TheyBuyForYou: Enabling Procurement Data Value Chains

Elena Simperl[1], Oscar Corcho[2], Marko Grobelnik[3], Dumitru Roman[4], Ahmet Soylu[4(✉)], María Jesús Fernández Ruíz[5], Stefano Gatti[6], Chris Taggart[7], Urška Skok Klima[8], Annie Ferrari Uliana[9], Ian Makgill[10], Philip Turk[4], and Till Christopher Lech[4]

[1] University of Southampton, Southampton, UK
[2] Universidad Politécnica de Madrid, Madrid, Spain
[3] Jožef Stefan Institute, Ljubljana, Slovenia
[4] SINTEF Digital, Oslo, Norway
ahmet.soylu@sintef.no
[5] Ayuntamiento de Zaragoza, Zaragoza, Spain
[6] Cerved Group Spa US, Milan, Italy
[7] OpenCorporates Ltd., London, UK
[8] Ministrstvo za javno upravo, Ljubljana, Slovenia
[9] OESIA Networks SL, Madrid, Spain
[10] OpenOpps Ltd., London, UK

Abstract. The release of a growing amount of open procurement data means that we are increasingly able, and even have the obligation, to scrutinize and analyse public spending for delivering better quality of public services, optimizing costs, preventing fraud and corruption, and building healthy and sustainable economies. The TheyBuyForYou project addresses this challenge by developing an integrated technology platform, with a cross-lingual and cross-border procurement knowledge graph, core services, open APIs, and online tools, and validating them in several business cases in public/corporate procurement in Slovenia, Spain and Italy. This paper gives an overview about the project's goals and challenges.

Keywords: Knowledge graph · Public procurement · Ontology · Interaction design · Data analytics · Cross-lingual document comparison

1 Introduction

The interaction between governments and their suppliers needs to be subjected to new levels of scrutiny to ensure the efficient delivery of public services and to protect the interests of taxpayers. With a spending in the range of trillions of euros[1], governments are facing a real responsibility to ensure that this money is

[1] http://ec.europa.eu/DocsRoom/documents/20679.

This work is funded by EU H2020 TheyBuyForYou project (780247).

© Springer Nature Switzerland AG 2020
M. Fazio and W. Zimmermann (Eds.): ESOCC 2018 Workshops, CCIS 1115, pp. 179–186, 2020.
https://doi.org/10.1007/978-3-030-63161-1_15

used in the best way possible and that decisions are made considering inclusive, long-term goals and strategies. The release of a growing amount of open procurement data means that we are increasingly able, and even have the obligation, to scrutinize and analyse public spending for delivering better quality of public services, optimizing costs, preventing fraud and corruption, and building healthy and sustainable economies [1].

The recently started TheyBuyForYou project[2] (Enabling procurement data value chains for economic development, demand management, competitive markets and vendor intelligence) addresses this challenge by developing an integrated technology platform, with a cross-lingual and cross-border procurement knowledge graph, core services, open APIs, and online tools, and validating them in several business cases in public/corporate procurement in Slovenia, Spain and Italy. This paper gives an overview about the project's goals and challenges.

The rest of the paper is structured as follows. Section 2 discusses the related work, while Sect. 3 presents the project background through project's objectives and challenges. Section 4 describes the TheyBuyForYou approach and Sect. 5 presents customer scenarios and business cases. Finally, Sect. 6 reports the current state of the project.

2 Related Work

Many of the tools in use by governments are often not optimised for government use, or are subject to restrictive contracts which unnecessarily complicate publishing open data. Other contracts, such as contracts for tender advertising portals are hampering the progress of transparency because the portals are claiming copyright over all data published in the portals, even though their public-sector clients are the authors and the data on tender opportunities are required to be published openly by law. The technical landscape for managing such contracts is very heterogeneous: for example, even in medium-sized cities, contracts are handled using different tools and formats across departments, including relational databases, Excel spreadsheets, and Lotus Notes. This makes it difficult to have a high-level overview of processes and decisions.

There are various initiatives whose purpose is to create de-jure and de-facto standards for electronic procurement, including such as Open Contracting Data Standard (OCDS)[3] and TED eSenders[4]. However, these are mostly oriented to achieve interoperability (i.e., addressing communication between systems), document oriented (i.e., the structure of the information is commonly provided by the content of the documents that are exchanged), and provide no standardised practices to refer to third parties, companies participating in the process, or even the main object of contracts. This at the end generates a lot of heterogeneity. Procurement domain can take advantage of applying the Semantic Web approach by

[2] https://theybuyforyou.eu.
[3] http://standard.open-contracting.org/latest/en/.
[4] http://simap.ted.europa.eu/.

reusing existing vocabularies, ontologies, and standards [1]. Specifically in the procurement domain, these include among others PPROC ontology [6] for describing public processes and contracts, LOTED2 ontology [3] for public procurement notices, PCO ontology [7] for contracts in public domain, and MOLDEAS ontology [9] for announcements about public tenders. LOTED2 is considered as a legal ontology and is comparatively more complex and detailed with respect to MOLDEAS, PCO, and PPROC. The latter is concerned on reaching a balance between usability and expressiveness.

3 Background

TheyBuyForYou explores how procurement and public spending data, paired with data management, analytics, and interaction design, could be used to innovate four key areas:

(i) economic development by delivering better economic outcomes from public spending, in particular for SMEs (to get better access to public tenders, competing with more established players etc.);
(ii) demand management by spotting trends in spending and supplier management to achieve long-term goals such as cost savings and efficiency gains;
(iii) competitive markets by identifying areas for cost cuts through healthier competition;
(iv) and, procurement intelligence by producing advanced analytics to inform decision support, risk monitoring and supply market analysis for procurement managers.

3.1 Objectives

Our first objective is to build a technology platform, consisting of a set of modular, Web-based services and APIs to publish, curate, integrate, analyse, and visualize an open, comprehensive, cross-border and cross-lingual procurement knowledge graph, including public spending and corporate data from multiple sources across the EU.

Our second objective is to support the realisation of the four innovation areas discussed, through a series of online tools and public portals, which allow suppliers, buyers, data journalists, data analysts, control authorities and regular citizens to explore and understand how public procurement decisions affect economic development, efficiencies, competitiveness and supply chains. For private buyers looking to overhaul their procurement and purchasing decisions, we will deliver vendor intelligence solutions with advanced analytics capabilities around risk monitoring, collusive tendering, and bespoke decision support.

Finally, our third objective is validation in the procurement market to investigate how the knowledge graph published can be used to support the four main innovation scenarios. While the economic development and procurement intelligence scenarios target SMEs and big industries looking for subcontractors, the other two offer decision support to public buyers and other parties interested in an analysis of the public spending market.

3.2 Challenges

The first challenge to meet is the heterogeneity of the underlying data, which covers structured (e.g., statistics, financial news) as well as unstructured (e.g., text, social media) sources in different languages and using their own terminology and formats (CSV, PDF, databases, websites, APIs etc.). To be truly useful, our technology will have to offer its services in real-time, including the thousands of new tenders published on official portals such as TED (Tenders Electronic Daily)[5] every week, as well as general-purpose corporate data streams such as business-centric social networks and stock market data.

The second challenge will be in turning this vast array of information into a semantic knowledge graph [12], an interconnected knowledge organization structure using Web URIs and linked data vocabularies, which can be analysed in depth to identify patterns and anomalies in procurement processes and networks. Finally, we need to find means to communicate the results of our analysis and inform decisions, which convey useful information while scaling well to complex data shapes and large volumes of data.

4 TheyBuyForYou Approach

The TheyBuyForYou approach has three layers. First, data and technology layer enables developers to create fully functional, robust, and scalable data integration pipelines, from sourcing the data; pre-processing, augmenting, and interlinking it; to learning patterns and anomalies; making predictions; and communicating the insights to specific audiences. Second, tools and products layer offers end-user tools and procurement APIs such as for visualisation, and document comparison. Finally, validation layer realises a set of selected business cases over online portals, one for suppliers and one for buyers, built on tools and products layer. The TheyBuyForYou approach is mapped to a high-level architecture as described in Fig. 1.

4.1 Procurement Knowledge Graph

The knowledge graph primarily integrates supplier data and procurement data. Core company data is provided by OpenCorporates[6]. Tenders and contracts data is provided by OpenOpps[7] in the OCDS (Open Contracting Data Standard) format[8], whose primary source of data is the TED data feed. The data is curated (e.g., missing and duplicate records), normalised, and integrated through a common ontology [5,6,11]. The entities in the knowledge graph (e.g., tenders and suppliers) are linked and reconciled. The data will be made available through SPARQL end-points, open APIs, and linked data interfaces.

[5] http://ted.europa.eu.
[6] https://opencorporates.com.
[7] https://openopps.com.
[8] http://standard.open-contracting.org/latest/en.

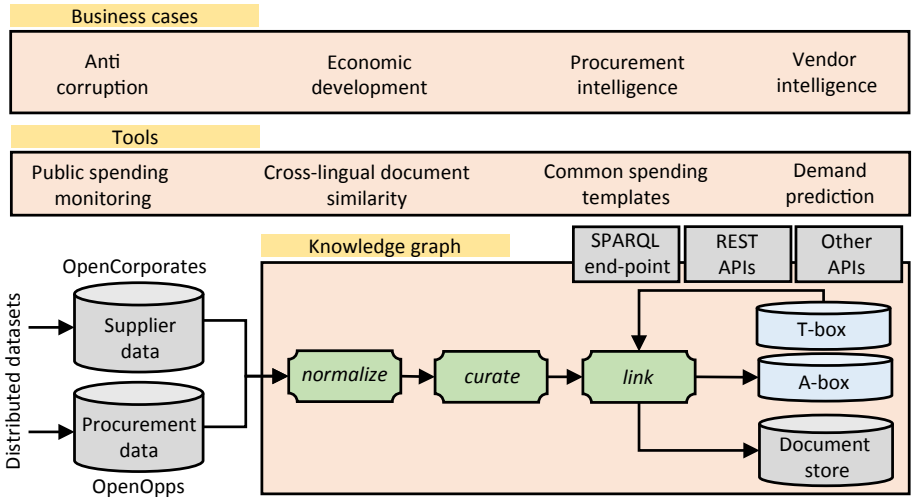

Fig. 1. TheyBuyForYou high level architecture.

4.2 Cross-lingual and Real-Time Analytics

A real-time monitoring and analysis framework is to be developed for public and private procurement information that is published over several different sources and in different languages. The framework needs to scale to the millions of documents available (calls, news, events etc.). Since most procurement documents are available in the native language of the issuing authority or organisation, cross-language support is a key [4]. A methodology for discovery of common order and spending patterns is also to be developed, which will offer additional insight into public procurement practices [2].

4.3 Data Interaction and Story Telling

TheyBuyForYou will develop tools, using such as interactive visualizations, configurable infographics, and automatically generated data stories, to explore procurement data and to support sense making and decision making based on this data. Three features will be at the core of the designs: encouraging exploration, as a means to master the high-dimensionality of the knowledge graph; emphasizing provenance trails for the data and accountability of analytics insights; and contextualizing visualizations and analytics findings via data narratives [8]. To make the creation of data narratives manageable, we will learn storytelling templates, which could be assembled into configurable infographics.

5 Customer Scenarios and Business Cases

We will develop a series of tools across four innovation scenarios (i.e., economic development, demand management, competitive markets, and procurement

intelligence), discussed in Sect. 3, to ensure the project impact. These scenarios will result in two open online portals: one for suppliers/bidders and a second one for buyers. Tools and APIs delivered will be used to implement initially three business cases (see Fig. 1) targeting the main customer segments in procurement. Two business cases will be with public administrations at the national and local levels in Spain and Slovenia, while the third one is a corporate business case and will result in a new commercial product.

5.1 Business Case 1: Slovenia

The first case targets competitive markets and advanced procurement intelligence scenarios and is led by Ministrstvo za javno upravo (The Ministry of Public Administration, Slovenia) with support from Jozef Stefan Institute. The Slovenian business case is centred around the theme of anti-corruption, including the following aspects: definition of selection criteria for tenders; monitoring diversity of requirements in similar procurement calls and issuing of an alert when relevant anomalies are identified; providing support in the post procurement process, so as to prevent unnecessary changes to the original contract; and offering transparency data to help ensure they are working with the most suitable and reliable suppliers.

5.2 Business Case 2: Spain

The second case targets economic development, demand management, and competitive markets scenarios and is led by Ayuntamiento de Zaragoza (City of Zaragoza, Spain) with support from Oesia. They together with other regional public administrations in Spain launched the PPROC ontology within W3C for the standardisation of public procurement information according to the Spanish legislation. City of Zaragoza and Oesia aim to improve their procurement business by focusing on easier access to smaller companies for specific types of tenders, better understanding of demand from public organizations both inside and outside Spain, and identifying opportunities to cut costs.

5.3 Business Case 3: Italy

The third case targets advanced procurement intelligence and is led by Cerved. Cerved will develop a new product for vendor intelligence targeting the entire procurement market. The product will use open data (e.g., knowledge graph created in this project); proprietary data (e.g., chamber of commerce data, balance sheet data etc.); and third-party data (e.g., website traffic statistics). An important goal is to provide a more nuanced supplier analysis and classify relevant companies into micro, small and medium-sized enterprises by cross-connecting multiple data sources. This will enable Cerved to target SMEs better and define appropriate similarity scores between tenders and potential bidders.

6 Current Status

At the time of writing, the project is in the middle of its first year and plan is to publish the first version of the knowledge graph at the end of first year. Currently data providers, i.e., OpenOpps and OpenCorporates, are expanding their data coverage by identifying, prioritising, and auditing new data sources with respect to some quality criteria (e.g., legal, practical, and technical) and the needs of the business cases. Some data curation and integration activities directly take place at the side of our data providers, while main integration tasks will be handled by DataGraft, a cloud-based platform for data transformation and publishing [10].

A cross-lingual document similarity service for automatic comparison of public orders and spending documents across different languages has been implemented and deployed through a RESTful API. A conceptual framework has been developed for describing dimensions of data visualisation, based on a review of background literature and media, for the purpose of informing the initial process of ideation prior to the creation of visualisation and narrative components. Finally, existing ontologies for suppliers and procurement data are being reviewed for re-use (cf. [1]) and requirements are being collected from business cases.

References

1. Alvarez-Rodríguez, J.M., et al.: New trends on e-Procurement applying semantic technologies: current status and future challenges. Comput. Ind. **65**(5), 800–820 (2014)
2. Chandola, V., et al.: Anomaly detection: a survey. ACM Comput. Surv. **41**(3), 15:1–15:58 (2009)
3. Distinto, I., et al.: LOTED2: an ontology of European public procurement notices. Semant. Web **7**(3), 267–293 (2016)
4. Fortuna, B., et al.: A kernel canonical correlation analysis for learning the semantics of text. In: Kernel Methods in Bioengineering, Communications and Image Processing (2006)
5. Kharlamov, E., et al.: Ontology based data access in Statoil. Web Semant. Sci. Serv. Agents World Wide Web **44**, 3–36 (2017)
6. Muñoz-Soro, J.F., et al.: PPROC, an ontology for transparency in public procurement. Semant. Web **7**(3), 295–309 (2016)
7. Necaský, M., et al.: Linked data support for filing public contracts. Comput. Ind. **65**(5), 862–877 (2014)
8. Portet, F., et al.: Automatic generation of textual summaries from neonatal intensive care data. Artif. Intell. **173**(7), 789–816 (2009)
9. Rodríguez, J.M.Á., et al.: Towards a pan-European e-procurement platform to aggregate, publish and search public procurement notices powered by linked open data: the MOLDEAS approach. Int. J. Softw. Eng. Knowl. Eng. **22**(3), 365–384 (2012)

10. Roman, D., et al.: DataGraft: one-stop-shop for open data management. Semant. Web **9**(4), 393–411 (2018)
11. Suchanek, F.M., et al.: Knowledge bases in the age of big data analytics. Proc. VLDB Endow. **7**(13), 1713–1714 (2014)
12. Yan, J., et al.: A retrospective of knowledge graphs. Front. Comput. Sci. **12**(1), 55–74 (2018)