, and connected with semantically related other resources by RDF links (Berners-Lee, 2006). A recent research by Çağdaş and Stubkjær (2015) applied the Linked Data approach into the domain of cadastre and land administration by introducing the notion of Linked Land Administration to the research community. The notion proposes use of Linked Data technologies for semantic management of datasets about the ownership, value and use of immovable property units; but also covers semantic management of domain resources, such as scholarly and legislative resources, for enhancing science and education. The former aspect has already been addressed by Çağdaş and Stubkjær (2015), as the eGovernment Core Vocabularies (EC ISA, 2012) were extended through a Core Immoveable Property Vocabulary,⁶ presented according to the RDF Schema. This enables the representation of datasets kept in land administration registries (i.e. cadastre, land registry, building and dwelling registry, address registry, and property tax registry) according to RDF, and thus integrates distributed land administration datasets through the Linked Data approach. The present research focuses on the latter aspect, semantic management of scholarly and legislative resources in the Linked Data environment. The scholarly resources cover scientific and educational resources organized in libraries, such as books, monographs, encyclopaedias, journal articles, conference papers, letters, manuscripts, theses and dissertations, as well as international standards issued by International Standardization Organization (ISO) and Open Geospatial Consortium (OGC). The legislative resource includes legislation, and all writings produced by the legislator explaining and justifying legislation (CEN, 2010, p. 5), such as statutory laws, statutes, executive orders, court rulings, code of conducts, and white papers.

The library and legal informatics research communities have provided a number of metadata schemas and ontologies that might be employed for RDF representation of metadata about scholarly and legislative resources in land administration domain (see *A brief introduction to semantic publishing and legal informatics artefacts* section). However, as pointed out by Keßler et al. (2013), the adaptation of a sole metadata schema or ontology is not sufficient to efficiently characterize resources of various domains. A formal vocabulary for the domain of cadastre and land administration supports more unambiguous identification of content of scholarly and legislative resources, as well as semantic annotation of document parts and terms used in documents. Vocabularies, such as 'Multilingual thesaurus on land tenure' (FAO, 2003) and 'Land Tenure Lexicon' (IIED, 2000), may partly respond to such requirements. These terminological resources, however, do not sufficiently address and relate core terms of the cadastral domain both from the legal and technical aspects, or are not available in Semantic Web formats. Moreover, staff of the EU Joint Research Centre have identified the importance of SKOS thesauri in the context of Spatial Data Infrastructures (SDIs) with a view of improving query functionalities in the INSPIRE Geoportal (Fugazza and Vaccari, 2011; Fugazza and Luraschi, 2012). This article in *The Cadastre and Land Administration Thesaurus* section, therefore, introduces a new domain vocabulary in terms of a thesaurus in SKOS format that may be used for specifying metadata of scholarly and legislative resources,

as well as enrichment of these resources through semantic annotations. The mutually related core domain terms of the proposed thesaurus are supposed to enhance information discovery and retrieval, and hereby organize domain knowledge.

The developed thesaurus, called the Cadastre and Land Administration Thesaurus (CaLaThe), is mainly derived from terms and their relationships of the LADM, as well as from reports of the project Property Formation in the Nordic countries (Kort og Matrikelstyrelsen, 2006) by employing the ANSI/NISO (2005) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. However, it includes also terms from other thesauri, including the AGROVOC,⁷ the GEMET⁸ with INSPIRE⁹ Spatial Data Themes, the STW Thesaurus for Economics,¹⁰ Cycorp's OpenCyc ontology,¹¹ and United Kingdom's Integrated Public Sector Vocabulary.¹² The CaLaThe could be employed by libraries or other organizations such as International Office of Cadastre and Land Records¹³ (OICRF), the International Federation of Surveyors¹⁴ (FIG), European Land Information Service¹⁵ (EULIS) that aims publishing land administration related scholarly and legislative resources as Linked Data. It also could be used for semantic enrichment of land administration legislation, e.g. in terms of a recently launched European Union (EU) Framework Programme 7 project, European and National CASE Law and Legislation Linked in Open Data Stack¹⁶ (EUCases).

The remaining part of the article is organized as follows: In the following section, we present a brief introduction to Semantic Web artefacts developed in the library and legal informatics fields. *The Cadastre and Land Administration Thesaurus* section summarizes the methodology applied for the development of the domain thesaurus, and provides an overview of the content of the thesaurus and its concept collections. *Evaluation of the developed land administration thesaurus* section presents results of an evaluation process applied for assessing the usefulness of the designed thesaurus, and the last section concludes the article.

⁷ The AGROVOC is a controlled vocabulary covering all areas of interest of the Food and Agriculture Organization (FAO), including food, nutrition, agriculture, fisheries, forestry, environment etc. It is published by FAO and edited by a community of experts, see: <http://fao.org/agrovoc/>.

⁸ The GEMET (GEneral Multilingual Environmental Thesaurus), has been developed as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data Sources, and the European Environment Agency, see: <http://eionet.europa.eu/gemet/>.

⁹ The INSPIRE is an EU initiative which aims at establishing an infrastructure for spatial information in Europe to support environmental policies and activities, see: <http://inspire.ec.europa.eu/>.

¹⁰ The Standard Thesaurus Wirtschaft, dubbed: STW Thesaurus for Economics, is a controlled, structured vocabulary for subject indexing and retrieval of economics literature, see: <http://zbw.eu/stw/>.

¹¹ <http://www.cyc.com/platform/opencyc>.

¹² The Integrated Public Sector Vocabulary is an encoding scheme for populating the e-Government Metadata Subject element of metadata, see: <http://www.esd.org.uk/standards/ipsv/2.00/>.

¹³ The International Office of Cadastre and Land Records (OICRF) is a study and documentation centre for cadastre, land administration and affiliated fields of interest, see: <http://www.oicrf.org/>.

¹⁴ The International Federation of Surveyors (FIG) is an international, non-government organisation whose purpose is to support international collaboration for the progress of surveying in all fields and applications, see: <http://www.fig.net/>.

¹⁵ The European Land Information Service (EULIS) is a European Economic Interest Group governed by members, each responsible for land and property information in its own country or regio. It owns a service that provides easy access to land and property information, see: <http://eulis.org/>.

¹⁶ EUCases aims at establishing a pan-European law and case law Linking Platform by transforming multilingual legal open data into linked open data after semantic and structural processes (e.g. structuring legislation with markup languages, classifying the legal texts with EuroVoc, and enrichment legal texts with ontologies, see <http://apis.bg/en/eucases/eucasesproject2.html>).

⁴ A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource in the Web, see: <http://tools.ietf.org/html/rfc3986>.

⁵ The Resource Description Framework (RDF) is a standard model for data interchange on the Web, see: <http://www.w3.org/RDF/>.

⁶ <https://joinup.ec.europa.eu/asset/cipv/description/>.

Download English Version:

<https://daneshyari.com/en/article/6547859>

Download Persian Version:

<https://daneshyari.com/article/6547859>

[Daneshyari.com](https://daneshyari.com)