Realising a Federation of Repositories of Reusable Metadata

Nikolaos Loutas, Stijn Goedertier, Christophe Colas,
Michiel De Keyzer, Debora Di Giacomo, Joao Rodrigues Frade
PwC EU Services
firstname.lastname@pwc.be

Makx Dekkers
AMI Consult
makx@makxdekkers.com

Vassilios Peristeras
EC, Directorate-General for Informatics, Interoperability Solutions for European Public Administrations
vassilios.peristeras@ec.europa.eu

Abstract

Semantic assets and the agreements associated with them are essential elements for organisations to understand the meaning of the information they exchange – without them this information would be of little use. In order to facilitate the access of public administrations in Europe to reusable semantic assets, the Interoperability Solutions for European Public Administrations Programme of the European Commission (ISA Programme) has been running for the last 3 years an action on syndicating content from different semantic asset repositories and making it available through a single point of access. In this paper we present the current state of the federation of semantic asset repositories on Joinup, namely a set of online collections of semantic assets maintained by public administrations, standardisation organisations and businesses, which currently counts more than 1500 semantic assets from 21 partner organisations, described using the Asset Description Metadata Schema (ADMS).

Keywords: ADMS; semantic asset; semantic interoperability; e-Government; federation.

1. Introduction

The EU Digital Agenda (European Commission, 2010a) identifies the lack of interoperability between electronic public services as one of the seven major obstacles to the Digital Single Market (European Commission, 2010b). One facet of interoperability – the focus of this paper – is semantic interoperability. The European Interoperability Framework (European Commission, 2010c) defines it as "the ability of organizations to process information from external sources in a meaningful manner, such that the precise meaning of exchanged information is understood and preserved throughout exchanges between parties".

Semantic interoperability requires consensus on data models, schemas and reference data – henceforth referred to as semantic assets – for data exchange to happen without semantic conflicts (Peristeras et al., 2008). Despite their importance, semantic assets are not easily discoverable on the Web, e.g. via traditional search engines, and metadata about their meaning is seldom available. Navigating on the websites of the different publishers of semantic assets and browsing available semantic assets is not efficient either. As a result, ICT projects tend to reinvent the wheel and redefine assets which are already available. This results in fragmentation, redundancy and unnecessary waste of resources in the development of Information Systems and interoperability conflicts when users of different semantic assets need to interconnect their Information Systems to exchange information. Ultimately this hampers semantic interoperability.

In this vein, numerous efforts originating in different domains have kicked off. Hundreds of ontologies from Health and Life Sciences have been registered in BioPortal (The National Center for Biomedical Ontology, 2011). The Metadata Registry of the Publication Office (OP) of the European Commission makes available a number of Named Authority Lists (NAL). NIEM

makes it possible for US organisation and public administrations to share critical data. Other related work includes the eGov-Share project, which has led to a CEN Workshop Agreement (CEN, 2009), and the Ontology Metadata Vocabulary (OMV) (Hartmann, et al., 2005) focusing on describing general e-Government resources and ontologies respectively. Finally, Linked Open Vocabularies (LOV) provides access to a collection of more than 300 reusable vocabularies.

In order to facilitate the access of public administrations in Europe to reusable semantic assets, the ISA Programme has been running for the last 3 years an action on syndicating content from different semantic asset repositories and making it available through a single point of access on Joinup. To make this possible, the ISA Programme has created a vocabulary to describe semantic assets, the Asset Description Metadata Schema (ADMS Working Group, 2012).

In the remainder of this paper we discuss the current state of the – so called – ADMS-based federation of semantic asset repositories on Joinup, namely a set of online collections of semantic assets maintained by public administrations, standardisation organisations and businesses.

2. The Asset Description Metadata Schema

ADMS is a standardised vocabulary which aims at helping publishers of semantic assets to document what their assets are about (their name, their status, theme, version, etc) and where they can be found on the Web (ADMS Working Group, 2012). ADMS was developed engaging a working group of experts from EU Member States and Institutions, industry and academia, and following a standardised process and methodology.

ADMS descriptions can then be published on different websites while the asset itself remains on the website of its publisher. ADMS embraces the multi-publisher environment and, at the same time, it provides the means for the creation of aggregated catalogues of semantic assets and single points of access to them based on ADMS descriptions.The primary classes of ADMS are semantic asset repository, semantic asset and distribution.

A semantic asset repository is a system or service that provides facilities for storage and maintenance of descriptions of semantic assets and distributions, and functionality that allows users to search and access these descriptions.

A semantic asset is an abstract entity that reflects the intellectual content of the asset and represents those characteristics of the asset that are independent of its physical embodiment. Ontologies, schemata, domain models, controlled lists, taxonomies and reference datasets may all be examples of semantic assets.

A distribution represents a particular physical embodiment of a semantic asset. A distribution is typically a downloadable computer file (but in principle it could also be a paper document) that implements the intellectual content of an asset.

ADMS is doing for semantic assets what Really Simple Syndication (RSS) has done for Web resources and, more in particular, for the publication of news on the Web by the many news publishers. Similar to RSS, once the ADMS description is created in RDF it can be published on the Web and understood by content aggregators everywhere. The syndication of semantic asset descriptions will improve their visibility and discoverability. As semantic assets become more visible and discoverable, more projects are likely to reuse them. This will improve interoperability as Information Systems will use similar semantic assets at interface level. The publisher of the semantic asset will benefit from a larger user base (Frade et al., 2012). Preliminary work towards this direction was carried out by Shukair et al. (2013) and Shukair et al. (2012). ADMS v1.00 was contributed to W3C's Government Linked Data (GLD) Working Group, which in turn published it as a W3C WG Note (2013).

3. Federation of Semantic Asset Repositories

In January 2013 the European Commission has launched the ADMS-based federation of semantic asset repositories on Joinup, namely an online aggregation service to make it easier for public administrations to find and reuse semantic assets (see FIG. 1). This service aggregates descriptions of more than 1500 semantic assets, described using ADMS, received from 21 partner organisations summarised in Table 1. The service, which features a simple and an advanced faceted search, will increase the visibility of semantic assets described with ADMS on the Web. This will also stimulate the reuse of semantic assets by national and pan-European initiatives, thus leading to cost savings.

TABLE 1: Organisations part of the ADMS-enabled federation

Type of organisation	Repository name
Public Administrations of EU Member States	BE - Belgian Interoperability Catalogue
	DE – Xrepository
	DK - Digitaliser.dk
	EE – RIHA
	FI - Yhteentoimivuus.fi
	GR - Greek Interoperability Catalogue
	ES - CENATIC - National Reference Centre of ICT applications based on open source
	NL - Dutch Standardisation Forum - "Comply or explain"-standards
	CZ - Information system of data elements
EU Institutions	EU Semantic Interoperability Catalogue
	EU Publications Office - Metadata Registry
	INSPIRE interoperability assets
Standardisation bodies	CEN - European Committee for Standardisation
	DCMI - Dublin Core Metadata Initiative
	GS1 in Europe eDox
	W3C Standards and Technical Reports
Other organisations	Internal Commission on Civil Status (CIEC/ICCS)
	Linked Open Vocabularies (LOV)
	ListPoint
	Wolters Kluwer Vocabularies
	ESD Standards

For an organisation to partner with the federation and make their semantic assets accessible via the ADMS-based federation the following steps have to be taken:

1. *Select semantic assets.* The organisation should first identify the semantic assets that they want to share and analyse them. It is highly likely that only one part of their collection will qualify for being reusable in a cross-sector and/or cross-border setting. The semantic assets with a high-reuse potential should be the ones to be made searchable via the federation.
2. *Create ADMS-RDF metadata descriptions.* The organisation should then describe the selected semantic assets following the ADMS v1.00 specification. Table 2 shows a part of the ADMS-RDF description of the Corporate Bodies NAL of the OP. Different options exist for developing the ADMS-RDF metadata descriptions:
    a. Organisations that wish to share a large number of semantic assets (e.g. hundreds) are advised to develop exporters that translate their native metadata format to ADMS. W3C and Digitaliser.dk have developed such exporters; or
    b. Organisations that have only few semantic assets to describe (e.g. a few tenths) can develop the ADMS-RDF metadata descriptions manually. In order to facilitate this, the ISA Programme (2012a) has made available spreadsheet template that follows the ADMS specification. Organisations can use the spreadsheet

template to describe their semantic assets and then create ADMS-RDF from it using the Open Refine RDF extension.

3. *Validate the ADMS-RDF*. After creating the ADMS-RDF metadata descriptions, we recommend checking their validity (in terms of completeness and conformance to the specification). An ADMS-RDF validator has been developed by the ISA Programme (2012b) and is freely available.
4. *Publish the ADMS-RDF metadata descriptions on Joinup.* The final step is to publish the ADMS-RDF metadata descriptions of semantic assets on Joinup, thus making them discoverable and retrievable through the ADMS-based federation. Two option exists for publishing the metadata descriptions on Joinup:
    a. Upload the ADMS-RDF file directly on Joinup; or
    b. Upload the ADMS-RDF file on a Web directory and configure Joinup's harvester, so that it can harvest the file on the remote location.

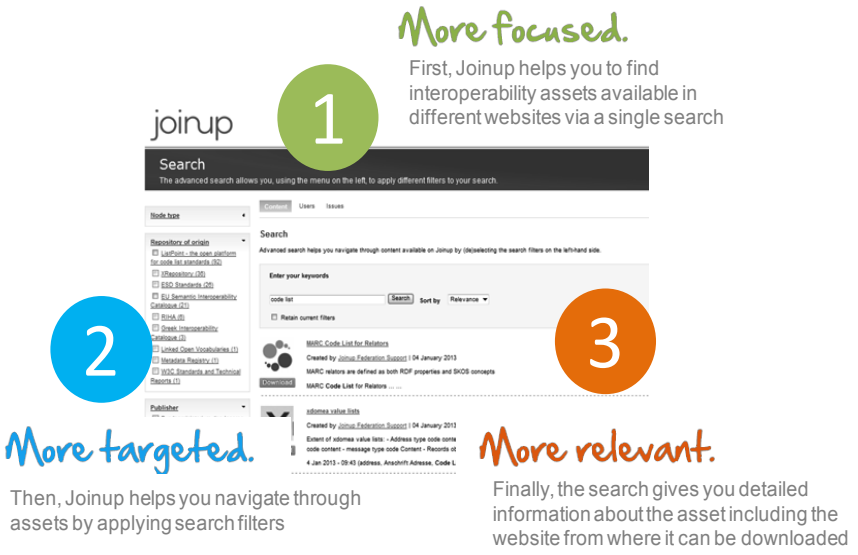

FIG. 1. Featuring the ADMS-based federation on Joinup. Available at https://joinup.ec.europa.eu/catalogue/all

TABLE 2: Example of the ADMS-RDF description of the Corporate Bodies NAL

<rdf:type rdf:resource="http://www.w3.org/ns/adms#SemanticAsset"/>
<adms:contactPoint rdf:resource="http://publications.europa.eu/mdr/contact.html"/>
<adms:interoperabilityLevel rdf:resource="http://purl.org/adms/interoperabilitylevel/Semantic"/>
<adms:metadataDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2013-03-07T00:00:00+0100</adms:metadataDate>
<adms:metadataLanguage rdf:resource="http://id.loc.gov/vocabulary/iso639-1/en"/>
<adms:metadataPublisher rdf:resource="http://publications.europa.eu/resource/authority/corporate-body/PUBL"/>
<adms:status rdf:resource="http://purl.org/adms/status/Completed"/>
<dcterms:temporal rdf:resource="http://publications.europa.eu/resource/authority/corporate-body/PeriodOfTime/1"/>
<dcterms:type rdf:resource="http://purl.org/adms/assettype/NameAuthorityList"/>
<dcterms:alternative xml:lang="en">Corporate bodies NAL</dcterms:alternative>
<dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2012-03-23T00:00:00+0100</dcterms:created>
<dcterms:description xml:lang="en">The Corporate bodies named authority list (NAL) or Common Authority Table (CAT) is a controlled vocabulary listing corporate entities such as European institutions and bodies with their authority code and label(s) in the 23 official languages of the EU (when available). The list covers as well subdivisions of the different institutions such as DG's. The Corporate bodies NAL is part of the Core Metadata (CM) used in the data exchange between the institutions involved in the legal decision making process and the Publications Office of the EU. The NAL is under governance of the Interinstitutional Metadata Maintenance Committee (IMMC) and maintained by the Publications Office of the EU in its Metadata Registry (MDR)</dcterms:description>

4. Conclusions and Future Work

Implementing the ADMS-based federation of semantic asset repositories comprises a social and a technical process. Coordinating the two processes has up to now been crucial for the success of this undertaking.

The social process entails challenges related to:

- Reaching out to and identifying potential partner repositories in Europe and beyond.
- Coordinating with the partner organisations and getting their buy-in.
- Understanding asset licensing, sharing and reuse conditions.
- Deciding on and managing the lifecycle of the metadata in the federation.

The main technical challenges include:

- Technical capacity building in the partner organisations so that they can create the metadata descriptions of their assets (see also step 2 in section 3).
- Cleansing, modelling and harmonising the metadata received from partner organisations, e.g. harmonising data types, replacing free text values with terms coming from controlled vocabularies, completing missing mandatory attributes, assigning Uniform Resource Identifiers (URIs), transforming the metadata into RDF.

The ISA Programme will continue to promote the adoption of ADMS by publishers of semantic assets so that a greater number of reusable semantic assets from a broader range of partner organisations would be searchable via Joinup.

References

ADMS Working Group. (2012, January). The Asset Description Metadata Schema (ADMS) - Specification. Retrieved April 04, 2013, from http://joinup.ec.europa.eu/asset/adms/release/all

CEN. (2009). CEN Workshop Agreement on 'Discovery of and Access to eGovernment Resources' CWA 15971:2009 part 1 to 6. Retrieved April 04, 2013, from http://www.cen.eu/CEN/sectors/sectors/isss/workshops/Pages/wsegovshare.aspx

European Commission. (2010a). Digital Agenda for Europe 2010 - 2020. Retrieved April 04, 2013, from European Commission: http://ec.europa.eu/digital-agenda/

European Commission. (2010b). Pillar I: Digital Single Market. Retrieved April 04, 2013, from http://ec.europa.eu/digital-agenda/en/our-goals/pillar-i-digital-single-market

European Commission. (2010c). European Interoperability Framework (EIF) for European public services. Retrieved April 04, 2013, from http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

Hartmann, J., Palma, R., Sure, Y., Haase, P., & del Carmen Suárez-Figueroa, M. (2005). OMV -- Ontology Metadata Vocabulary. Chris Welty, ISWC 2005 - In Ontology Patterns for the Semantic Web.

ISA Programme (2013), Process and methodology for developing semantic agreements. Retrieved July 03, 2013, from https://joinup.ec.europa.eu/community/core_vocabularies/document/process-and-methodology-developing-semantic-agreements

ISA Programme (2012a). Generate ADMS asset descriptions from a spreadsheet with Google Refine RDF. Retrieved April 04, 2013, from https://joinup.ec.europa.eu/asset/adms/document/generate-adms-asset-descriptions-spreadsheet-refine-rdf

ISA Programme (2012b) ADMS Validator for the Joinup Application Profile. Retrieved April 04, 2013, from https://joinup.ec.europa.eu/svn/adms/ADMS_validator/

Frade J.M, Goedertier S., Di Giacomo D., Peristeras V., Loutas N.(2012). Building Semantic Interoperability through the Federation of Semantic Asset Repositories. In I-SEMANTICS 2012 – 8th Int. Conference on Semantic Systems, Graz, Austria.

Linked Open Vocabularies (LOV). Retrieved April 04, 2013, from http://lov.okfn.org/dataset/lov/

Peristeras, V., Loutas, N., Goudos, S. K., & Tarabanis, K. A. (2008). A conceptual analysis of semantic conflicts in pan-European e-government services. J. Information Science , 34 (2), 877-891.

Shukair G. Loutas N., Peristeras V. (2012). Integrating Linked Metadata Repositories into the Web of Data. In Proceedings of COLD 2012 – 11th International Semantic Web Conference, Boston, USA.

Shukair G. Loutas N., Peristeras V., Skarlss, S. (2013). Towards semantically interoperable metadata repositories: The Asset Description Metadata Schema. In Computers in Industry, vol. 64 (1), pp. 10–18.

The National Center for Biomedical Ontology. (2011). Bioportal. Retrieved April 04, 2013, from http://bioportal.bioontology.org/

W3C Government Linked Data Group (2013). Asset Description Metadata Schema (ADMS), Working Group Note. Retrieved July 03, 2013, from http://www.w3.org/TR/vocab-adms/