Towards Linked Open Business Registers:
The Application of the Registered Organization Vocabulary in Greece

Anastasia Varytimou, International Hellenic University, Kentriki Makedonia, Greece
Nikolaos Loutas, PwC, Zaventem, Belgium
Vassilios Peristeras, EC, DG for Informatics, Interoperability Solutions for European Public Administrations, Brussels, Belgium

ABSTRACT

Business data is high-value data and has a high reuse potential both in national and in cross-border settings. National business registers, tax authorities and others are currently holding such data. Opening up basic business data and making it available in machine-readable formats will definitely increase its reuse. In this work, the authors present, following a stepwise approach, how this can be achieved. They used a part of the dataset of legal entities, comprising approximately 29000 entities, maintained by the Greek Tax Authorities and described it with the Registered Organization vocabulary (RegOrg). The authors followed a publishing and consumption pipeline and made this data available as Linked Open Government Data (LOGD). They conclude this work with a number of findings and lessons learnt which can be useful for base registers and other owners of authoritative data that are considering moving towards LOGD.

Keywords: Business Registers, Core Vocabulary, Linked Data, Open Government Data, Publishing Pipeline, Semantic Interoperability, Semantic Web

INTRODUCTION

Business data is high-value data (Policy paper G8 Open Data Charter and Technical Annex, 2013; Spanish Ministry of Finance et al., 2012) and has a high reuse potential both in national and in cross-border settings. Business data may be of different types, such as basic data about a company (e.g. legal name, address, representative, establishment date and company type), company identifiers and annual balance sheets. Despite its great potential (Deloitte Analytics paper, 2013; Yiu, C., 2012), business data – similar to other types of valuable data – is still locked in business registers and company databases.

In this vein, the European Commission published the directive for interconnection of business registers of the Member States (2012/17/EU) (European Parliament & Council of the European

DOI: 10.4018/IJSWIS.2015040103
Union, 2012) which calls the European public administrations to open up basic business data, such as the name and legal form of the company, the registered office of the company and the Member State where it is registered, the registration number of the company and information on winding-up or insolvency proceedings.

The real value of Open Government Data (OGD) (Study on Business Models for Linked Open Government Data - BM4LOGD, 2013; Advisory Board on Public Sector Information, 2006; Australian Government & Office of the Australian Information Commissioner, 2011; Pollock R., 2006; Dekkers, M., et al. 2006; Lathrop, D., & Ruma, L., 2010) including business data, is revealed when different datasets are integrated. But data integration is far from straight-forward. Three main problems are commonly come across: different, often incompatible, licenses under which OGD is published, incompatible file formats (Bizer, C., 2009), and semantic interoperability conflicts both at the schema and at the data level (Peristeras, V., et al., 2008).

In this work, we focus on the last two. Basic business data coming from different registers is most likely not to be semantically interoperable and can therefore not be integrated and reused within and across borders. This happens due to the lack of common identifiers and semantics, e.g. vocabularies and controlled lists used for describing the data, the absence of commonly agreed metadata and the multilingualism issue (Ding, L., et al. 2012; Alani, H., et al., 2007). Additionally, different registers may publish their datasets in different, incompatible, file formats, e.g. CSV and PDF (Sunlight Foundation, 2013).

We therefore concentrate on the use of widely-accepted, reusable and extendible vocabularies in publishing basic business data. We employ Linked Data technologies as a means of publishing and retrieving this data in both human- and machine-readable formats (Berners-Lee, T., 2006; Berners-Lee, T., 2009; Heath, T., & Bizer, C., 2011). This way, Linked Data acts as an enabler of data interoperability and integration, and also provides open standards for data identification and representation (Shadbolt, N., et al., 2012, Sheridan, J., & Tennison, J., 2010).

Our experience has shown that most literature on publishing linked data up to now dealt with less complex, non-operational data (or statistical data which is thoroughly covered), usually in English. From discussions with practitioners we saw that there is a need to approach the topic also from an engineering point of view, to identify the practical challenges, the organisational limitations, but also the new research contributions and innovation needed for taking the publishing of linked data to a production-ready level. Our work aims at building a bridge between linked data research and new potential adopters of the technology, hence increasing its use.

In that purpose we provide a stepwise introduction to researchers, practitioners and policy makers on how to move from basic business data described in various formats housed in government registries towards LOGD, exposing challenges and emphasizing on the points of attention. A walkthrough example is used, demonstrating the publishing and linking of basic business data of 29000 organizations registered in the Greek Tax Authorities.

The remainder of the paper is structured as follows. In the following section we present the case study of the Greek Tax Authorities Business Registry and the problems that need to be tackled. Next, we describe the current state-of-art in opening-up business registers. We then introduce the Core Vocabularies that have been developed from the European Commission’s Interoperability Solutions for European Public Administrations (ISA) Programme. We focus mainly on how the Registered Organization (RegOrg) vocabulary (previous called Core Business Vocabulary) can be used to achieve interoperability among business registers. We continue with a description of the process that we followed to address the problems of our case study, while in the next section we detail each step of the process. Finally in the last section we discuss our conclusions and lessons learnt.
Challenges on Semantic Web Services
www.igi-global.com/chapter/challenges-semantic-web-services/35721?camid=4v1a