



eGovernment Core Vocabularies

February 2012

v.2





EUROPEAN COMMISSION

DIRECTORATE-GENERAL

INFORMATICS

B-1049 Brussels – Belgium



Interoperability solutions for European public administrations (ISA)

Brussels, 17 February 2012

eGovernment Core Vocabularies

1. Background – Problem statement

The Joinup collaborative platform¹, hosted by the European Commission's ISA (Interoperability Solutions for European Public Administrations) programme², promotes semantic interoperability in the European Union Member States by collecting, evaluating, indexing and making available a large number of semantic interoperability assets from a single point of access. In this way, developers can easily discover and re-use assets such as 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 divergent cultural, legal, and organisational environments that characterise the EU's 27 Member States. 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 Vocabularies' to promote European semantic interoperability

Following a similar trend to vocabularies used as metadata on the web, definitions should first be agreed on fundamental concepts, where divergent and/or conflicting views can be handled.

These concepts are what we describe as 'Core Concepts'³. Examples of existing Core Concepts include FOAF (Friend of a Friend), a vocabulary describing persons; SKOS (Simple Knowledge Organisation System), to represent classification schemes; DC (Dublin Core), a metadata vocabulary; and SIOC (Semantically-Interlinked Online Communities Project), for connecting Internet blogs, forums and mailing lists.

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, for example XML (Extensible Markup Language), RDF (Resource Description Framework), and JSON (JavaScript Object Notation).

¹ <http://joinup.ec.europa.eu/>

² <http://ec.europa.eu/isa/>

³ http://joinup.ec.europa.eu/community/core_vocabularies/description

There are two main ideas in the ‘Core’ notion:

- 1) 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; and
- 2) The Core is extensible: Domain-specific specialisations can be drafted on top of the Core representation. For example, the ‘Core Person’ could have a minimum set of characteristics that describe a person regardless of his/her country or context (e.g. place of birth, date of birth and gender). Later, certain specialisations of this Core Person could be drafted as extensions to describe in more details types of persons including voters, employees, passengers and patients.

Their compliance to the Core Person specification guarantees a minimum degree of cross-domain interoperability, while providing domain-specific communities both with the freedom and a common starting point for drafting their own specialisations by adding metadata to the Core.

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

Core Concepts are identified, discussed and specified through a process of iterative and collaborative design, coupled with pilot implementations, with the participation of the broader community, and feedback from the ISA Technical Working Groups and National Interoperability Framework owners. This process will result in a small library of recommended vocabularies, e.g. 20-30, for reuse in National Interoperability Frameworks, Metadata Standards Catalogues, and developers’ communities.

The conceptual models remain technology-neutral. For usability purposes these models are encoded as technical vocabularies in various formats including UML, XML, RDF and Linked Data. In this way, Core Concepts provide the conceptual models for the development of technical ‘Core Vocabularies’.

The main phases of this work include the following:

- a) *Conceptualisation phase / Drafting Core Concepts*: Social process of consensus building with stakeholders, primarily being Member States’ representatives. Harmonisation of conflicting models, abstractions, modularisation, logic and knowledge engineering.
- b) *Encoding phase / Drafting Core Vocabularies*: Concrete representations of Core Vocabularies are drafted in UML, XML, RDF and as Linked Data⁴.

3. Work so far and next steps

The Core Vocabularies line of work was discussed with the Member State representatives in the ISA Trusted Information Exchange Working Group in March 2011. The feedback has been positive and supportive. Effectively, work has started since then.

At first, over 20 candidate Core Vocabularies have been identified. From this superset, three Core Vocabularies have been selected, and Task Forces (TFs) have been set up to draft the specifications. European Commission experts have been also invited to chair the TFs. More specifically:

- The Core Person TF⁵ is chaired by Eurojust experts.
- The Core Business TF⁶ is chaired experts from the European Commission’s Internal Market and Services Directorate-General.

⁴ <http://lab.linkeddata.deri.ie/2010/star-scheme-by-example/>

⁵ http://joinup.ec.europa.eu/asset/core_person/description

- The Core Location TF⁷ is chaired by experts from the Joint Research Centre (JRC), Ispra, and more specifically by the team responsible for the EU INSPIRE Directive.

W3C is actively involved and supports this work providing the overall methodology and coordination of the TFs.

With a broad participation of 67 experts from 21 EU Member States, including specialists from the European Commission, national governments, international standards organisations, academics and experts, the TFs prepared the draft specifications and made them available for public comments via the Joinup platform for a period of one month (mid-February 2012 – mid-March 2012). Comments received will be processed by mid-April 2012, when the final version of the three Core Vocabularies will become available.

These final Core Vocabularies will be proposed for endorsement by EU Member States' representatives in the context of the ISA Programme. The specifications will be further promoted for adoption by National Interoperability Frameworks and/or national Metadata Catalogues later in 2012. Implementation opportunities will be also sought especially in connection to promoting interoperability amongst European base registries (e.g. population registries, business registries).

Based on the results and lessons learnt, further development of Core Vocabularies is foreseen in 2013.

4. Conclusion

A set of commonly agreed Core Vocabularies supported by the EU Member States provides a concrete starting point for promoting semantic interoperability amongst European public administrations. The ISA Programme has initiated this work, promoting the collaborative development of these vocabularies in an open and inclusive way. There is already a lot of interest and active participation among a large community of relevant stakeholders. The first three Core Vocabularies are under development and concrete results are expected already in 2012. In this way, the European Commission pioneers and seeks results in the fast-evolving area of semantic vocabularies and metadata standards for public administrations.

⁶ http://joinup.ec.europa.eu/asset/core_business/description

⁷ http://joinup.ec.europa.eu/asset/core_location/description