Metadata in Akoma Ntoso: an introduction

Prof. Fabio Vitali
University of Bologna

Data and metadata (again)

- Authors' contribution: data
  - The words and punctuation and breaks, exactly as have been written and accepted by the original author (with legislation, the legislative body)
  - Very little control in terms of content, structure, organization
- Editors' contribution: metadata
  - Anything that the publisher or any intermediate content producer feels appropriate to add or complement the content.
  - Publication data. Lifecycle information. Provisions' analysis. Editors' footnotes.
  - Publishing organization can impose tight control on metadata, and provide much information.

Metadata in Akoma Ntoso

- We define metadata all editorial content (i.e. content added by the editorial process out of Parliament rooms)
- Vice versa, all actual content of the document needs to have a place in the appropriate content sections.
  - ***Descriptors***: i.e., a set of meta information providing info about the document and its publication and edition details, including its official promulgation date, its official network address, and so on.
  - ***Lifecycle***: the lifecycle element provides information about the events that the document has undergone, and references to the documents that have caused these events..
  - ***Notes***: the text of the editorial notes that might be produced to comment and expand the actual text of the document. Note references inside the text point to notes contained here.
  - ***Proprietary***: this subsection allows any additional metadata to be specified in any order and vocabulary.
- The development of the meta section is not finished yet. Support for Dublin Core and Eurovoc-Africa is granted.

Metadata schemas

- Metadata on a document is collected in a number of assertions:
  - The publication of the act happened on 21/06/2006
  - The publication of the act happened in the Official Gazette #130
  - The enactment date can be computed to be 10/07/2006
  - …
- These assertions assign values to a fixed set of metadata elements, that are defined in a metadata schema. E.g., Dublin Core
- The limitation of metadata schemas is that values are just plain strings, not references to entities.

Ontologies (1)

- The assertion “The author of these slides is Fabio Vitali” is more formally represented as “The value of the author property of these slides is “Fabio Vitali”.
- In a metadata schema, “Fabio Vitali” is just a string. All inferences based on this value (correctness, identification, association with related information, association with other documents authored by “Fabio Vitali”, etc.) are impossible or out of the picture.
- An ontology defines classes, properties and relations. The value associated to a property can be either a string, or a reference to an instance of a class.
- For instance, in an ontology “Fabio Vitali” would be a reference to an instance of the class “person”, and this justifies inferences based on this value.

Advantages of ontologies over metadata schemas

- We are not using strings but references to entities
  - the name of the person writing these slides can be presented as "*Fabio Vitali*", "*Vitali Fabio*", "*F. Vitali*" and/or "*Prof. Fabio Vitali*",
  - Also any other variation, including spelling errors, name changes, etc.
- We can associate other properties to the entity being referred
  - The e-mail address, the telephone number, the birth date, the VAT code, etc.
  - E.g., it becomes possible to ask for the telephone number of the person writing these slides

Referring to instances in ontologies

- Based on identifiers called URIs (Universal Resource Identifiers)
- A list of fully described instances are provided somewhere (e.g. on the parliamentary server), and each is provided with an identifier
- Assertions connecting a document to an instance would just use that identifier.
- If, for instance
  - "**`www.site.com/slide4.ppt`**" is the URI for these slides
  - "**`#FabioVitali`**" is the URI for the person whose name in Fabio Vitali, then

`www.site.com/slide4.ppt authoredBy #FabioVitali`

is the above assertion using a reference, instead of a string

References to ontologies in Akoma Ntoso

- Always based on an intermediary in the metadata section
- Always referring to an instance placed on an external ontology database, that is part of Bungeni
- Still, always containing enough information for readability and understanding even without access to the external ontology database.

```
<akomantoso>
  <meta>
    <person id="123"
            href="FV"/>
  </meta>

  <debate>
    <question by="123">
       Blah blah
    <question>
```

```
<person id="FV">
  <name>Fabio Vitali</name>
  <tel>+39 051 2094872</tel>
</person>
<person id="MP">
  <name>Monica Palmirani</name>
  ...
```

Next: Example 1: documents

Example 1: documents

```
<lifecycle>
  <event id="evn1" source="org01" date="2005-08-29"/>

  <event id="evn2" source="pas01" date="2005-12-31"/>
</lifecycle>

<references>
  <Original   href="/bill/2003/76_0904_en.xml”
              id="org01” showAs="Bill #76/2003" />
  <PassiveRef href="/act/2005/97_1231_en.xml”
              id="pas01” showAs="Act #97/2005"/>
</references>
```

Next: Example 2: people and roles

Example 2: people and roles

```
<akomantoso>
  <minutes>
    <metadata>
      <references>
        <Person href="/ontology/person#MakOnyango"
                id="MakOnyango"  showAs="Mr. Mak'Onyango" />
        <Role   href="/ontology/role#MFA" id="MFA”
                showAs="Minister for Foreign Affairs and
                International Co-operation" />
      </references>
    </metadata>
    <debate>
      <subdivision id="que824">
        <title>REVIVAL OF E.A. COMMUNITY</title>
        <question by="MakOnyango" to="MFA">
          <p>Mr. Mak'Onyango asked the Minister for Foreign
  Affairs and International Co-operation</p>
          ...
      </subdivision>
```