 isa
Interoperability Solutions for
European Public Administrations

SEMIC.EU Semantic
Interoperability Centre
Europe

case study

SEMIC.EU Core Vocabularies: the Core Person Case

August 2011

Simplified, reusable, and extensible data models

SEMIC.EU Core Vocabularies: the Core Person Case

This case study gives an outlook to the upcoming Core Vocabularies initiative of SEMIC.EU, which will be launched in the second half of 2011. It also describes the early achievements of an example, namely the SEMIC.EU Core Person specification.

The need for harmonization through Core Vocabularies

Delivering European public services to European citizens and businesses will be difficult, if not impossible, without interoperability among European public administrations. (Digital Agenda for Europe 2010 - 2020)

The need for effective interoperability is a central part of the Digital Agenda, one of the flagship initiatives in the Europe 2020 Strategy. In this context, the SEMIC.EU project continues, currently supported by the ISA Programme, to support public administrations in Europe addressing semantic interoperability concerns related to the **cross-border** and/or **cross-sector** electronic interactions between public administrations that result from the implementation of EU policies and activities. To this end, SEMIC.EU has set up a collaborative platform, including a so-called **semantic asset repository**, and a set of services around this platform to support Public Administrations to become aware about interoperability requirements and find, choose, design, and implement semantic interoperability assets.

The assets in the SEMIC.EU asset repository mostly target one specific interoperability context, such as for example the exchange of a person's CV data containing a summary of his or her skills and experiences. However, to foster cross-border and cross-sector semantic interoperability, harmonization should also start at a higher level of abstraction that surpasses the context of individual interoperability domains, and thus the cultural, legal, organisational differences of individual Member States. In particular, harmonisation requires agreeing on common semantics of fundamental, *Core Vocabularies* across Europe. Such Core Vocabularies are the starting point for agreeing on new semantic interoperability assets and defining mappings between existing assets. Semantic interoperability assets that map to or extend such Core Vocabularies guarantee a minimum level of cross-domain and cross-border interoperability.

Core Concepts and Core Vocabularies: Definition

Core Concepts are simplified, reusable, and extensible data models that capture the fundamental characteristics of an entity in a context-neutral fashion (European Commission, ISA Programme, 2011). Core Concepts can be represented as **Core Vocabularies** using different formalisms such as RDFS, XMLS, JSON Schema. The above-mentioned definition of Core Concepts incorporates the following ideas:

- **Context-neutral**: Core Concepts make abstraction from the contextual specificities that are linked to the usage of a concept in a particular country of domain.
- **Simplified**: Core Concepts only focus on *fundamental* characteristics of an entity. This *minimalistic* approach facilitates consensus building. In contrast, complexity and over-specification usually result in diverged views where consensus is hard to reach, especially at the European level (European Commission, ISA Programme, 2011).
- **Reusable and extensible**: The SEMIC.EU Core Concepts are capture basic and generic characteristics of an information entity and enable domain specific specializations to be drafted on top of the core representation (European Commission, ISA Programme, 2011).

The idea of Core Concepts can be related to existing work, such as among others the Core Components in the Core Component Library (CCL) (UN/CEFACT, 2011) and work on Foundational Ontologies such as DFO, DOLCE, GFO, REA, and SUMO (Guizzardi & Wagner, 2005). The recent schemas published by schema.org, an initiative of Google, Microsoft, and Yahoo are also relevant in this context.

An early example: the SEMIC.EU Core Person

The SEMIC.EU Core Person is a first pilot for the specification of Core Concepts by SEMIC.EU. The lessons learned from its development will be used to gain insight in the general development process of Core Concepts. The Core Person data model describes the minimum characteristics of a person regardless of the context in which this entity is involved. When other interoperability assets – such as Voter, Passenger, Employee, or Patient Data Model – extend the Core Person data model or provide mappings to the Core Person, they guarantee a minimum of cross-border and/or cross-domain interoperability.

SEMIC.EU Core Person

Full Name [0..1]
Given Name [0..1]
Family Name [0..1]
Gender [0..1]
Date of Birth [0..1]
Place of Birth [0..1]
Country of Birth [0..1]
Citizenship [0..*]

The SEMIC.EU Core Person model contains eight elements. These elements are almost always present when dealing with personal data. The model has been restricted to these eight elements in order to minimise the potential of semantic conflicts and to increase the possibility for reuse.

An important aspect to creation of the SEMIC.EU Core Person is the social process of consensus building that resulted in its specification. In particular, more than 100 SEMIC.EU users have contributed to the specification of this core data model for personal data.

Once finalised, a Core Person concept and vocabulary has the potential to be used as a core element for any person-related data exchange in e-Government projects

and public administrations in Europe. Information systems of public administrations that would reuse a finalised and agreed Core Person vocabulary would be able to more easily exchange person data across borders and sectors, hereby minimising potential semantic conflicts. The reuse of the Core Person vocabulary can be established in the following ways:

- **By extending the Core Person vocabulary**: in this case, the reuse of the elements of the SEMIC.EU Core Person, and their semantics, is natively taken-onboard.
- **Through a mapping to the SEMIC.EU Core Person**: in this case, the reuse is realised by mapping an existing data model to the SEMIC.EU Core Person. The SEMIC.EU Core Person can ultimately be used as a pivot canonical data model and therefore facilitate the exchange of data among information systems.

Potential applications

The following initiatives are practical examples of person-related data which can potentially be mapped to the SEMIC.EU Core Person:

- **GBA Civil Register, The Netherlands**: The register of all residents of the Netherlands used at municipal level and reused for all other purposes including person-related data.
- **RNI Register of Non-Inhabitants, The Netherlands**: Persons not resident in the Netherlands but having a relationship with the Dutch Government.
- **Students Register, Croatia**: A data model for student records at the core of which is a unique ID for each student.
- **Voters Register, Russia**: The "Vybory" register in Russia addresses the challenges of interoperability of systems at different administrative levels with a focus on accurate identification of individuals as voters.
- **BAA Airports Limited, United Kingdom**: The information standards authority in BAA models Persons, Groups and Organisations as parties of interest in the daily operations of the airport.
- **National Health Service, United Kingdom**: Data model for the national demographics database used by the UK's National Health Service (NHS).

The next steps

SEMIC.EU will continue its work on the Core Concepts and derived Core Vocabularies, including the finalisation of the SEMIC.EU Core Person. The main objective of this work is to reach agreement among the EU Member States on a set of 20-30 Core Concepts, such as the Core Person, the Core Organisation, the Core Vehicle, et cetera. The work will consist of the following steps (European Commission, ISA Programme, 2011):

1. **Design of the Specification Process**: The approach to specify and reach agreement on the Core Concepts and Vocabularies will be designed in advance. This approach will include indentifying the appropriate stakeholder representatives and a clarification of their roles and responsibilities.
2. **Specification of and agreement on Core Concepts**: The Member States will be invited to propose Core Concepts. A process of consensus building will lead to the resolution of differences and agreement on the Core Concept Specifications. The SEMIC.EU Core Person specification will also undergo this process.

"My first and foremost concern is that interoperability assets remain simple, in order to become practice-ready, i.e. (re-)usable, and at the same time powerful in making data and information exchange possible in different contexts."

Vassilios Peristeras, SEMIC.EU project Officer

3. **Specification of Core Vocabularies**: For each Core Concept a representation in XML and RDF will be specified. The RDF Vocabularies will have concrete mappings to existing Web vocabularies such as FOAF, DC, SIOK, or these of schema.org.
4. **Publication**: The Core Concepts will be published online as part of the SEMIC.EU asset repository, allowing to access, search, browse, and navigate the set of vocabularies.
5. **Promotion and usage**: The promotion and usage of the Core Concepts and Vocabularies by the Member States' public administrations will be fostered amongst others via the mediation of their National Interoperability Frameworks and the European Commission.

SEMIC will actively seek to align its actions with stakeholders outside the e-Government domain, or event outside the EU. Initial contacts and discussions with third parties including European Commission DGs and standardization organisations have been very promising and revealed great interest and willingness to support the action.

To know more

European Commission, DG Information Society. (2010, May). *Digital Agenda for Europe 2010 - 2020*. Retrieved January 2011, from European Commission Information Society: http://ec.europa.eu/information_society/digital-agenda/index_en.htm

European Commission. (2010, December 15). *The European eGovernment Action Plan 2011-2015 - Harnessing ICT to promote smart, sustainable & innovative Government*. Retrieved April 2011, from European Comission: http://ec.europa.eu/information_society/activities/egovernment/action_plan_2011_2015/docs/action_plan_en_act_part1_v2.pdf

European Commission, ISA Programme. (2010). *A new approach for semantic asset management - Project Officer's vision*. Retrieved from SEMIC.EU: http://www.semic.eu/semic/view/snav/About_SEMIC/interview-peristeras.xhtml

European Commission, ISA Programme. (2011, May). *e-Government Core Vocabularies: The SEMIC.EU approach*. Retrieved from European Commission - Directorate-General Informatics: http://www.semic.eu/semic/view/documents/egov-core-vocabularies.pdf

European Commission, ISA Programme. (2011). *SEMIC.EU Core Concepts*. Retrieved from SEMIC.EU: http://www.semic.eu/semic/view/snav/shared-development/core-concepts.xhtml

European Commission, ISA Programme. (2011). *SEMIC.EU Core Person: A pan-European schema for personal data*. Retrieved from SEMIC.EU: http://www.semic.eu/semic/view/snav/Conformance.xhtml

European Union. (2010, December 16). *Towards interoperability for European public services*. Retrieved from Europa Press Release - RAPID: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/10/1734&forma

Guizzardi, G., & Wagner, G. (2005). Some Applications of a Unified Foundational Ontology in Business Modeling. In M. Rosemann, & P. Green, Ontologies and Business Systems Analysis. DEA Publisher.

Ministerial e-Government Conference. (2009, November 18). *Ministerial Declaration on e-Government*. Retrieved from ICT for Government and Public Services: http://ec.europa.eu/information_society/activities/egovernment/events/past/malmo_2009/press/ministerial-declaration-on-egovernment.pdf

UN/CEFACT. (2011). Core Component Library (UN/CCL). Retrieved from UNECE - United Nations Economic Commission for Europe: http://www.unece.org/cefact/codesfortrade/unccl/CCL_index.htm

Weill, P., Woerner, S. L., & McDonald, M. (2010, May). Reuse is more important and harder than we thought. *Research Briefing –MITSloan Center for Information Systems Research , 2010* (5).