Initiative for developing eProcurement Ontology

eProcurement ontology architecture and formalisation specifications

Deliverable WP 1.1

Eugeniu Costetchi

17 April 2020

Disclaimer

Project acronym	ePO
Project title	Initiative for developing eProcurement Ontology
Document reference	eProcurement ontology architecture and formalisation specifications
Author(s)	Eugeniu Costetchi
Editor(s)	Eugeniu Costetchi
Contractor	Infeurope S.A.
Framework contract	10688/35368
Actual delivery date	17 April 2020
Delivery nature	Report (R)
Dissemination licence	Creative Commons Attribution 4.0 International
Filename	wp1-1-ontology-architecture
Suggested readers	project partners, future users, legal practitioners, software architects

Abstract

TBD

Contents

1 Introduction

This document provides a working definition of what is the architectural stance and the design decisions that shall be adopted for the eProcurement ontology along with the description of how to generate it.

1.1 Background considerations

Public procurement is undergoing a digital transformation. The EU supports the process of rethinking public procurement process with digital technologies in mind. This goes beyond simply moving to electronic tools; it rethinks various pre-award and post-award phases. The aim is to make them simpler for businesses to participate in and for the public sector to manage. It also allows for the integration of data-based approaches at various stages of the procurement process.

Digital procurement is deeply linked to eGovernment. It is one of the key drivers toward the implementation of the "once-only principle" in public administrations – a cornerstone of the EU's Digital Single Market strategy. In the age of big data, digital procurement is also crucial in enabling governments to make data-driven decisions about public spending.

With digital tools, public spending should become more transparent, evidence-oriented, optimised, streamlined and integrated with market conditions. This puts eProcurement at the heart of other changes introduced to public procurement in new EU directives.

PSI directive [10] across the EU is calling for open, unobstructed access to public data in order to improve transparency and to boost innovation via the reuse of public data. Procurement data has been identified as data with a high-reuse potential [2]. Therefore, making this data available in machine-readable formats, following the data as a service paradigm, is required in order to maximise its reuse.

Given the increasing importance of data standards for eProcurement, a number of initiatives driven by the public sector, the industry and academia have been kick started in the recent years. Some have grown organically, while others are the result of standardisation work. The vocabularies and the semantics that they are introducing, the phases of public procurement that they are covering, and the technologies that they are using all differ. These differences hamper data interoperability and thus its reuse by them or by the wider public. This creates the need for a common data standard for publishing public procurement data, hence allowing data from different sources to be easily accessed and linked, and consequently reused.

In this context, the Publications Office of the European Union aims to develop an eProcurement ontology.

The objective of the eProcurement ontology is to act as this common standard on the conceptual level, based on consensus of the main stakeholders and designed to encompass the major requirements of the eProcurement process in conformance with the Directives and Regulations [11, 12, 13, 14].

1.2 Target audience

The target audience of the eProcurement ontology, defined in [20], comprises the following groups of stakeholders:

- Contracting authorities and entities, i.e. buyers, such as public administrations in the EU Member States or EU institutions;
- Economic operators, i.e. suppliers of goods and services such as businesses, entrepreneurs and financial institutions;
- Academia and researchers;
- Media and journalists;
- Auditors and regulators;
- Members of parliaments at regional, national and EU level;
- Standardisation organisations;
- NGOs; and
- Citizens [20].

1.3 Context and scope

In the past years much effort was invested into the eProcurement ontology initiative, including definition of requirements, provision of general specifications, identification of the main use cases, and laborious development of a preliminary shared conceptual model expressed using Unified Modelling Language (UML) [4, 5].

The general methodology for developing the eProcurement ontology is described in [9, 3–15]. It describes a process comprising the following steps:

1. Define use cases
2. Define the requirements for the use cases
3. Develop a conceptual data model
4. Consider reusing existing ontologies
5. Define and implement an OWL ontology

The ultimate objective of the eProcurement ontology project is to put forth a commonly agreed OWL ontology that will conceptualise, formally encode and make available in an open, structured and machine-readable format data about public procurement, covering end-to-end procurement, i.e. from notification, through tendering to awarding, ordering, invoicing and payment [20].

Work so far has concentrated on the conceptual modelling of the eNotification phase, taking into consideration the needs of other phases. The UML conceptual model has been created with the forthcoming procurement standard forms (eForms) in mind; the model has not been mapped to the current standard forms.

In the 2020 ISA^2 work programme a new project has been set up to analyse existing procurement data through the lens of the newly developed conceptual model. This means that the conceptual model needs to be transposed into a formal ontology and a subset of the existing eProcurement data (initially only the notification phase is considered) must be transformed into RDF format such that they instantiate the eProcurement ontology and are conform to a set of predefined data shapes.

Working under the assumption that Steps 1–4 have been completed, the current efforts channel on designing, implementing and executing the necessary tasks in order to accomplish Step 5 from the above process.

Once the formal ontology is created and the XML data is transformed into RDF representation, the data can be queried in order to validate the suitability to satisfy the business use cases defined in Section 3 of [9].

This document comprises of architectural specification and implementation guidelines that shall be taken into consideration when developing the formal ontology. Other related artefacts (i.e. documents, scripts and datasets) are presented in Section 3, where is described, in detail, the process for accomplishing the generation the formal eProcurement ontology, transformation of XML data and the ontology validation.

There is a number of aspects that are excluded from the scope of this project stage:

- Change management and maintenance of the ontology content.
- Content authoring and conceptual design of the domain model.
- Practical implementation of systems that implement the ontology.

Currently in scope are the following items:

- deigning an ontology architecture (this document),
- a set of guidelines and conventions for the UML conceptual model [6],
- a set of transformations scripts from the UML model into a formal ontology
- a set of scripts to transform for the existent XML eProcurement data into RDF format,
- a method to validate the generated formal ontology using the current eProcurement data.

2 Requirements

2.1 Functional requirements

This section provides the main functionalities and use cases that the ontology should support. These requirements are derived from the use cases identified in the report on policy support for eProcurement [20] and outlined in the in the eProcurement project chapter proposal [8].

1. *Transparency and monitoring*: to enable verification that public procurement is conducted according to the rules set by the Directives and Regulation [11, 12, 13, 14].
    (a) Public understandability
    (b) Data Journalism
    (c) Monitor the money flow

(d) Detect fraud and compliance with procurement criteria

(e) Audit procurement process

(f) Cross-validate data from different parts of the procurement process

2. *Innovation & value added services*: to allow the emergence of new applications and services on the basis of the availability of procurement data.

    (a) Automated matchmaking of procured services and products with

    (b) businesses

    (c) Automated validation of procurement criteria

    (d) Alerting services

    (e) Data analytics on public procurement data

3. *Interconnection of public procurement systems*: to support increased interoperability across procurement systems.

    (a) Increase cross-domain interoperability among Member States

    (b) Introduce automated classification systems in public procurement systems

2.2 Non-functional requirements

This section provides the characteristics, qualities and general aspects that the eProcurement ontology should satisfy.

- The practices, technologies and standards must be aligned with the European Directive on open data and the reuse of public sector information [15] and European Publications Office standards and practices.
- The terminology used in the ontology should be reused from established core vocabularies [27] and domain ontologies as long as their meaning fits into the description of the eProcurement domain.
- The concept and relation labels must be multilingual covering at least the official European Languages [26].

- The formal ontology, and the related artefacts, must be generated from the eProcurement UML conceptual model, serving as the single source of truth, through a set of predefined transformation rules [7].
- The content of the ontology must be consistent with the predefined set of UML conceptual model conventions [6].
- The ontology identifiers must follow a strict URI policy defined in Section **??**.
- The ontology, and the related artefacts, must be layered in order to support different degrees of ontological commitment and levels formal specification stacked on each other (see Section 4.3).
- The ontology, and the related artefacts, must be sliced in order to support a modular organisation of the domain in terms of self contained or semi-dependent modules (see Section **??**).
- The ontology design must commit long term URI persistence.

2.3 General design criteria

For the purpose of knowledge sharing and interoperation between programs based on a shared conceptualisation, Gruber [16] proposes a set of preliminary *design criteria* a formal ontology should follow:

- *Clarity.* An ontology should communicate the purpose and meaning of defined terms. Definitions should be objective and independent of social and computational context even if the underlying motivations arise from them. Formalism is the means to this end, and when possible the logical formulation should be provided.
- *Coherence.* Ontology should permit inferences that are consistent with the definitions. At the least, the defining axioms should be logically consistent. Coherence should also apply to the concepts that are defined informally, such as those described in natural language documentation and examples.
- *Extensibility.* Ontology should be designed to anticipate the uses of the shared vocabulary. It should offer a conceptual foundation for a range of anticipated tasks and the representation should be crafted so that one can extend and specialize the ontology monotonically. This feature supports and encourages

reuse and further specialisations of ontologies and creation of the application profiles.

- *Minimal encoding bias.* The conceptualization should be specified at the knowledge level without depending on a particular symbol-level encoding.
- *Minimal ontological commitment.* Ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities. Ontology should make as few claims as possible about the world being modeled, allowing the parties committed to the ontology freedom specialize and instantiate the ontology as needed. An ontological commitment is an agreement to use the shared vocabulary, with which queries and assertions are exchanged between agents, in a coherent and consistent manner. We say that an agent commits to an ontology if its behaviour is consistent with the definitions in the ontology [16].

3 Process and methodology

The main effort of the current stage of the proiject is to develop a formal ontology. This corresponds to Step 5 of the process described in [9, 3–15] and repeated in Section 1.3.

This section expands and addresses in detail the process of defining and implementing an OWL eProcurement ontology. The underlying assumption is that the conceptual data model developed at Step 3 serves as an input for the creation of the ontology, and that this process shall be automatic.

In addition to producing the ontology as an artefact, we also needs to validates its fitness to represent existing data and test whether the functional and non-functional requirements are respected. Figure 1 depicts the sequence of steps as a BPMN process diagram [30].

The conceptual model serves as the single source of truth, the process starts with development of a series of transformations scripts. The conceptual model needs to be adjusted in order to fit a set of UML modelling conventions [6] making it suitable input for the transformation scripts. Provided that the conceptual model is conform, the transformation can be executed. Finally the validation of the formal ontology can be performed using the existing eProcurement data.

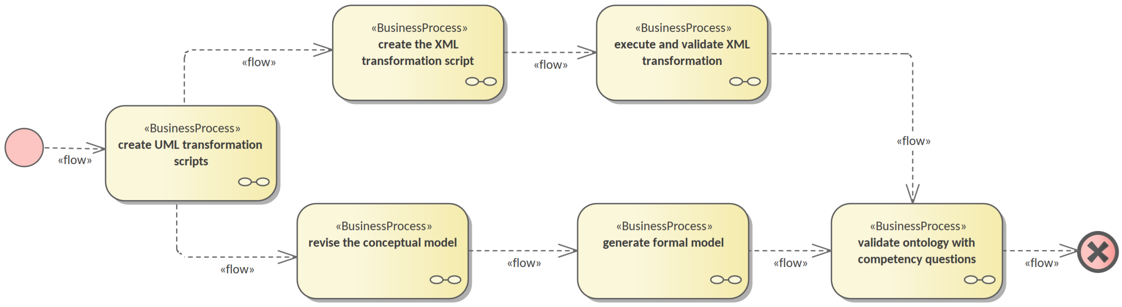


Figure 1: The main steps to implementing and validating the formal eProcurement ontology

The existing eProcurement data needs to be transformed from XML into RDF format. So, in parallel, after the UML transformations are created and along with them, the ontology architecture and UML conventions, then a set of XML transformation scripts can be developed. Once they are ready, they need to be executed on previously selected datasets, to convert them into RDF data instantiating the formal ontology. Only then, when the datasets are available, the ontology can be validated.

The next subsections describe each of these six steps in more detail in order to provide rationale and introduce each artefact in part.

3.1 UML transformation scripts

The process start with authoring two documents laying the foundations of the entire process: the ontology architecture and the UML modelling conventions [6]. The main purpose of the ontology architecture (this document) specifications is to describe why the ontology is being built, what its intended uses are, who the end-users are, and which requirements the ontology should fulfil. Moreover, it states how the ontology should be structured in order to facilitate maintenance and usage patterns.

The conceptual model must comply with a set of UML modelling conventions making it suitable input for the transformation scripts, which implement the same conventions. The two parallel actions staring the process are depicted in Figure 2.

The UML conventions document serves, at large, as requirements specification for the XSLT script that checks the UML conceptual model whether it conforms to the

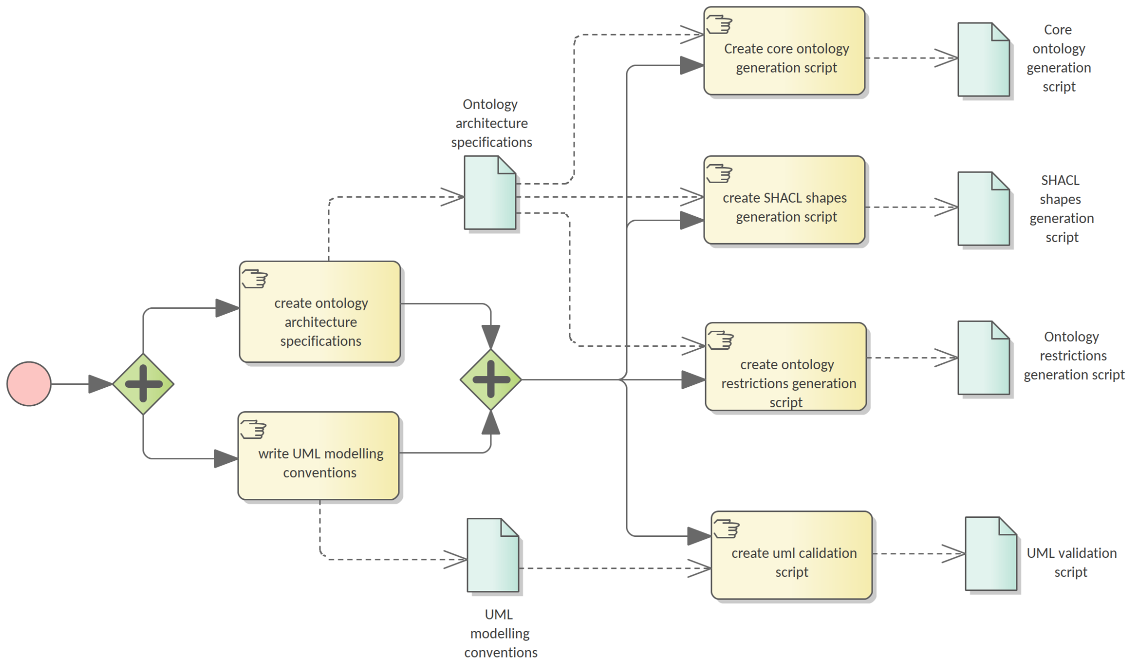


Figure 2: Creation of the specifications documents and the UML transformation scripts

conventions.

The ontology architecture specification (this document) serves, at large, as requirement specifications for the development of three XSLT scripts to generate the formal ontology. These scripts can be developed independent of each other as they refer to different aspects of the formal ontology as described in Section **??**.

The input for these scripts is the UML conceptual model serialised in XMI 2.5.1 format [1].

3.2 Conceptual model revision

The conceptual model revision is an iterative process. The validation script is execution outputs a report indicating if there are any deviations from the conventions, and detailing which are they and eventually what are the necessary actions to resolve them.

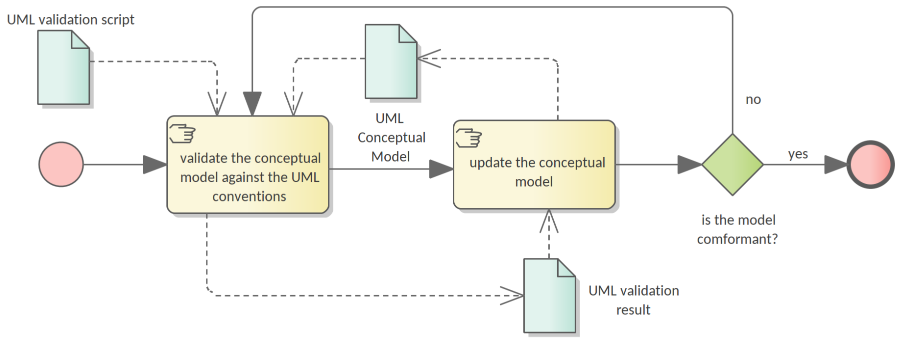


Figure 3: Adjustment of the UML conceptual model guided by the validation script

3.3 Formal ontology generation

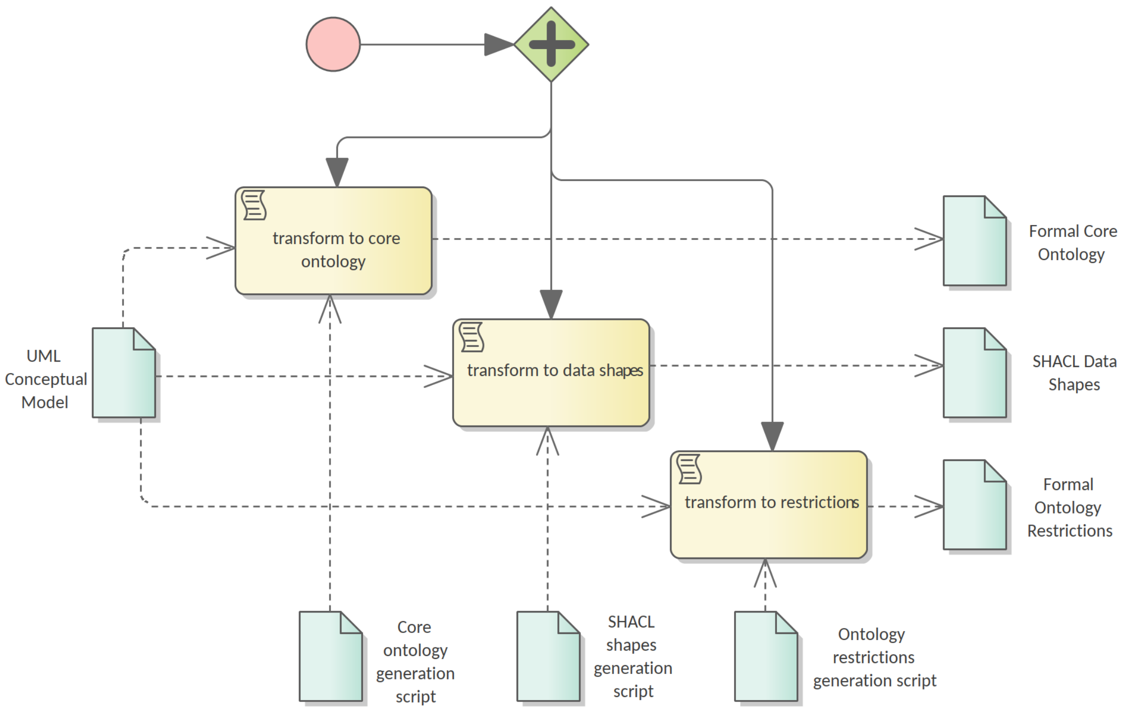
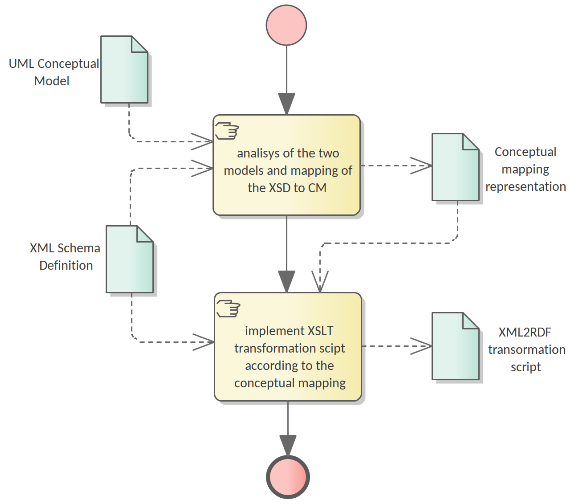
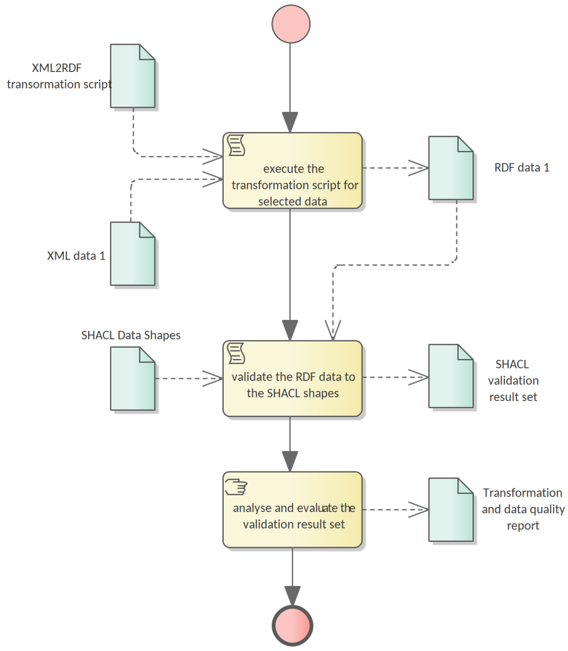


Figure 4: Generation of the formal ontology from the UML conceptual model



(a) Implementation of the XML transformations script based on the mappings from XSD schemas onto the conceptual model



(b) Transformation of the existent eProcurement XML data into RDF representation and validating the end result conformance

Figure 5: XML to RDF data transformation

3.4 XML to RDF data transformation

4 Architectural considerations

4.1 Separation of concerns

The successful application of an ontology or the development of an ontology-based system depends not just on building a good ontology but also on fitting this into an appropriate development process. All computing information models suffer from a semantic schizophrenia. On the one hand, the model represents the domain; on the other hand, it represents the implemented system, which then represents the domain. These different representation requirements place different demands upon its structure [23].

One of the common ways to manage this problem is a separation of concerns. OMG's Model Driven Architecture (MDA) [29] is a well documented structure where a model is built for each concern and this is transformed into a different model for a different concern.

Transformation deals with producing different models, viewpoints, or artefacts from a model based on a transformation pattern. In general, transformation can be used to produce one representation from another, or to cross levels of abstraction or architectural layers [28].

The process described in Section 3 adopted some of these principle and employs model transformation to achieve the project objectives.

4.2 Semantics

Users of OWL [22] can actually select between two slightly different semantics: *direct semantics* that corresponds to the Description Logics (DL) [3], and *RDF-based semantics* that is based on translation of the OWL axioms into directed graphs. In this document we assume by default the direct semantics. In particular cases (i.e. SPARQL entailments and SHACL data shapes) RDF-based semantics is adopted and is explicitly mentioned in the document.

Description logics provide a concise language for OWL axioms and expressions. DLs are characterised by their expressive features. The description logic that supports all class expressions with $>, \bot, \sqcap, \sqcup, \neg, \exists$ and $\forall$ is known as $\mathcal{ALC}$ (which originally used to be an abbreviation for Attribute Language with Complement). For a formal introduction into DL please consult Baader et al. [3].

Inverse properties are not supported by $\mathcal{ALC}$, and the DL we have introduced above is actually called $\mathcal{ALCI}$ (for $\mathcal{ALC}$ with inverses) [19]. Many description logics can be defined by simply listing their supported features. We will use this notation when discussing degrees of expressivity for the ontology layers in Section **??**.

Computing all interesting logical conclusions of an OWL ontology can be a challenging problem, and reasoning is typically multi-exponential or even undecidable. To address this problem, the recent update OWL 2 of the W3C standard [25, 22] introduced three profiles: *OWL EL*, *OWL RL*, and *OWL QL*. These lightweight sublanguages of OWL restrict the available modelling features in order to simplify reasoning. This has led to large improvements in performance and scalability, which has made the OWL 2 profiles very attractive for practitioners [19].

On the other hand, the validation data shapes are expressed using Shapes Constraint Language (SHACL) [18]. Its semantics is based on RDF graphs but full RDFS inferencing is not required. SHACL processors may operate on RDF graphs that include RDF entailments [24] and SPARQL specific entailments [21]. The entailment regime specifies conditions that address the fourth condition on extensions of basic graph pattern matching [17, 24].

This architecture delimits different concerns in Section **??** in a stack of layers and assigns levels of expressivity to each of the layers in Section **??**.

4.3 Layering

5 Final word

Bibliography

[1] Xml metadata interchange (xmi) specification: Version 2.5.1. Standard formal/2015-06-07, Object Management Group (OMG), 2015. URL `http://www.omg.org/spec/XMI/2.5.1`.

[2] P. Archer, L. Bargiotti, M. D. Keyzer, S. Goedertier, N. Loutas, and F. V. Geel. Report on high-value datasets from eu institutions. Deliverable SC17DI06692, European Commission , 2014.

[3] F. Baader, I. Horrocks, and U. Sattler. Description logics. In *Handbook on ontologies*, pages 3–28. Springer, 2004.

[4] G. Booch, J. Rumbaugh, and I. Jacobson. *Unified Modeling Language User Guide, The (2nd Edition) (Addison-Wesley Object Technology Series)*. Addison-Wesley Professional, 2005. ISBN 0321267974.

[5] S. Cook, C. Bock, P. Rivett, T. Rutt, E. Seidewitz, B. Selic, and D. Tolbert. Unified modeling language (UML) version 2.5.1. Standard formal/2017-12-05, Object Management Group (OMG), Dec. 2017. URL `https://www.omg.org/spec/UML/2.5.1`.

[6] E. Costetchi. eProcurement uml conceptual model conventions. Recommendation, Publications Office of the European Union, April 2020.

[7] E. Costetchi. eProcurement uml conceptual model to owl ontology transformation. Recommendation, Publications Office of the European Union, April 2020.

[8] M. Dekkers, E. Stani, and F. Barthelemy. D02.02 - project charter proposal. Deliverable SC378DI07171, Publications Office of the European Union, 2017.

[9] M. Dekkers, E. Stani, B. Wyns, and F. Barthelemy. D02.01 - specification of the process and methodology to develop the eprocurement ontology with initial draft of the eprocurement ontology for 3 use cases. Deliverable SC378DI07171, Publications Office of the European Union, 2017.

[10] European Parliament and the Council. Directive 2013/37/eu of the european parliament and of the council of 26 june 2013 amending directive 2003/98/ec on the re-use of public sector information text with eea relevance. *OJ*, L 175: 1–8, 2013. CELEX:32013L0037.

[11] European Parliament and the Council. Directive 2014/23/eu of the european parliament and of the council of 26 february 2014 on the award of concession contracts text with eea relevance. *OJ*, L 94(2014/23/EU):1–64, 2014. CELEX:32014L0024.

[12] European Parliament and the Council. Directive 2014/24/eu of the european parliament and of the council of 26 february 2014 on public procurement and repealing directive 2004/18/ec text with eea relevance. *OJ*, L 94:65–242, 2014. CELEX:32014L0024.

[13] European Parliament and the Council. Directive 2014/25/eu of the european parliament and of the council of 26 february 2014 on procurement by entities operating in the water, energy, transport and postal services sectors and repealing directive 2004/17/ec text with eea relevance. *OJ*, L 94:243–374, 2014. URL `http://data.europa.eu/eli/dir/2014/25/oj`. CELEX:32014L0025.

[14] European Parliament and the Council. Directive 2014/55/eu of the european parliament and of the council of 16 april 2014 on electronic invoicing in public procurement text with eea relevance. *OJ*, L 133:1–11, 2014. CELEX:32014L0055.

[15] European Parliament and the Council. Directive (eu) 2019/1024 of the european parliament and of the council of 20 june 2019 on open data and the re-use of public sector information. *OJ*, L 172:56–83, 2019. CELEX:32019L1024.

[16] T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies*, 43(5-6):907–928, 1995.

[17] P. Hayes. RDF semantics. W3C recommendation, W3C, Feb. 2004. http://www.w3.org/TR/2004/REC-rdf-mt-20040210/.

[18] H. Knublauch and D. Kontokostas. Shapes constraint language (SHACL). W3C recommendation, W3C, July 2017. https://www.w3.org/TR/2017/REC-shacl-20170720/.

[19] M. Krötzsch. Owl 2 profiles: An introduction to lightweight ontology languages. In *Reasoning Web International Summer School*, pages 112–183. Springer, 2012.

[20] N. Loutas, N. Loutas, S. Kotoglou, and D. Hytiroglou. D04.07 - report on policy support for eprocurement. Deliverable SC245DI07171, ISA programme of the European Comission, 2016.

[21] C. Ogbuji and B. Glimm. SPARQL 1.1 entailment regimes. W3C recommendation, W3C, Mar. 2013. http://www.w3.org/TR/2013/REC-sparql11-entailment-20130321/.

[22] B. Parsia, P. Patel-Schneider, and B. Motik. OWL 2 web ontology language structural specification and functional-style syntax (second edition). W3C recommendation, W3C, Dec. 2012. http://www.w3.org/TR/2012/REC-owl2-syntax-20121211/.

[23] C. Partridge, A. Mitchell, and S. de Cesare. Guidelines for developing ontological architectures in modelling and simulation. In *Ontology, Epistemology, and Teleology for Modeling and Simulation*, pages 27–57. Springer, 2013.

[24] P. Patel-Schneider and P. Hayes. RDF 1.1 semantics. W3C recommendation, W3C, Feb. 2014. http://www.w3.org/TR/2014/REC-rdf11-mt-20140225/.

[25] P. Patel-Schneider, B. Parsia, and B. Motik. OWL 2 web ontology language structural specification and functional-style syntax. W3C recommendation, W3C, Oct. 2009. http://www.w3.org/TR/2009/REC-owl2-syntax-20091027/.

[26] Publications Office of the European Union. Interinstitutional style guide 2011, 2011. doi: doi:10.2830/36616.

[27] PwC EU Services. e-government core vocabularies handbook. Report, ISA programme of the European Comission, 2015. URL `https://ec.europa.eu/isa2/library/e-government-core-vocabularies-handbook_en`.

[28] J. Siegel. Object management group model driven architecture (mda) mda guide rev. 2.0. 2014. URL `http://www.omg.org/cgi-bin/doc?ormsc/14-06-01`.

[29] R. Soley et al. Model driven architecture. *OMG white paper*, 308(308):5, 2000.

[30] S. A. White. Introduction to bpmn. *Ibm Cooperation*, 2(0):0, 2004.