eGovernment Core Vocabularies

The SEMIC.EU approach

May 2011

v.1

European Commission
Directorate-General for Informatics

EUROPEAN COMMISSION
DIRECTORATE-GENERAL
INFORMATICS

Interoperability solutions for European public administrations (ISA)

Brussels, May 5, 2011

eGovernment Core Vocabularies: the SEMIC.EU approach

1. Background – Problem statement

The SEMIC.EU platform promotes semantic interoperability in EU MSs by collecting, evaluating, indexing and making available a large number of semantic assets from a single point of access. In this way, developers can easily discover and re-use assets like data models, taxonomies, codelists and vocabularies developed by others facing similar use cases.

Agreeing on common semantic assets across Europe is a critical step towards semantic interoperability. However, severe difficulties stem from the diverged cultural, legal, organizational environments that characterize the 27 EU MSs. Despite that, if semantic interoperability is to become a reality, common agreements on semantics supported by a social process of consensus building should start to evolve from a certain starting point.

2. Core Concepts as a first step towards European semantic interoperability

Following a similar trend to vocabularies used as metadata on the web (e.g. FOAF, SKOS, DC, SIOC), definitions should first be agreed on fundamental concepts, where diverged and/or conflicting views can be handled. These concepts are what we describe as *Core Concepts*[1].

A *Core Concept* is a simplified data model that captures the minimal, global characteristics/attributes of an entity in a generic, country and domain neutral fashion. It can be represented as *Core Vocabulary* using different formalisms (e.g. XML, RDF, JSON)

There are two main ideas in the *Core* notion: a) the *Core* is highly reusable: the specification is simple and captures basic and generic characteristics of an information entity, regardless of the context this entity is used, b) the *Core* is extensible: domain specific specializations can be drafted on top of the core representation. As an example, the Core Person could have a minimum set of characteristics that describe a Person regardless of the country or the context this entity is involved (e.g. Place of Birth, Date of Birth, Gender, etc). Later, certain specializations of this Core Person could be drafted as extensions to describe in more details special types of Persons like Voter, Employee, Passenger, Patient etc. Their compliance to the Core Person specification guarantees a minimum of cross-domain interoperability while it provides domain-specific communities both with the freedom and a common starting point for drafting their own specializations by adding metadata to the Core.

This approach increases possibilities for reuse, as complexity and overspecification usually result in diverged views where consensus is hard to reach, especially at the European level.

Core Concepts should be identified, discussed and specified through a process of iterative design, coupled by pilot implementations, with the participation of the broader SEMIC

[1] http://bit.ly/i5kGa9

community, and feedback from the ISA Technical Working Groups and National Interoperability Framework owners. This process will result in a small library of recommended conceptual schemas/vocabularies (e.g. 20-30, +/-6 properties each) for reuse in the National Interoperability Frameworks, Metadata Standards Catalogues, developers' communities, etc.

These conceptual models should at first be technology-neutral. For usability purposes these models should be then encoded as technical vocabularies in various formats including XML, RDF and Linked Data. In this way Core Concepts provide the conceptual models for the development of technical *Core Vocabularies*.

3. Work so far and next steps

The Core Concept line of work was discussed with the MSs representatives during the last ISA Trusted Information Exchange WG, 3rd March 2011. The feedback has been positive and supportive. Work has already started to define the *Core Person* concept[2] and an *Asset Description Metadata Schema*[3]. Both Core Concepts will be used as pilots for gaining insights on the overall development process. As a next step, MSs will be invited to propose candidate Core Concepts. After a consultation, MSs will agree on the final list of Core Concepts. A WG per Concept is then foreseen to draft each specification. As soon as a Core Concepts become available, its technical representation through Core Vocabularies will take place.

All these Core Vocabularies will be stored in a platform. This platform for vocabulary development, hosting and reuse will be developed as an extension of the SEMIC infrastructure, providing an interface and APIs for accessing, querying, browsing, downloading these Core Vocabularies in various formats.

The main phases of this work include the following:

a) Conceptualization phase / Drafting Core Concepts: Social process of consensus building with our stakeholders, that primarily being MSs representatives. Harmonization of conflicting models, abstractions, modularization, logic and knowledge engineering.

b) Encoding phase / Drafting Core Vocabularies: based on the Core Concepts to be agreed at the previous phase, concrete representations of Core Vocabularies are drafted in XML, RDF and as Linked Data[4]. Active involvement of external parties from the *standardization community* to be sought. Initial contacts and discussions with DG Enterprise, ETSI, and W3C have been very promising and revealed great interest and willingness to support the action.

c) Platform development phase: A platform will be developed as part of the SEMIC.EU infrastructure to provide hosting, access, search, browsing, and navigation of the linked open metadata cloud of vocabularies.

4. Benefits

- A set of Core Concepts supported by all EU MSs provides a concrete starting point for promoting semantic interoperability amongst European Public Administrations.
- A platform of eGovernment Core Vocabularies hosted by EC and supported by W3C has very good chances to attract enormous visibility not only in the 27 MSs but globally. The EC could pioneer the fast evolving area of hosting and maintenance of semantic vocabularies and metadata standards for public administrations, where trust and provenance are considered of critical importance.

[2] http://www.semic.eu/semic/view/snav/Conformance/specification.xhtml?cid=389703
[3] http://www.semic.eu/semic/view/snav/network/Communities/ADMS.xhtml?cid=389730
[4] http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/

5. Draft Development plan

a) ADMS and/or Core Person first draft to be discussed with the MSs, Core Vocabularies to be drafted and initial implementation to be tested by Q3 2011

b) The list of Core Concepts to be developed should be agreed by Q4 2011

c) All Core Concepts to be drafted by WGs and endorsed by MSs by Q3 2012

d) In parallel, the vocabulary infrastructure will be developed and tested with mock examples. It will be gradually populated with actual Core Vocabularies as soon as these become available.

e) By Q4 2012 the eGovernment Core Vocabulary Platform goes live providing access to a set of 20-30 Core Vocabularies

ISA Contact:

Vassilios Peristeras, DG DIGIT, ISA Unit, vassilios.peristeras@ec.europa.eu

Montoyer (MO) 34, Floor: 3, Office: 103; B-1049 Brussels, Belgium

Tel: +32 (0) 2 29 81014