Estonian Semantic Interoperability Initiative

**Martin Luts**, EE

The Estonian Semantic Interoperability Architecture

eGovernment services are built upon the exchange of information. Semantic or content interoperability is about ensuring that the meaning of the information exchanged is not lost in the process, that it is readable and understood by the involved people, applications, and institutions. In order to deliver cross-border services, the European institutions and Member States will have to agree on a multitude of semantic specifications, such as descriptions of people, products, processes, forms, etc. They will also have to agree on how to formulate these descriptions and where to store them for public use.

Many Member States of the European Union (MS) currently have some Semantic Interoperability initiatives ongoing, see for example Germany’s initiative Deutschland Online, Italian’ initiative in public administration, Finnish semantic initiative FinnONT0, and Semantic Latvia project. The scope and accent of these initiatives are quite different – some MS focus on consolidating semantic assets in several governmental institutions already in place into semantic portals, some on building full-scale national semantic web infrastructures, others target syntactic or semantic descriptions of data schemas, some are on the level of human-oriented descriptions of assets, others try to reach automatic use.

There are also pan-European initiatives, which include:

- **SEMIC.EU** (SEMantic Interoperability Centre Europe), led by the European Commission’s IDABC programme. SEMIC.EU is designed as a brokerage platform for third party semantic assets – like classification lists, ontologies, etc. It is not meant for creating/maintaining nor standardising the format of semantic assets.
- **semanticGov**, fully titled “Providing Integrated Public Services to Citizens at the National and pan-European level with the use of Emerging Semantic Web Technologies”. SemanticGov is an EU-funded (FP6) research and development project that aims at building the infrastructure (software, models, services, etc) necessary for enabling the offering of semantic web services by public administration.

This paper outlines the Estonian Semantic Interoperability initiative in the public sector, describes its architecture and components used in architecture, present state and future developments. The organisational, time, budget and project management views are not addressed in this paper.

Semantic Interoperability Architecture for State Information System and Registries in Estonia

Interoperability Platform for State Information System in Estonia

The state information systems' data transport layer X-Road (X-tee in Estonian) is a technical and organisational environment that enables secure data transfer between digital government databases and enables secure data transfer between individuals and government institutions. It also coordinates the access of individuals to information being processed in government databases.

X-Road platform supports the technical interoperability of the components and registries of state information system. Currently, X-Road does not support Semantic Interoperability. In the following we describe the extension of X-Road – the semantic layer to be built on top of X-Road infrastructure.

Overview of Semantic Interoperability Architecture for State Information System
Semantic Interoperability requires a system for the semantic description of several types of objects (infoware or information resources), like:

- operations (e.g. web services): the input-output data, preconditions, effects, logical relation between input and output structures;
- business processes;
- data-structures in databases or data-structures to be exchanged;
- official and other documents, web-pages;
- other types of objects, in public or private sector.

Estonia

The Semantic Interoperability architecture for state information system and registries of Estonia (see Figure 1) consists of the following interrelated components: ontologies, semantically annotated objects (contained in infoware's metadata), supported by policies and guidelines, several processes and workflows, tools, educational activities, PR, among others.

The Administration System for the State Information System (RIHA) is the central tool in the Semantic Interoperability architecture for state information system. RIHA fulfils the
following tasks in the Semantic Interoperability architecture:

- Hosting and publishing of ontologies.
- Hosting and publishing of infoware’s metadata, including semantics.
- Serving as semantic search engine for semantic assets (resources).

In the following sections the most important components of the Semantic Interoperability architecture are considered in detail.

**Ontologies**

**Ontology Support for SemIO**

One of the main components of the Semantic Interoperability architecture – as designed for Estonian state registries – are ontologies. We use the term ‘ontology’ meaning “a formal explicit specification of a shared conceptualization for a domain of interest” (Gruber 1993: 199-220).

The ontology component in our architectural framework is not a monolithic structure – for the purposes of easier, domain-expert driven maintenance, it is divided into domains, e.g. “Environment”, “Social Affairs”. The initial tree for categorising ontologies is based on the official naming of EU activities. The language used in the semantic description of ontology objects draws from W3C recommendation OWL (Web Ontology Language). The ontologies are developed and maintained in a distributed manner, the tools to be used are not prescribed as long as certain standards are followed (for example, Collaborative Protégé, Semantic Media Wiki, and others may be in use), but stored and published centrally – in the administration system for the state information system.

Ontologies that support the Semantic Interoperability architecture are classified into 3 layers depending on how general they are i.e. how general concepts they cover (see Figure 2).

![Figure 2. Modularity and layering of ontologies component in interoperability architecture](image-url)
The reason for such a segregation of ontologies is that we foresee many different use-case scenarios of ontologies that go beyond the needs and borders of the state information system. For example, top ontology can be used by many applications, including commercial.

**Ontology Creation and Maintenance**

From the methodological point of view, development of ontologies is a complex task and a lot of ontology building experiences are needed in order to participate in the production of good ontologies. In this respect, Estonia is at the initial level of education and experience, both in the universities as well as in software companies. However, the current Semantic Interoperability initiative is going to improve the situation.

The ontology creation process is planned to be data-driven, following a mixture of bottom-up and middle-out approaches – we start with data-object used as input-output in web-services offered by main state registries (i.e. population registry, land cadastre, vehicle registry). As a preliminary study shows, some 200 concepts will cover about 50% of the uses of data-object used as input-output in web-services. We are trying to partially automate this ontology-learning task, including extracting concept relations from the so-called query-type web-services.

We plan to use release-based publishing of ontologies, i.e. every single or minor change does not reflect on officially published ontologies. The number of the year and release number are encoded into URLs which are used to point to the elements of ontologies, so the versioning of ontologies is achieved.

**Semantically Enriching State Registries**

**Semantic Description of Operations Performed by Databases**

We see our first priority to semantically annotate operations performed by state registries – as it promises quick return of investment and we also consider it to solve the complex task of integrating state registries – and creating e-services more efficiently – to some extent.

Semantic annotation of operations includes the description of every single operation and its input-output data-structures, preconditions, effects, relation between input-output, in a form of reference to the respective entry in the domain ontology. The languages used for description are WSDL and SA-WSDL (Semantic Annotations for
WSDL and XML Schema). For example, describing the “cadastral unit” input data element of a web service, instead of describing it in human-readable form in the WSDL document, a SA-WSDL pointer, which refers to the appropriate record in land survey ontology is used.

Excerpt from semantically annotated WSDL follows:

```xml
<wsdl:types>

  <xs:simpleType name="immovable"
  
  "http://www.ee/onto/land/2008/release2/cadastralUnit">

  </xs:simpleType>
</wsdl:types>
```

**Semantic Description of Databases and Data Structures**

Semantic annotation of databases and data structure includes the description of every single data element as a reference to the respective entry in the domain ontology. The languages used for description are XML for data structures, XMI for all types of database (relational, object-oriented or other) schemas combined with SA-WSDL (Semantic Annotations for WSDL and XML Schema). For XMI we have created a special UML profile for attaching SA-WSDL pointers into serialised UML (i.e. XMI).

**Semantic Description of Other Assets**

The architecture outlined above is not closed for annotating only web-services or data structures. Adding semantics to business processes, documents, web pages and other objects, also beyond public sector, can be carried out with moderate effort. The reuse of ontologies, knowledge, tools is in the vision of the developers of this architecture.

**Guidelines**

Department of State Information Systems, Ministry of Economic Affairs and Communications of Estonia, has released “Instructions for the Semantic Description of Databases and Operations Performed by Databases” (Ministry of Economic Affairs and Communications of Estonia 2007a) and “Methodology for the Semantic Interoperability of Databases and Operations
Performed by Databases” (Ministry of Economic Affairs and Communications of Estonia 2007b) (available in English at http://www.riso.ee/en/information-policy/interoperability). Whereas the European Semantic Interoperability Strategy is quite general, the “Instructions for the Semantic Description” provide specific rules for application owners, developers and auditors.

Martin Luts, EE, received a MSc in IT Governance from Tallinn University in 2004 and is currently a Ph.D. student at Tallinn University of Technology. Since 2005 Martin has actively lead the Estonian public sector's semantic initiative and works for the IDABC programme as a member of the SEMIC.EU Advisory Group.