Introduction

- The analysis of European (and other) standards have shown some similarities, many differences
- Three options:
  - Creating: scrap all of them, and invent a new one.
  - Adopting: one of them is assumed better than the others, and chosen.
  - Adapting: define a common ground whereby all existing standards can be reflected and used. A common ground.
- Suggestion: Adapt. Easier to understand, accept, convert. Difficult to interoperate, but we can work on that.

An XML Language for Estrella

- Basic principles proposed
- Meaningful semantic blocks
- Identification of text elements
- Knowledge representation
- Functionalities of the Content Management System

Basic Principles (1/4)

1. Reliance on existing standards
   - Web-related: XML, XML Schema, XML Namespace, RDF, OWL, URIs, etc
   - Legal markup oriented: NIR, MetaLex, Akoma Ntoso, Lex Dania, etc.
2. Distinction between
   - ***Content***: All and only the text that has been approved by the promulgating body. Strict interpretation of text.
   - ***Presentation***: style and layout of actual publication body (e.g. Official Gazette). Assumed to be derived from meaning, expressing meaning, but not, per se, part of content approved by emanating body.
   - ***Metadata***: any further content added by editorial board before publication. Includes markup and proper metadata.

Basic Principles (2/4)

3. Strong naming policies
   - At the document level, adopt systematically a syntax based on URIs.
   - At the internal level, adopt systematic ids to refer to parts and fragments.
4. Self containment
   - Fundamental for long-term preservation of documents
   - All XML documents contain all the information needed to use them, and such information is kept with the at all times
   - Even when referring to external elements (e.g., a name database), sufficient information (i.e., not just the record id) must be stored in the document

Basic Principles (3/4)

5. Strong distinction between
   - ***data types***: the content modes need to be few and well understood and very flexible
   - ***element names***: the actual name need to be rich and fully describing the actual nature of marked up fragments
6. Strong ontological structure for metadata
   - A full ontology of classes, with clear distinction and no overlap of relations and literal values
   - Full use of OWL and RDF for formalization
   - TBD: a syntax for placing such metadata in XML documents

Basic Principles (4/4)

7. Extensive but constrained extensibility (***genericity***)
   - We will not be able to capture all possible semantically-relevant elements in a single vocabulary.
   - Some extension need to be provided
   - Complete extensibility gives too much freedom and prevents interoperability
     - Containers for metadata
     - Generic elements (assuming meaning through a class attribute)
     - Reliance on foreign namespace for other kinds of extension (e.g., MathM for Math formulas, SVG for drawings, etc.)
8. Modular organization for schemas
   - A single schema with many subcases rather than many schemas

Meaningful Semantic blocks

- ***Containers***: list of different elements
- ***Hierarchical elements***: Russian dolls-style
- ***Paragraphs***: containers of text and inline elements, organized vertically
- ***Inline elements***: containers of text within paragraphs (no breaking of lines), with special meaning - typographical or semantic
- ***Milestones***: empty elements
  - ***Placeholders***: individual locations within content
  - ***Metadata elements***: outside of content, values specified as attributes

N.B.: Separation of metadata and content markup

Identification of structures

- Identifiers must be used ***everywhere***
- For documents and all class instances, URIs:
  - Permanent, readable, hierarchical, understandable
  - URNs (as in NIR) or PURL (as in Akoma Ntoso)
- For elements in XML documents, *id*s:
  1. ***Individuals*** (at most one instance: preamble, conclusion, etc.)
  2. ***Unnumbered repeatables***: no explicit numbers, many instances. E.g., paragraphs, references, etc.
  3. ***Globally numbered repeatables***: numbers exist in documents, and start at beginning of document regardless of hierarchical structure. E.g., articles, attachments
  4. ***Locally numbered repeatables***: numbers exist in documents, and start at the beginning of containing element. E.g., art, section, etc.

Knowledge representation

- Accompanying metadata must be
  - rich,
  - complete,
  - ontologically sound,
  - extensible,
  - authored,
  - versioned,
  - (digitally signed?)
- A full set of classes need to be devised (LKIF)
- Careful separation between classes and properties
- Careful subclassing of master classes

We need a mechanism for expressing metadata info in documents (remember self containment!)

An initial set of classes

- A basic set of classes describing our documents, loosely based on FRBR structure:
  - Source of law (work)
  - Version (expression)
  - Variant (manifestation)
  - File (item)
- Other possible classes:
  - Normative system
  - Folder
  - Content Components (at all levels)
  - Roles and Actions
  - Agents (individuals and organizations)
  - Places and times

For instance… (1/2)

- ***Addressee***: Someone to whom a source of law is addressed. A property of work and/or expression
- ***Public body***: A body created by an act of law. A subclass of organization.
- ***Public Decision***: A written decision of a public body. A type of document (a subclass of work)
- ***Public Act***: An act that can only be performed using a public competence ??? Act as in source of law, a subclass of document. Act as in action, a subclass of action.
- ***Competence:*** the power/right to perform certain acts. A property of organization and role.

For instance… (2/2)

- ***Public Competence***: the power/right to perform certain public acts. A subproperty of competence
- ***Legislative Competence***: the power/right to legislate. A subproperty of competence
- ***Legislator***: A public body with legislative competence. A subclass of role
- ***Assignment, Delegation, Subdelegation, Mandate (of Competence)***: A public act that assigns, transfers, lends the competence to perform a public act to a public body. ??? Again: Act as in source of law, subclasses of document. Act as in action, subclasses of action.

Content Management System

- A first organization in four separate scopes
  - ***Editing***: for both official and non-official drafting (e.g., drafting offices in Parliaments as well as private publishing houses)
  - ***Workflow support***: for the tracking and control of document generation flows in Public Administration
  - ***Consolidation***: automatic and/or semi-automatic generation of current law text
  - ***Publication***: both on paper, web, and new and unforeseeable media

Conclusions

- Strong support for correct content markup
- Strong separation of metadata and content
- Support for foreseeable and unforeseeable tasks
- Strong ontology for description of document relationships (and legal content?)