2007-02-22
DRAFT

STANDARDIZATION STRATEGIES
for
SEMANTIC INTEROPERABILITY
in the field of
RISK GOVERNANCE

from the perspective of
WIN (IP in 6FP-ICT for the Environment)
WP 2200 - MULTH

Gerhard Budin, Univ. Vienna

I Introduction, perspectives, and scoping

The following observations are made from my personal perspective due to my horizontal activities over many years in the field of standardization for semantic interoperability. In addition to the WIN project, where semantic interoperability is a crucial topic seen from within the multi-topic cluster domain of risk management, there are more generic activities or activities in related domains that deserve more attention in this context. **Semantic Interoperability Standards (SIS)** cover a broad spectrum of ICT standards, language standards, organizational and procedural standards.

II Taxonomic Overview of relevant organizations and initiatives for SIS

1. European Standardization in the CEN context (Comité européen de normalisation) means the production of European standards (EN). The ISSS within CEN (Information Society Standards System) produces so-called CEN Workshop Agreements (CWAs) as publicly available specifications that might later become full European standards. Within the CEN/ISSS framework there are several thematic areas that are relevant for the domain cluster of risk management:

   1.1. **eGovernment** -> I was part of a project team that was asked to set up an eGovernment Focus Group (EGFG) that will further study the necessities of standardization in this field. Risk management is very often carried out by public authorities at local, regional, national, European and international levels, and in the context of their horizontal eGovernment infrastructures and procedures. Thus the topic of risk management will definitely have to be a topic for eGovernment standards initiatives. As a member of the steering committee of this eGovernment Focus Group I will propose at our first meeting on the 5th of March to include this topic in the work programme and scope of this Focus Group, since a project team will be asked to produce a detailed report on the topics to be specified

   1.2. From a conceptual point of view, there are many "**administrative nomenclatures**" to be standardized or that need a specific semantic interoperability framework to become operational across all domains: "ADNOM" stands for a CEN/ISSS Workshop

(that I happen to chair) that has produced a CWA designed to build an operational infrastructure in Europe for reaching semantic interoperability for all kinds of administrative terminologies. DG Enterprise had sponsored the work of the project team in order to produce CWA 15526:

1.2.1. The main deliverable was CWA 15526 "The Establishment of a European Network for Administrative Nomenclature" with the following main sections

1.2.1.1.: a Survey on existing administrative nomenclatures and similar terminological resources in Europe as well as on the organizations managing these resources

1.2.1.2.a Procedural methodology with principles and recommendations for efficient, collaborative, and sustainable terminology management

1.2.1.3.a Demonstrator implementing the ADNOM approach (using ebXML, Topic Maps (XTM), ISO 16642 (Terminology Markup Framework), ISO 11179 (Data modelling), and other standards), with an "ADNOM Seamless Knowledge Core model" with federated registries, navigation services, etc. an Emerging organizational network of relevant institutions (stake holders), integrating existing networks; dissemination efforts, with ongoing consortium building.

1.2.2. The CWA document also contains a concrete example of mapping multilingual administrative nomenclatures, using a simplified example on names of government ministries and agencies in Austria, France, and Germany linked to the different scopes and responsibilities of these administrative units as far as pension schemes are concerned. In the second phase of the ADNOM project in 2007-2008 risk management terminologies should be treated more specifically. The work performed in WP 2200 of the WIN project (MULTH) directly contributes to this standards-related work by providing a bibliography of relevant language resources in multiple languages. ADNOM in turn can assist the WIN project by assisting in co-locating and inter-linking such resources across ministries and government agencies all over Europe.

1.3. **Cultural Diversity Focus Group** (CDFG): This group focuses on ICT standards and their role in preserving and promoting Europe's cultural and linguistic diversity by representing their languages and cultural specificities on the web and in any IT infrastructure and digital resource (e.g. by a European locale registry, language and script coding standards, etc. As cultural diversity manifests itself obviously in any domain and sphere of life, Risk governance is also an important topic there, similar to all other domains. As a member of this group I will propose to establish a relevant and useful link among WIN and this standards group. ETSI is also involved in this group with its guidelines for cultural diversity management.

1.4. There are numerous other groups in the CEN/ISSS framework that are relevant to the WIN project, together with John Ketchell and his colleagues at CEN HQ we will systematically scan all relevant groups and help create useful links for the WIN community

2. **COPRAS – Cooperation Platform for Research and Standards:** This platform was created under the 6. Framework Programme (FP 6) by CEN, CENELEC and ETSI as European standards bodies, the World Wide Web Consortium (W3C), and the Open Group, and has been operational in helping R&D projects in the ICT sector to contribute effectively to standards production and consistent use of standards. In January the closing conference took place. There is the hope to make use of the results as much as possible. I have started to discuss future options with John Ketchell from CEN for initiating comparable initiatives in FP 7 in order to more specifically integrate research, pre-

normative research, research & development, and basic research in an operational innovation framework. The guidelines that the COPRAS consortium has produced for help in the interaction between standards and research prove very useful in designing more generalized frameworks. The broad spectrum of domains under the risk governance umbrella will certainly greatly benefit from a systematic, horizontal, cross-project co-operation and concertation efforts in order to use available standards where relevant, and to contribute to Semantic Interoperability Standards, e.g. by extending metadata schemata, etc. The WIN project can contribute to COPRAS or its follow-up initiative, but WIN in turn can also profit from using the results of the COPRAS project and should actively promote a suitable follow-up initiative.

3. **Semantic Interoperability Standards in the WIN project and related ICT for the Environment projects:** in all work packages and work areas of the WIN project, relevant standards are used:

3.1. The complete SOA (Service-oriented architecture) as originally specified by WIN together with sister projects ORCHESTRA, OASIS, and co-operating organizations such as ESA (see "Towards an open risk management service architecture for INSPIRE and GMES" includes open ICT and geospatial standards from W3C, OMG, OGC, ISO, OASIS (the Standards Group), WSI, ISO RM-ODP, etc. is a common foundation. (for details see the relevant documents from the websites of these projects and organizations)

3.2. In WP 2200, for instance, W3C standards such as XML, RDF, OWL, and emerging recommendations such as SKOS are actively used. In addition, ISO standards such as ISO 11179, ISO 16642, ISO 12620, and many other relevant language resource management and terminology engineering standards are being actively used. In fact, current work in task 3 of WP 2200 (ontology) is currently contributing to newly emerging Semantic Interoperability Standards (SIS) by joint work between the MULTH team and CRONOS for dynamically linking multilingual terminological resources to WIN data models through XMLS – a promising road to new implementation schemata and patterns for multilingual tagging and annotation. Also the lexicographical aspect of the work in WP 2200 is supported by an official standard – ISO 1951 (I am one of the co-authors of this standard) that lays down XML-based standard principles for the design and layout of (electronic) dictionaries in a markup language called LexML. The Frame-semantic aspect of the semantic modelling work by the MULTH team (A. Rothkegel) is also based on linguistic standards and the ontology-compatible FrameNet initiative. From the language engineering perspective we also use WordNet, a world wide famous lexical semantic net database (Princeton) that is also linked to an upper ontology.

3.3. In the environmental field at large, there is the same awareness of the need of SIS – organizations such as EEA, EPA, UNEP, FAO, JRC, OECD, IPRC, etc. have been working together at regional and international levels to harmonize classification systems, thesauri, terminologies, ontologies, and other knowledge organization systems by operationalizing standards such as XML, RDF, OWL, ISO 11179, ISO 16642, ISO 19112, etc. in the same, harmonized way

4. **SIS at ISO level and in specific support projects for pre-normative research:** semantic interoperability has become of highest concern, not only within ISO, but also among international co-ordination among ISO, United Nations, and many other international and intergovernmental organizations, see for instance the Minutes of the 17th Meeting of the ISO-IEC-ITU-UN/ECE MEMORANDUM OF UNDERSTANDING

MANAGEMENT GROUP (MoU/MG), 16-17 October 2006, at ITU-T headquarters (Montbrillant building), Geneva, Switzerland, where it was agreed to harmonize parallel standards development in terms of "core components" of semantic interoperability. ISO/TC 37 (Terminology and other language and content resources) has been very active in pursuing this high level agreement, since its work focusing on language standards and terminology standards is essentially a horizontal effort to reach global, cross-language, multi-domain semantic interoperability. In addition to this horizontal interoperability there is also what we call vertical interoperability across different levels of abstraction, across the different layers of the Semantic Web, across the different levels of interoperability (syntactic, semantic, pragrmatic), etc. The following groups in the ISO/IEC system, among others, co-operate in pursuing SIS:

4.1. ISO/TC 37 (Terminology and other language and content resources): (the author of this document is chair of Sub-Committee 2 of this group and co-author of many of the standards listed below:)

ISO 704:2000:   Terminology work -- Principles and methods
ISO 860:1996:   Terminology work -- Harmonization of concepts and terms
ISO 1087-1:2000: Terminology work -- Vocabulary -- Part 1: Theory and application
ISO 639-1:2002:   Codes for the representation of names of languages -- Part 1: Alpha-2 code
ISO 639-2:1998:   Codes for the representation of names of languages -- Part 2: Alpha-3 code
ISO 639-3:2007:   Codes for the representation of names of languages -- Part 3: Alpha-3 code
for comprehensive coverage of languages
ISO 1951:2007:   Presentation/representation of entries in dictionaries -- Requirements, recommendations and information
ISO 10241:1992:   International terminology standards -- Preparation and layout
ISO 12199:2000:   Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet
ISO 12615:2004:   Bibliographic references and source identifiers for terminology work
ISO 12616:2002:   Translation-oriented terminography
ISO 15188:2001:   Project management guidelines for terminology standardization
ISO 1087-2:2000: Terminology work -- Vocabulary -- Part 2: Computer applications
ISO 12200:1999:   Computer applications in terminology -- Machine-readable terminology interchange format (MARTIF) -- Negotiated interchange
ISO 12620:1999:   Computer applications in terminology -- Data categories
ISO 16642:2003:   Computer applications in terminology -- Terminological markup framework
ISO 24610-1:2006: Language resource management -- Feature structures -- Part 1: Feature structure representation

Ongoing Projects in the field of language engineering:
ISO/WD 21829:   Terminology for language resources
ISO/WD 24610-2: Language resource management -- Feature structures -- Part 2: Feature system declaration
ISO/CD 24611:   Language resource management -- Morphosyntactic annotation framework
ISO/CD 24613:   Language resource management - Lexical markup framework (LMF)
ISO/WD 24614-1: Word segmentation of written texts for mono-lingual and multi-lingual information processing -- Part 1: General principles and methods
ISO/AWI 24614-2: Word segmentation of written texts for mono-lingual and multi-lingual information processing -- Part 2: Word segmentation for Chinese, Japanese and Korean
ISO/WD 24615:   Language resource management -- Syntactic annotation framework
ISO/AWI 24616:   Multi lingual information framework -- Multi lingual resource management
ISO/AWI 24617-1: Language resource management -- Semantic annotation framework (SemAF) -- Part 1: Time and events

4.2. JTC 1 SC 32/WG 2 (meta data and meta data modelling):

ISO/IEC 11179-1:2004 Information Technology -- Metadata Registries (MDR) - Part 1: Framework

ISO/IEC 11179-2:2005 Information technology -- Metadata Registries (MDR) - Part 2: Classification for administered items (Revision of ISO/IEC 11179-2:2000)

ISO/IEC 11179-3:2003 Information technology -- Metadata Registries (MDR) - Part 3, Registry Metamodel and basic

attributes

ISO/IEC 11179-4:2004 Information Technology -- Metadata Registries (MDR) - Part 4: Formulation of data definitions

ISO/IEC 11179-5:2005 Information Technology --Metadata Registries (MDR) - Part 5: Naming and identification principles

ISO/IEC 11179-6:2005 Information Technology -- Metadata Registries (MDR) - Part 6: Registration

4.3. ISO TC 184/SC 4

4.4. ISO TC 211: with the following standards:
    ISO 6709:1983:   Standard representation of latitude, longitude and altitude for geographic point locations
    ISO 19101:2002: Geographic information -- Reference model
    ISO/TS 19103:2005: Geographic information -- Conceptual schema language
    ISO 19105:2000: Geographic information -- Conformance and testing
    ISO 19106:2004: Geographic information -- Profiles
    ISO 19107:2003: Geographic information -- Spatial schema
    ISO 19108:2002: Geographic information -- Temporal schema
    ISO 19108:2002/Cor 1:2006
    ISO 19109:2005: Geographic information -- Rules for application schema
    ISO 19110:2005: Geographic information -- Methodology for feature cataloguing
    ISO 19111:2003: Geographic information -- Spatial referencing by coordinates
    ISO 19112:2003: Geographic information -- Spatial referencing by geographic identifiers
    ISO 19113:2002: Geographic information -- Quality principles
    ISO 19114:2003: Geographic information -- Quality evaluation procedures
    ISO 19114:2003/Cor 1:2005
    ISO 19115:2003: Geographic information -- Metadata
    ISO 19115:2003/Cor 1:2006
    ISO 19116:2004: Geographic information -- Positioning services
    ISO 19117:2005: Geographic information -- Portrayal
    ISO 19118:2005: Geographic information -- Encoding
    ISO 19119:2005: Geographic information -- Services
    ISO/TR 19120:2001: Geographic information -- Functional standards
    ISO/TR 19121:2000: Geographic information -- Imagery and gridded data
    ISO/TR 19122:2004: Geographic information / Geomatics -- Qualification and certification of personnel
    ISO 19123:2005: Geographic information -- Schema for coverage geometry and functions
    ISO 19125-1:2004: Geographic information -- Simple feature access -- Part 1: Common architecture
    ISO 19125-2:2004: Geographic information -- Simple feature access -- Part 2: SQL option
    ISO/TS 19127:2005: Geographic information -- Geodetic codes and parameters

4.5. Coordination work. These groups regularly meet in the Open Forum for Metadata Registries (OFMR) in order to co-ordinate the ongoing standards work and to harmonize incompatible standards in order to contribute to the common goal, i.e. SIS! An example of pre-normative research carried out by a research consortium is the LIRICS project, co-financed by the European Commission in the eContent Programme. Many of the standards and standards projects under ISO TC 37 are currently being promoted or developed by the LIRICS consortium (http://lirics.loria.fr), the author is a member of this consortium on behalf of his home university. LIRICS includes an Industry Advisory Board with more than 20 company experts from language industry in order to advise the development team to focus on standards that are actually needed and that are directly applicable in various implementations.
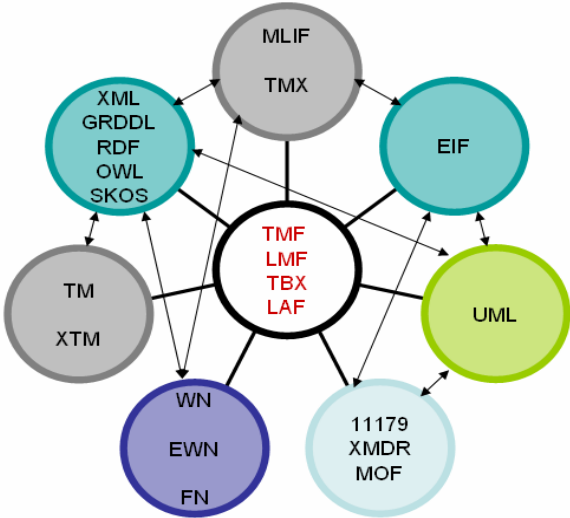
III Conclusions: Semantic Interoperability Standards (SIS) - a global view

The following figure visualizes the multiple connections between many of these (but not all) standards, standards initiatives, organisations, etc. in their common endeavour to reach what I call a terminological language resource interoperability framework, in order to have a leverage structure for creating SIS standards.

These standards are grouped according to their thematic links and their organisational contexts of origin. The figure following this figure is a list with all the acronyms that are used to denote these standards. EIF, for instance, stands for the European Interoperability Framework, a generic architecture with basic principles for interoperability, that was created by DG Enterprise, IDABC, and related departments or DGs of the European Commission. UML, on the other hand, is a very technical standard (Unified Modelling Language) by OMG.

Terminological Language Resource Interoperability Modeling Framework

Acronyms

- XML: Extensible Markup Language
- RDF: Resource Description Framework
- SKOS: Simple Knowledge Organization System
- GRDDL: Gleaning Resource Descriptions from Dialects of Languages
- OWL: Web Ontology Language
- TM: Topic Maps
- XTM: XML Topic Maps
- WN: WordNet
- EWN: EuroWordNet
- FN: FrameNet
- TBX: Termbase Exchange Format
- TMX: Translation Memory Exchange Format
- MLIF: Multilingual Information Framework
- TMF: Terminology Markup Framework
- LMF: Lexical Markup Framework
- LAF: Linguistic Annotation Framework
- EIF: European Interoperability Framework
- ISO 11179: Metadata Registry (MDR) standards framework
- XMDR: Extended Metadata Registry
- MOF: Metamodel Object Facility
- UML: Unified Modeling Language

Concertation work on SIS

Concertation work among ongoing FP 6 projects in the ICT for the Environment sector should now focus on SIS, on implementing them in SOA, and all other architectures such as INSPIRE, GMES, etc.

In FP7 we are all starting now with new initiatives, follow-up projects or spin-off projects in various sectors of FP 7 to further pursue this common goal.