Building Semantic Interoperability through the Federation of Semantic Asset Repositories

João Rodrigues Frade
PwC EU Services
Woluwedal 18
B-1932 Brussels, Belgium
joao.frade@pwc.be

Stijn Goedertier
PwC EU Services
Woluwedal 18
B-1932 Brussels, Belgium
stijn.goedertier@pwc.be

Vassilios Peristeras
European Commission, Directorate-General for Informatics,
Interoperability Solutions for European Public Administrations
Rue Montoyer 34
B-1000 Brussels, Belgium
vassilios.peristeras@ec.europa.eu

Debora Di Giacomo
PwC EU Services
Woluwedal 18
B-1932 Brussels, Belgium
debora.di.giacomo@pwc.be

Nikolaos Loutas
DERI, NUI Galway
IDA Business Park, Lower Dangan
Galway, Ireland
nikos.loutas@deri.org

ABSTRACT
According to the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission, interoperability relates to the ability of disparate organisations to interact towards mutually beneficial and agreed goals, involving the sharing of information and knowledge [1]. 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 this paper, semantic assets are defined as ontologies, data models, data dictionaries, code lists, XML and RDF schemas which are used for information exchange and that can be reused by implementers of Information Systems, in particular, as part of machine-to-machine interfaces. However, field research has shown that developers, practitioners and researchers working in the field of semantic interoperability in eGovernment tend to reinvent the wheel and semantic assets are rarely reused. In order to encourage and bootstrap the reuse of semantic assets in the EU and beyond, this paper introduces two important initiatives driven by the ISA Programme: the Asset Description Metadata Schema (ADMS) and the ADMS-enabled federation of semantic asset repositories on Joinup.

Categories and Subject Descriptors
D.2.12 [**Interoperability**]

General Terms
Standardization

Keywords
ADMS, Semantic Interoperability, Metadata management, e-Government, syndication of content, RDF, XML

I-SEMANTICS 2012, 8th Int. Conf. on Semantic Systems, Sept. 5-7, 2012, Graz, Austria

1. INTRODUCTION
The EU Digital Agenda [2] identifies the lack of interoperability between electronic public services as one of the seven major obstacles to the Digital Single Market [3]. One facet of interoperability – the focus of this paper – is semantic interoperability. The European Interoperability Framework [4] 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".

In the EU e-Government Action Plan Plan [5] semantic interoperability is put forward as "[…] an essential precondition for an open, flexible delivery Framework of e-Government services". It is a necessary – but not sufficient – condition for the electronic exchange and usage of information across the boundaries of information systems.

Semantic interoperability requires consensus on data models and reference data – henceforth referred to as semantic assets – for data exchange to happen without semantic conflicts. The unnecessary redundancy, variants, lack of awareness and management of semantic assets leads to increasing costs and overhead in the development and maintenance of Information Systems.

As they are enablers of interoperable information exchange, these assets are usually created, published and maintained by standardisation bodies. Nonetheless, ICT projects and groups of experts also create such assets. There are therefore many publishers of semantic assets with different degrees of formalism. Many of these assets are developed for a specific business domain, others are multi-purpose and cross sector. Some have a very well clear geographical context, others are international.

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 finding them is not straightforward, ICT projects tend to reinvent the wheel and recreate semantic 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 [6]. Bioportal is an instance of the Open Ontology Repository (OOR) Initiative. Other instances of OOR include Spatial Ontology Community of Practice (SOCoP), COLORE, Ontohub. However, unlike Bioportal, their success is questionable as only few ontologies have been registered in them [7].

The eXtended MetaData Registry (XMDR) focused on the development of improved standards and technology for storing and retrieving the semantics of data elements, terminologies, and concept structures in metadata registries [8]. XMDR aimed to Propose revisions to ISO/IEC 11179 Metadata Registry (MDR) Standard [9]. However, this effort is nowadays not active.

Other previous work includes the eGov-Share project, which has led to a CEN Workshop Agreement [10], and the Ontology Metadata Vocabulary (OMV) [11] focusing on describing general e-Government resources and ontologies respectively.

Experience has shown that obliging all publishers of semantic assets to make their semantic assets available on a centralised repository is in the majority of cases not a viable approach for two basic reasons: first, there are simply too many publishers of semantic assets; second, there is no organisation which can sanction them to do this. Syndication of content seems to be the most efficient way forward. To make this possible, the ISA Programme has created a vocabulary to describe semantic assets, the *Asset Description Metadata Schema* (ADMS) [12].

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 descriptions can then be published on different websites while the asset itself remains on the website of its publisher (a.k.a. syndication of content). 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.

ADMS will do 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 or XML 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.

2. ADMS: A STANDARD VOCABULARY TO DESCRIBE SEMANTIC ASSETS

ADMS [12] was conceived to overcome the barriers related to the discovery and sharing of semantic assets (see Figure 1).

According to the ADMS Working Group any of the following may be a semantic asset [12]:

- Syntax specifications, e.g. W3C Date and Time format [13], ISO 8601 [14];
- Code lists, e.g. language codes, country and region codes such as ISO 639 [15], ISO 3166 [16], NUTS [17];
- Controlled vocabularies or taxonomies, e.g. subjects, domains, document types, quality and security levels, such as Eurovoc [18], Agrovoc [19] and the UK Integrated Public Sector Vocabulary [20];
- Reference collections of organizations and persons, e.g. registers of government agencies, business registers, citizen databases;
- References to geographic locations and jurisdictions, e.g. cadastre, spatial planning, map information;
- Metadata specifications, standards and schemas, e.g. UK e-GMS [21], NL OWMS [22], INSPIRE [23]; and
- Mapping specifications, e.g. conversions across controlled vocabularies and metadata schemas.

The ADMS vocabulary is expressed in UML, RDF and XML Schema. The last two can be used for exchanging descriptions of semantic assets over the Web.

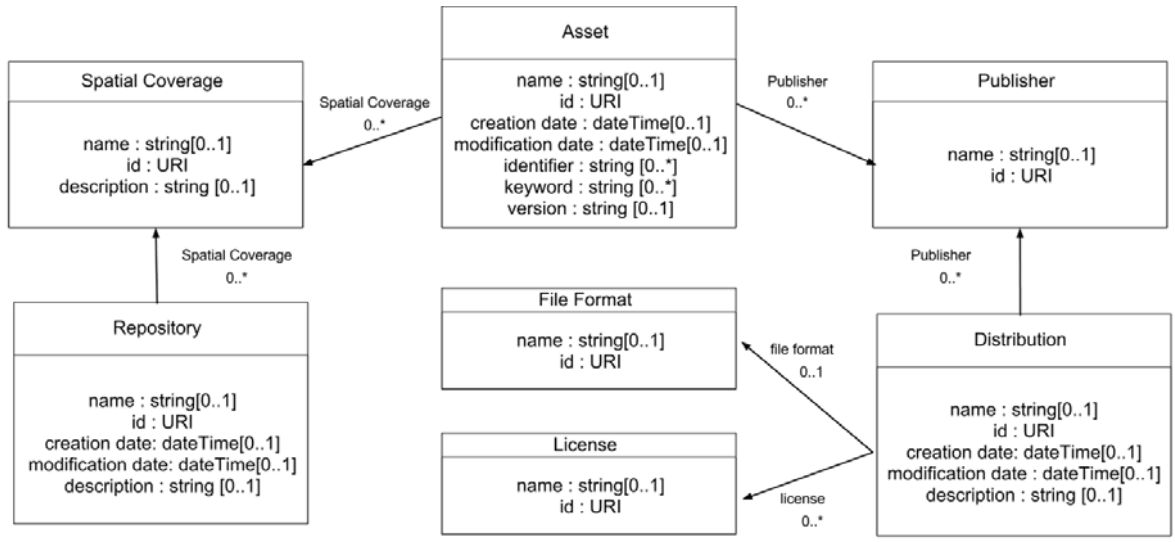

Figure 1 Core concepts of ADMS

2.1 ADMS RDF

ADMS is expressed in RDF Vocabulary Description Language, a XSLT was developed to render it into an HTML page for human readers.

The primary classes of ADMS are Semantic Repository, Semantic Asset and Semantic Asset Distribution. They are all subclasses of concepts defined in a generic model called RADion [24].

A Semantic Asset Repository is a system or service that provides facilities for storage and maintenance of descriptions of Semantic Assets and Semantic Asset 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

A Semantic Asset 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

The majority of terms used to express ADMS in RDF are reused from existing vocabularies, notably Dublin Core [25] and FOAF. Domain [26], and range restrictions have not been defined for terms reused from existing vocabularies and have only been defined for ADMS's own terms. More specifically, attributes such as dcterms:identifier is used for expressing the id of a semantic asset and foaf:homepage is used for linking to a Web page that is fully dedicated to a specific semantic asset.

2.2 ADMS XML SCHEMA

The backbone of ADMS in XML Schema (XSD) is a Common Library of information elements provided by the Universal Business Language (UBL) standard [27].

The philosophy behind this design is to achieve reusability of information elements defined by the Core Component Technical Specification (CCTS) of UN/CEFACT [28] (the basis of UBL).

3. FEDERATION OF SEMANTIC ASSET REPOSITORIES

The ISA programme in partnership with the DERI, NUI Galway[1] developed a prototype aggregator of ADMS descriptions [29] to test the concept of syndication of semantic asset descriptions on the web. This prototype used ADMS v0.6 and implemented a SPARQL endpoint to federate a few publishers of semantic assets; the Digitalisér.dk (DK), the XRepository (DE), and the European repository; former SEMIC.EU currently migrated to Joinup.

Following the success of this prototype, the Commission has set up a multidisciplinary working group to finalise ADMS. 43 people of 20 EU Member States as well as from the US and Australia have participated in the finalisation of ADMS. Most of them were experts from standardisation bodies, research centres and the EU Commission. The working group used a methodology based on W3C's processes and methods. ADMS v.1 was officially released in April 2012.

The Commission will offer a single point of access to semantic assets described using ADMS via its collaborative platform, Joinup. This service will increase the visibility of semantic assets described with ADMS on the web. This will also stimulate the reuse of semantic assets by pan-European initiatives, including the Commission leading to cost savings.

For a semantic asset to be made available through the federation, three steps have to be taken as illustrated in Figure 2. First, publishers should identify the semantic assets that they want to share and analyze them. The selected semantic assets should then be described following the ADMS specification (currently ADMS v1). Finally, the descriptions of the semantic assets should be published in the federation, thus making them discoverable and retrievable.

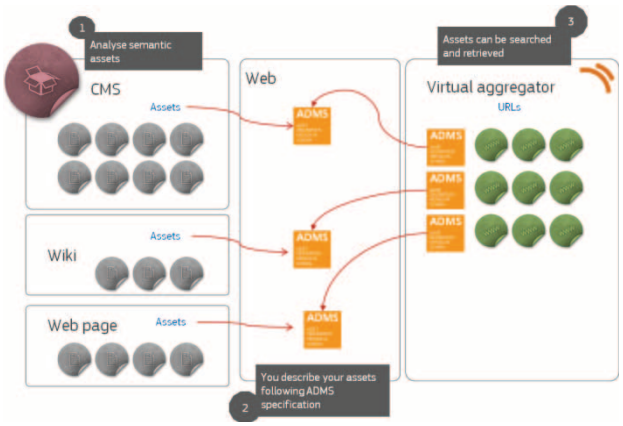

Figure 2 Typical adoption Use Case

4. CONCLUSION AND FUTURE WORK

The Commission will continue to promote the adoption of ADMS by publishers of semantic assets and will launch the federation of semantic asset repositories in June 2012.

ADMS specification will be contributed to the W3C's Government Linked Data (GLD) Working Group. The GLD Working Group was set up in 2011 to provide standards and other information which help governments around the world to publish their data as effective and usable Linked Data using Semantic Web technologies. ADMS will be published by the GLD WG as First Public Working Drafts for further consultation within the context of the typical W3C standardization process. The desired outcome of that process will be the publication of these vocabularies as open Web standards available under W3C's Royalty-Free License.

Finally, it is worth mentioning that other major standardization bodies such as OASIS, The Open Group and ETSI have also shown great interest in the ADMS initiative.

5. REFERENCES

[1] European Commission, "ISA - Interoperability Solutions for European Public Admnistrations," February 2012. [Online]. Available: http://ec.europa.eu/isa/.

[2] European Commission, "Digital Agenda for Europe 2010 - 2020," May 2010. [Online]. Available: http://ec.europa.eu/information_society/digital-agenda/index_en.htm. [Accessed January 2011].

[3] European Commission, "Digital Single Market," [Online]. Available: http://ec.europa.eu/information_society/newsroom/cf/pillar.cfm. [Accessed February 2012].

[4] European Commission, "European Interoperability

[1] http://www.deri.ie

Framework (EIF) for European public services," 12 2010. [Online]. Available: http://ec.europa.eu/isa/strategy/doc/annex_ii_eif_en.pdf.

[5] European Commission, "eGovernment Action Plan 2011 - 2015," DG INFSO, [Online]. Available: http://ec.europa.eu/information_society/activities/egovernment/action_plan_2011_2015/index_en.htm. [Accessed February 2012].

[6] Bioportal, [Online]. Available: http://bioportal.bioontology.org/. [Accessed 12 June 2012].

[7] Open Ontology Repository (OOR) Initiative, [Online]. Available: http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository. [Accessed 12 June 2012].

[8] eXtended MetaData Registry (XMDR) Project, [Online]. Available: https://hpcrd.lbl.gov/SDM/XMDR/. [Accessed 12 June 2012].

[9] International Standardization Organization, "ISO/IEC 11179, Information Technology -- Metadata registries (MDR)," [Online]. Available: http://metadata-standards.org/11179/. [Accessed 12 June 2012].

[10] CEN, "CEN Workshop Agreement on 'Discovery of and Access to eGovernment Resources' CWA 15971:2009 part 1 to 6," 2009. [Online]. Available: http://www.cen.eu/CEN/sectors/sectors/isss/workshops/Pages/wsegovshare.aspx.

[11] J. Hartmann, R. Palma, Y. Sure, P. Haase and M. del Carmen Suárez-Figueroa, "OMV -- Ontology Metadata Vocabulary," in *Chris Welty, ISWC 2005 - In Ontology Patterns for the Semantic Web*, 2005.

[12] ADMS Working Group, "The Asset Description Metadata Schema (ADMS) - Specification," January 2012. [Online]. Available: http://joinup.ec.europa.eu/asset/adms/release/all. [Accessed February 2012].

[13] W3C, "Date and Time Formats," September 1997. [Online]. Available: http://www.w3.org/TR/NOTE-datetime. [Accessed March 2012].

[14] International Organization for Standardization, Data elements and interchange formats, "Information interchange -- Representation of dates and times," 2004. [Online]. Available: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=4087. [Accessed March 2012].

[15] International Organization for Standardization, Codes for the representation of names of languages, "Part 4: General principles of coding of the representation of names of languages and related entities, and application guidelines," 2010. [Online]. Available: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39535. [Accessed March 2012].

[16] International Organization for Standardization, Codes for the representation of names of countries and their subdivisions, “Part 2: Country subdivision code,” 2007. [Online]. Available: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39718. [Accessed March 2012].

[17] Eurostat, "NUTS classification (Nomenclature of territorial units for statistics) is a hierarchical system," 2008. [Online]. Available: http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature. [Accessed March 2012].

[18] EC Publications Office, "EuroVoc," [Online]. Available: http://eurovoc.europa.eu/drupal/. [Accessed March 2012].

[19] Food and Agriculture Organization of the United Nations, "AGROVOC," [Online]. Available: http://aims.fao.org/website/AGROVOC-Thesaurus/sub. [Accessed March 2012].

[20] UK Cabinet Office, e-Government Unit, "IPSV-Integrated Public Sector Vocabulary," [Online]. Available: http://www.esd.org.uk/standards/ipsv. [Accessed March 2012].

[21] UK Cabinet Office, e-Government Unit, "ESD Standards, e-GMS-e-Government Metadata Standard," 2004. [Online]. Available: http://www.esd.org.uk/standards/egms/. [Accessed March 2012].

[22] Standaarden Overheid.nl, "OWMS - Overheid.nl Web Metadata Standaard," [Online]. Available: http://standaarden.overheid.nl/owms/4.0/doc/. [Accessed March 2012].

[23] EC Joint Research Centre, , " Infrastructure for Spatial Information in the European Community – Inspire," 2007. [Online]. Available: http://inspire.jrc.ec.europa.eu/index.cfm. [Accessed March 2012].

[24] ADMS Working Group, "RADion - Repository, Asset, Distribution - Specification," March 2012. [Online]. Available: https://joinup.ec.europa.eu/asset/radion/description. [Accessed April 2012].

[25] Dublin Core, "DCMI - Dublin Core Metadata Initiatives," [Online]. Available: http://dublincore.org/specifications/. [Accessed February 2012].

[26] D. Brickley and L. Miller, "FOAF Vocabulary Specification," August 2010. [Online]. Available: http://xmlns.com/foaf/spec/. [Accessed April 2012].

[27] OASIS, "UBL - Universal Business Language," 2006 October. [Online]. Available: http://docs.oasis-open.org/ubl/cs-UBL-2.0/UBL-2.0.html. [Accessed April 2012].

[28] UN/CEFACT, "CCTS - Core Components Technical Specification Version 3.0," September 2009. [Online]. Available: http://www.unece.org/fileadmin/DAM/cefact/codesfortrade/CCTS/CCTS-Version3.pdf. [Accessed April 2012].

[29] G. Shukair, "Federated Portal of Repositories on Semantic Interoperability Solutions," 2011. [Online]. Available: http://vmudi205.deri.ie/elda/index.html. [Accessed February 2012].