Towards an Ontology for Public Procurement Based on the Open Contracting Data Standard

Ahmet Soylu[1(✉)], Brian Elvesæter[1], Philip Turk[1], Dumitru Roman[1],
Oscar Corcho[2], Elena Simperl[3], George Konstantinidis[3],
and Till Christopher Lech[1]

[1] SINTEF Digital, Oslo, Norway
ahmet.soylu@sintef.no
[2] Universidad Politécnica de Madrid, Madrid, Spain
[3] University of Southampton, Southampton, UK

Abstract. The release of a growing amount of open procurement data led to various initiatives for harmonising the data being provided. Among others, the Open Contracting Data Standard (OCDS) is highly relevant due to its high practical value and increasing traction. OCDS defines a common data model for publishing structured data throughout most of the stages of a contracting process. OCDS is document-oriented and focuses on packaging and delivering relevant data in an iterative and event-driven manner through a series of releases. Ontologies, beyond providing uniform access to heterogeneous procurement data, could enable integration with related data sets such as with supplier data for advanced analytics and insight extraction. Therefore, we developed an ontology, the "OCDS ontology", by using OCDS' main domain perspective and vocabulary, since it is an essential source of domain knowledge. In this paper, we provide an overview of the developed ontology.

Keywords: Procurement · OCDS · Ontology

1 Introduction

Public entities worldwide are increasingly required to publish information about their procurement processes (e.g., in Europe, with EU directives 2003/98/EC and 2014/24/EU8) in order to improve effectiveness, efficiency, transparency, and accountability of public services [8]. As a result, the release of a growing amount of open procurement data led to various initiatives (e.g., OpenPEPPOL[1], CEN BII[2], TED eSenders[3], CODICE[4], Open Contracting Data Standard (OCDS)[5])

---

[1] https://peppol.eu.

[2] http://cenbii.eu.

[3] https://simap.ted.europa.eu/web/simap/sending-electronic-notices.

[4] https://contrataciondelestado.es/wps/portal/codice.

[5] http://standard.open-contracting.org.

for harmonising the data being provided. XML formats and file templates are defined within these standards to make it possible to structure the messages exchanged by the various agents involved in electronic procurement. These standards are mostly oriented to achieve interoperability, addressing communication between systems, and hence they usually focus on the type of information that is transmitted between the various organizations involved in the process. The structure of the information is commonly provided by the content of the documents that are exchanged. Furthermore, there are no generalised standardised practices to refer to third parties, companies participating in the process, or even the main object of contracts. In sum, this still generates a lot of heterogeneity. Ontologies have been proposed to alleviate this problem [1,10]. Several ontologies (e.g., PPROC [5], LOTED2 [3], MOLDEAS [7], PCO [6]) have recently emerged, with different levels of detail and focus (e.g., legal, process-oriented, pragmatic). However, none of them has had a wide adoption so far.

In this context, OCDS is highly relevant due to its high practical value and increasing traction. It defines a common data model for publishing structured data throughout all the stages of a contracting process. It is document-oriented and focuses on packaging and delivering relevant data in an iterative and event-driven manner through a series of releases. However, in its current form, OCDS is a mere data structure. An ontology, beyond providing uniform access to heterogeneous procurement data, could enable integration with related data sets for advanced analytics and insight extraction. For instance, in the context of the EU project TheyBuyForYou[6] [8,10], we aim to integrate procurement and supplier data for improving effectiveness, efficiency, transparency, and accountability of public procurement through analytics and integrated data access.

To this end, in this paper, we report on the design and development of an ontology—the "OCDS ontology"—that uses the main perspective and vocabulary of OCDS, since it is an essential source of domain knowledge with high adoption.

2   Ontology-Based Approach

An ontology is a formal and shared specification of a domain of interest in a machine-readable format. One could consider the reasons behind choosing an ontology-based approach from knowledge representation and logical programming perspectives [2]. For the former, firstly, ontologies provide a commonly agreed terminology, that is a vocabulary and the semantic interpretation of the terms provided. Using a well-specified and unambiguous terminology enables the sharing and integration of data between disparate systems. Secondly, one could use a network of ontologies in a modular way to integrate different but related data sets without implementing yet another information model. For the latter, due to logical foundations of ontologies, one could infer new facts (i.e., implied information) through logical reasoning from the existing data (i.e., explicit information) and check the consistency of data.

---

[6] https://theybuyforyou.eu.

The adoption of ontology-based approaches is gaining momentum due to raising paradigms such as knowledge graphs (KG) [11] and ontology-based data access (OBDA) [4]. A KG represents real-world entities and their interrelations. Semantic KGs created by using ontologies and related technologies such as RDF and OWL benefit from high expressive and logical capabilities of ontologies for representing, exchanging, and querying data. OBDA approach is complementary in this respect, where an ontology defines a high-level global schema for existing data sources. For example, it could enable virtualising multiple heterogeneous data sources to a semantic KG without needing to alter original data sources through mappings between the underlying data sources and an ontology [9].

Our main interest and challenge in this context is to provide an ontology to allow uniform access to procurement data through a common terminology, which is based on a well-established standard, and to allow integrating procurement data with other related data sets. An example could be linking procurement data with company data for fraud analysis by detecting abnormal patterns in data.

3   OCDS Ontology

OCDS' data model is organised around a contracting process by gathering all the relevant information associated with a single initiation process in a structured form. Phases in this process include mainly planning, tender, award, contract and implementation information. An OCDS document may be one of two kinds: a release or a record. A release is basically associated to an event in the lifetime of a contracting process and presents related information, while a record compiles all the known information about a contracting process. A contracting process may have many releases associated but should have only one record.

Each release provides new information and may also repeat the previous information which still holds. A release document is composed of a number of sections. These are mainly: parties, planning, tender, awards, contract, and implementation. OCDS defines data fields for each section, and for some of those fields, it uses "open" or "closed" code lists, providing fixed and recommended lists of values respectively. There are also extensions defined by the OCDS or third parties for enabling publishing external data. We refer interested readers to the full OCDS specification for more details.

3.1   Development Process

We went through the reference specification of OCDS release and interpreted each of the sections and extensions including structured and unstructured information. The result is the first set of classes and properties forming an ontology for OCDS as depicted briefly in Fig. 1. Each ontology element is annotated with `rdfs:comment` and `rdfs:isDefinedBy` in order to provide a mapping to the corresponding OCDS fragment.
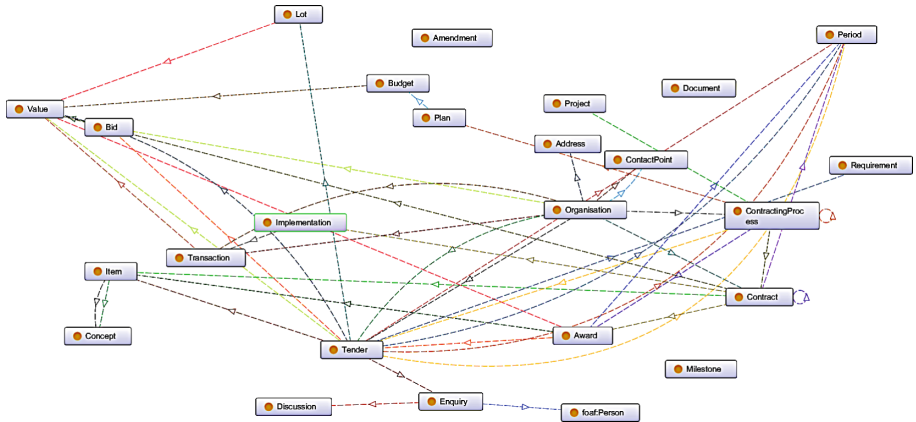
Fig. 1. OCDS ontology visualised with Protege OntoGraph plugin.

Though release reference seems to have a process-oriented perspective in some parts of the text, we avoided including process classes (except for the core class `ContractingProcess`), since the amount of information describing different processes is rather limited. We also avoided describing the release events for simplicity and assume that the ontology reflects the latest data available.

In line with Linked Data principles, we reused terms from external vocabularies and ontologies when appropriate. These include Dublin Core[7], FOAF[8], Schema.org[9], SKOS[10], and the W3C Organization ontology[11]. At this point, we have not used any cardinality restrictions, although OCDS provides some information on this. Finally, we used OWL annotation properties for specifying domain and ranges for generic properties in order to avoid any over restriction (i.e., `ocds:usesDataProperty`).

This edition of the OCDS ontology is available online in GitHub in two versions[12]: one version only with core OCDS terms and a second version with extensions (e.g., enquiries, lots, etc.).

3.2   Ontology Topology

In total, there are currently 24 classes, 62 object properties, and 80 datatype properties created from the four main OCDS sections and 11 extensions. In what follows, we zoom into each core class (i.e., directly mapping one of the OCDS sections) and discuss related classes and properties. The core classes are
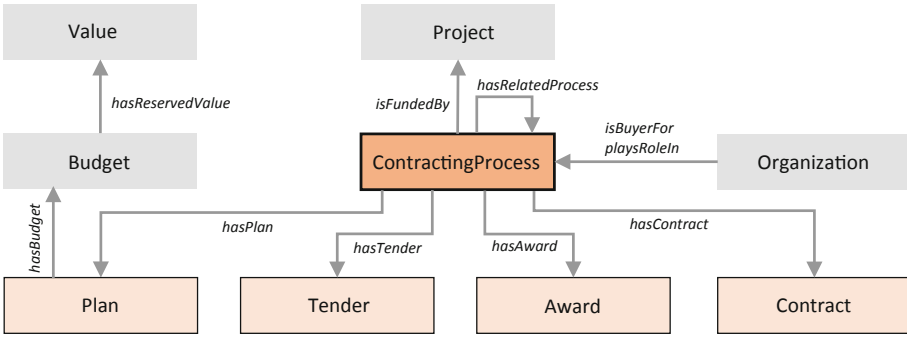
---

[7] http://dublincore.org.

[8] http://xmlns.com/foaf/spec.

[9] https://schema.org.

[10] https://www.w3.org/2004/02/skos.

[11] https://www.w3.org/TR/vocab-org.

[12] https://github.com/TBFY/ocds-ontology/tree/master/model.

Fig. 2. ContractingProcess class and its neighbourhood.

`ContractingProcess`, `Plan`, `Tender`, `Award`, and `Contract`. Classes emerging from extensions are marked with "e" in the figures.

In Fig. 2, the neighbourhood of the `ContractingProcess` class is shown. A contracting process may have one planning and tender stage and multiple awards and contracts. The object property `hasRelatedProcess` has some subproperties originating from related process code list (e.g., `hasRelatedFrameworkProcess`, `hasRelatedParentProcess`, etc.). Process level title and description extension is used in the `ContractingProcess` class.
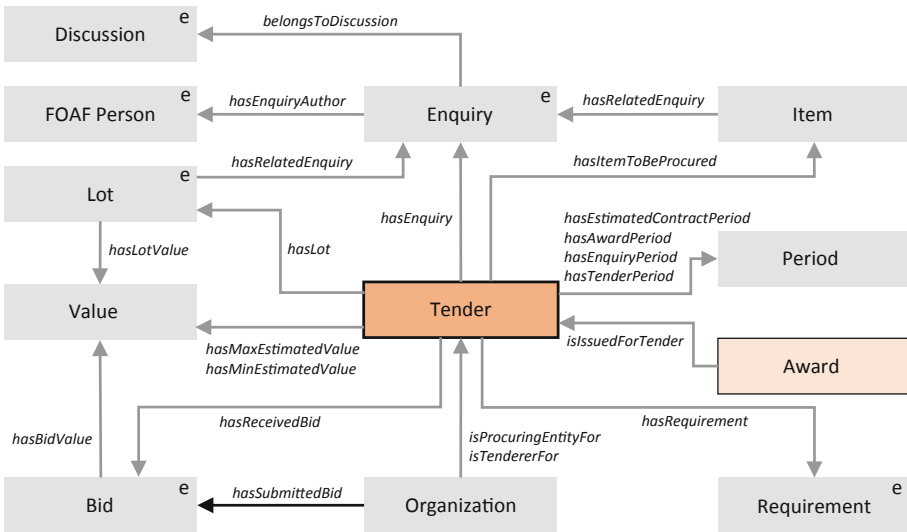
Fig. 3. Tender class and its neighbourhood.

The neighbourhood of the `Tender` class is shown in Fig. 3. Each tender may have multiple awards issued. The classes emerging from the tender
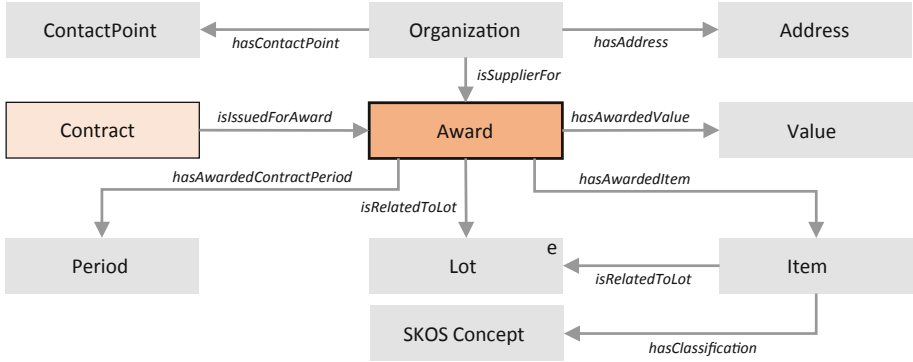
Fig. 4. Award class and its neighbourhood.

section and connected to the `Tender` class are `Value`, `Period`, `Award`, `Item`, and `Organization`. Several extensions are implemented, such as enquiries, lots, requirements, and bids, which map to the following classes: `Lot`, `Bid`, `Discussion`, `Enquiry`, `Requirement`, and `Person`.

In Fig. 4, the neighbourhood of the `Award` class is shown. There may be only one contract issued for each award. Other classes emerging from the award section and connected to `Award` class are `Value` and `Organization`. Lot extension also involves the `Award` class through the `isRelatedToLot` property. Item classifications (such as CPV – common procurement vocabulary) are realised through the use of SKOS `Concept`.
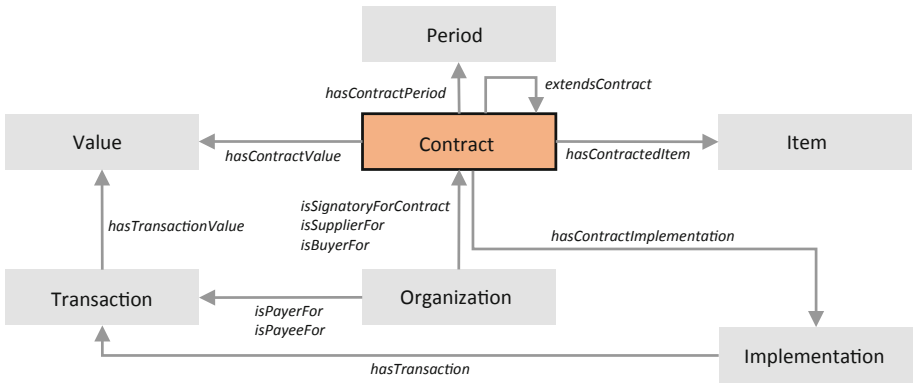

Fig. 5. Contract class and its neighbourhood.

Finally, the neighbourhood of the `Contract` class is shown in Fig. 5. It includes `Value`, `Item`, `Period`, and `Organization`. Multiple buyers extension is applied through extending the range of `isBuyerFor` property to the `Contract`.

Not all the classes and extensions implemented in the ontology are mentioned here. We refer readers to the OCDS specification and to the ontology documentation for more information regarding each class, OCDS mappings, datatype properties, extensions, and external vocabularies and ontologies used.

4 Conclusions

We developed an ontology derived from OCDS' schema and its extensions. We expect this to have a practical value and will be an important contribution for ongoing ontology development efforts, such as the upcoming eProcurement ontology[13]. Regarding future work, possible directions include the use of the Shapes Constraint Language (SHACL)[14] for validating RDF graphs based on the OCDS ontology (including cardinalities), development of a process ontology for procurement in combination with the OCDS ontology from a modular perspective, and extending the OCDS ontology to capture history (i.e., release events).

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. Baader, F., et al. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, New York (2003)
3. Distinto, I., et al.: LOTED2: an ontology of European public procurement notices. Semant. Web **7**(3), 267–293 (2016)
4. Giese, M., et al.: Optique: zooming in on big data. IEEE Comput. **48**(3), 60–67 (2015)
5. Muñoz-Soro, J.F., et al.: PPROC, an ontology for transparency in public procurement. Semant. Web **7**(3), 295–309 (2016)
6. Necaský, M., et al.: Linked data support for filing public contracts. Comput. Ind. **65**(5), 862–877 (2014)
7. 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)
8. Simperl, E., et al.: Towards a knowledge graph based platform for public procurement. In: Garoufallou, E., Sartori, F., Siatri, R., Zervas, M. (eds.) MTSR 2018. CCIS, vol. 846, pp. 317–323. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-14401-2_29
9. Soylu, A., et al.: Querying industrial stream-temporal data: an ontology-based visual approach. J. Ambient Intell. Smart Environ. **9**(1), 77–95 (2017)

---

10. Soylu, A., et al.: Towards integrating public procurement data into a semantic knowledge graph. In: Proceedings of the EKAW 2018 Posters and Demonstrations Session Co-Located with 21st International Conference on Knowledge Engineering and Knowledge Management (EKAW 2018), CEUR Workshop Proceedings, vol. 2262, pp. 1–4. ceur-ws.org (2018)
11. Yan, J., et al.: A retrospective of knowledge graphs. Front. Comput. Sci. **12**(1), 55–74 (2018)