CEN MetaLex
Facilitating Interchange in E-Government

Alexander Boer
aboer@uva.nl
MetaLex

- Initiative taken by us in 2002
- Workshop on an open XML interchange format for legal and legislative resources
  - [www.metalex.eu](http://www.metalex.eu)
  - [www.cen.eu](http://www.cen.eu)
- Common document format, metadata set, ontology, naming convention, and processing model for documents and metadata
- CEN/ISSS publicly available specification (2006 & 2010)
Problem context

- Bibliographic information about legislation has to meet very specific & peculiar requirements
- Jurisdiction-specific & producer-initiated XML standardization
- User organizations & software developers increasingly deal with legislation from various publishers & jurisdictions
- No standardized view on legislation means no specialized legal software
Current uses of MetaLex

- Norme in rete (Italy) implements it
- Akoma Ntoso (Pan-African Parliamentary Information) implements it
- Single legislation service (UK) is a compatible metadata delivery framework
- European Parliament will use it internally (AT4LEX)
- `doc.metalex.eu` makes Netherlands corpus available as generic MetaLex (showcase)
- Various authoring tools and CMS support it
What does using MetaLex mean?

- XML standards that extend MetaLex XML schema & conform to MetaLex specification
- XML standard transformation conformance
- Metadata standards extending MetaLex metadata
- For tools:
  - MetaLex aware metadata processing & citation dereferencing
  - Importing generic MetaLex XML or an implementation of it
Lessons learned: important concepts in MetaLex

- Content models
- Bibliographic identity
- Naming conventions
- Legislative events and actions
- Linked open data
Content models

- MetaLex schema defines generic content models and generic elements
- Content models are purely syntactic formulations: only the structure of contained matter should determine content model
- E.g. containers, blocks, inlines, milestones, metadata carriers & arbitrary quoted structures
- MetaLex schema conformance usually requires no changes to XML structures
Naming conventions

- Adherence to an IRI reference-based, open, persistent, meaningful, memorizable and “guessable” naming convention based on provenance information

- Provenance information is either encoded into the IRI, or associated to it as a set of key-value pairs (metadata)
  - E.g. PURL, relative URI, URN, OpenURL
Bibliographic identity

- The naming convention must reliably distinguish (FRBR) works, expressions, manifestations, and items
  - sets and subsets of key-value pairs
- In law, original expressions are temporal consolidations, language variants, and authorized translations
- In law, retroactive corrections and annulments of modifications lead to *ex tunc* consolidations (two different branches)
Stratification of bibliographic entities
Why works, expressions, manifestations, items?

- Knowledge representation is about expressions
- Legal citation is usually to works, sometimes expressions
- Inclusion of expression or manifestation or item is a different thing
- Editing takes place on manifestations
- HTTP dereferencing is of items
- Metadata confusion
  - E.g. all have been created at some date
Simple bibliographic identity example

- `/ukpga/1985/67` is a work
- `/ukpga/1985/67/2003-04-01` is an expression
MetaLex compliant naming conventions

- **Names are used for**
  - Document self-identification
  - Citation of other documents
  - Inclusion of document components

- **Name is either formatted as an**
  - Internationalized Resource Identifier (IRI) reference, or
  - As a set of subject-predicate-object triples
  - A GRDDL translator extracts name RDF metadata triples *and* constructs a name IRI reference for each name encountered in the source
MetaLex compliant naming conventions

- **GRDDL translator requirement:**
  1. \( N(i) \) is a name set of triples about opaque IRI \( i \)
  2. \( n \) is the naming IRI reference that encodes the same name as \( N(i) \)
  3. From \( n \) construct \( N(n) \)
  4. From \( N(i) \) construct \( n \)
  5. Enforce that \( N \) is equivalent to \( \{n\} \)
    - Hence if \( N(i) \) then \( i = n \) etc.
    - Use OWL2 for the purpose
    - \( \{n\} \) is a so-called nominal concept
Constraint enforcement (in OWL2)

\{v_1/v_2/ \ldots /v_n\} \equiv \exists P_1.\{v_1\} \cap \exists P_2.\{v_2\} \cap \ldots \cap \exists P_n.\{v_n\}

Let for instance \(i_e\) be /\(v_1/v_2/ \ldots /v_n/v_1/v_2/ \ldots /v_n\), the IRI reference to an expression, and \(i_w\) be /\(v_1/v_2/ \ldots /v_n\), the IRI reference to the work it realizes.

1. \(\{i_e\} \equiv \exists \text{metalex:realizes.}\{i_w\} \cap \exists \text{metalex:resultOf.}(\exists P_{e_1.}\{v_1\} \cap \exists P_{e_2.}\{v_2\} \cap \ldots \cap \exists P_{e_n.}\{v_n\})\)

2. \(\{i_w\} \equiv \exists \text{metalex:resultOf.}(\exists P_{w_1.}\{v_1\} \cap \exists P_{w_2.}\{v_2\} \cap \ldots \cap \exists P_{w_n.}\{v_n\})\)
Common conformance issues in IRI names

- The name of a work, expression, or manifestation is a proper subset of the name of another entity within the same ontological level;
- The name of an item is equal to that of a manifestation, or the manifestation claims to identify itself as an item;
- The name of a manifestation is equal to that of an expression or does a bad job of uniquely identifying the manifestation among other manifestations of that expression;
- The name of the work and the initial expression are the same.
Fragment identification

frbr:realizes

owl:sameAs

AE6

B9C

3F5

...
Event descriptions

- Organize metadata about legal documents around events and acts
  - **Events**: interesting changes
  - **Acts**: events intentionally brought about by agents.

- This makes sense from both the point of view of
  - knowledge representation tactics, and
  - legal theory.
Argument from Legal Theory

- Institution/constitutiveness interpretation of legislation (or contracts, or driver’s licenses, tax statement forms:
- One undertakes a legal act on the institutional level by producing a written statement in accordance with a certain procedure.
- The document is the physical residue of the intentional act: it functions as **physical evidence that a legislative act** that modified institutional reality **happened**, and
- it **declares the intent** of that act.
- An “Act of Parliament” is the **physical result** of that **act** of Parliament.
A particular metadata description is usually about a snapshot of some entity taken in a particular state ➞ perceived stasis of the entity over a time interval that does not take account of changes that are outside the domain of interest.

The granularity of that snapshot varies across metadata vocabularies, depending on the targeted community.
To a community that works with certain legislation daily, the insertion of a new provision is for instance an important event to be noted, and even to prepare for;

For the casual reader it just happens to be one of the many constituting parts of that document at the moment of consulting, a casual confrontation with its current state.

The more you know of something, the more interesting events you take note of...
Argument from Knowledge Representation Tactics

- For establishing semantic interoperability between metadata vocabularies exploit the fact that people, organizations, places, dates, events, etc., exist in all domains.
- The event/act plays the mediating role between these entities and the resource the metadata description is about.
Argument from Knowledge Representation Tactics

- E.g. author, publication date, and publication channel information are participants in the publication (promulgation) event.
- Coherence between the old consolidation, the new consolidation, the modifying legislation, the modifying authority, and the modification date.
- Modification date is
  - Date of enactment of modifying source
  - End date of old consolidated version
  - Start date of new consolidation version
Argument from Knowledge Representation Tactics

- Unnecessary duplication
- More opportunities for error
- Incompatible descriptions (perspectives)?
- Unnecessary maintenance
- Events as a conceptual coat rack for missing information that is for preparation for future legislation.
  - Date of enactment of draft versions is unknown, but ordering time constraints is essential for simulation
Events in doc.metalex.eu

The date at which the expression was created
"2009-10-23"^^xsd:date

rdf:value
sem:hasTimeStamp
time:inXSDDateTime
time:Instant

rdf:type
sem:timeType

ml:Date

sem:EventType

ml:LegislativeModification

opmv:Process

http://doc.metalex.eu/id/date/2009-10-23


The process that generated the expression

opmv:wasGeneratedAt

opmv:wasGeneratedBy


The expression (version) URI of a regulation

http://doc.metalex.eu/id/event/BWBR0017869/2009-10-23

The creation event of the regulation

opmv:Artifact

ml:BibliographicExpression

ml:resultOf

The event type of the regulation

sem:hasTime

The creation event of the event
Linked open data

Recommended (W3C) best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF.
Linked open data
MetaLex as linked open data

- Storing signed printouts of legislation in a safe ➔ syntactic criterium for authenticity
- Only specification + authentic manifestation required to know the expression?
- Linked open data depends on the availability of Internet infrastructure for dereferencing and logical inference
- Authentic manifestation as envelope for all relevant metadata?
100,222,683 triples

This service hosts (almost) all Dutch national regulations in CEN MetaLex XML and as RDF Linked Data. Current coverage is 28,752 document versions, covering practically all regulations available through http://wetten.overheid.nl

Information

This service hosts (almost) all Dutch national regulations in CEN MetaLex XML and as RDF Linked Data. Current coverage is 28,752 document versions, covering practically all regulations available through http://wetten.overheid.nl

This server was developed by Rinke Hoekstra and is maintained by the Leibniz Center for Law of the Universiteit van Amsterdam.

The regulations were automatically converted by means of a (hopefully) generic conversion script, that is able to perform a similar conversion for new legislative.

MetaLex and Linked Data

CEN MetaLex is a standard for how sources of law and references to sources of law are to be represented in XML. It is an interchange format, a lowest common denominator for other standards, intended not to replace jurisdiction-specific standards and vendor-specific formats in the publications process but to impose a standardized view on legal documents for the purposes of information exchange and interoperability in the context of software development.

Linked Data is a W3C sanctioned approach to publishing metadata on the Web.
Lessons learned: important concepts in MetaLex

- Content models
- Bibliographic identity
- Naming conventions
- Legislative events and actions
- Linked open data