Interoperability of Metadata

Semantic Interoperability

ISO TC 37 Secretariat
c/o Infoterm

Christian Galinski

Bamako, Mali 2005-05-06/07
Overview - semantic interoperability - how?

- The problem: Data integration and interoperability
- Metadata + data models + metamodels ...
- **Content** = representation object
- Content entities
- Content development
- ISO/ TC 37
- ISO/ TC 37/ SC 1 ~ 4
- ISO/ TC 37 Outlook
- Metadata and metadata repositories
- Content repositories
- Workflow methodology
- System of federated repositories/ registries
Metadata approach

- Metadata = data about data (types)
- Metadata scheme of a DB = ontology / terminology
- Data (values) = re-interpretable information
- Data / values = digital content
- (metadata) Names ➔ naming conventions
- Metadata description = definition, attributes, ...
- Metadata must be normalized (standardized)
- Values can be standardized
- Common vs. local needs ➔ core elements
- Core elements ➔ ontology*

ISO/ TC 37 - Bamako 2005-05-06/ 07
... thus eCl@ss supports the entire process chain of material management, including eCommerce.
### State-of-the-art

#### METHODOLOGY

<table>
<thead>
<tr>
<th>ISO 16642*</th>
<th>(family of) metamodels*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 12200**</td>
<td>Datamodels**</td>
</tr>
<tr>
<td></td>
<td>eBusiness</td>
</tr>
<tr>
<td>ISO 12620***</td>
<td>Domain data dictionaries***</td>
</tr>
<tr>
<td></td>
<td>Data categories</td>
</tr>
</tbody>
</table>

#### APPLICATIONS

<table>
<thead>
<tr>
<th></th>
<th>DDDs</th>
<th>DDDs</th>
<th>DDDs</th>
<th>DDDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain data dictionaries***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>DDDs</td>
<td>DDDs</td>
<td>DDDs</td>
<td>DDDs</td>
<td></td>
</tr>
</tbody>
</table>

<Basic principles and requirements concerning multilingual e/m-content development, data categories/metadata, data modelling, rules for repositories (maintained in MAs/RAs/Reg’s)

*ISO 16642 TMF; ISO 10303-11 EXPRESS; ISO 10303-21 SDAI; …
**ISO 12200 MARTIF; ISO 13584-42 PLIB ~ IEC 61360-2
***ISO 12620 Data categories; ISO 13584-511 Fastener dictionary; IEC 61360-4 Core dictionary; …

ISO/TC 37 - Bamako 2005-05-06/07
CONTENT = representation objects?

content (at the level of concepts): representation of knowledge/meaning

Any content entity / unit / item (=representation object) can be:

• representation of ... tangible/intangible “information object”

• content produce digital or non-digital product

• meta-content for describing/ordering/accessing content ...
CONTENT ENTITIES

e.g. entry in a product catalogue

- **Name of company** (@ enterprise)
- **Name of product** (model) (™ enterprise)
- **Generic name of product** (e.g. © Harmonized System)
- **Class** (name under which the product falls) (e.g. © eCl@ss)
- **Verbal/ textual description** (© enterprise)
- **Picture** (© enterprise)
- **Technical data**
  - (unified) branch properties (e.g. © OAGi)
  - Standardized characteristics (e.g. © DIN)
  - Enterprise product specific data (e.g. for collaborative business)
  - Enterprise internal data (maybe confidential/secret)

ISO/TC 37 - Bamako 2005-05-06/07
Workflow management for content development:
- net-based, distributed, cooperative creation of structured content

- multilingual
- multimodal
- multimedia

Re-use in applications:
(based on the “single-source” principle)
- eLearning
- eGovernment
- eHealth
- eBusiness
- other e...s

- multi-channel output
- accessibility requirements
eContent → mContent

- too many standards e.g. in eBusiness, e...?
- standards lacking / insufficient?
- → standards concerning content & semantic IO:
  - content entities / items / units
    - metadata
    - + data modelling principles
    - + working methodology
  - repositories / registries (of metadata and data)
    - organization
    - operation (incl. maintenance procedures)
    - technology (incl. design of support tools)
    - business models (incl. IPR/copyright management / DRM)
DATA MODELS & METADATA

1. ISO/TC 37 ➔ multilingual mContent development
   (incl. language resources and LR management)

2. application areas: eBusiness, eHealth, eLearing,
   eGovernment, product data management, ...
   JTC 1/SC 2, 7, 22, 29, 31, 32, 34, 35, 36; ISO/TC 154;
   ISO/TC 184/SC 4; ISO/TC 215; IEC/TC 3, 93

3. other initiatives and consortia:
   OIDDI - Open and Interoperable Data Dictionaries Initiative
   OASIS, OMG, OAGi, ...

 ➔ (informal) coordination & harmonization: MoU/MG
ISO/TC 37 ↔ ISO/IEC-JTC 17SC 32

- ISO/TC 37: Terminology and language and content resources
- ISO/IEC-JTC 1/SC 32/WG 2: Metadata

Terminology in ISO/TC 37 → metadata methodology
- Terminology as knowledge representation
- Terminologies as means of domain communication
- Terminologies as means of access to other kinds of information (objects)
- Terminologies as means of knowledge ordering at micro-level

ISO/TC 37 - Bamako 2005-05-06/07
Meta-model

Terminological Data Collection (TDC)

Global Information (GI)  Complementary Information (CI)

Terminological Entry (TE)

Language Section (LS)

Term Section (TS)

Term Component Section (TCS)
<termEntry id="ID67">
  <descrip type='subjectField'>hydrography</descrip>
  <descrip type='definition'>Bouys, beacons, fog signals ...</descrip>
  <langSet lang='en'>
    <tig>
      <term>aid to navigation</term>
      <termNote type='termType'>fullForm</termNote>
    </tig>
  </langSet>
  <langSet lang='es'>
    <tig>
      <term>Ayuda a la navegacion</term>
    </tig>
  </langSet>
</termEntry>
STRATEGIC PARTNERSHIP (1)

Fundamental methodology standards of TC 37:

- **ISO 704:2000** Terminology work – Principles and methods
- **ISO 10241:1992** International terminology standards – Preparation and layout
- **ISO 12620:1999** Computer applications in terminology – Data categories
- **ISO 16642:2003** Computer applications in terminology – Terminological markup framework
- **ISO 15188:2001** Project management guidelines for terminology standardization
STRATEGIC PARTNERSHIP (2)

Basic standards of JTC 1/ SC 32/ WG 2:

- **ISO/IEC 11179** Information technology – Metadata registries (MDR)
- **ISO/IEC 14957** Information technology – Notation of format for data elements
- **ISO/IEC 19763** Information technology – Framework for metamodel interoperability
- **ISO/IEC 19773** Information technology – Metadata modules (MM)
- **ISO/IEC 20943** Information technology – Procedures for achieving metadata registry content consistency
- **ISO/IEC 20944** Information technology – Metadata registry interoperability and bindings (MDRIB)
Metadata (&) Interoperability

metadata = types of data?

- **Terminology of semantic interoperability**
  - “attribute”, “value”, “characteristic”, “property”, etc.

- **Interoperability of metadata (registries)**
  - requirements, rules, guidelines, ...

- **Role change**: e.g. attribute → metadata, attribute → value, ...

- **Relations between attributes, etc.**

- **Core elements / components / ...**
  - OWL, Dublin Core, ...

- **Definition of metadata = terminology?**
  - naming principles, definition, etc.

- **Data models & formats**
  - XML, UML/UBL, etc.

- **(multiple) Classification**

- **Multilinguality, multimodality, ...**
SEMANTIC INTEROPERABILITY

Standardization – Top-down

- Harmonization of metadata & repositories
- Unification of principles and methods of data modelling
- Standardization of meta-models
- Standardization of workflow methodology

Standardization – Bottom-up

- Product classification - ontologies
- Product identification - terminologies
- eCatalogue data - LRs

by using net-based distributed cooperative working methods – peer2peer
TYPES OF REPOSITORIES

- Classification etc. → multiple
- Properties → acc. to types
- Data dictionaries → acc. to types
- Metadata → acc. to types
- Terminologies → acc. to types & domains
- Ontologies → acc. to types & domains

→ huge amounts of repository items
to be taken care of in federated registries
### MAs, RAs, Registries (for repositories)

<table>
<thead>
<tr>
<th>Maintenance Agencies</th>
<th>Registration Authorities</th>
<th>other kinds of registries</th>
</tr>
</thead>
<tbody>
<tr>
<td>software components</td>
<td>ontologies</td>
<td>data structures/ datamodels &amp; XML schemas</td>
</tr>
<tr>
<td>communication protocols, messages</td>
<td>data structures/ datamodels &amp; XML schemas</td>
<td>metadata/ data categories</td>
</tr>
<tr>
<td>interfaces, interface elements</td>
<td>metamodels</td>
<td></td>
</tr>
<tr>
<td>data dictionaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td>certain types of content items: authority data, attributes, values, proper names, ...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- high degree of authority and high stability over time
- high degree of consistency and some stability over time
- lower degree of consistency and stability over time

- networking of distributed repositories e.g. in the form of **federated DBs**
- and net-based distributed cooperative content development & maintenance
federation of repositories

CONTENT ENTITY + ATTRIBUTES

content entity

ADMIN + CLASS

GRM

SRC

+IPR
Different representations for one conceptual entity

<table>
<thead>
<tr>
<th>ADMIN + CLASS</th>
<th>Representation 1</th>
<th>Representation 2</th>
<th>Representation 3</th>
<th>Representation 4</th>
<th>Representation 5</th>
</tr>
</thead>
</table>

ISO/TC 37 - Bamako 2005-05-06/07
Link to grammar repositories

GRM = grammar
- Word stem → other entities …
- Endings
- Term elements → other entities …
- Gender
- Collocation
- UK pronunciation → speech data
- US pronunciation → speech data
Link to source data repositories (1)

SRC = published source (and usage conditions)
(e.g. according to ISO 12615)

- Coded author ➔ bibl. DB …
- Coded title
- Coded publisher ➔ publ. DB …
- Year of publication
- Page of citation
Link to source data repositories (2)

SRC = expert as source

- Coded expert
- Coded subject field
- Coded institution
- Date of information
- ...

→ expert + address DB …
→ institution + address DB …

- Title(s), position
- Privacy requirements
- Confidentiality
- …
Federated repositories

Federation of repositories:

- **Repositories of defined types of content**
  - based on metadata (registries)

- **Defined types of content → content repositories**
  - maintained by the community best fit for

- **Different language versions of such repositories**
  - maintained by the language community interested

- **Federating different types of content**
  - semantic interoperability?

**semantic interoperability ↔ federated repositories**
Content and content management

Re-usability (by single source) across technical platforms, (organization, language) barriers, …

Resource sharing → efficient use of WWW horizontal / diagonal re-use of content entities / units / items in all kinds of e…s:

- eLearning
- eScience
- eGovernment
- eHealth
- eBusiness
- eProcurement

→ONE methodology
EU Interoperability Framework (EIF)

- **Technical interoperability**
  - Hardware-related: ICT hardware
  - Software-related: software/network architecture, ...

- **Semantic interoperability**
  - Syntactic interoperability (*ICT approaches*)
  - Conceptual interoperability (*content point of view*)
  - Pragmatic interoperability (*content point of view*)

- **Organisational interoperability**
Semantic Interoperability

- Technical interoperability
- **Semantic interoperability**
  - Syntactic interoperability *(ICT approaches)*
    - basic principles and requirements }standard(s) 3
    - syntactic communication protocols }{(maintenance) 2 }
    - messages, interfaces, etc. }{(maintenance) }
    - data dictionaries (1) }{(maintenance) 1 }
    - metamodels }MAs }
    - data modelling, XML schemas }RAs }
    - metadata/ data categories }RAs }
    - certain data elements, data dictionaries (2) }registries }
    - basic principles and requirements }standard(s) }
  - Conceptual and pragmatic interoperability *(content point of view)*
    - Organisational interoperability
    - Organisational interoperability

ISO/ TC 37 - Bamako 2005-05-06/ 07
Semantic interoperability standards

- Content-related requirements
- Workflow methodology
- Metadata
- Metadata repositories
- Data modelling principles and requirements
- Micro data models
- Metamodels
- Content repositories
- Federation of repositories
- ...

ISO/TC 37 - Bamako 2005-05-06/07
ISO/IEC policy concerning repositories

- Typology of content entities / units / items
- Different kinds of repositories
- Different kinds of MAs, RAs and other registries for content/data (according to type of content etc.)
- Different business models incl.
  - Maintenance, distribution, payment etc.
- Central ISO/TDB for standardized terminology
- New workflow methodology for net-based distributed cooperative content development (and maintenance)

ISO & IEC to develop a comprehensive policy and a coherent strategic framework of federated repositories for the sake of global semantic interoperability
CONTENT POLICIES

**USER position**: wants content → NOT hardware & NOT software

**UN position**: facilitation of trade

**UNESCO position** (since 2003): Recommendation 32C/R41
- cultural and linguistic diversity
- bridging the digital divide (incl. computer/media/content literacy)
- universal accessibility (incl. people with special needs)

**EU position**:
- e-content → m-content (incl. multilinguality)
- e-business: technical, legal, financial, applications, business, ... aspects
  + multilinguality, + terminology, + eCataloguing, + eClassification
- standardization (in/for eBusiness): INTEROPERABILITY
- EU Interoperability Framework (EIF)
CONFERENCES

- Terminology Summer School
  - Cologne (Germany) 2005-07-18/22

- TAMA 2005 “Terminology in Advanced Management Applications”
  - Wiesbaden (Germany) 2005-11-09

- TKE 2005 “Terminology and Knowledge Engineering”
  - Copenhagen (Denmark) 2005-08-15/19

- OFMR 2006 “Open Forum on Metadata Registries”
  - Japan 2006-03-20/22
Thank you for your attention

ISO/TC 37 Secretariat:
SECRETARY: Christian Galinski
CHAIRMAN: Håvard Hjulstad (SN)
ADDRESS: Aichholzgasse 6/12
A-1120 Vienna - Austria
Tel: +43-1-817 44 99
Fax: +43-1-817 44 99-44
infopoint@infoterm.org
http://www.infoterm.info