Empowering the access to public procurement opportunities by means of linking controlled vocabularies. A case study of Product Scheme Classifications in the European e-Procurement sector

Jose María Alvarez-Rodríguez [a,*], José Emilio Labra-Gayo [b], Alejandro Rodríguez-González [c], Patricia Ordoñez De Pablos [d]

[a] The South East European Research Center, Thessaloniki, Greece
[b] WESO Research Group, Department of Computer Science, University of Oviedo, 33007 Oviedo, Spain
[c] Bioinformatics at Centre for Plant Biotechnology and Genomics UPM-INIA, Polytechnic University of Madrid, Madrid, Spain
[d] WESO Research Group, Department of Business Administration, University of Oviedo, 33007 Oviedo, Spain

A B S T R A C T

The present paper introduces a method to promote existing controlled vocabularies to the Linked Data initiative. A common data model and an enclosed conversion method for knowledge organization systems based on semantic web technologies and vocabularies such as SKOS are presented. This method is applied to well-known taxonomies and controlled vocabularies in the business sector, more specifically to Product Scheme Classifications created by governmental institutions such as the European Union or the United Nations. Since these product schemes are available in a common and shared data model, the needs of the European e-Procurement sector are outlined to finally demonstrate how Linked Data can address some of the challenges for publishing and retrieving information resources. As a consequence, two experiments are also provided in order to validate the gain, in terms of expressivity, and the exploitation of this emerging approach to help both expert and end-users to make decisions on the selection of descriptors for public procurement notices.

Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Government bodies and public institutions as a whole are the most important buyers in the European Union (EU), since public procurement spending represents around 19% of EU Gross Domestic Product (GDP) (European Commission, 2010b). Given this situation there is a growing interest and commitment (European Commission, 2010a) to ensure that these funds are well managed and most of inefficiencies are eliminated. Electronic public procurement or e-Procurement (Podlogar, 2007) emerges as an alternative to link and integrate inter-organizational business processes and systems with the automation of the requisitioning, the approval purchase order management, accounting processes among others through the Internet-based protocol.

In this context the European Commission (EC) is trying to unlock this potential, the 2004 Public Procurement Directives 2004/17/EC and 2004/18/EC introduced several provisions and

projects (European Commission, 2013c; European Commission, 2013b) aimed at enabling e-Procurement uptake in all Member States. In this light of modernizing the European public procurement sector to support growth and employment the EC also identified (European Commission Internal Market Directorate-General, 2010) both regulatory and natural barriers in the access to public procurement in the EU context, especially for SMEs (European Commission, 2008). According to this evaluation, a real European Single Market (European Commission, 2011) has not yet been achieved causing losses, being cost-inefficient, missing opportunities for society and leading to a situation where more than 27 national markets co-exist instead of an EU-wide public market (Monti, 2010). In this sense, one relevant action to ease the inter-connectivity and interoperability in this landscape was the creation of the Tenders Electronic Daily (European Commission, 2013a; European Commission, 2013d) (TED) by the EC. It is the on-line version of the "Supplement to the Official Journal of the European Union", dedicated to European public procurement notices (1500 new announcements every day). Taking into account that the type of contract and the geographical information are two of the main variables that serve to be aware of new business opportunities, the EU has established the use of the "Common Procurement Vocabulary" (hereafter CVP refers to CPV 2008) and

the "Nomenclature of territorial units for statistics" (NUTS) as mandatory to annotate public procurement notices. Nevertheless depending on the country and the scope of use distinct classifications have been also created. In the specific case of e-Procurement, the "United Nations Standard Products and Services Code" (UNSPSC) in Australia, the "North American Industry Classification System" (NAICS) in United States or the "Integrated Tariff of the European Communities" (TARIC) in the EU are examples of similar efforts to unify and model procurement-related data. As a consequence a real, standardized and integrated environment for e-Procurement (and business) data to encourage the creation of knowledge-based services cannot be easily deployed. That is why some of the potential advantages outlined by the EC (European Commission, 2010b) such as increased accessibility and transparency, benefits for individual procedures, benefits in terms of more efficient procurement administration and potential for integration of EU procurement markets and systems cannot be reached in a short-term.

Additionally public bodies are also defining knowledge organization systems (KOS), such as the classifications in the Eurostat's metadata server (RAMON), to enable users to annotate information objects providing an agile mechanism for performing tasks such as exploration, searching, automatic classification or reasoning. The structure and features of these systems are usually very heterogeneous, although some common aspects can be found in all of them: hierarchical relationships between terms or concepts and multilingual character of the information. However the lack of mappings among them makes almost impossible a proper reuse of these systems out of their scope. As motivating and explanatory examples of the necessity of linking knowledge organization systems in the e-Procurement context, four common situations are presented below:

- An American SME that usually manages the NAICS vocabulary for being aware of public procurement notices in United States is looking for new business opportunities in Europe. Although they perfectly know that the CPV is used in European public procurement notices they are not able to retrieve any announcement due to the fact there is not mapping between the CPV and NAICS classification systems.
- A Spanish family-owned company is also looking for new business opportunities in Bulgaria. They have been informed that the national government is requiring providers for low-value procurement opportunities that match their services but they are finding three main issues: (1) Bulgarian civil servants annotate these procurement opportunities with different codes as Spanish ones do; (2) these notices are not published in TED (because of their value) and (3) the language is a simple barrier they cannot overcome.
- A department of statistics in some international organism, such as the WorldBank, is gathering data and information from different public institutions and scopes to compare some economical indicators. Due to the fact that data and information are managed using distinct knowledge organization systems, a simple task of integrating this information to create a report is becoming a major challenge.
- A company has implemented a commercial alert service that enables the possibility of getting latest public procurement opportunities according to a customer profile. Although business users perfectly know how contracts are described using some classification scheme, sometimes they are not able to translate a client query to contract descriptors. If a service could suggest these descriptors exploiting some kind of knowledge-base, business users decisions would be supported by a tool, the alert service would be more accurate and customer satisfaction would be increased.

On the other hand and following the principles of the Open Data initiative, the vice-president Neelie Kroes is leading the Open Data Strategy (European Parliament & European Council, 2003) for Europe. The strategy is strongly focused on the commercial value of the reuse of Public Sector Information (PSI). In this case e-Procurement data is considered to be a key-enabler of the new data-based economy enabling a greater transparency; delivers more efficient public services; and encourages greater public and commercial use. Therefore an enriched version of public procurement data to ease the access to business opportunities can be the first action to encourage SME participation and create a real and competitive public market of procurement opportunities.

Moreover, the emerging Web of Data and the sheer mass of information now available make it possible the deployment of new added-value services and applications based on the reuse of existing vocabularies and datasets. The popular diagram of the Linked Data Cloud (Cyganiak & Jentzsch, 2011), generated from metadata extracted from the Comprehensive Knowledge Archive Network (CKAN) out, contains 336 datasets, with more than 25 billion RDF triples and 395 million links. With regards to e-Procurement, a new group of 130 CKAN datasets have been released from the "OpenSpending.org" site. In this realm, data coming from different sources and domains have been promoted following the principles of the Linked Data initiative (Berners-Lee, 2006) to improve the access to large documental databases, e.g. e-Government resources, scientific publications or e-Health records. In the case of KOS, such as thesauri, taxonomies or classification systems are developed by specific communities and institutions in order to organize huge collections of information objects. These vocabularies allow users to annotate (Leukel, 2004; Leukel, 2004; Leukel & Maniatopoulos, 2005) the information objects and easily retrieve them, promoting lightweight reasoning in the Semantic Web. Topic or subject indexing is an easy way to introduce machine-readable metadata for a resource's content description. Indeed, Product Scheme Classifications (also known as PSCs), such as the CPV, are a kind of KOS that have been built to solve specific problems of interoperability and communication between e-Commerce agents (Omelayenko & Fensel, 2001; Schmitz & Leukel, 2005). In fact PSCs are considered to be a key-enabler of the next European e-Procurement domain and other supply chain processes (Alor-Hernández, Gómez Berbís, & Rodríguez González, 2010).

Obviously the public information published by governmental contracting authorities, more specifically PSCs, are a suitable candidate to apply the Linked (Open) Data (LOD) principles and semantic web technologies providing a new data environment. The conjunction of these initiatives provides the adequate building blocks for an innovative and unified pan-European information system that encourages standardization of key processes and systems. In fact economic operators can take advantage of this approach to overcome some technical interoperability issues and make a better reuse of public sector information. According to the aforementioned points main contributions of this paper can be summarized as follows:

- Modeling, promotion and inter-linkage of PSCs, structure and data, following the LOD principles.
- Publication of all information via a SPARQL endpoint providing a public procurement Linked Data node and, more specifically, a new PSCs catalogue as Linked Data.
- Exploitation of the PSCS catalogue and an existing dataset of public procurement notices via a contract descriptor recommendation service.
- Demonstration of the gain (in terms of number of descriptors to retrieve public procurement notices) and the semantic exploitation capabilities of the new PSCs catalogue.