Ministry of Economic Affairs and Communications
Department of State Information Systems

Methodology for the Semantic Interoperability of Databases and Operations Performed by Databases

Version 1.2

2007-04-08

This document is open for proposals from public, private and third sector organisations as well as from other interested parties. Please send your proposals to koosvoime@riso.ee.

The methodology is reviewed and, if necessary, updated annually. In new versions proposals made during the previous period will be taken into account. The “Instructions for the Semantic Description of Databases and Operations Performed by Databases” and the “Methodology for the Semantic Interoperability of Databases and Operations Performed by Databases” can be downloaded at http://www.riso.ee/en/information-policy/interoperability.

Contents

1 Introduction

Background
This methodology forms a part of the "Estonian IT Interoperability Framework"[1] and draws from the "Semantic Interoperability Strategy" (Version 0.5, 2005-09-15). The methodology itself serves as the basis for the "Instructions for the Semantic Description of Databases and Operations Performed by Databases" (v1.1, 2007-02-25).

Purpose of methodology
The purpose of this methodology is to maintain records of the decisions concerning the technologies chosen for enhancing the semantic interoperability of state information systems.

Contents of methodology
The "Methodology for the Semantic Interoperability of Databases and Operations Performed by Databases" (further referred to as the "methodology") outlines the principles for compiling the "Instructions for the Semantic Description of Databases and Operations Performed by Databases" (further referred to as the "instructions").

The methodology answers the following questions:

- WHAT? – i.e. what is semantic interoperability within the scope of this methodology and these instructions;
- WHO? – i.e. who are the target groups of semantic interoperability and relevant instructions, and what are their interests, tasks and objectives;
- WHY? – i.e. why to take the trouble to maintain semantic records;
- HOW? – i.e. what are the options to enhance semantic interoperability and grounds for choosing the appropriate methods and techniques.

The methodology explains the choices made with regard to the architecture and standards. It proceeds from the mapped principles and risks and on the basis of that opts for the best alternative.

The methodology provides reasons for the rules (requirements) outlined in the instructions.

It includes a list of problems and ideas to be tackled and considered in the next versions of the methodology and instructions, ranked by priority.

Target groups
The primary target groups of the methodology include:

- Compilers of instructions;
- Improvers of methodology;
- Trainers for the implementers of the instructions.

[1] See http://www.riso.ee/en/information-policy/interoperability.

2 Feasibility of the semantic description of databases and operations

Feasibility

Organisations are able to invest less in the production of data by integrating external data sources. However, integration does not come for free – the reuse of data has its price. It is necessary to integrate different data structures and solve conflicts arising from semantic diversity (e.g. terms with different fields of meaning).

Example of semantic diversity and potential semantic conflict. Local government X1 defines “street” in the local register of roads as a hard-surface road without sidewalks, whereas local government X2 defines it as a road with any kind of surface and with sidewalks. The task is to allocate funds for the winter maintenance of roads depending on their surface area. Obviously, the different definitions (classification systems) of roads by X1 and X2 do not allow for proper allocation of funds.

In addition to exchange of data, semantic interoperability calls for exchange of metadata between organisations, more precisely knowledge of the semantics of data. Semantic interoperability entails exchange of information about the context of data, i.e. relations, operations and functioning in general. Semantics determines how the elements of the data structures exchanged are related to real world objects, relations and events.

An appropriate way to achieve semantic interoperability is to describe explicitly and in detail the data exchanged between organisations, i.e. to give definitions of data. Within the next three years we recommend defining operations in a way that would enable information system developers (architects, analysts, designers, programmers, project managers etc.) to find the operation necessary for fulfilling their task, establish its suitability and integrate it with their application. The way of describing operations semantically must also allow for partial automatic generation of information system documentation (e.g. description of interfaces). It need not allow for automatic identification and integration of operations.

Example

State agency X orders from company Y the design and implementation of a citizen-oriented service A (e.g. subsistence benefit). The task contains the business process described in business terms (domain terms) and the test conditions necessary for providing service A. The use of test conditions requires sending an inquiry to the information system of state agency Z. If state agency Z has compiled and published the documentation (including the semantics) of the operations provided, integrator Y can make decisions with regard to operations that are compliant with the test conditions without having to bother the person acquainted with the operations of state agency Z[2]. The implementation of service A includes also maintenance of the semantic records of that operation, thus allowing for its future use as a component of a third operation.

Costs and benefits

[2] A confirmation on the suitability of the operation from the person acquainted with the operations of state agency Z might still be required.

Benefits:

1. The quality of data will improve, data from different sources can be integrated, there will be less errors and inaccuracies upon using the data and making decisions (related to misinterpretation of data as well as discrepancies arising from duplication of data);
2. Less investment in the production (acquisition) of data will be needed since it will be easier to reuse the existing data. Moreover, such data will enhance the work quality of the coordinators establishing new and updating existing databases;
3. Parties will have to spend less time on integrating the information systems of different organisations;
4. Sustainability of the application or information system will increase and it will be easier to make further developments. The knowledge base related to information systems will be preserved also after key persons have left the organisation;
5. The necessity of data can be established, i.e. retention periods can be determined.

Costs:

1. Resources for maintaining semantic records and reusable resources (e.g. work equipment, training);
2. Resources for training the maintainers and users of semantic records.

Balance of costs and benefits

Presuming that the time spent on defining the operations and publishing the definitions is approximately two or three times the time spent on replying to the questions of every system developer wishing to connect an information system, it is feasible to compile the description if at least two or three other information systems will be using the respective operation.

3 Parties interested in semantic interoperability

The parties interested in semantic interoperability include:

- Database owners, entities contracting the application (including semantic description);
- Implementers of operations (including compilers of semantic descriptions);
- Users and integrators of operations;
- Coordinators and developers of state information systems.

As regards the choice of parties and semantic description level(s), we proceed from the Zachman Framework[3] and the pragmatic presupposition that an agency is able to compile and manage a semantic description on one level only (if at all).

The Zachman Framework and several more recent frameworks drawing from that describe the following levels of abstraction for descriptions:

1. Scope, contextual level, target group – planner;
2. Business model, conceptual level, target group – owner (the "business side");
3. System model, logical level, target group – architect, designer;
4. Technology model, physical level, target group – implementer;
5. Detailed representations, out of context level, target group – sub-contractor.

With respect to semantic interoperability, we recommend to choose the third level, i.e. the system model.

4 Knowledge and skills necessary for the implementation of instructions

Description of operations performed by databases requires knowledge of the respective domain and information technology, as well as the processes of description and documentation.

In reality, there are few persons that have all the necessary knowledge – some know and handle one thing better and others other things. The knowledge of semantic interoperability is usually interdisciplinary and concerns the fields of information technology and information science. Semantic description of operations provided by databases cannot be carried out successfully without knowing the information science principles for compiling descriptions or any other records, or having IT skills to implement these principles.

The skills and knowledge as well as resources for training of the implementers of instructions currently available in Estonia have to be taken into consideration.

The user roles of semantic records may generally be divided as follows:

1. consumers of operations have the role of a "reader". Semantic description of operations should be in a format easy to find, read and comprehend without any need for specific knowledge. The role of the "reader" may be attributed to both domain experts (the "business side") as well as information system developers.
2. Operation providers, creators and record-keepers have the role of a "writer". That role requires specific knowledge of:
   - web services and related standards WSDL and SA-WSDL;
   - XML, including namespaces, XML Schema;

[3] See e.g. http://en.wikipedia.org/wiki/Zachman_framework.

- OWL and RDF/S;
- Naming and defining terms, data elements and operations;
- Compiling UML class diagrams, using domain terms and relations between them;
- Compiling other UML diagrams for the purpose of domain description (e.g. UML activity diagram for describing business processes, or UML state diagram for describing the lifespan of a class object).

In Estonia, this specific knowledge can be obtained either independently or through in-service training. Figure 1. Knowledge necessary for the implementation of instructions

shows the knowledge required for implementing Version 1 of the instructions that developers typically also have nowadays (lower gray level): skills in data modelling in an object-oriented (OO) paradigm, using e.g. UML, and skills with XML structures and in describing services in WSDL. The green colour stands for the knowledge that an average developer should acquire (i.e. how to compile and format a domain glossary in RDFS/OWL or how to compile a semantic description of a web service, using SA-WSDL). Further elaboration of the present methodology and instructions probably requires additional knowledge and skills that cannot be currently identified by the compilers of methodology, as these depend on the future developments in the field of semantic interoperability.

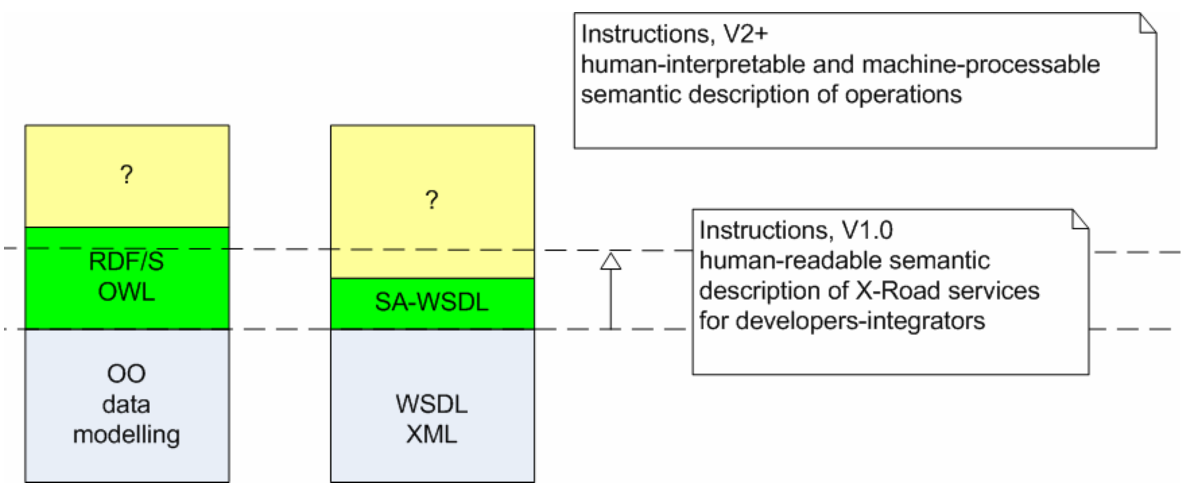

Figure 1. Knowledge necessary for the implementation of instructions

5 Semantic interoperability architecture

Semantic interoperability – the ability to understand the exchanged data in a similar way and make adequate use of it – requires a system for the semantic description of data in databases and the input and output data used in performing necessary operations. The architecture for such a system proceeds from the following principles:

1. Principle of reuse – one data element or operation should be described only once and in any other use the initial description should be referred to;

2. Simple and standard[4] compilation, management and search of the descriptions of data elements and operations.

The semantic interoperability architecture (see below below) consists of the following components:

1. **Domain glossary** – central description of the semantics of data elements and operations. The description is divided into domains, e.g. "land survey". The language used in the semantic description is OWL. The description is stored in the administration system for the state information system (RIHA);
2. **Semantic description of databases and operations** – includes the description of every single database or operation and its components and, if available, a reference to the respective entry in the domain glossary. The languages used for description are WSDL, SA-WSDL etc. The description is stored in the administration system for the state information system (RIHA).

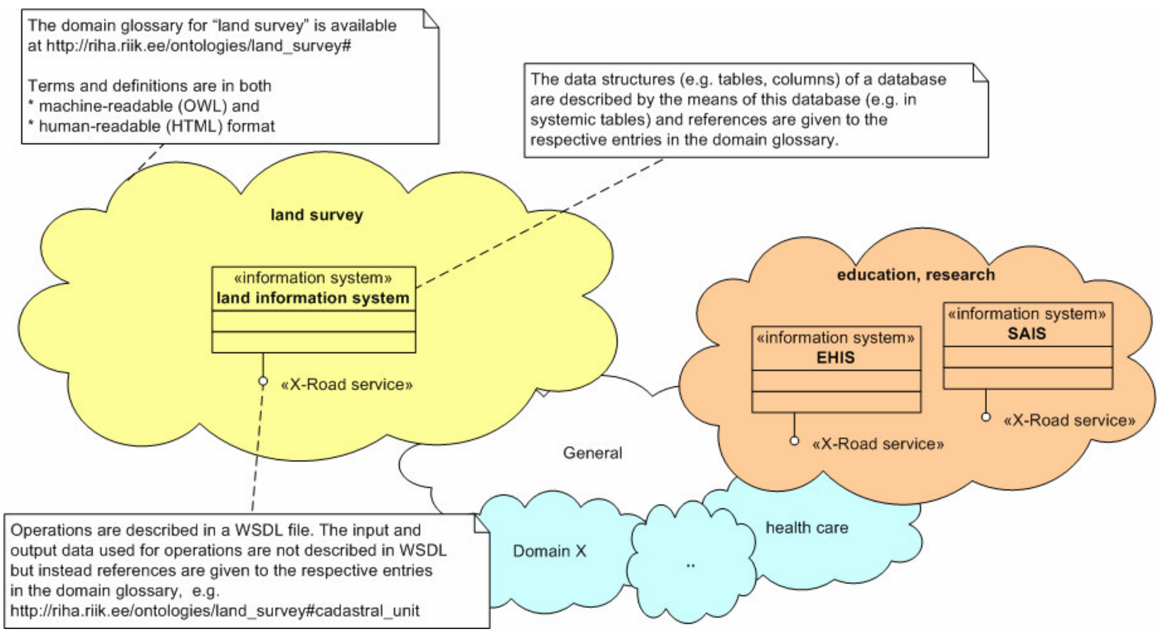

Figure 2. Semantic interoperability architecture

Possible locations for descriptions of web services and domain glossaries:

1. distributed architecture, i.e. at every separate database;
2. centralised architecture.

[4] Compliance with W3C standards and recommendations ensures semantic interoperability at international level, including with regard to relevant EU programmes. See "Interoperability between national administrations for pan-European eGovernment services" (IP/06/216; 2006-02-23).

Given that every operation should be unique, including the description, it is reasonable to describe operations in a WSDL file with the respective database as well as store the data centrally for easier access. The terms of a domain should be described centrally. Descriptions of databases and input/output data elements of operations should include references to a central glossary.

Semantic descriptions are necessary for the following information resources:

- information services;
- databases.

Another thing to consider is whether to store the metadata within the information resource (e.g. WSDL includes semantic description) or outside it (the semantic description of a database field is located in the domain glossary; the resource described includes a link to the metadata entry and not vice versa). Generally, it is recommended to store recurrent descriptions (usually data elements) outside the resource (in a domain glossary) and operation-related descriptions inside it, presuming that operations are typically unique.

Storing descriptions outside the information resource requires more work: creation and maintenance of repositories as well as motivation and supervision of describers. For further information on creating repositories see standard ISO/IEC11179. On the other hand, with that approach describers might be more motivated, because there are more supervisors as the description is available for everyone.

5.1 General compilation and publication process of semantic descriptions

Figure 3. General compilation and publication process of semantic descriptions

gives a general overview of the business process of compiling and publishing general descriptions as an UML activity diagram. The comments include preconditions and postconditions.

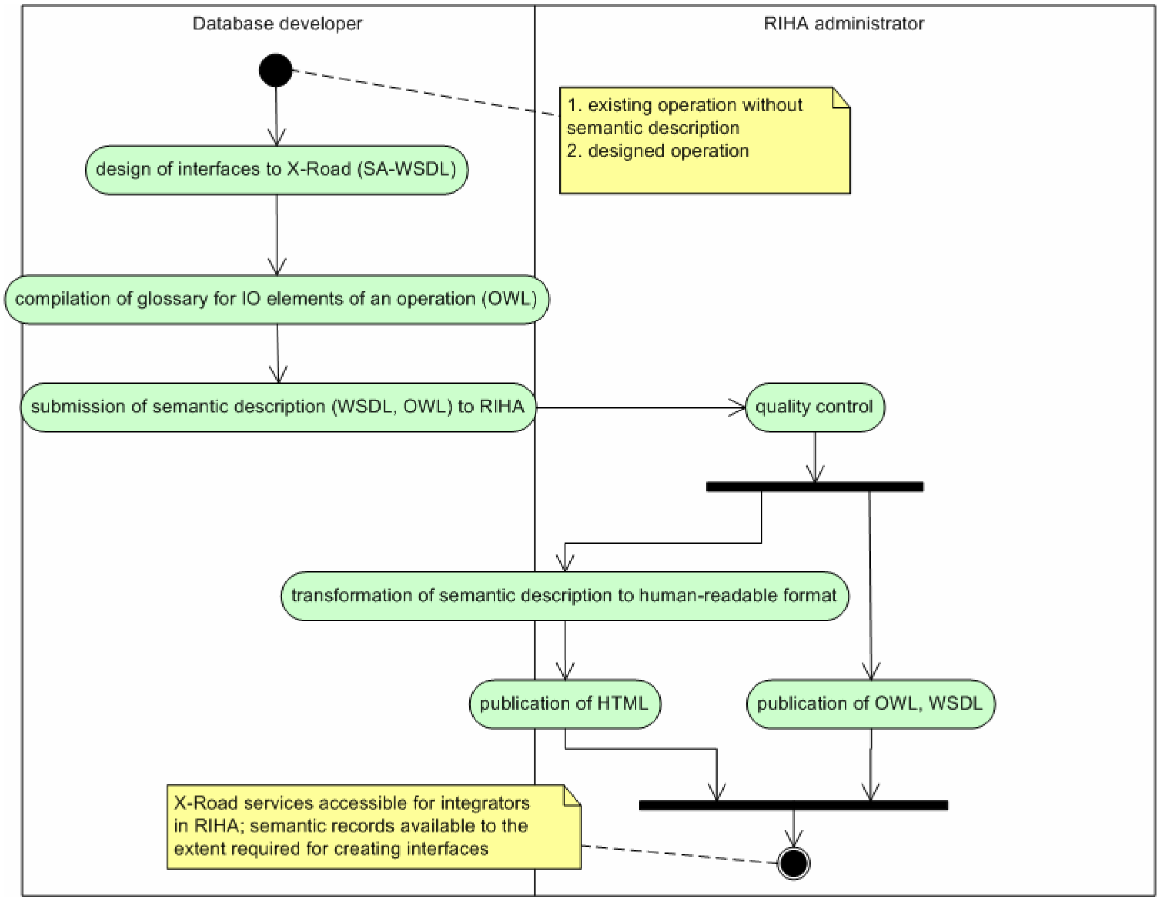

Figure 3. General compilation and publication process of semantic descriptions

6 Underlying reasons for the rules and recommendations included in the instructions

6.1 Principles

The compilation and revision of the methodology and instructions should proceed from the following principles (listed in priority order):

1. Diversity is constrained only to a reasonable extent:
    a. The instructions do not constrain the developer in revising the development process (choosing the type of semantic description). The implementer adds operations to the existing IS development process. The results of operations are described under the rules included in the instructions;
    b. The instructions and training (as far as possible) are independent of products (development resources);
2. The instructions must be understandable, acceptable and easily implemented also for “programmer Mark, and today” [see Chapter 8, Risk ID1] and not only for more able state agencies;

3. Sustainability – the human-interpretable semantic description created on the basis of the instructions must also "function" for machine-processable purposes.

Naturally, the above principles can be revised on reasonable grounds.

6.2 Section on the distribution of glossaries into domains

Piloting should establish whether the distribution of a glossary into domains and the basis for defining domains as proposed in the instructions is altogether necessary (see Chapter 9 "Next steps for the implementation of instructions").

See also the principles for distribution into domains in the Work Planning Information System of ministries.

6.3 Section on the semantic description of web services

Within the scope of Version 1 of the methodology and instructions, operations are limited to web services. See also idea ID7 under Chapter 10.

The semantic description of web services draws from W3C recommendations:

1. WSDL (Web Services Description Language) together with
2. SA-WSDL[5] (Semantic Annotations for WSDL and XML Schema) http://www.w3.org/TR/2007/CR-sawsdl-20070126.

Reasoning

1. Lack of alternatives, i.e. standards of the appropriate level;
2. The above set is very likely to become a global de facto standard with regard to semantic description of web services.

6.4 Section on terminology

Terms and concepts are treated here because even the different context of use of the key concepts complicates understanding between different domains.

A term is a verbal designation of a general concept in a specific subject field or domain. A concept is a unit of knowledge that is created by a unique combination of characteristics and is verifiable. A conception is approximately the same as a concept, but in a broader sense, it is an understanding or a realisation of something – like in a conceptual model. For the purposes of this document, "concept" is used in its general sense.

As regards the practical use of terms, it is important that different users understand the concept (characteristics) of a term in a similar way. Unfortunately, the semantic fields of

[5] SA-WSDL had the status of Candidate Recommendation as at 27 February 2007.

domain terms vary and therefore users of different domains do not understand the meaning of words in the same way. Since the use of this methodology requires cooperation between parties with different roles, mutual understanding is of utmost importance.

The semantic description of web services draws from W3C recommendations:

1. RDF (Resource Description Framework) and RDFS (RDF Schema);
2. OWL (Web Ontology Language) http://www.w3.org/TR/2004/REC-owl-features-20040210/. Levels up to OWL DL (included) should be used in implementation.

Reasoning

1. The above set is very likely to become a global de facto standard with regard to the description of ontologies.

Alternative 1 is to use the UML ontology profile of OMG.

+ UML is more common among Estonian developers.

+ UML is supported by a greater variety of tools compared to RDFS/OWL editors and companies have invested in them.

- UML (and its profile(s)) is not a very common format for ontology.

In short, global developments remain to be seen in that respect. It might happen that the ontologies created in XML-based RDFS/OWL and those based on UML can be automatically interpreted/converted to MOF.

What will change for IS developers? The current (best) practice is that a domain glossary is created as part of the documentation in the course of IS development, e.g. in the form of a table within the specification. In implementing the instructions, developers have to compile the glossary in another format – RDFS/OWL – which can later be transformed into a human-readable glossary (e.g. transformation of XSL into HTML, a plain text etc.).

7 Risks related to the implementation of instructions

Risk ID and description	Likelihood of risk occurrence [0...1]	Significance of consequences of risk occurrence [0…3]	Preventive measures for risk management
[1] “Ontology fever” – excessive concentration on the creation, improvement and tools of ontology	0.3	1	Limited resources of developers
[2] Wrong decisions regarding the use/integration of services and arising from quality problems of semantic descriptions	Nearly 1	0–3	Relevant warning included in descriptions

8 Risks related to the compilation and revision of the methodology and instructions

The following includes a preliminary list of risks related to the compilation and revision of the methodology/instructions and a risk management plan. In case new circumstances should appear, the list of risks and/or risk management plan will be revised accordingly in the course of updating the methodology/instructions.

Risk ID and description	**Likelihood of risk occurrence [0...1]**	**Significance of consequences of risk occurrence with respect to the project objective [0…3]**	**Preventive measures for risk management**
[1] The methodology or instructions are too complicated, impracticable and/or academic – an average "programmer Mark" cannot use them.	0.2	2	Involvement of practicians in the compilation process Iterative testing of the instructions in practice
[2] No use of the methodology or instructions – the trouble of compiling semantic records outweighs the potential benefits	**0.4**	**2**	Following the principle of ease in the compilation of the methodology and instructions Changing users' attitude through training and informing Control mechanisms for the implementation of the instructions
[3] The methodology or instructions are of no help for the contracting entity in evaluating the performing entity's compliance with the methodology or instructions upon developing a database/operation.	0.1	1	Control methods will be developed in the course of practical implementation
[4] The methodology or instructions do not consider the needs and skills of different user groups (programmers/architects/business parties; developers not proficient in Estonian etc.)	**0.3**	**3**	Involvement of practicians in the compilation process Public discussion of the methodology and instructions Iterative testing of the instructions in practice
[5] The scope of the methodology or instructions gets diffused or extended (e.g. from semantic description to general documentation)	0.2	2	Strong contracting entity Internal quality control of the performing entity
[6] The methodology or instructions do not comply with other instructions, standards and practices of IS documentation	0.1	2	Public discussion Internal quality control of the performing entity

9 Next steps for the implementation of instructions

The primary task for achieving semantic interoperability is to compile a human-readable glossary with examples for the input and output data of the system as well as a description of the operation. Such adequate description is necessary for proceeding to the next level:

- Domain and inter-domain ontology;
- Machine-processability: automatic identification, launch and composition of the operation;
- …

The first stage of implementation includes setting the objective of creating a human-readable and interpretable semantic description and not a machine-processable one. The main way to enhance semantic interoperability between organisations or information systems is to:

- Describe semantically the data exchanged and operations performed, i.e. to provide the terms and definitions of these data and operations by opening their meaning;
- Name the data structures and operations in a way that would reveal their nature or function as explicitly as possible.

The activities of the first stage could include the following:

1. Pilot implementation of operations of 2-3 databases by:
   - Choosing databases that provide operations of great interest;
   - Taking an existing databases and operations and updating their semantic description;
   - Applying the bottom-up approach by describing the terms in the glossary serving as the input/output for existing (or designed) web services and not creating an a priori ontology;
   - Establishing roles: domain expert, ontology designer(?), IS developer (architect/designer, programmer);
   - Consultation from the compilers of the instructions, if required;
   - Testing the descriptions and using interested parties who want data from the pilot database, i.e. test the suitability of descriptions for connecting to the database.
2. Presentation of results and lessons learnt;
3. Discussion of possible changes in the next version of the instructions or methodology;
4. Making a decision concerning the publication and adoption of the instructions;
5. PR work related to the instructions: introduction at seminars, trainings;

6. Implementation of the instructions;
7. Regular updating of the instructions.

10 Problems and ideas to be tackled and considered in the next versions of the methodology and instructions

ID	Name and description of the idea or problem	Priority (1–5, 1 = most important)
1	Multilingualism	3
2	Administration of overlapping parts in common domain glossaries	1
3	Description of business processes (WS-BPEL, BPEL4WS etc.), business rules (XSTONE RQL, OCL etc.)	3
4	Semantic description of the compensation mechanism for launching operations	5
5	Machine-processable description of the preconditions, postconditions etc. of operations; description of the technical, i.e. the quality parameters of operations	4
6	Besides the semantic description of operations also the semantics of launching	4
7	Extension of the scope of operations to which the methodology and instructions apply (v1 – web services)	1
8	Administration of changes in the ontology and other administrative processes	2
9	Addition of an example of classifications to the instructions	2
10	Consideration of connecting operations to the document management system or functions of an agency and extending that idea also to the semantics/glossaries. Risks become unfocused and are not directly related to databases	4
11	Connection of the general description of databases with the semantics (of operations)	
12		

Annexes

Annex 1. Overview of source materials on the semantic interoperability of databases and operations

10.1.1 Previous documents drafted in Estonia

Estonian IT Interoperability Framework (Version 2.0; 15.09.2005)

- The Framework outlines the principles of interoperability and the aspect of interoperability between document management systems and geoinformation systems.
- According to Chapter 2.4 “Descriptions and quality of services”, services should preferably contain the following information: (1) the syntax and the protocol of a service (e.g. in case of X-Road services these are given in the WSDL file format); (2) service provision policy (based on which principles, to whom and for which purposes the service is provided), and (3) quality indicators of the service (functionality, reliability and efficiency).
- The service provider must describe the service in a WSDL file.

Legal framework, organisation and action plan of enhancing semantic interoperability XML strategy

- Has no direct impact on the methodology or instructions.

Semantic Interoperability Strategy (Version 0.5; 2005)

- The strategy outlines the key points for semantic interoperability; gives a short description of human- and machine-readable means for presenting semantic description, and envisages an action plan for enhancing semantic interoperability in Estonia. The strategy is meant to be an introduction to and analysis of the field of semantics.
- It does not include a chapter on implementation – this is the task of the present methodology and instructions.
- The strategy considers achieving interoperability to be largely a matter of organisation. Organisation ensures sustainability and quality. Therefore, it is recommended to maintain records about data structures and protocols in accordance with specific guidelines.
- Annex 2 to the strategy, covering the semantic interoperability and action plan, focuses on the semantic description of data and recommends using RDF(S) and OWL as languages.

10.1.2 European Union policies and applications

European Interoperability Framework for pan-European eGovernment services Framework (2004) http://ec.europa.eu/idabc/en/document/2319/5644

- The Framework presents a semantic interoperability strategy: principles, objectives and an action plan. It also outlines means of achieving semantic interoperability and gives examples from the European context. The document does not include a detailed methodology or instructions for achieving semantic interoperability.

- The “Estonian Semantic Interoperability Strategy (Version 0.5)” is generally in line with IDABC’s documentation on the semantic interoperability strategy.

Documentation on the Content Interoperability Strategy:
(1) Working paper V2; (2) Feasibility study for an XML-Clearinghouse
http://ec.europa.eu/idabc/en/document/3875/5644

1. IDABC Content Interoperability Strategy. Working paper (September, 2005)

- The working paper envisages the future of content interoperability and points out that the most important success factors are related to organisational performance. This ensures: (1) Long-term sustainability; (2) User feedback; (3) Open process for making changes; (4) Quality; (5) Use; (6) Focus on process; (7) Introduction of semantics resources; (8) Joint use (EU and Member States); (9) Central role of registers, and (10) Organisational support from the EU: XML Clearinghouse for Semantic Interoperability.

2. IDABC Semantic Interoperability Strategy: The European XML Clearinghouse. Feasibility Study. (December 2005)

- The study explains the role of a clearinghouse. It is similar to the idea of creating an Estonian repository. The objectives are to: (1) Provide for seamless data exchange; (2) Tackle the challenge of multilingualism; (3) Create synergies through mutual exploitation of existing assets, and (4) Support collaboration.
- The study also describes the functioning principles.

Documentation on the infrastructure for cross-border eGovernment services

http://europa.eu.int/idabc/en/document/3760/5585

- The web site defines the architecture, technology and infrastructure necessary for delivering pan-European services. The site contains three documents: (1) ”Architecture for European eGovernment services 1.0” (November 2004) 54 p; (2) ”Technology and market trends for European eGovernment services 1.2” (November 2004) 36 p; (3) ”Infrastructure requirements for European eGovernment services 3.1” (November 2004) 34 p.
- It would serve as a useful input in creating a repository.

Study on Interoperability at Local and Regional Level. Long list with case studies that exist in the good practice framework and are related to the subject of interoperability (May 2005)

http://europa.eu.int/information_society/activities/egovernment_research/doc/case_studies_interoperability.pdf

- The study describes the use of services provided by databases in various EU countries including those where semantic interoperability has been applied.

10.1.3 Other sources

US Government initiative: The Data Reference Model (Version 2.0, November 2005) http://www.whitehouse.gov/omb/egov/documents/DRM_2_0_Final.pdf.

- The Data Reference Model (DRM) is one of the five reference models of the Federal Enterprise Architecture (FEA). It aims at providing cross-usage of data. The DRM covers three areas: (1) data description; (2) data sharing, and (3) data context. Work proceeds from the standard ISO/IEC 11179.
- The third chapter of the DRM deals with data description. The model is divided into different parts: (1) data schema; (2) entity; (3) data type, and (4) attribute.
- Traditional data description is related to semantic interoperability.

Semantic Grid Community Portal http://www.semanticgrid.org/

Pan-European Semantic Interoperability Community of Practice (SICoP) http://web-services.gov/

- This site focuses on improving semantic interoperability and linking semantic data. Among other issues, upper-level ontologies area addressed (candidates include SUMO, DOLCE, OpenCyc).

10.1.4 Overview of the Metadata Registries Standard (ISO/IEC 11179)

ISO/IEC 11179, Information Technology – Metadata Registries (MDR) is a six-part standard that covers the semantics and presentation of data as well as registering data descriptions. These descriptions help to understand and use the meaning of data properly.

ISO/IEC aims at:

- Standardisation of the structure and content of metadata registries;
- Providing access to metadata and searching options through content semantics;
- Promoting understanding and use of data standards;
- Facilitating greater interoperability through the use of standards.

Part 1: Framework – discusses fundamental ideas and concepts of data elements.

Part 2: Classification

Part 3: Registry metamodel and basic attributes

Part 4: Formulation of data definitions – provides guidance on how to develop data definitions.

Five requirements are listed. A data definition shall:

(a) be stated in the singular;

(b) state what the concept is, not only what it is not;

(c) be stated as a descriptive phrase or sentence(s);

(d) contain only commonly understood abbreviations;

(e) be expressed without embedding definitions of other data or underlying concepts.

The document also includes a number of recommendations, examples and counterexamples.

Part 5: Naming and identification principles – provides guidance for the identification of administered items. That part of the standard describes the international registration data identifier (IRDI). Basically, the IRDI is meant for semantic data exchange between information systems. Moreover, the standard states that the administered items described on the basis of this principle can be exploited also between different organisations. Thus, in principle it meets the requirements regarding semantic description.

Part 6: Registration.

In brief

Consistent application of the Metadata Registries Standard would considerably improve the options for documenting (incl. describing) data and thus also semantic interoperability. As the standard dates back before the era of web services, it would be reasonable to adopt some parts of it (e.g. rules for naming and defining; linking the level of data elements to higher levels).

10.1.5 Dublin Core Metadata Element Set (ISO-EVS 15386:2004)

Dublin Core is a standard for the description of metadata elements of any type of information resource. It is quite widespread. For instance, it is integrated to the semantic description of OpenOffice documents. The standard is becoming popular also in Estonia.

The following list includes Dublin Core metadata that might prove useful in compiling descriptions.

Name	Identifier	Definition	Comment
Title	Title	A name given to the resource	Typically, a Title will be a name by which the resource is formally known.
Creator	Creator	An entity primarily responsible for making the resource	Examples of a Creator include a person, an organisation, or a service. Typically, the name of a Creator should be used to indicate the entity.
Subject	Subject and Keywords	The topic of the resource	Typically, Subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.
Description	Description	An account of the resource	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.

Name	Identifier	Definition	Comment
Publisher	Publisher	An entity responsible for making the resource available	Examples of a Publisher include a person, an organisation, or a service. Typically, the name of a Publisher should be used to indicate the entity.
Contributor	Contributor	An entity responsible for making contributions to the resource	Examples of a Contributor include a person, an organisation, or a service. Typically, the name of a Contributor should be used to indicate the entity.
Date	Date	A point or period of time associated with an event in the lifecycle of the resource	Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format.
Type	Resource type	The nature or genre of the resource	Types include general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (e.g. DCMI Type Vocabulary [DCT]). To describe the physical or digital manifestation of the resource, the element “format” is used.
Format	Format	The physical or digital manifestation of the resource	Typically, Format may include the media-type or dimension of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Examples of dimensions include size and duration. Recommended best practice is to select a value from a controlled vocabulary (e.g. Internet Media Types [MIME]).
Identifier	Resource identifier	An unambiguous reference to the resource within a given context	Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system. Examples of formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

Name	Identifier	Definition	Comment
Source	Source	A reference to a resource from which the present resource is derived	The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string or number conforming to a formal identification system.
Language	Language	A language of the intellectual content of the resource	Recommended best practice for the values of the Language element is defined by RFC 3066 [RFC3066] which, in accordance with the standard ISO 639 [ISO639], includes two- and three-letter language codes and sub-codes when necessary.
Relation	Relation	A reference to a related resource	Recommended best practice is to identify the related resource by means of a string or number conforming to a formal identification system.
Coverage	Coverage	The extent or scope of the resource	Coverage will typically include spatial location (a place or geographic coordinates), temporal period (a period label, date, or date range), or jurisdiction (name of an administrative entity). Recommended best practice is to select a value from a controlled vocabulary (e.g. the Thesaurus of Geographic Names [TGN]), and, where appropriate, use named places or periods in preference to numeric identifiers, such as sets of coordinates or date ranges.

When using the Dublin Core set, it is reasonable to create an application profile for the semantic description of the metadata of resources. The application profile of metadata is a set of metadata elements gathered from one or more metadata sets for the purpose of joining these elements into a new entity. The aim of the application profile is to adjust the existing standard metadata set to a particular domain (objective) while preserving the interoperability of different metadata sets.

In addition to Dublin Core, the available set of metadata elements included in the X-Road instructions has been covered.

The following Dublin Core elements are defined as mandatory:

- **Title**
  - Title
- **Description**
  - A description with annotation specifying the element

- **Identifier** (URI)
  - Resource identifier

In future, the introduction of the following Dublin Core elements could be considered:

- **Subject** – a topic preferably chosen from the Estonian Universal Thesaurus[6]
  - E.g. <dc:subject nr1:type="nr2:EÜM">
- **Date**
- May be useful in managing changes but that role should be taken by, for example, a version in WSDL (see also "Relation")
- **Relation**
- May be useful in managing versions, substitutions and relations

These elements are allowed to be used with the following Dublin Core qualifier:

- **isPartOf**

They can be related to elements from other name spaces, such as the X-Road:

- xtee:title = dc:title
- xtee:technotes + xtee:notes = dc.description

The Dublin Core set may be used in descriptions regardless of its own syntax, namely by using the concept of Dublin Core by means of labels developed in Estonia (e.g. <xroad>).

Annex 2. Additional tools and resources for the implementation of instructions and training

10.1.6 Tools used for the description of ontologies and web services:

- SemanticWorks (Altova – XMLSpy)
- Ontology editor Protégé (tutorials available at http://www.co-ode.org)
- OntoEdit

10.1.7 Training materials

- www.w3schools.com/rdf/default.asp
- www.w3schools.com/rdf/rdf_owl.asp

[6] http://eum.nlib.ee/ (2007-01-11).

Annex 3. Background information on the key concepts of the methodology

Definitions of terms are available in the instructions. The following includes background information on more significant concepts.

Service

"Service" is a term used in the service-oriented architecture (SOA).
It is an approach allowing for better use of information resources. The concept of "service" is much narrower than "public service". Furthermore, the meaning of "service" does not overlap in these two terms. Thus, the Ministry of Finance has mainly focused on the quality of service in its instructions on public service standards and compilation guidelines. In principle, services are regarded as tasks contracted out In the SOA, however, services rather stand for the business logic and IT-related options for using the data. Public services or services in general may but need not be implemented as IT solutions in the form of web services.
The main difference is still not conceptual but rather perceptual. In terms of IT, service is perceived in a narrower sense as a single procedure or inquiry which has an input and an output. In reality, it is seen as a process with an operational content carrying a "real" meaning.

For instance, in terms of IT we can talk about a service of personal identification. In real life, no one cares to think that the personal identification carried out in a state agency while tending to some business there is actually a service provided to them. Another extreme case is to talk about services as something very general. For example, schools are providing a schooling service. Obviously, no one regards it as one single service. It gets more complicated when it comes to the standard of services. The free movement of services is yet another thing.

Semantic interoperability

Semantic interoperability is the ability of software systems to make adequate use of data received from other software systems. "Adequate" here means that the data of another system used for connecting software systems is used as the source data in terms of meaning.

The "Estonian IT Interoperability Framework" defines "semantic interoperability" as follows: Semantic interoperability means that operations (events) are comprehensible and reliable also for those not responsible for creating them. For example, a decision must be understood as a decision, a permit must be understood as a permit etc. To this end, semantic interoperability is required at a different level, namely at the level of semantics of data. A prerequisite for semantic interoperability is the reliability of the online service provider in terms of data quality. Semantic interoperability can be divided into (1) the semantics of operations performed, and (2) the semantics of data.