Publications Office
of the European Union

Initiative for digital transformation in the Metadata and Reference Data Sector of the Publications Office of the European Union

# Asset Publication Lifecycle Enterprise Architecture

# Disclaimer

The views expressed in this report are purely those of the Author(s) and may not, in any circumstances, be interpreted as stating an official position of the European Union. The European Union does not guarantee the accuracy of the information included in this study, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Union.

**This report was prepared for the Publications Office of the European Union by Infeurope.**

## Document metadata

| | |
|---|---|
| **Reference** | WP 1.1: Asset Publication Lifecycle Enterprise Architecture |
| **Corporate Author** | Publications Office of the European Union |
| **Author** | Eugeniu Costetchi |
| **Reviewers** | Denis Dechandon and Willem Van Gemert |
| **Contractor** | Infeurope S.A. |
| **Framework contract** | 10688/35368 |
| **Work package** | WP 1.1 |
| **Delivery date** | 25 September 2020 |
| **Suggested readers** | sector management, business and technical staff, stakeholders, enterprise architects, software developers, project partners |

© European Union, 2020

# Abstract

Public procurement is undergoing a digital transformation. The EU supports the rethinking of the 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.

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.

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.

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.

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

# 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 reference asset management lifecycle materialised as the publishing workflow used by the Standardisation Unit (SU) at the Publications Office of the European Union (PO).

In this document we (a) establish the baseline architecture, supported by the strategic and motivational information; (b) develop a target architecture guiding the transitional processes of implementing new technologies. This constitutes a natural evolution in response to changing mission needs defined by the SU management, and also takes into consideration the strategic directions proposed by PO and European Commission (EC) and European Parliament (EP).

## 1.1 Background considerations

Given the increasing importance of data standards for the European institutions, 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 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 reference data and models, hence allowing data from different sources to be easily accessed and linked, and consequently reused.

PSI directive [11] 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. The reference data maintained and published by the PO have 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.

In this context the Publications Office of the European Union maintains and publishes an ever growing number of reference data assets vital in the context of inter-institutional information exchange. With regards to reference data, PO provides an ever growing number of services to the main institutional stakeholders and with the aim to extend them to a broader public; enabling active or passive participation in the reference data life cycle, standardisation and harmonisation.

## 1.2 EU trajectory towards semantic and linked data

European institutions set sail to adopt Semantic Web and Linked Data technologies as part of the vision to become data centric e-government bodies [cite find directives].

Many of the legacy systems in the institutions use XML data format for exchange and document formats governed by the XSD schemas. The aim is to evolve so that the existing and the new systems are capable to operate with semantic data representations using RDF [cite], OWL [cite], SHACL [cite] and other representations and serialised at least in RDF/XML [cite], Turtle [cite] and JSON-LD [cite] formats.

For this reason, the PO is already publishing the data in RDF format for over a decade using Cellar repository [cite]. And the SU, in particular, is committed to publish and disseminate the reference data in semantic formats. Next we describe the state of affairs of the SU to describe the context of the current work.

## 1.3 Target audience

The target audience for this document comprises the following groups of stakeholders:

- Management of the SU
- Enterprise architects and data governance specialists
- Documentalists involved in the reference data lifecycle
- Technical staff in charge of operating the workflow components
- Developers in charge of the workflow implementation

## 1.4 Document scope

This document aims to support SU in the transition towards the semantic technologies focusing on the architecture of the publishing workflow. The central use case is to support the asset management lifecycle presented in Section (below). It includes managing the incoming requests, editing the reference assets in VocBench3 system, then exporting the RDF data and passing them as input to a set of processes that validate, assess, transform, package and finally publish the assets in Cellar, the main dissemination platform.

This document will provide a motivational, business and application account of the workflow. Each of these accounts is limited strictly to the success scenario of the above mentioned use case and does not include possible extensions and variations which may be.

# 2 Publication workflow digital transformation

## 2.1 State of play

The SU publishes reference data in several formats, most important being XML, XSD and RDF/XML. On the technological side, the SU currently employs a legacy custom-built system for controlling and executing, in part, the asset management lifecycle operations (legacy workflow system). The system was developed using a mixture of XSLT technology, Perl and Bash scripting languages. The system was developed to execute a wide variety of conversions and transformations based on XML source files into various other formats including human readable documents.

The source data representation (XML in this case) has the primary role to serve as the only source of truth, and additionally, maintaining non-redundancy and rich expressivity. All other data forms and representations are secondary and are generated by transformation and conversion processes from the source representation.

One peculiarity of the legacy system setup is that the editing of the asset content is performed using Microsoft Excel. This is done by transformation of the content from XML representation into Excel stylesheets, which are edited by the SU documentalists, and then the stylesheets are converted back into XML form. This way a circular transformation is achieved which also serves as an integrity checking and validation mechanism. In addition XSD schema definitions are used to validate the XML source representation.

The legacy workflow system uses the file system for data persistence. In addition, this functionality is aided by a version controlling system, SVN [cite], to trace the temporal evolution of data.

Some steps in the legacy workflow are automated. The automation is based on cron tasks and SVN hooks that, upon changes in the source XML or Excel files trigger a set of conversion mechanisms. Some other steps require manual triggering and eventually parameterization intervention. The execution of the automated steps often requires assistance of technical staff or IT skills above average which represents an impediment for the non technical documentalists and a hindrance for the IT staff.

Moreover, the maintenance of this system is burdened by a technical debt that accumulated over time, because the system evolved organically based on ever flowing requests.

## 2.2 Towards semantic technology workflow

SU's mission regarding the technological evolution is to migrate towards Semantic Web and Linked Data technologies and representations. The maintenance of the reference data is currently done based on XML source representation and the desired transition is towards RDF based representation. For that purpose Excel and XML sources are no longer suitable and a dedicated editor is necessary.

To solve this issue, SU took the development flagship of the VocBench3[cite] system - a web-based, multilingual, vocabulary editing tool based on the SKOS [cite] model, which is built on top of RDF/S standard. Later, VocBench3 was developed to support authoring of RDFS and OWL vocabularies.

Switching to RDF-based sources and adoption of VocBench3 system implies a technological and business process disruption. The main reason being the legacy workflow system, which operates with XML based sources only and does not support RDF sources. RDF representation being only a by-product derived from XML.

VocBench3 naturally adopted a persistence based on triple stores, which are NoSQL database systems implementing the directed graph data model instead of the hierarchical or relational model of data. The relational data model is mentioned here because the Excel worksheets are based on the tabular data organisation; the hierarchical data model is mentioned because the XML is fundamentally a hierarchically organised data structure; and each of them is only partially compatible with the graph paradigm present in semantic data models.

Migration towards a new workflow that integrated VocBench3 thus requires reconciliation between file-system and database approaches to persistence. Also a paradigmatic transition to graph based data representation, from the hierarchical models of source representation, and tabular models used for source authoring is necessary.

The legacy workflow system is also lacking in validation, structural analysis (fingerprinting) and content comparison capabilities (calculating the difference between two versions of an asset). A transition would imply development of at least these new capabilities in order to maintain the current business processes.

# 3 Architecture building blocks

## 3.1 Methodology

In this document we take an enterprise architecture perspective and aim to provide several architecture aspects which are necessary and sufficient to describe the publishing process.

In developing this architecture, we adopt parts of the TOGAF methodology, which is a framework for enterprise architecture that provides an approach for designing, planning, implementing, and governing an enterprise information technology architecture.

For the architecture representation, we adopt ArchiMate language, which is an open and independent enterprise architecture modeling language to support the description, analysis and visualization of architecture within and across business domains in an unambiguous way [what is archimate]. The motivation view is developed by implementing the following steps. First, TOGAF & ArchiMate were chosen as frameworks. Then some interviews were conducted with the SU management, having as a goal to elicit who are the main stakeholders and their motivations. The interview notes have been distilled and organised in ArchiMate diagrams (using EnterpriseArchitect Tool for drawing the models). Finally this report was written presenting the resulting motivation structure.

# 4 Motivation architecture

# 5 Business architecture

# 6 Application architecture

# 7 Technical architecture

# 8 Conclusions

# 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] 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.

[4] S. O. Bradner. Key words for use in RFCs to Indicate Requirement Levels. RFC 2119, Mar. 1997. URL `https://rfc-editor.org/rfc/rfc2119.txt`.

[5] D. Brickley and R. Guha. RDF schema 1.1. W3C recommendation, W3C, Feb. 2014. http://www.w3.org/TR/2014/REC-rdf-schema-20140225/.

[6] S. Cirulli. Xspec v0. 5.0. *XML LONDON 2017*, 2017.

[7] 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`.

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

[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.

[11] 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.

[12] 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.

[13] 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.

[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] M. Kay. XSL transformations (XSLT) version 3.0. W3C recommendation, W3C, June 2017. https://www.w3.org/TR/2017/REC-xslt-30-20170608/.

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

[17] M. Krötzsch, I. Horrocks, M. Smith, and B. Glimm. OWL 2 web ontology language conformance (second edition). W3C recommendation, W3C, Dec. 2012. http://www.w3.org/TR/2012/REC-owl2-conformance-20121211/.

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

[19] E. Rescorla, R. Barnes, and S. Kent. Further Key Words for Use in RFCs to Indicate Requirement Levels. RFC 6919, Apr. 2013. URL `https://rfc-editor.org/rfc/rfc6919.txt`.